

A close-up photograph of a person's face, focusing on their eyes and forehead. They are wearing a virtual reality headset with a bright blue light emanating from the center. From this light, several thin, glowing blue lines radiate outwards towards the bottom of the frame, creating a sense of depth and perspective. The background is a dark, solid color.

Fostering Artificial Intelligence at Schools

Inclusion guide for (non-formal) educators in the context of Artificial Intelligence education

Table of Contents

Introduction	4
About the guide	5
FAIaS Guide objectives	5
1. Developing inclusive non-formal education programs on artificial intelligence	6
1.1. The advantages of learning about AI in a non-formal setting	7
1.2. Promoting gender, diversity, and inclusion in the learning environment	8
1.3. Inclusive teaching practices in non-formal education	10
2. Diversity and inclusion in artificial intelligence	14
2.1. Artificial intelligence – what is it?	15
2.2. Bias in artificial intelligence	17
2.3. Why is gender, diversity and inclusion important for AI?	23
2.4. Female role models changing the field of artificial intelligence	24
3. Guides for activities promoting an inclusive approach to artificial intelligence	28
3.1 Activity 1: The tag game	30
3.2 Activity 2: Draw the description	33
3.3 Activity 3: Guess the person	37
3.4 Activity 4: Gender bias using LearningML with dataset 1	41
3.5 Activity 5: Gender bias using LearningML with dataset 2	47
3.6. Activity 6: Identifying bias in AI	52
Conclusions	57
Definitions and glossary	59
References	60
Image Credits	62

Imprint

2023: CollectiveUP info@collectiveup.be

Editors: CollectiveUP (BE), King Juan Carlos University (SP), Teatro Circo Braga (PT), Vrije Universiteit Brussel (BE).

Authors: Liliana Carrillo, Chrysanthi Katrini, Paul Fenton, Loredana Bucșeanu (CollectiveUP), Marjon Blondeel (Vrije Universiteit Brussel), Meritxell Diaz, Antonio Romero, Gregorio Robles (King Juan Carlos University), Sara Borges, Joana Miranda (Teatro Circo de Braga).

Layout by: Jurate Laugalyte (CollectiveUP), Carlos Nivia Otero (CollectiveUP).

Translated by: Frederick Ducatelle (Dutch), Ana Oliveira Monteiro (Portuguese), Gregorio Gobles and Meritxell Díaz Coque (Spanish).

We suggest citing this report as follows: Carrillo, L., Katrini, C., Fenton, P., ... (2023). *Inclusion guide for (non-formal) educators in the context of Artificial Intelligence education.*

www.fosteringai.net

This publication has been written within the Erasmus + project Fostering AI at Schools, which aims to support educators to integrate and use Artificial Intelligence concepts in their lessons by creating and sharing resources about AI that can be used in the classroom.

CollectiveUP thanks the Digital Belgium Skills Fund (DBSF) for co-funding the AI4InclusiveEducation project (www.ai4inclusiveeducation.be) with a grant number DBSF2023-14. DBSF co-funded and partially supported the achievement of the Erasmus+ FAIaS project by creating an impact in Belgium.

This publication (in English, Spanish, Dutch, and Portuguese) and further publications and tools produced in the project can be downloaded free of charge at: www.fosteringai.net

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>



Disclaimer

Fostering AI at Schools is a project partnership, co-funded through the European Commission. This publication reflects the views only of the authors, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

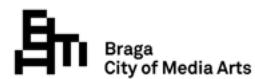
Fostering AI at Schools – Erasmus+ Project, Action type “Cooperation partnerships in school Education”, Grant number: 2020-1-ES01-KA201-083047.



Co-funded by the
Erasmus+ Programme
of the European Union



Teatro Circo de Braga
EM, S.A.



Introduction

In an era where artificial intelligence (AI) is becoming increasingly prevalent in our daily lives, the ethical and responsible development of this transformative technology is of paramount importance [1]. Fostering Artificial Intelligence at Schools (FAiaS) is an innovative and transformative project dedicated to nurturing ethical and responsible artificial intelligence (AI). Its ultimate goal is to improve AI knowledge in children and youngsters. By doing so, FAiaS aspires to cultivate a generation that is not only knowledgeable about AI, but also actively engaged in critical thinking and responsible decision-making. With the increasing prevalence of AI in our society, this project recognizes the urgent need to address the complex ethical challenges and societal implications – in particular the need for young people to be more knowledgeable and skilled regards AI usage.

Through thought-provoking articles, case studies, and practical guidelines, the project explores various facets of AI ethics and the societal impact of AI [2] for use in non-formal education. By examining real-world examples and drawing from the experiences of experts, FAiaS provides valuable insights and frameworks to guide the responsible development and deployment of AI systems.

This book serves as a guiding light, providing a framework for navigating the complex ethical dilemmas that arise as AI becomes increasingly integrated into various aspects of our lives, including healthcare, finance, education, and more. “Inclusive guide for (non-formal) educators” tackles topics such as algorithmic bias, privacy concerns, accountability, transparency, and the impact of AI on the job market. The book advocates for a holistic approach to AI development—one that prioritizes the well-being and empowerment of individuals and communities, while ensuring that AI systems adhere to principles of fairness, equity, and justice.

About the guide

This guide is a useful resource for non-formal educators who want to incorporate concepts of artificial intelligence (AI) into their curriculum and teaching, with a focus on diversity and inclusion practices. In this way, it complements the “Guideline for teachers in formal education” developed in the framework of the FAIaS project.

Chapter 1 Sets the context for the opportunities created by non-formal education to learn about AI, introduces a series of concepts related to the promotion of gender, diversity, and inclusion in the learning environment, and gives examples of inclusive teaching practices in non-formal education.

Chapter 2 Dives deeper on the challenges of developing inclusive AI systems, introducing the concept of bias in AI, its causes, impact and possible solutions. It also discusses why gender, diversity and inclusion is important when creating AI technologies, and gives examples of women role models changing the field of AI.

Chapter 3 Consists of practical guides for activities promoting an inclusive approach to artificial intelligence. It includes a description and step-by-step instructions for each activity.

Finally, this guide presents a glossary of terms and further resources to be used in the teaching practice.

FAIaS Guide objectives

- Understanding the basics of AI and its impact on society.
- Learning about practices that promote diversity and inclusion in the non-formal educational context;
- Understanding the role of non-formal educators in promoting a more inclusive and equitable future for AI.
- Recognizing the importance of inclusivity in AI and its applications.
- Developing skills to design and deliver engaging and inclusive AI lessons.
- Acquiring the ability to facilitate discussions and debates on AI and its implications.
- Engaging students to become critical and reflective learners in the field of AI.



1

Developing inclusive non-formal education programs on artificial intelligence

1.1. The advantages of learning about artificial intelligence in a non-formal setting

AI has the potential to revolutionize the way we live, work, and learn. As AI continues to grow in popularity and importance, it is becoming increasingly important for individuals to understand and be able to work with AI. Non-formal education programs can play an important role in teaching AI to a wide range of individuals, from students and professionals to those who are just starting to explore the world of technology.

One of the key advantages of non-formal education is its flexibility and adaptability. Non-formal education programs can be tailored to meet the specific needs and interests of individual learners, making them an ideal platform for teaching AI. For example, online courses and workshops can be designed to provide hands-on experiences with AI, while community-based programs can help individuals build their AI skills and knowledge in a supportive and collaborative environment.

Another advantage of non-formal education is its ability to reach underrepresented communities. AI is a rapidly growing field, and it is important to ensure that individuals from all backgrounds and experiences have access to education and training in AI. Non-formal education programs can help bridge the gap between formal and informal education, providing opportunities for individuals to learn about AI who may not have access to traditional education or training programs.

In order to teach AI effectively in non-formal education programs, it is important to focus on both the technical and ethical aspects of AI. Technical skills, such as coding and programming, are essential for individuals who want to work with AI. However, it is also important to teach individuals about the ethical and social implications of AI, including issues related to privacy, data protection, and bias in AI systems. By including both technical and ethical aspects of AI in non-formal education programs, individuals can develop a well-rounded understanding of AI and be better prepared to make informed decisions about its use.

Therefore, non-formal education programs have the potential to play an important role in teaching AI to a wide range of individuals and complementing the existing programs in formal education. By leveraging the advantages of non-formal education, such as flexibility and adaptability, and by focusing on both the technical and ethical aspects of AI, individuals can develop the skills and knowledge they need to succeed in the rapidly changing technological landscape. By providing access to AI education and training, non-formal education programs can help individuals from all backgrounds and experiences reach their full potential and contribute to the development of a more inclusive and equitable world [3].

1. Developing inclusive non-formal education programs on AI

Some of the key ways in which non-formal education can support the learning about technology and AI for youth include:

- **Providing hands-on experiences:** non-formal education programs can provide youth with hands-on experiences with technology, such as coding workshops, robotics programs, and maker spaces.
- **Fostering creativity and innovation:** by providing learners with a variety of tools and resources to experiment and explore with technology, non-formal education can foster their creativity and innovation competencies.
- **Supporting underrepresented communities:** non-formal education can help address the digital divide by providing access to technology and technology education to communities that are typically underrepresented in technology fields.
- **Encouraging lifelong learning:** non-formal education can help youth develop a lifelong love of learning about technology and prepare them for the rapidly changing technological landscape.
- **Building practical skills:** Non-formal education programs can provide youth with practical skills in areas such as computer programming, web design, and app development, which can help them build marketable skills and prepare for careers in technology.

In sum, the role of non-formal education in learning about AI and technology, more broadly, is to provide a complementary, learner-centered and supportive environment to help young people develop the skills and knowledge they need to thrive in the digital age [4].

1.2. Promoting gender, diversity, and inclusion in the learning environment

Promoting gender, diversity, and inclusion in non-formal education programs on AI is crucial for achieving a more equitable technology industry. AI is a rapidly growing field, and it is essential that individuals from all backgrounds and experiences have access to education and training in AI. Non-formal education programs can play an important role in promoting gender, diversity, and inclusion in AI by creating opportunities for individuals from underrepresented communities to learn about AI and develop the skills they need to succeed in this field.

One of the key advantages of non-formal education programs is their flexibility and adaptability. This allows programs to be designed to meet the specific needs and interests of individual learners, including those from underrepresented communities.

1. Developing inclusive non-formal education programs on AI

In order to promote diversity, and inclusion in non-formal education programs on AI, it is important to ensure that the programs are accessible and inclusive. This can be achieved by offering flexible and affordable program options, and by providing support services to ensure that individuals from underrepresented communities are able to participate in the programs. It is also important to ensure that the programs are taught by instructors who reflect the diversity of the community, and who have experience working with individuals from underrepresented communities.

Another important aspect of promoting diversity, and inclusion in non-formal education programs on AI is to focus on both the technical and ethical aspects of AI. Technical skills, such as coding and programming, are essential for individuals who want to work with AI. However, it is also important to teach individuals about the ethical and social implications of AI, including issues related to privacy, data protection, and bias in AI systems. By including both technical and ethical aspects of AI in non-formal education programs, individuals can develop a well-rounded understanding of AI and be better prepared to make informed decisions about its use [5].

By creating an inclusive learning environment, we can improve the quality of education for all learners. **But what does this mean?**

Gender, Diversity and Inclusion in education refers to the recognition and appreciation of the diverse identities, experiences, and perspectives of students, educators, and other stakeholders within the education system. It involves creating an inclusive and equitable learning environment where all individuals feel valued and supported, regardless of their gender, race, ethnicity, religion, sexual orientation, ability, age, or socio-economic background.

This includes addressing issues of discrimination, bias, and marginalization, and implementing strategies to support the academic and personal success of all students. Gender, Diversity and Inclusion in education also involves actively promoting equity and social justice, and challenging systems and structures that perpetuate inequality and exclusion [6].

1.3. Inclusive teaching practices in non-formal education

As an educator, your role is key in helping learners build a strong foundation for their futures. Promoting an inclusive environment means recognizing the differences between students, as well as their similarities. It means making sure that all students are comfortable about who they are, and giving them opportunities to explore their interests outside of school. By doing so, we help create a more inclusive environment for everyone involved. Here are a few ways to do that:

1. Make sure that you are aware of your own biases and stereotypes

If you find yourself making assumptions about students based on their appearance or behavior, take a step back and try to see them through a different lens! Make an effort to check your assumptions. You may not even realize that you have made certain assumptions about students, but this is a good way to start building awareness of your own biases.

All people harbor beliefs and attitudes about groups of people based on their race or ethnicity, gender, body weight and other traits. Those beliefs and attitudes about social groups are known as **biases**. Biases are beliefs that are not founded by known facts about someone or about a particular group of individuals [7].

Why do biases exist? In most cases, biases form because of the human brain's tendency to categorize new people and new information. To learn quickly, the brain connects new people or ideas to past experiences. Once the new thing has been put into a category, the brain responds to it the same way it does to other things in that category. More precisely:

- **Too much information:** our world has so much information that we need shortcuts and tricks to pick out the important bits;
- **Not enough meaning:** we fill in gaps in information using what we think we already know;
- **Need to act fast:** we need to assess and apply info, try to predict the future, and act on new insights – but don't have much time;
- **What should we remember?** We can not remember everything so we try to retain what is important for the future and forget the rest!

1. Developing inclusive non-formal education programs on AI

Here are some examples of biases that educators may have and that can hinder the education process:

- **Racial Bias:** An educator may unconsciously hold negative stereotypes about certain racial groups and this may affect their teaching style and behavior towards students from these groups [8].
- **Gender Bias:** An educator may have gender stereotypes that impact their interactions with male and female students. For example, they may assume that boys are better at math and science, while girls are better at language and arts [9].
- **Age Bias:** An educator may have biases against students of different ages. For example, they may assume that younger students are immature and lack the ability to concentrate, while older students are too set in their ways to learn new things [10].
- **Ability Bias:** An educator may have biases against students with disabilities or special needs, assuming that they are unable to learn or participate in the same way as other students [11].
- **Socioeconomic Bias:** An educator may unconsciously hold biases based on a student's socioeconomic status, assuming that students from lower-income families are less intelligent or motivated [12].

2. Be mindful of the language that you use

Being inclusive means using words like "everyone" and "no matter what" instead of phrases like "boys and girls" or "those who identify as boys or girls." Try to avoid making assumptions about gender identity by asking students what pronouns they prefer (for example: he/him, she/her, they/them) rather than assuming that everyone uses masculine ones.

3. Make sure that you are aware of the needs of all students

This means taking time to ask questions about how they feel and what they want to learn during your class. Be sure to include everyone in discussions and activities, as well as provide opportunities for those who may feel left out by default.

4. Make an effort to create a safe space for them

This may mean setting clear expectations for behavior and enforcing them consistently. It also means being mindful of what you say and do in front of your students, as well as making sure that you don't unintentionally leave them out or make them feel uncomfortable.

5. If you are an educator and find yourself having difficulty with any of these steps, it is important to reach out for support

Many organizations have at least one staff member who is trained in helping educators who struggle with diversity issues. If yours does not, try searching for local partners that provide this kind of support [13].

Non-formal education programs have the potential to play a vital role in promoting gender, diversity, and inclusion in AI, both when it comes to creating an inclusive learning environment and emphasizing the importance of diversity in developing AI systems. By leveraging the advantages of non-formal education, such as flexibility and adaptability, and by focusing on both the technical and ethical aspects of AI, individuals from underrepresented communities can develop the skills and knowledge they need to succeed in this rapidly growing field. By providing access to AI education and training, non-formal education programs can help to build a more inclusive and equitable technology industry, where individuals from all backgrounds and experiences are able to reach their full potential and contribute to the development of a more inclusive and equitable world.

Inclusive teaching practices are essential in creating a supportive learning environment for all students. Inclusivity means that all students are valued and respected, regardless of their backgrounds, experiences, and abilities. This approach to education helps to promote student engagement, academic success, and social and emotional well-being.

One of the key principles of inclusive teaching is to recognize and respect the diversity of students in the classroom. This includes acknowledging and valuing the different backgrounds, cultures, and experiences that students bring to the learning environment. By embracing this diversity, teachers can help students to feel seen and valued, and can create a sense of community in the classroom.

Another important aspect of inclusive teaching is to provide accommodations for students with diverse needs. This might include using technology, such as closed captions or audio descriptions, to support students with hearing or visual impairments, or providing alternative assessment methods for students with learning difficulties. Teachers should also be flexible in their teaching approaches, adapting their methods to suit the needs of individual students.

Inclusive teaching also involves creating a safe and supportive learning environment, where all students feel comfortable expressing themselves and asking questions. This means creating a classroom culture that is free from discrimination and harassment,

1. Developing inclusive non-formal education programs on AI

and actively promoting respect and acceptance for all students. Teachers can achieve this by establishing clear expectations for behavior, fostering positive relationships between students, and by addressing any incidents of bullying or discrimination promptly and effectively.

In addition, inclusive teaching practices involve engaging students in the learning process and encouraging them to take an active role in their own education. This means providing opportunities for students to participate in class discussions, ask questions, and share their ideas and perspectives. Teachers should also encourage students to work collaboratively, helping them to develop important social and communication skills [14].

Here are some examples of inclusive teaching practices that can be used in AI education programs:

- **Diverse representation in material:** Including examples of diverse people, cultures, and perspectives in course material and case studies.
- **Accessibility:** Using technology and alternative learning methods to make the course accessible to all students, regardless of ability.
- **Inclusive language:** Avoiding language that may be discriminatory, such as gendered or stereotypical language, and using inclusive language in class discussions and activities.
- **Encouraging diverse perspectives:** Creating opportunities for students to share their unique perspectives and experiences, and encouraging diverse perspectives in class discussions.
- **Group work:** Encouraging students to work in diverse groups, helping them to build positive relationships and develop important social and communication skills.
- **Flexible teaching approaches:** Being flexible in teaching approaches and adapting to the needs of individual students, rather than using a one-size-fits-all approach.
- **Providing context:** Including historical and cultural context in the teaching of AI to help students understand its development and impact on different communities.
- **Encouraging ethical discussions:** Encouraging discussions about the ethical and societal implications of AI and how it may impact different communities and individuals.
- **Providing opportunities for reflection:** Providing opportunities for students to reflect on their own biases and experiences, and encouraging them to become active advocates for diversity, equity, and inclusion in the field of AI.

By incorporating these inclusive teaching practices, educators can help students to develop a more diverse and inclusive understanding of AI, and equip them with the skills and knowledge needed to contribute to a more equitable future [15].



2

Diversity and
inclusion in artificial
intelligence

2.1. Artificial intelligence – what is it?

For the purpose of our guide we will not go deep into the technical aspects of AI, but rather look at the public perception of AI, how it shapes the world around us and why it matters.

More detailed information, instructions on teaching AI, and lessons plans can be found in the “Guidelines for teachers in formal education” developed as part of the “Fostering artificial intelligence at School” project and that can be downloaded for free here: <https://fosteringai.github.io/project/result1/>

Artificial intelligence (AI) is a discipline from computer science that tries to replicate and develop human intelligence and its implicit processes through computers. The term was coined in 1956 at the Dartmouth Conference by the renowned computer scientist John McCarthy, although some first experiences go back to shortly after World War II.

Today, AI is considered one of the defining technologies of the last decade, and perhaps also the next. It is currently a “hot topic”: media coverage and public discussion about AI is almost impossible to avoid. Google, Facebook and Apple are developing intelligent bots and AI assistants that promise to change how we work, play and learn. Wherever you look, there is evidence of it: your smartphone listens to your commands and understands what you are saying; self-driving vehicles are becoming a reality; robotics continue to evolve in factories across the globe. Cloud computing has made AI more accessible than ever before. As a result, developers are creating a wide range of products with AI at their core: interacting with customers using chatbots; helping retailers predict what customers want and when they will buy it; handling medical diagnosis automatically within minutes instead of days or weeks [16].

Although there is no agreement on a unique definition of AI, there is a consensus on four main approaches: two of them based on humans (systems that think like them and systems that act like them) and two based on rational aspects (systems that think rationally and systems that act rationally).

Some examples of AI systems include:

- Speech recognition software, which can understand and transcribe spoken words;
- Image recognition systems, which can identify objects or people in images or video;
- Natural language processing (NLP), which allows machines to understand and generate human-like language;
- Decision-making algorithms, which can analyze data and make decisions based on that analysis.

2. Diversity and inclusion in artificial intelligence

In education, there are many ways in which AI is being used to enhance the learning experience and improve outcomes. Some examples include [17]:

- **Adaptive learning systems:** these use AI to personalize learning experiences for each student based on their strengths, weaknesses, and learning style. The system can adjust the content and pace of the learning material to better suit the individual student's needs.
- **Intelligent tutoring systems:** these use AI to provide personalized tutoring and feedback to students as they work through assignments and practice problems. These systems can help students identify areas where they need additional support and provide guidance and resources to help them learn more effectively.
- **Language translation and language learning tools:** AI can be used to translate text and speech from one language to another, making it easier for students to learn and communicate with people who speak different languages.
- **Grading and assessment tools:** AI can be used to grade assignments and assessments, providing students with immediate feedback and freeing up educators to focus on more complex tasks.
- **Virtual assistants:** AI-powered virtual assistants can help students find information, complete tasks, and answer questions, freeing up more time for educators to focus on teaching and learning.
- **Personalized recommendations:** AI can analyze a student's past performance and preferences to provide recommendations for learning materials and resources that are most likely to be helpful and engaging for that student.

Because AI is increasingly shaping the world we live in, it is important not only to learn about its impact and potential implications, but also about how to be active in developing the technology in a way that benefits the entire society. Here are some reasons why learning about AI is important:

- **Preparing for the future:** AI is rapidly becoming integrated into many industries, and learning about AI can help individuals prepare for future job opportunities and ensure they are equipped to thrive in a rapidly changing workforce.
- **Understanding its impacts:** AI has the potential to have significant impacts on society, both positive and negative. By learning about AI, individuals can gain a better understanding of its potential implications, including the potential for job displacement, privacy concerns, and ethical considerations.
- **Making informed decisions:** AI is increasingly being used to make decisions that affect people's lives, from healthcare to criminal justice. Learning about AI can help individuals to make informed decisions about these applications and advocate for responsible and ethical use.
- **Fostering innovation:** AI has the potential to drive technological innovation and create new products and services. By learning about AI, individuals can better understand its potential applications and contribute to shaping its development in ways that align with their values and priorities.

→ **Developing critical thinking skills:** Learning about AI requires critical thinking and the ability to evaluate complex information.

Learning about AI is essential for individuals to understand its impact and be equipped to navigate its potential implications. By investing in education about AI, individuals can ensure that they are prepared to participate in shaping a more equitable and responsible future [18].

2.2. Bias in artificial intelligence

Automated systems are not inherently neutral. Computers must be coded by persons and as a consequence they reflect prejudices and preferences of those who code the AI algorithms behind these systems.

Bias in AI refers to the tendency of AI systems to exhibit certain prejudices or stereotypes in their decision-making processes and outputs. This can occur due to the data that is used to train the AI, as well as the algorithms and models that are used to build and operate the system. AI bias occurs because human beings both choose the data that algorithms use and decide how the results of those algorithms will be applied. Without extensive testing and diverse teams, it is easy for unconscious biases to enter machine learning models. Then AI systems automate and perpetuate those biased models.

One common way bias is introduced in such systems is because of **biased data**. For instance, suppose you want to automate the decision process of whether an applicant gets accepted into your university or college. Suppose that in the past you have admitted mostly men. If you use this historical data to train the new system, it is likely this gender bias is replicated.

When using real-world data, like news articles or social media posts, AI systems are likely to reinforce the existing prejudices. Translate apps for instance tend to be biased against women when translating from languages that use gender neutral pronouns. Typical examples are “she takes care of the children” and “he invests” [19].

When looking at classification of data, **the way the data is collected is very important. Both under and oversampling may lead to bias**. An example of undersampling is when face detection systems are mostly trained on white male faces and as a consequence do not recognize darker female faces as good as lighter male faces. Oversampling on the other hand may lead to over representation of certain groups. For instance, crimes committed in areas frequented by the police will be more present in the data set and the AI model trained on the data will probably reflect this disproportion and these areas will be labelled as more dangerous.

People can also **reinforce bias in already deployed AI models**. For example, an AI system might already be showing ads for a high-interest credit card to people with a lower level of education. They might click on this ad without realizing this will reinforce the already existing bias and they will continue to receive such suggestions.

Bias in AI can have significant negative impacts on society and individuals. The potential impacts of bias in AI include:

Discriminatory outcomes

AI systems that are biased may produce results that unfairly disadvantage certain groups of people, such as women or minority groups. This can lead to discrimination and inequality in areas such as employment, lending, and healthcare.

Facebook's Ads Algorithm

In 2019, Facebook [20] was found to be in contravention of the U.S. constitution, by allowing its advertisers to deliberately target adverts according to gender, race, and religion, all of which are protected classes under the country's legal system. Job adverts for roles in nursing or secretarial work were suggested primarily to women, whereas job ads for janitors and taxi drivers had been shown to a higher number of men, in particular men from minority backgrounds. The algorithm learned that ads for real estate were likely to attain better engagement stats when shown to white people, resulting in them no longer being shown to other minority groups.

This issue stems from how the AI machine learns. As is the nature of machine learning algorithms, the ad platform formed a pattern from the data it was given, but the pattern reflected existing societal inequalities and left unchecked would have helped to propagate them further. This practice is known as "discrimination-based advertising," and it has been found to perpetuate harmful stereotypes and perpetuate discrimination in society. This issue highlights the importance of ethical considerations in the development and use of AI. It is essential that algorithms are designed and monitored in a way that does not perpetuate or amplify existing biases and discrimination.

Limited opportunities

Bias in AI systems can limit opportunities for certain groups of people. For example, if an AI system is biased against women, it may not recommend them for certain job positions or education programs, leading to a lack of advancement and growth for these individuals.

Amazon's biased recruiting tool

In 2018, Reuters [21] reported that Amazon had been working on an AI recruiting system designed to streamline the recruitment process by reading resumes and selecting the best-qualified candidate. Unfortunately, the AI seemed to have a serious problem with women, and it emerged that the algorithm had been programmed to replicate existing hiring practices, meaning it also replicated their biases. That is because Amazon's computer models were trained to vet applicants by observing patterns in résumés submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry.

In effect, Amazon's system taught itself that male candidates were preferable. It penalized résumés that included the word "women's", as in "women's chess club captain". And it downgraded graduates of two all-women's colleges. Rather than helping to iron out the biases present in the recruitment process, the algorithm simply automated them. Amazon confirmed that they had scrapped the system.

Bias in hiring and recruitment can occur when certain groups of people are unfairly disadvantaged or excluded from job opportunities based on factors such as their race, gender, age, religion, sexual orientation, or disability status. This can lead to a lack of diversity in the workplace and can have negative consequences for both the affected individuals and the company as a whole.

It is important for companies to be aware of the potential for bias in their hiring and recruitment processes and to take steps to mitigate it. This can include using diverse recruitment sources, implementing fair and objective evaluation criteria, and training employees on diversity, equity, and inclusion.

Misinformation

Bias in AI systems can lead to the dissemination of false or misleading information. For example, if an AI system is biased against a certain group of people, it may spread misinformation about that group, leading to negative stereotypes and discrimination.

News recommendation systems that use AI algorithms to personalize content based on a user's previous reading habits and preferences can also perpetuate misinformation. If these algorithms are biased towards particular perspectives, they can reinforce existing beliefs and spread false or misleading information.

Ineffective decision-making

Bias in AI systems can lead to poor decision-making, as the AI system may not be accurately considering all relevant factors. This can lead to negative consequences in areas such as healthcare, criminal justice, and finance.

Racism in U.S. healthcare provision

In 2018, a team from the University of California [22], Berkeley discovered a problem with an AI system that was being used to allocate care to 200 million patients in the U.S., which resulted in black patients receiving a lower standard of care. The problem stemmed from the fact that the system was allocating risk values using the predicted cost of healthcare as the determining variable, and because black patients were often less able to pay or were perceived as less able to pay for the higher standard of care, the AI system essentially learned that they were not entitled to such a standard. Across the board, black people were assigned lower risk scores than white people, despite the fact that the black patients were also statistically more likely to have comorbid conditions and thus in fact experience higher levels of risk. This in turn meant that black patients were less likely to be able to access the necessary standard of care, and more likely to experience adverse effects as a result of having been denied the proper care. Having made this discovery, the UC Berkeley team worked with the company responsible for developing the tool to find variables other than cost through which to assign the Expected risk scores, reducing bias by 84 percent.

Numerous instances exist where bias in AI has led to racism in healthcare

provision. One example is the use of algorithms to predict which patients are at risk of developing certain conditions or diseases. These algorithms are often trained on data that reflects the healthcare experiences of certain populations, which means that they may not accurately predict the risks of individuals from other racial or ethnic groups. This can result in healthcare providers disproportionately allocating resources and treatments to certain racial or ethnic groups, rather than providing equal care to all patients.

Overall, it is crucial for healthcare providers and technology companies to be aware of the potential for bias in AI and to take steps to mitigate it. This may involve using diverse datasets to train algorithms, implementing safeguards to prevent discrimination, and ensuring that individuals have the opportunity to provide input and feedback on the use of AI in healthcare.

Bias in AI can have serious consequences, as it can lead to unfair treatment of certain groups and can perpetuate existing social inequalities. It is important to recognize and address bias in AI systems.

What can we do about the biases in artificial intelligence? [23]

Change the way people are educated about science and technology

One way to reduce bias in AI is to change the way people are educated about science and technology. This could include providing more diverse representation in science and technology curricula, highlighting the impact of historical and current biases in the field, and emphasizing the importance of ethical considerations in the development and use of AI.

Another approach could be to incorporate more hands-on, experiential learning opportunities that allow students to directly engage with and understand the potential biases and implications of AI. This could include projects or case studies that challenge students to identify and address biases in AI systems or to consider the ethical implications of their use.

Additionally, providing more resources and support for educators to better understand and address these issues can help to create a more inclusive and equitable learning environment for all students. This could include professional development opportunities, access to research and resources, and ongoing support and guidance from experts in the field.

Be aware of the contexts in which AI can help correct for bias as well as where there is a high risk that AI could exacerbate bias

When deploying AI, it is important to anticipate domains potentially prone to unfair bias, such as those with previous examples of biased systems or with skewed data. As both individuals and companies have some sort of social responsibility, we have an obligation to regulate our modeling processes to ensure that we are ethical in our practices. Organizations will need to stay up to date to see how and where AI can improve fairness—and where AI systems have struggled. Companies should establish a debiasing strategy within their overall AI strategy that contains a portfolio of technical, operational and organizational actions:

- **Technical strategy** involves tools that can help them identify potential sources of bias and reveal the traits in the data that affects the accuracy of the model
- **Operational strategies** include improving data collection processes using internal “red teams” and third party auditors.
- **Organizational strategy** includes establishing a workplace where metrics and processes are transparently presented.

Consider Human-in-the-Loop systems

The goal of Human-in-the-Loop technology is to do what neither a human being nor a computer can accomplish on their own. When a machine cannot solve an issue, humans must interfere and solve the problem for them. As a consequence of this procedure, a continuous feedback loop is created. With continuous feedback, the system learns and improves its performance with each subsequent run. As a result, human-in-the-loop leads to more accurate rare datasets and improved safety and precision.

Most popular machine learning algorithms require large amounts of labeled data to produce accurate results. However, there are many cases where there is not enough unlabeled data to draw from. For example, if you are looking for examples of fake news in a language with only a few thousand speakers, there might not yet be any examples of fake news in that language. Therefore, the algorithm will have nothing to learn from. In this case, keeping humans in the loop can ensure the same level of accuracy even for rarer types of data. This is the case of Facebook, which still keeps humans in the loop when it comes to monitoring and addressing activity on the social media platform.

2.3. Why is gender, diversity and inclusion important for artificial intelligence?

AI systems are designed and built by humans, and they often reflect the biases, prejudices, and stereotypes of the people who design them. If the AI development team is not diverse, it is more likely that the AI system will perpetuate harmful biases and stereotypes. Teams that achieve diversity across race, gender, sexual orientation, age or economic conditions develop better AI systems because diverse people reason differently and can come up with more and new ideas, are better at spotting bias and drive more creative thinking. It is a matter of cognitive diversity.

When YouTube [24] launched their video upload app for iOS, between 5 and 10 percent of videos uploaded by users were upside-down. Were people shooting videos incorrectly? No. Their early design was the problem. It was designed for right-handed users, but phones are usually rotated 180 degrees when held in left hands. Without realizing it, YouTube created an app that worked best for their almost exclusively right-handed developer team.

Data is giving us information for making informed decisions. But the reality is that data can be interpreted depending on who reads that data. Artificial intelligence interpretation of data can be as biased as the human intelligence behind it. What can happen when you have an excellent AI team but it is not cognitively diverse? As shown in the previous example, Amazon's recruiting algorithm was gender-biased, discarding competitive female candidates for roles after learning that the percentage of women in those positions was lower.

Promoting gender, diversity and inclusion in AI can help to ensure that AI technologies are designed and developed with a human-centered approach. Women and other underrepresented groups bring unique perspectives, experiences, and skills to the table, which can lead to the development of more creative, innovative, and effective AI technologies that better serve the needs of society. When AI systems are not diverse and inclusive, they may not adequately serve the needs and interests of all members of society. This can lead to a lack of trust in AI systems and a lack of adoption by certain groups.

2.4. Female role models changing the field of artificial intelligence

Women play an important role in shaping the future of Tech and AI. By increasing the visibility of female role models in Tech and AI and promoting them, we aim to inspire non-formal education professionals to create more inclusive educational programs where gender, diversity and inclusion play a central role in the development of educational activities and environments.

It must be pointed out that the list is not comprehensive, it is an attempt to give an overview of the incredible women that are a driving force in the field of AI across the world. By pushing the boundaries of what is possible, and by expanding our knowledge of the numerous and diverse potential of AI in health, infrastructure, climate, energy and language (amongst others), these women play an important role as role models for the future generation of female scientists, engineers and entrepreneurs.

Carolina Barcenas [25]

Senior Vice President for Research at Visa

She leads a team driving innovation in 4 areas of emphasis: Artificial Intelligence & Machine Learning where she is responsible for exploring and developing advanced ways for leveraging data to create business value for Visa through AI techniques; Blockchain; Security and Commerce of the Future. The team focuses on both product-driven and foundational research. Carolina has served as a co-leader of Visa Women in Technology in Austin and often participates in events that support the advancement of women, such as the Texas Women Conference, Latinas in Tech summits and Girlstart. She is the organizing force behind the community college intern program that focuses on non-traditional candidates. She has worked both in industry as well as academia and has over 20 years of experience in machine learning solutions in fintech. Prior to joining Visa, she spent 7 years at PayPal where she was responsible for managing the risk of small and medium e-commerce sellers. Her expertise is in deep data mining techniques. She holds a Ph.D. in Applied Statistics from the Georgia Institute of Technology as a Fulbright Scholar.

In my case, I started thinking that I would have a career as a physicist; however, I discovered that my passion was problem solving and creating mathematical abstractions to be utilized in decision making. After realizing my professional passion, I readjusted my goals and pursued a career in machine learning. Goals change and, as you progress, don't be afraid to adapt. Ultimately, it's key to work on something you are passionate about.

Stephanie Lampkin [26]

Technical founder and CEO of Blendoor

In her role as technical founder and CEO of Blendoor, Stephanie Lampkin is passionate about leveraging augmented intelligence and people analytics to match a diverse workforce of candidates to companies, ensuring candidate anonymity in the process. This, she believes, can mitigate unconscious bias, giving qualified people more opportunities and creating better companies and economies as a result. Stephanie also collects data for BlendScore, which is a report published by Blendoor. BlendScore rates tech companies on their recruitment, compensation, retention, and promotion of women and underrepresented minorities.

I started Blendoor three months after I had a pretty surprising interview experience with Google. I was living in New York at the time and had just graduated from MIT Sloan. I was interviewing for an Analytical Lead position at Google, and the hiring team came back to me with the consensus that I would be better suited for a Sales or Marketing position because they didn't see me as quite technical enough for the role. Prior to this interview, I had been coding and exposed to CS from an early age, got an engineering degree from Stanford, worked at Microsoft for 5 and a half years, and had just completed an MBA. Something about their feedback didn't seem quite right to me, but I politely declined the Sales/Marketing position and moved on. Later, I found out that Google at the time was only 2% African American, 3% Latino, and about 25% women. The narrative they were telling the media was that it was a pipeline problem, they just can't find enough qualified women and people of color. This led to an aha moment for me — to create a platform that makes it really easy for both candidates and companies to connect so that companies can get access to a broader pool of diverse talent. That was how Blendoor came to be.

Freyja Jørgensen [27]

Innovation Manager at Simula/Gründergarasjen, and Leader of the NORA.startup Steering Group

Freyja Jørgensen has a lot of energy and is the kind of person that simply makes things happen. One of the most active groups within the Norwegian artificial intelligence Research Consortium (NORA) is NORA.startup. The NORA.startup initiative will be a gateway to greater cooperation between academia, incubators and startup companies in the field of AI, machine learning and robotics. Jørgensen has an educational background combining molecular biology and psychology. She has extensive experience from working at the intersection between academia and startups, and has been involved in several educational and innovation projects. Formerly Jørgensen has

worked to promote science and technology in education but now she dedicates her time to supporting early stage technology companies in the Gründergarasjen incubator. She is also working to further build the NORA.startup network and support the AI startups there. Jørgensen is responsible for much of the initiative's internal organization and event management that will enrich the AI ecosystem of Norway and beyond.

From a young age I knew that somehow, I wanted to work with people and science. I dreamt of becoming a children's doctor, or a researcher. Having a grandfather with a chemistry PhD and a dedicated biology teacher in high school, inspired me further. Later, being involved in several educational and innovation projects at Simula, I quickly became very eager to work in a dynamic and interdisciplinary environment. That eventually led me to the extremely inspiring and quickly-growing field of technology startups and AI. AI and other technologies will continue to play a central role in our lives, and I find it important to support and understand this field. Each day I learn something new and meet new challenges and possibilities in the work I do – that is one of the main things that motivates me in my job.

Rebekah Agwunobi [28]

Student Choate Rosemary Hall and winner of the Caroline D. Bradley scholarship

Rebekah's interest in computers was supported by her family from a young age, starting with a JavaScript class she took in 3rd grade. Since then, she discovered AI and its applications for social good at Stanford AI4ALL, spent two summers as a research intern at USC, created a machine learning directed study at her high school, and more. Alongside her interest in tech, she's a passionate activist who cares about issues including mass incarceration, prison reform, and diversity in STEM. She's hopeful about using AI and computer science as effective tools to solve real-world problems.

"I wanted something built into my course schedule that I found empowering and allowed me to explore things I was interested in. That's why last term I did a machine learning directed study. In the class, we talked about some of the general techniques I learned at Stanford AI4ALL, and we also talked about some more advanced concepts. We did Kaggle competitions, and I learned a lot about different applications of machine learning in areas like art generation and music. I learned a lot of theory behind the math I was doing."

Abbey Lin [27]

Controls & Compliance Lead at Microsoft Norway,
Co-founder and Board Member Oslo.ai, Founder of Ethical
AI Resources

Abbey Lin has worked in the field of Ethics and Compliance over the last 15 years, most recently with Microsoft. It was through these experiences that she saw the emerging ethical challenges and opportunities shaping AI and wanted more people to be aware. During her maternity leave a few years ago, she created Ethical AI Resources, a website and news feed that gathers information and the latest news about responsible AI, which has become a popular resource on twitter www.twitter.com/alethical (@aithical). In addition Abbey is the co-founder of Oslo.AI, which is helping build and support a community of AI enthusiasts in Oslo through events and networking opportunities. Abbey is also a board member of Visualyst.co, an AI startup focused on Video compliance.

If you care about diversity and inclusion, you need to also care about responsible AI. AI has the ability to make a positive impact on the world, but can also, and already has, further negatively impacted already marginalized communities. Responsible AI needs to be considered at all stages of product development and consumers of AI need to question and demand transparency of how the product was created.



3

**Guides for activities
promoting an inclusive
approach to artificial
intelligence**

This section includes a collection of practical activities that non-formal education institutions can include in their educational programs to teach students about AI.

The proposed activities introduce different concepts of AI and include step-by-step descriptions of each activity, materials to be used, and complementary educational resources to support educators to plan and deliver engaging and meaningful activities to their students.

Before getting started, here are a few tips on creating a welcoming and inclusive learning environment for all students, regardless of their background or prior knowledge of AI:

- Begin by introducing the concept of AI to your students and explaining its various applications. You can use videos, texts, and other materials to provide an overview of AI and its potential impacts on society.
- Make sure to emphasize the importance of diversity, equity, and inclusion in the field of AI. This can include discussing the ways in which AI has been used to perpetuate biases and discussing strategies for creating more inclusive and equitable AI systems.
- Use case studies and real-world examples to illustrate the various ways in which AI is being used in different industries and sectors. This can help students see the practical applications of AI and understand how it is impacting the world around them.
- Encourage students to think critically about the ethical implications of AI and discuss the potential consequences of using AI in different contexts. You can use prompts and discussion questions to facilitate this conversation.
- Consider incorporating hands-on activities, such as coding exercises or projects, to give students the opportunity to apply what they have learned and explore AI concepts in a more interactive way.
- Encourage students to stay up-to-date on the latest developments in AI and to continue learning about the field even after the course is over. This can include staying informed about relevant news, attending conferences and workshops, and joining relevant online communities and organizations.

3.1 Activity 1: The tag game [29]

Learning Objectives

1. **Enhance non-verbal communication:** Participants will practice non-verbal cues and gestures to form groups without using verbal communication, fostering effective communication skills.
2. **Foster critical thinking and decision-making:** Participants will consider various factors, such as tag colors, shapes, patterns, and sizes, to make decisions about group formations, promoting critical thinking and decision-making skills.
3. **Promote collaboration and teamwork:** Participants will work together to form groups based on common characteristics, encouraging cooperation and teamwork without relying on verbal communication.
4. **Explore inclusivity and diversity:** Participants will reflect on their group formations and discuss whether they considered factors beyond the tags, fostering an understanding and appreciation of diversity.
5. **Facilitate reflective and open discussion:** The final discussion initiated by the leader encourages participants to share their decision-making processes and insights, promoting reflection, active listening, and the exchange of ideas among participants.
6. **Raise awareness about bias:** Participants will come into a realization of their own biases, mostly what affinity bias is, and what everyone share in common even though we are all different

Materials:

Scrap paper

Sticky tape

Pens or pencils

Coloured paper

Description of the activity

Get ready with tags or stickers for each participant. You have the option to create your own by cutting out shapes from colored paper or utilize pre-made stickers. Ensure that the tags come in a variety of colors, shapes, patterns, and sizes. For instance, you might have several purple tags, but they could be in the form of circles, squares, and triangles, with patterns such as squares, dots, and stripes.

The facilitator/leader begins by distributing individual tags to all participants. Each person should stick their tag to their clothing, placing it between their waist and shoulders. Once everyone is tagged, the participants start moving around the designated space, being mindful of others around them.

The leader then introduces variety by instructing everyone to move in a different way, such as hopping, skipping, or jumping. After a few minutes of movement, the leader prompts the participants to form small groups without using verbal communication. Observing how they split themselves into groups becomes crucial. Participants may consider their tags and organize themselves based on color or shape. Once the groups have formed, everyone spreads again, continuing to move around the space.

The process is repeated multiple times, with participants regrouping without talking. As the game progresses and more groups are formed, the participants eventually gather in a large circle. At this point, the leader initiates a discussion, encouraging everyone to share how they decided on their group formations. The conversation explores whether anyone looked beyond the tags or intentionally created diverse groups encompassing different shapes, colors, patterns, and sizes.

A step-by-step guide

Step 1 The person leading the game should give everyone their own tag. Everyone should stick their tag to their top between their waist and their shoulders.

Step 2 Everyone should move around the space and stay at a safe distance from each.

Step 3 After a minute or two, the person leading the game should ask everyone to get into small groups without talking.

Step 4 Once the groups have formed, everyone should break apart again and continue to move around the space.

Step 5 Everyone should repeat steps three and four so they have made three or four different groups without talking.

Step 6 Everyone should gather in a large circle at a safe distance from each other.

Discuss

The person leading the activity should ask:

- How did they decide the groups they'd split into?
Take it in turns to share their thoughts.
- Did anyone look beyond the tags?
- Did anyone intentionally form a diverse group with a variety of shapes, colors, patterns, and sizes?

Reflection

Through this activity, participants were introduced to the concept of diversity and specifically explored affinity bias, which is a form of unconscious bias.

- Were individuals surprised to learn about affinity bias?
- How did they feel when reflecting on their group formations during the tag game and their list of trusted individuals?
- What are some advantages and challenges associated with collaborating in groups comprising both similar and diverse individuals?

Participants are encouraged to reflect on the various groups they belong to, such as a school class, sports team, or circle of friends. They should consider whether these groups consist mainly of individuals who are similar or different from themselves. Subsequently, participants should form pairs and engage in conversations about strategies for enhancing their understanding of diversity in their day-to-day lives and interactions. Taking turns, participants can share their ideas if they choose to do so.

3.2 Activity 2: Draw the description [30]

Learning Objectives

- Developing observation skills:** Participants will enhance their ability to observe and pay attention to details.
- Enhancing creativity:** Participants will exercise their creativity and imagination by interpreting descriptions in a visual form.
- Improving communication skills:** Participants will practice effective communication by accurately conveying descriptions through their drawings.
- Encouraging critical thinking:** Participants will engage in critical thinking by analyzing descriptions and translating them into visual representations.
- Fostering teamwork:** Participants will collaborate with others by sharing their interpretations and working together to solve challenges.

Materials

Coloured pens or pencils

Scrap paper

Prints of the photos and descriptions of the chosen people object, scene, person or concept

Description of the Activity

To begin the activity, participants gather in a group and are provided with drawing materials such as paper and pencils. The facilitator or a designated participant takes on the role of the describer.

They will read out a description of an object, scene, person or concept, without revealing what it is. The descriptions can be specific or abstract, allowing for a range of interpretations. As the facilitator/describer reads the description, participants listen carefully and visualize the details in their minds. They then have a set amount of time to translate their mental images into drawings on their papers. The drawings should aim to capture the essence of the described object or scene.

Once the drawing time is up, participants share their drawings with the group. The describer reveals the original object or scene, and participants compare their drawings to see how accurately they captured the description. This encourages reflection and discussion, allowing participants to explore the differences and similarities in their interpretations.

After each round, the roles of describer and participants can be rotated, ensuring everyone gets a chance to practice describing and drawing. The activity can be customized by varying the complexity of descriptions, introducing different themes, or incorporating time limits to add an element of challenge.

A step-by-step guide

Step 1

The person leading the activity should read the description of the object, scene, person or concept.. You can use the example below, of Katherine Johnson, who was one of the first African American women working for NASA, receiving her presidential award [31].

Example

Description of the image: In this scene, we see a mathematician from NASA receiving a presidential award in 2015. This mathematician was born in 1918 and made important contributions in the early days of NASA, including pioneering the use of computers for many tasks.



Step 2

Everyone should draw the object, scene, person or concept based on the description.

Step 3

Give everyone a few minutes to complete their drawings.

Step 4

Encourage everyone to share what they drew. Compare the similarities and differences between the drawings. For the example image used some of the questions can be:

- What are they wearing?
- What size are they (tall, short, fat, thin)?
- What color is their skin?
- Do they have a gender?
- How old are they?

Step 5

Show everyone the picture of the first person and reveal their name, maybe some details about who they are or what they do/done and chat about whose drawing is most similar to the actual image of the person. What are the most significant differences in people's drawings and the actual image?

Discuss

Ask people why they drew the description the way they did. What about the description made them think the object, scene, person or concept looked a certain way? What was it that made them choose a certain age, race, gender, size, or look for the person? What made them choose the colors or shapes they chose?

Reflection

In this activity, everyone drew object, scene, person or concept based on descriptions of the work they've done and the impact they've made. Ask everyone to think about the drawings which differed a lot from the actual image. Often we associate certain traits with certain jobs, personalities, achievements, placing or usability. This is unconscious bias and is caused by lots of things, like only learning about history from one perspective, or watching TV shows and movies where characters and scenarios are built on ignorant or harmful stereotypes.

1. Individual reflection: after completing their drawings, participants can take a moment to reflect on their own interpretations, considering such questions as:

- How did I interpret the description?
- What elements did I focus on and why?
- Did I capture the essence of the described object, scene, person or concept?
- What challenges did I face during the drawing process?
- What could I have done differently to improve my drawing?

Encouraging participants to reflect individually allows them to gain insights into their own thinking process, artistic choices, and areas for improvement.

2. Group reflection: after sharing their drawings with the group, participants can engage in a group reflection session facilitated by the describer or a designated leader. This session can include discussion and exploration of the different interpretations and variations in the drawings. Some reflection prompts for the group discussion may include:

- What similarities and differences do we notice in our drawings?

3. Guides for activities promoting an inclusive approach to AI

- What aspects of the description were challenging to capture visually?
- How did the variations in our interpretations contribute to a richer understanding of the description?
- Did any drawings provide unique perspectives or insights?

Group reflection allows participants to gain a broader perspective on the diverse interpretations and encourages open dialogue and appreciation for different approaches. It also fosters a sense of collective learning and shared experiences.

Explore various approaches to reduce unconscious bias. Inquire if anyone can propose suggestions. For instance on regards of the example image, you may:

- Immerse yourself in positive portrayals of individuals from diverse genders, races, or abilities by engaging with literature such as "Little Leaders: Exceptional Men in Black History" and "Exceptional Women in Black History" by Vashti Harrison or films like "The Peanut Butter Falcon."
- Initiate discussions about instances where people have faced discrimination based on factors like race, gender, sexuality, ability, and age.
- Developing an understanding of the issue enables better resolution.
- Make it a habit to seek multiple perspectives on historical events. Different individuals possess distinct viewpoints and experiences, influencing how they perceive and recount past occurrences.

Note: The image used in this activity is acting as an example. Feel free to use any kind of image that fits your course or curriculum.

3.3 Activity 3: Guess the person

Learning Objectives

- 1. Recognize biases:** Participants will become aware of the existence of biases, specifically the halo effect, which influences our perceptions and judgments of individuals based on limited information or preconceived notions.
- 2. Understand the halo effect:** Participants will gain an understanding of the halo effect as a cognitive bias that leads us to attribute positive qualities to individuals based on a single characteristic or trait, such as physical appearance or fame.
- 3. Challenge assumptions:** Participants will be encouraged to question their assumptions and biases by examining their initial perceptions of the famous engineer woman. They will explore how the halo effect can lead to inaccurate judgments and stereotypes about someone's profession or abilities.
- 4. Promote awareness of diversity in professions:** Participants will broaden their understanding of the diverse range of professions by realizing that individuals can excel in fields that may not align with common stereotypes or expectations. They will appreciate that talents and skills are not limited by gender or appearance.
- 5. Foster critical thinking:** Participants will develop critical thinking skills by analyzing their own thought processes and biases when making judgments. They will learn to question the influence of the halo effect and strive for more objective and unbiased assessments of individuals.
- 6. Encourage open dialogue:** The activity will promote open discussions among participants, allowing them to share their initial assumptions, reflect on their biases, and explore the impact of the halo effect. This dialogue will foster a deeper understanding of biases and their implications in various contexts.
- 7. Empower participants to challenge biases:** Through this activity, participants will be empowered to challenge biases and stereotypes they encounter in their daily lives. They will develop strategies for recognizing and overcoming the halo effect, promoting fairer and more equitable judgments and decision-making processes.
- 8. Enhance empathy and inclusivity:** By exploring biases and the halo effect, participants will develop empathy and a greater appreciation for diversity. They will understand the importance of creating inclusive environments that value individuals for their unique skills and contributions rather than superficial attributes.
- 9. Apply insights to real-world situations:** Participants will be encouraged to transfer their learnings from the activity to real-world situations, such as workplace interactions, hiring processes, or personal relationships. They will strive to make more informed judgments and decisions based on a comprehensive evaluation of individuals rather than relying solely on the halo effect.

10. Foster a culture of inclusion: The activity aims to contribute to the creation of a culture of inclusion, where biases are challenged, diversity is celebrated, and individuals are valued for their abilities and achievements rather than superficial characteristics. Participants will be encouraged to be advocates for fairness and equality in their personal and professional lives.

Materials

Picture of people you would like to talk about that are “victims” of halo effect

Description of the Activity

In this activity, the facilitator presents participants with an image of a famous person without disclosing her/his profession. The participants are then asked to guess what profession they believe she/he belongs to based solely on the image.

The purpose of this activity is to raise awareness around biases, specifically focusing on the halo effect. The halo effect is a cognitive bias where individuals form overall judgments or impressions of others based on limited information or a single positive trait. In this case, participants may be influenced by the image of a famous person, assuming a profession that aligns with common stereotypes or preconceived notions.

After participants have shared their initial guesses, the facilitator initiates a discussion on biases and the impact of the halo effect. Participants are encouraged to reflect on their assumptions and consider how the halo effect can lead to inaccurate judgments about someone's profession or abilities. The facilitator guides the conversation to explore the diversity of professions and the importance of recognizing individual talents and skills beyond superficial characteristics.

A step-by-step guide

Step 1

The person leading the activity should show the image of the famous person (ie Hedy Lamarr) and ask the participants to guess her/his profession.

Example

Who is Hedy Lamarr?

- Worked in films in Hollywood;
- Inventor who pioneered the technology that would one day form the basis for today's WiFi, GPS, and Bluetooth communication systems;
- Received the Invention Convention's Bulbie Gnass Spirit of Achievement Award.



Step 2

Give everyone a few minutes to think about their answer.

Step 3

Ask the participants to share their answers. In this step you can be as creative as you want. For example if you are in a physical place such as a classroom you can ask the participants to either use post-it to write their answer and use then a white board or a wall to display the results, or if you are on an online session then you can use Mural or Miro board.

Step 4

Encourage everyone to share what they wrote. Compare the similarities and differences between the answers:

- What was your initial assumption about the profession of that person in the image? What factors influenced your perception?
- Did the fame or appearance of that person in the image affect your judgment about her/his profession? How so?
- Have you ever encountered situations where you made assumptions about someone's abilities or profession based on superficial characteristics? Can you share any examples?

Discuss

Ask people questions about their own biases based on the example above such as Can you think of any real-life instances where the halo effect may have led to biased judgments or missed opportunities? How do you think the halo effect can impact decision-making processes, such as hiring or promotions? What strategies can we adopt to encourage diversity and inclusion, both in our personal lives and in broader societal contexts?

Reflection

Through this activity, participants realize how easily we can fall prey to biases, making snap judgments about someone's profession or abilities without considering the full scope of their skills and qualifications. The initial guess they made about the profession of the famous person in the image was influenced by her/his appearance, which made them question the potential impact of the halo effect on their own thinking.

Engaging in an open discussion with fellow participants further deepened their understanding of biases and their implications. Hearing others share their initial assumptions and experiences reinforced the pervasiveness of biases in various aspects of life. It was eye-opening to recognize that they too have at times made judgments based on superficial characteristics, overlooking the true abilities and potential of individuals.

The activity prompts them to reflect on the broader societal impact of biases and the importance of creating environments that value diversity and inclusivity. It made them realize that biases can hinder fair decision-making processes, affecting areas such as hiring, promotions, and overall equality.

One aspect that stands out during the discussion is the need for self-awareness and introspection. Recognizing our own biases is crucial to challenging and overcoming them. The activity encouraged us to question our assumptions, to be conscious of the potential influence of the halo effect, and to strive for more objective and inclusive judgments.

Note: The image used in this activity is acting as an example. Feel free to use any kind of image that fits your course or curriculum.

3.4 Activity 4: Gender bias using LearningML with dataset 1

Learning Objectives

1. **Practice recognizing and mitigating gender bias:** Participants will actively practice identifying and addressing gender bias by engaging with the LearningML website and utilizing computer-based activities. They will develop skills in critically analyzing algorithms, models, and data to minimize the impact of bias in technology-driven processes.
2. **Increase awareness of unconscious bias:** Participants will gain a deeper understanding of unconscious bias and its implications. They will explore how biases can influence decision-making and learn strategies to recognize and counteract unconscious biases in various contexts.
3. **Reflect on the use of automated software for hiring:** Participants will reflect on the ethical considerations and potential biases associated with using automated software for hiring purposes. They will critically analyze the advantages, limitations, and potential risks of relying on technology-driven systems for hiring decisions. Participants will develop a refined understanding of the role of human judgment in the recruitment process.
4. **Enhance critical thinking and decision-making skills:** Engaging with computer-based activities and reflecting on bias in technology and automated systems will foster participants' critical thinking abilities. They will develop skills to evaluate the reliability and validity of automated software and make informed decisions regarding its implementation in various domains.
5. **Foster awareness of the ethical implications of technology:** Participants will gain insights into the ethical considerations surrounding the use of technology, particularly in relation to bias and discrimination. They will reflect on the responsibility of individuals and organizations to ensure fairness, transparency, and inclusivity when utilizing automated systems.
6. **Promote digital literacy and responsible technology use:** Participants will develop digital literacy skills and understand the potential impact of technology on society. They will explore the importance of responsible technology use, considering the potential consequences of biased algorithms and the need for ongoing evaluation and improvement.
7. **Encourage self-reflection and continuous learning:** Participants will be encouraged to reflect on their own biases, assumptions, and perspectives throughout the learning process. They will develop a growth mindset, embracing the value of continuous learning and actively seeking opportunities to challenge and expand their understanding of bias and technology.

Materials

Laptop with internet access

[LearningML](#)

Description of the Activity

In today's recruitment and selection processes, websites like LinkedIn have become the initial step. Because these platforms use AI models for candidate selection, it is crucial to ensure that these models are designed in a way that provides equal opportunities for individuals of all races, genders, religions, and nationalities. However, if the AI models are created with unconscious bias, there is a risk of introducing discrimination during the hiring process.

In this activity, we will demonstrate an example of how gender bias can be introduced into an AI model specifically designed to assist startups in hiring computer science experts. The model will be developed using biased data based on 'gender,' and it will categorize candidates into two groups: "Hired" or "Not Hired." Through this demonstration, participants will witness how the model makes hiring decisions and observe the impact of this introduced bias on the results.

The dataset used to train this model is relatively small, consisting of 10 images for the "Hired" class and 10 images for the "Not Hired" class. The images representing the "Hired" class predominantly depict men working with laptops, as expected from computer scientists. In contrast, for the "Not Hired" class, we intentionally introduced bias by selecting images of women engaged in unrelated activities such as cooking or taking care of children. There are only a few images of women working with laptops, reinforcing stereotypical gender roles.

Through this activity, participants will gain a firsthand understanding of the potential consequences of biased data and its influence on AI models. They will recognize the importance of creating fair and unbiased AI models to ensure equitable hiring practices and challenge the existing biases present in recruitment processes.

A step-by-step guide

Step 1

- Click [here](#) in order to navigate through the website which contains the resources (images) you will need for the activity
- Download the file: **source_activity1.zip** on your Desktop or working folder
- Unzip the downloaded file. The file source_activity1.zip contains two files:
 1. activity1_gender_ub.json
 2. test_image.jpeg

Braga FAIas LTAA Activities

Activity guides:

- [Activity 1 Guide](#)
- [Activity 2 Guide](#)
- [Activity 3 Guide](#)

Activity resources:

- 1 [Resources for Activity 1](#)
• [Resources for Activity 2](#)

Step 2 Launch the [LearningML website](#)

Step 3 Click on: **LearningML v1.3 (advanced)**

The screenshot shows the homepage of the LearningML website. At the top, there's a navigation bar with links for WHAT IS, WHY, LEARN, DEVELOP, BLOG (ENGLISH), DOWNLOADS, ABOUT, and CONTACT. Below the navigation is a large banner featuring a cartoon character interacting with a control panel. Underneath the banner, there are four main options:

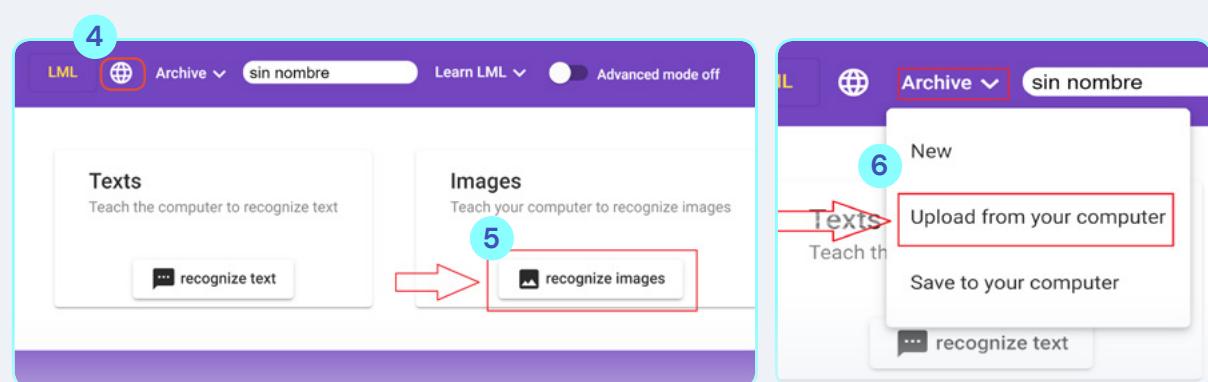
- LearningML v1.2 (original)**: Ideal for primary school students.
- LearningML v1.3 (advanced)**: This is the highlighted option, intended for high school, professional training, and university courses. It adds classification of sets of numbers and advanced mode.
- Imi-Snap!**: For those who want more power programming applications.
- LearningML-Desktop**: For those who prefer to have LearningML installed locally.

3. Guides for activities promoting an inclusive approach to AI

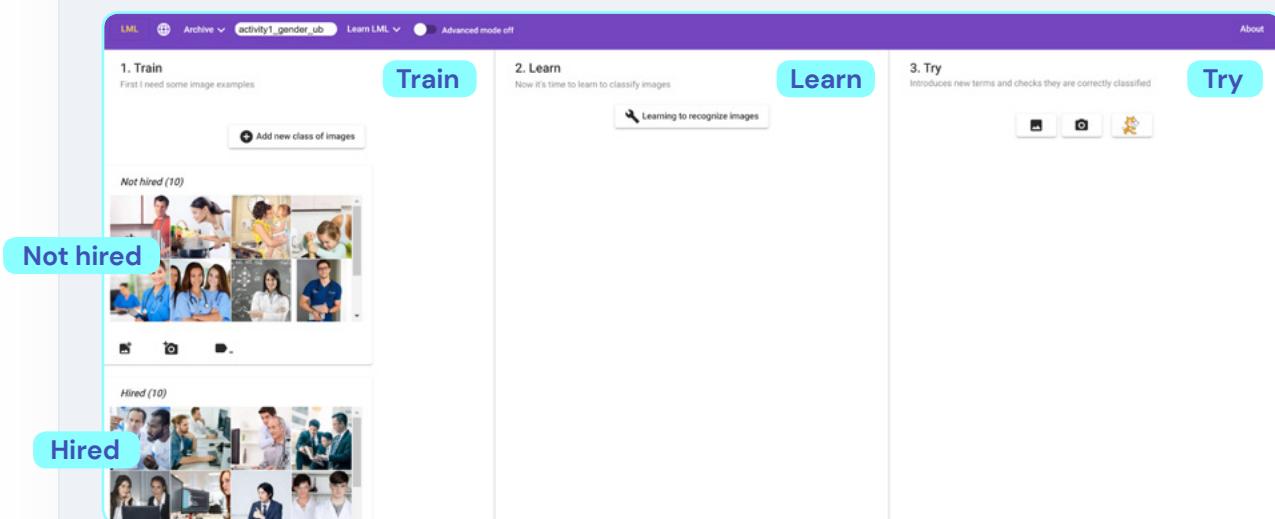
Step 4 Select your language: click on the circle

Step 5 Click on recognize images

Step 6 Upload file: **activity1_gender_ub.json**



Note: As you can see in the following picture the two classes (**Hired** and **Not hired**) have been added with their respective images.



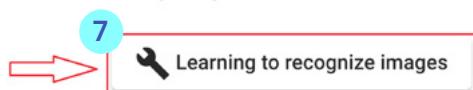
LearningML has 3 stages as any AI system:

- 1. Train:** "In this stage we feed our system or model the data that we have".
- 2. Learn:** In this stage the system is working on the backend and finds patterns among data and joins the things that are similar in order for it to be able to recognize any unknown information.
- 3. Try:** In this stage the system is trained and ready to recognize and categorize the new image. Careful : your test image shouldn't be part of your input/train data.

Step 7 Click on learning to recognize images:

2. Learn

Now it's time to learn to classify images



2. Learn

Now it's time to learn to classify images

Learning to recognize images

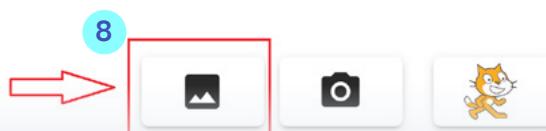
Learning from data
wait please



Step 8 Upload the test image:

3. Try

Introduces new terms and checks they are correctly classified

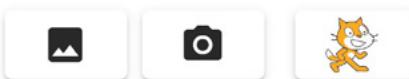


Results

In this activity, in the test_image there are two developer women. As result, the model classifies a candidate as follows: **Not hired**

3. Try

Introduces new terms and checks they are correctly classified



- Not hired (88.16 %)
- Hired (11.84 %)

Discuss

- Do you think that there is any additional bias in the image dataset used to train the model?
- Why do you think the result is Not Hired?
- What would you change on the data in order to change the result?

Note for the discussion

In order for the model to classify the female candidate as Hired, it needs to be trained without gender bias. If we add the same number of male and female pictures in similar circumstances, the model will rank a candidate equally: in other words, the gender of that person would be irrelevant to his or her eligibility for the position.

Reflection

The activity exploring the gender biased hiring system has been eye-opening and thought-provoking. It shed light on the potential implications of unconscious bias in recruitment processes, specifically related to gender.

Examining the AI model designed to assist startups in hiring computer science experts, we observed how bias can be introduced through the dataset used for training. The limited representation of women in the “Hired” class, compared to stereotypical images of women in non-computer science-related roles in the “Not Hired” class, highlighted the perpetuation of gender stereotypes and unequal opportunities.

It underscored the need for greater diversity and inclusivity in training datasets to ensure fair decision-making. By relying on biased data, we risk perpetuating discriminatory practices and hindering the progress towards a more inclusive society.

Moreover, this activity emphasized the importance of critical thinking and ethical considerations in the development and use of AI models for hiring purposes. It revealed the potential consequences of relying solely on automated systems without human intervention and oversight.

Reference: <https://fosteringai.github.io/project/result3/>

3.5 Activity 5: Gender bias using LearningML with dataset 2

Learning Objectives & Materials

Same as in Activity 4

Detailed description

In Activity 4, we created a classification model to help a startup to hire a Computer Science expert. The model in Activity 4 was trained with 'gender' biased data and as a result, the model classified men as more likely to be Hired for the job than women.

In Activity 5, we are going to create a new classification model to help startups to hire a Computer Science expert, but in this case we will train the model without 'gender' bias in the data. We added the same number of male and female pictures and both in similar situations.

As a result, you can see how important is the type and amount of data that we use for creating an AI model to avoid racial and gender biases or even to unintentionally create bias towards another group by trying to fix bias.any type of discrimination.

A step-by-step guide

Step 1

- Click [here](#) in order to navigate through the website which contains the resources (images) you will need for the activity
- Download the file: **source_activity2.zip** on your Desktop or working folder
- Unzip the downloaded file. The file source_activity2.zip contains two files:
 1. activity2_gender_ub.json
 2. test_image.jpeg

Braga FAIaS LTAA Activities

Activity guides:

- [Activity 1 Guide](#)
- [Activity 2 Guide](#)
- [Activity 3 Guide](#)

Activity resources:

- [Resources for Activity 1](#)
- 1 • [Resources for Activity 2](#)

3. Guides for activities promoting an inclusive approach to AI

Step 2 Click [here](#) to visit the LearningML website:

Step 3 Click on: **LearningML v1.3 (advanced)**

The screenshot shows the LearningML website with a cartoon character in the background. There are four main options displayed:

- LearningML v1.2 (original)**: Ideal to enter the world of Machine Learning. It can be used from the last years of primary school.
- LearningML v1.3 (advanced)**: This version adds the classification of sets of numbers and the advanced mode, with which you can explore the behavior of ML algorithms.
- Iml-Snap!**: For those who want more power programming applications. All phases of ML are done by programming. Ideal for high school, professional training and first university courses.
- LearningML-Desktop**: For those who prefer to have LearningML v1.3 installed on their computer (Linux, Windows, Mac) and skip the Internet. Ideal to be incorporated into educational Linux distributions and for schools that have internet connection problems.

Step 4 Select your language: click on the circle

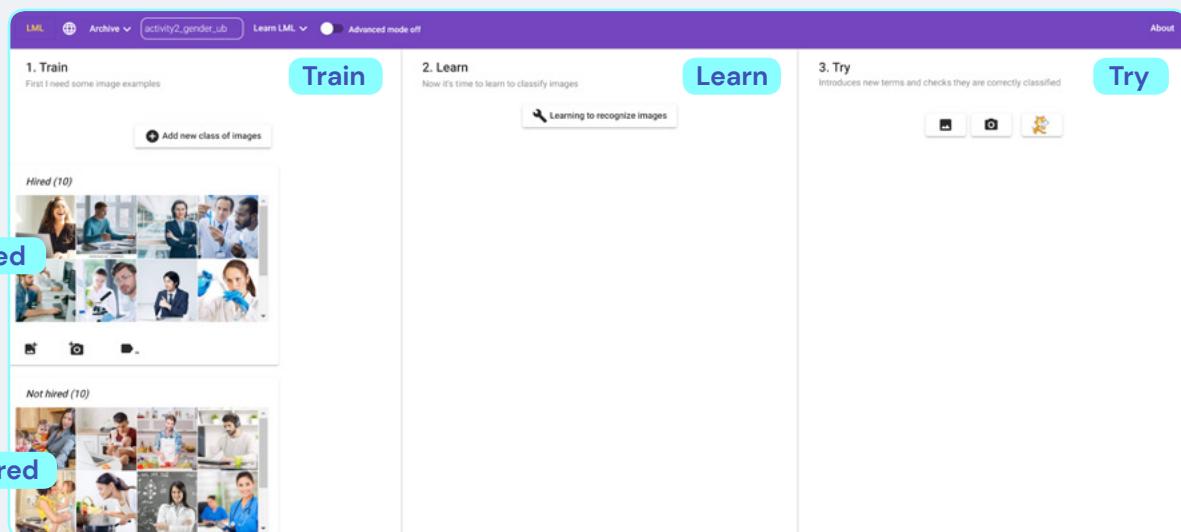
Step 5 Click on recognize images

Step 6 Upload file: **activity2_gender_ub.json**

The screenshot shows the LearningML interface with the 'Images' section selected. A red arrow points to the 'recognize images' button. A file upload dialog is open, showing the file 'activity2_gender_ub.json' with options to 'Upload from your computer' and 'Save to your computer'.

3. Guides for activities promoting an inclusive approach to AI

Note: As you can see in the following picture the two classes (**Hired and Not hired**) have been added with their respective images.



LearningML has 3 stages as any AI system:

- 1. Train:** "In this stage we feed our system or model the data that we have".
- 2. Learn:** In this stage the system is working on the backend and finds patterns among data and joins the things that are similar in order for it to be able to recognize any unknown information.
- 3. Try:** In this stage the system is trained and ready to recognize and categorize the new image. Careful : your test image shouldn't be part of your input/train data.

We have two categories for classifying a candidate: Hired and Not Hired. Each class needs 10 images, but we added images without gender bias in this example. We added the same number of images of both women and men in the same type of situations.

Step 7 Click on learning to recognize images:

2. Learn
Now it's time to learn to classify images

7  **Learning to recognize images**

2. Learn
Now it's time to learn to classify images

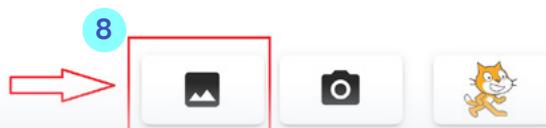
 **Learning to recognize images**

Learning from data
wait please

Step 8 Upload the test image:

3. Try

Introduces new terms and checks they are correctly classified



Results

In this activity, in the test_image there are two developer women. As result, the model classifies a candidate as follows: **Hired**

3. Try

Introduces new terms and checks they are correctly classified



- Hired (90.51 %)
- Not hired (9.49 %)

Discuss

Do you think that there is any additional bias in the image dataset used to train the model?

Reflection

This activity presented a valuable lesson by emphasizing the significance of training AI models without bias. By consciously incorporating a balanced representation of both genders in the data, in similar situations, the goal was to create a fairer model for assisting startups in their hiring process for Computer Science experts.

It underscores the reality that the type and quantity of data used in training AI models play a pivotal role in determining the model's behavior and potential biases. Biased data can perpetuate and even amplify existing social biases and discriminatory practices, while unbiased data can help counteract such biases and promote fairness.

It is essential to acknowledge that addressing bias in AI models goes beyond just gender biases. We must remain vigilant and ensure that racial, ethnic, and other forms of biases are also taken into account. The aim should be to create models that not only avoid preserving existing biases, but actively work towards mitigating discrimination of any kind.

Activities like the above serve as a powerful reminder of the ethical responsibility we bear when developing AI systems. They highlight the need for diverse and representative datasets, meticulous preprocessing techniques, and continuous evaluation to identify and rectify biases that may emerge.

To foster unbiased AI models, it is crucial to foster interdisciplinary collaboration among data scientists, ethicists, social scientists, and domain experts. By working together, we can develop robust methodologies and guidelines that prioritize fairness, transparency, and accountability in AI development.

In conclusion, the reflection on these activities emphasizes the significant role that data and bias play in the creation of AI models. By recognizing and actively addressing biases, we can strive to build AI systems that promote equality, fairness, and inclusion, helping us to create a future where AI technologies truly serve and benefit all members of society.

3.6. Activity 6: Identifying bias in AI

Learning Objectives

1. **Understand the concept of bias in AI:** Participants will gain a comprehensive understanding of bias in artificial intelligence and its potential impact on decision-making processes. They will explore different forms of bias, including data bias, algorithmic bias, and societal bias.
2. **Analyze bias identification techniques:** Participants will learn various methods and techniques used to identify bias in AI systems. They will explore approaches such as data analysis, evaluation metrics, and fairness measures to assess and detect bias in machine learning models.
3. **Evaluate the ethical implications of biased AI:** Participants will critically examine the ethical considerations associated with biased AI systems. They will reflect on the potential consequences of biased decision-making in real-world applications and understand the importance of fairness, transparency, and accountability in AI development and deployment.
4. **Enhance data analysis skills:** Through the exercise, participants will enhance their data analysis skills by exploring datasets and identifying potential biases. They will develop proficiency in data preprocessing, feature selection, and data visualization techniques to analyze and interpret biases present in AI systems.
5. **Foster awareness of societal impact:** Participants will develop an understanding of how biased AI systems can perpetuate existing societal biases and reinforce inequality. They will explore case studies and examples to grasp the wider societal impact of biased AI in areas such as hiring, criminal justice, and healthcare.
6. **Promote responsible AI development:** Participants will learn the importance of responsible AI development and the need to incorporate fairness, diversity, and inclusivity in the design and implementation of AI systems. They will consider strategies to mitigate bias and ensure that AI technologies benefit all individuals and communities.
7. **Encourage critical thinking and problem-solving:** Through the exercise, participants will enhance their critical thinking and problem-solving skills. They will learn to approach complex issues related to bias in AI with a systematic and analytical mindset, evaluating different perspectives and proposing effective solutions.
8. **Foster collaboration and interdisciplinary thinking:** Participants will engage in discussions and collaborative activities to explore bias in AI from multiple viewpoints. They will recognize the interdisciplinary nature of addressing bias, fostering collaboration between data scientists, ethicists, policymakers, and other stakeholders.

Materials

Laptop with internet access

Google account for using Google Colab

No prior coding experience is required

Description of the activity

Throughout the exercise, participants will work with a dataset from Kaggle, a renowned platform for data science and machine learning. They will be introduced to different forms of bias, such as data bias, algorithmic bias, and societal bias, and learn how these biases can manifest in AI systems. Participants will have the opportunity to analyze the dataset, preprocess the data, and employ data analysis techniques to identify potential biases.

Using evaluation metrics and fairness measures, participants will assess the performance and fairness of machine learning models trained on the dataset. They will gain insights into the challenges of identifying and measuring bias in AI systems and learn how to interpret the results of bias assessment.

The exercise also emphasizes the ethical implications of biased AI systems. Participants will explore case studies and real-world examples that demonstrate the societal impact of biased AI in areas like hiring, criminal justice, and healthcare. They will engage in discussions on the importance of fairness, transparency, and accountability in AI development and deployment.

A step-by-step guide

Step 1 Download the resources for this exercise by visiting [the website](#) and downloading the **Identifying_bias_in_AI.ipynb**, **toxic_data.csv** and **Instructions_notebook_bias.pdf** and save them in a folder dedicated to this activity.

IO4 Handbook Activities ([Download](#))

- 1 • **Identifying_bias_in_AI.ipynb:** a jupyter notebook
- **toxic_data.csv:** a csv file containing the data to be used in the notebook
- **Instructions_notebook_bias.pdf:** a pdf containing instructions on how to use the notebook

Note regarding data

At the end of 2017 the [Civil Comments](#) platform, a full-featured commenting plugin for independent news sites, shut down and chose to make their ~2m public comments from their platform available in a lasting open archive so that researchers could understand and improve civility in online conversations for years to come. Jigsaw, a unit within Google that explores threats to open societies and builds technology that inspires scalable solutions, sponsored this effort and extended annotation of this data by human raters for various toxic conversational attributes. In this exercise you will work with a small subset of the data that was used in the [Jigsaw Unintended Bias in Toxicity Classification](#) competition.

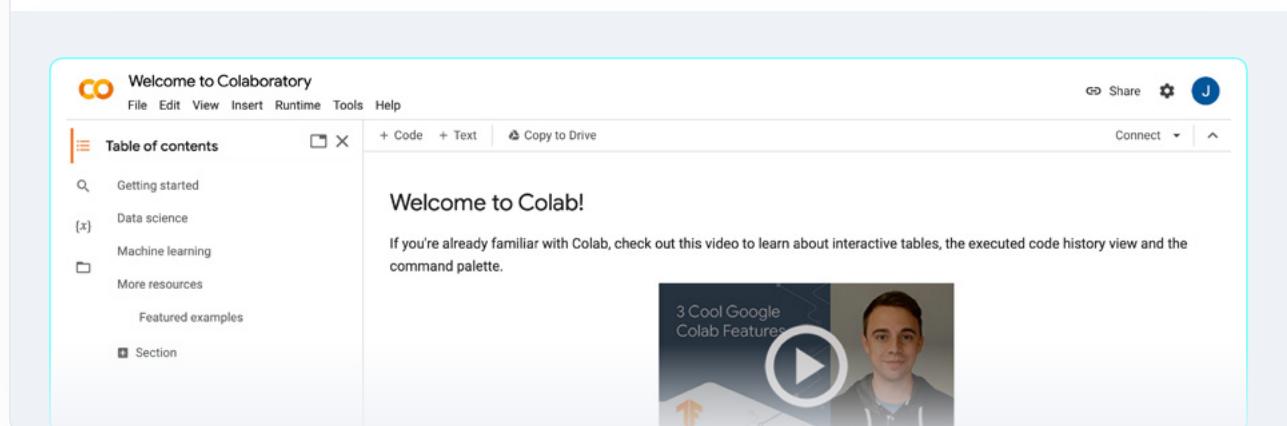
Step 2 Connect in your **gmail** account (if you aren't already).

Step 3 Open any internet browser you are using and type **Google Colab**.

Note: Google Colab is a user-friendly online platform that allows you to write and run code without needing any prior coding experience. It's like having a virtual notebook where you can write and execute code using Python, a popular programming language.

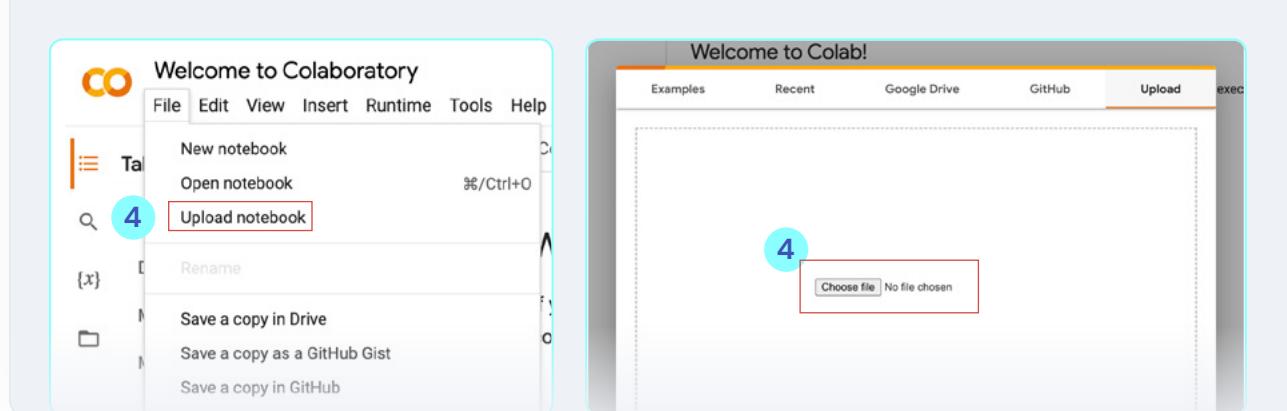
With Google Colab, you can create and edit code cells, which are like individual sections where you can write and run your code. It provides a simple and interactive interface where you can type your code and see the output right away.

One of the great advantages of Google Colab is that it comes pre-installed with many popular libraries and frameworks, such as TensorFlow and PyTorch, which are commonly used in data analysis, machine learning, and other fields. This means you can easily explore and work with advanced functionalities without the need for complex setup or installations.



3. Guides for activities promoting an inclusive approach to AI

Step 4 Upload the Identifying bias in **AI.ipynb** file you have initially downloaded.



Step 5 Once it is loaded, this is what you will see. On the left side there is the list of content of the notebook for easier navigation and on the main window (center right of the screen) you can see the actual notebook and its content by scrolling down.

A screenshot of a Google Colab notebook titled 'Identifying_bias_in_AI.ipynb'. The left sidebar shows a 'Table of contents' with sections such as 'Identifying Bias in AI', 'Prepare data and train model', 'Importing all packages and load data', 'Data exploration', 'Data preprocessing', 'Working with words and vocabularies', 'Train model', 'How does the model make decisions?', 'Deep dive', and 'Other possible types of biases'. The main window displays the first section, 'Identifying Bias in AI', which includes a brief introduction, instructions for running the exercise, and information about the code cells.

Step 6 Follow the instructions of the notebook in order to get the output.

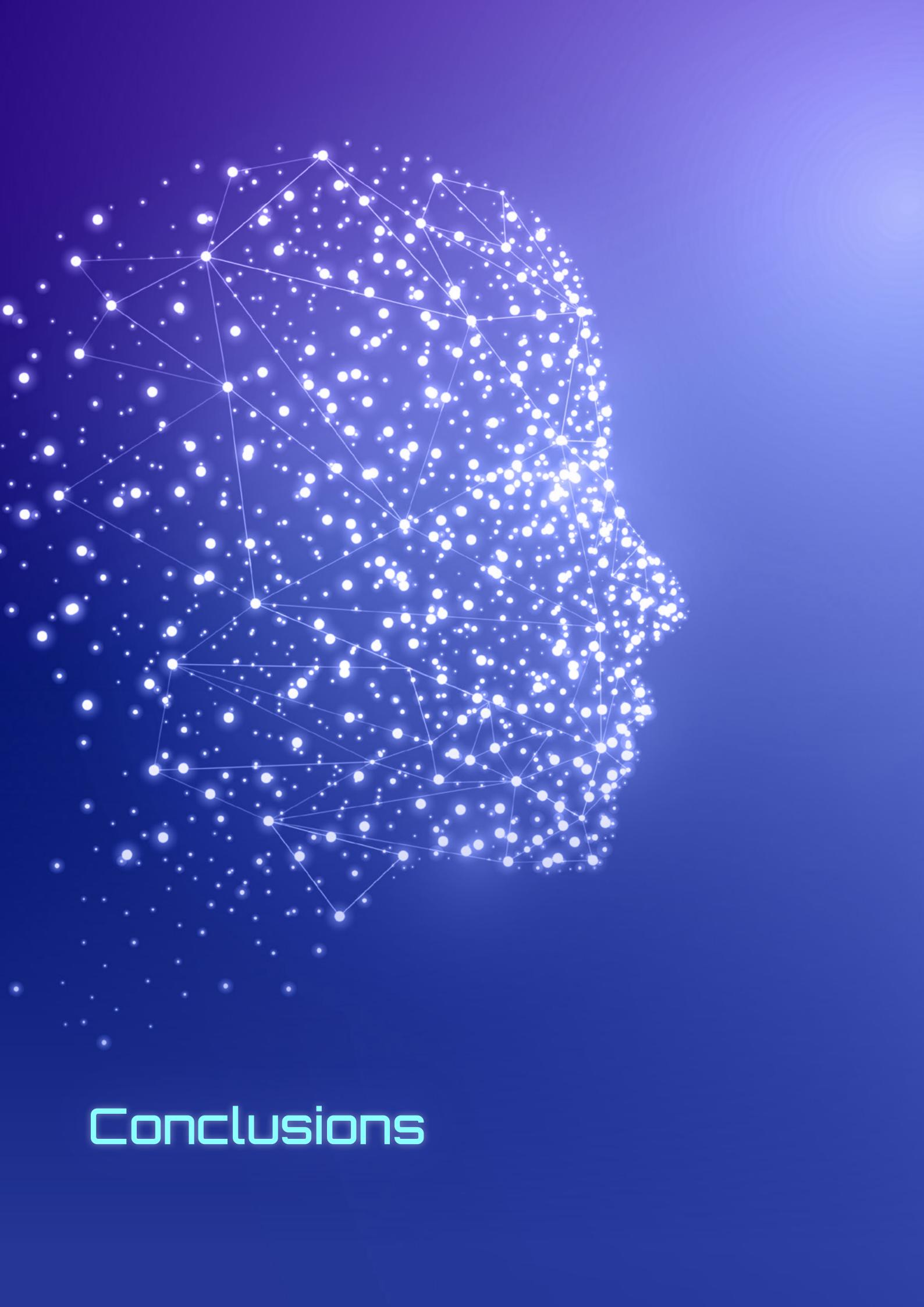
Reflection

Analyzing real-world datasets and exploring various techniques to detect bias in AI models allow the participants to grasp the complexity and challenges associated with this issue. It highlighted the importance of critically evaluating the data used to train AI models and considering the potential consequences of biased outcomes.

The activity also emphasized the ethical considerations surrounding biased AI systems. It prompted to reflect on the societal impact of such biases, particularly in areas like hiring, criminal justice, and healthcare. It served as a reminder of the responsibility we have as developers and users of AI to ensure fairness, transparency, and accountability in its design and implementation.

Through this exercise, participants further develop their data analysis skills, learning valuable techniques for data preprocessing, feature selection, and data visualization. These skills enabled them to identify and interpret potential biases within the AI models, strengthening their ability to analyze and address bias in future projects.

Collaborating with fellow participants and engaging in discussions on bias in AI provided a broader perspective on the issue. Hearing different viewpoints and considering diverse perspectives enhanced my understanding of the multifaceted nature of bias and its implications in AI systems.



Conclusions

In conclusion, the Fostering Artificial Intelligence at Schools (FAIaS) project and the accompanying guide for non-formal educators provide valuable resources and insights to navigate the complex landscape of AI ethics, responsible development, and inclusive teaching practices.

As AI continues to shape our society, it is crucial to prioritize the ethical and responsible deployment of AI systems while ensuring diversity, equity, and inclusion.

The guide recognizes the advantages of non-formal education in teaching AI and highlights its flexibility and adaptability to cater to the needs of diverse learners. It emphasizes the importance of incorporating both technical and ethical aspects of AI into education programs, equipping individuals with the skills and knowledge necessary to understand and navigate the implications of AI systems. By promoting hands-on experiences, fostering creativity and innovation, supporting underrepresented communities, encouraging lifelong learning, and building practical skills, non-formal education can empower individuals to thrive in the digital age.

Moreover, the guide underscores the significance of promoting gender, diversity, and inclusion in non-formal education programs on AI. By creating inclusive learning environments that value and respect the diverse identities and experiences of students, educators can foster a sense of belonging and support the academic and personal success of all learners. Inclusive teaching practices, such as being aware of biases, using inclusive language, addressing individual needs, creating safe spaces, and seeking support when needed, are crucial in creating supportive and engaging learning environments.

By leveraging the potential of non-formal education and embracing inclusive teaching practices, we can build a more equitable and inclusive future for AI. The FAIaS project and the guide for non-formal educators serve as important resources in this journey, providing educators with the tools and knowledge to empower the next generation of AI practitioners and users. By prioritizing ethics, responsible development, and inclusivity, we can ensure that AI technology serves the best interests of individuals and communities, while upholding principles of fairness, equity, and justice. Together, we can shape a future where AI is harnessed for the benefit of all.

Definitions and glossary [33], [34], [35]

Diversity	Socially, it refers to the wide range of identities. It broadly includes race, ethnicity, gender, age, national origin, religion, disability, sexual orientation, socioeconomic status, educational level, marital status, language, veteran status, physical appearance, etc. It also involves different ideas, perspectives and values.
Bias	Prejudice in favor of or against one thing, person, or group compared with another, usually in an unfair or negative way. Unconscious bias, also known as implicit bias, is defined as "attitudes and stereotypes that influence judgment, decision-making, and behavior in ways that are outside of conscious awareness and/or control".
Inclusion	The act of creating an environment in which any individual or group will be welcomed, respected, supported and valued as a fully participating member. An inclusive and welcoming climate embraces and respects differences.
Non-formal education	Education that is institutionalized, intentional and planned by an education provider, with the defining characteristic that it is an addition, alternative and/or a complement to formal education within the process of the lifelong learning of individuals. It is often provided to guarantee the right of access to education for all. It caters to people of all ages, but does not necessarily apply a continuous pathway-structure. It may be short in duration and/or low intensity, and is typically provided in the form of short courses, workshops or seminars. Non-formal education mostly leads to qualifications that are not recognized as formal qualifications by the relevant national educational authorities or to no qualifications at all. Non-formal education can cover programmes contributing to adult and youth literacy and education for out-of-school children, as well as programmes on life skills, work skills, and social or cultural development.
Prejudice	A preconceived judgment or preference, especially one that interferes with impartial judgment and can be rooted in stereotypes, that denies the right of individual members of certain groups to be recognized.
Stereotype	A form of generalization rooted in blanket beliefs and false assumptions, a product of processes of categorization that can result in a prejudiced attitude, critical judgment and intentional or unintentional discrimination. Stereotypes are typically negative, based on little information and do not recognize individualism and personal agency.

References

1. Rosenfeld, Hannah, Danni Yu, and Abhishek Gupta. 2023. "Ethical AI requires future-ready, inclusive education system | World Economic Forum." The World Economic Forum, June 5, 2023. <https://www.weforum.org/agenda/2023/06/ethical-ai-future-ready-inclusive-education-system/>
2. Blackman, Reid. 2020. "A Practical Guide to Building Ethical AI." Harvard Business Review, October 15, 2020. <https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai>
3. Burton, Emanuelle, Judy Goldsmith, Sven Koenig, Benjamin Kuipers, Nicholas Mattei, and Toby Walsh. 2017. "Ethical Considerations in Artificial Intelligence Courses." AI Magazine 38(2) (July): pp22–34. <https://doi.org/10.1609/aimag.v38i2.2731>
4. Norqvist, Lars, and Eva Leffler. 2017. "Learning in non-formal education: Is it "youthful" for youth in action?" Springer link 63 (March): pp235–256. <https://doi.org/10.1007/s11159-017-9631-8>
5. Zowghi, Didar, and Francesca da Rimini. 2023. Responsible AI: Best Practices for Creating Trustworthy AI Systems. N.p.: Pearson Addison Wesley. <https://doi.org/10.48550/arXiv.2305.12728>
6. Lang, Michael, Mark Freeman, Gaye Kiely, and Amy B. Woszczynski. 2022. "Special Issue Editorial: Equality, Diversity, and Inclusion in IS Education." Journal of Information Systems Education 33, no. 1 (February): pp1–6.
7. Perry, Sylvia P., Mary C. Murphy, and John F. Dovidio. 2015. "Modern prejudice: Subtle, but unconscious? The role of Bias Awareness in Whites' perceptions of personal and others' biases." Journal of Experimental Social Psychology 61 (November): pp64–78.
8. Warikoo, Natasha, Stacey Sinclair, Jessica Fei, and Drew Jacoby-Senghor. n.d. "Examining Racial Bias in Education: A New Approach." Educational Researcher 45 (9). <https://doi.org/10.3102/0013189X16683408>
9. Zittleman, Karen, and David Sadker. 2022. "Gender Bias in Teacher Education Texts: New (and Old) Lessons." Journal of Teacher Education 53, no. 2 (March): 168–180. <https://doi.org/10.1177/0022487102053002008>
10. Spencer, Barbara. 1980. "Overcoming the Age Bias in Continuing Education." New Directions for Continuing Education 8: pp71–86.
11. Willis, Robert J., and Sherwin Rosen. 1979. "Education and Self-Selection." Journal of Political Economy 87, no. 5 (October). <https://doi.org/10.1086/260821>
12. Glewwe, Paul. 2002. "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes." Journal of Economic Literature 40, no. 2 (June): pp436–482. <https://doi.org/10.1257/002205102320161258>.
13. Derman-Sparks, Louise, and Julie O. Edwards. 2019. "Understanding Anti-Bias Education." National Association for the Education of Young Children 74, no. 5 (November): pp6–13. <https://www.jstor.org/stable/26842300>
14. Stubbs, Sue. 2008. Inclusive Education. N.p.: The Atlas Alliance, pp69–106.

15. Mohammed, Phaedra S., and Eleanor N. Watson. 2019. "Towards Inclusive Education in the Age of Artificial Intelligence: Perspectives, Challenges, and Opportunities." In *Artificial Intelligence and Inclusive Education: Speculative Futures and Emerging Practices*, edited by Jeremy Knox, Yuchen Wang, and Michael Gallagher, 17–37. N.p.: Springer Nature Singapore. https://doi.org/10.1007/978-981-13-8161-4_2
16. Xu, Feiyu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. 2019. "Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges." In *Lecture Notes in Computer Science*, pp563–574. Vol. 11839. N.p.: Springer. https://doi.org/10.1007/978-3-030-32236-6_51
17. Gururaj, Tejasri. 2023. "10 examples of how artificial intelligence is improving education." *Interesting Engineering*. <https://interestingengineering.com/lists/examples-how-artificial-intelligence-improving-education>
18. Morron, Mirriam. 2023. "The Impact of AI on Learning and Creativity: A Critical Look at the Future." LinkedIn. <https://www.linkedin.com/pulse/impact-ai-learning-creativity-critical-look-future-mirriam-morron/>
19. Ullman, Stefanie. 2022. "Gender Bias in Machine Translation Systems." In *Artificial Intelligence and Its Discontents*, pp123–144. 10.1007/978-3-030-88615-8_7.
20. Hao, Karen. 2019. "Facebook's ad-serving algorithm discriminates by gender and race." MIT Technology Review. <https://www.technologyreview.com/2019/04/05/1175/facebook-algorithm-discriminates-ai-bias/>
21. Dastin, Jeffrey. 2018. "Amazon scraps secret AI recruiting tool that showed bias against women." Reuters. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
22. Haag, Christian, and Alexander Smith. 2019. "5 Examples of Biased Artificial Intelligence." Logically. <https://www.logically.ai/articles/5-examples-of-biased-ai>
23. Manyika, James, Jake Silberg, and Britanny Presten. 2019. "What Do We Do About the Biases in AI?" Harvard Business Review. <https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai>
24. Gonzalez, Glen. 2020. "Breaking Bias." Smith Communication Partners. <https://smithcp.com/breaking-bias/>
25. "Visa SVP Carolina Barcenas: "Educating the general public on AI is key to ethical use and smart" 2019. Medium. <https://medium.com/authority-magazine/visa-svp-carolina-barcenas-educating-the-general-public-on-ai-is-key-to-ethical-use-and-smart-ea080f37966c>
26. "Role Models in AI: Stephanie Lampkin | by AI4ALL Team | AI4ALL." 2019. Medium. <https://medium.com/ai4allorg/role-models-in-ai-stephanie-lampkin-f50e6885afa4>
27. "Female role models changing the field of Artificial Intelligence in Norway – NORA." n.d. Norwegian Artificial Intelligence Research Consortium. Accessed June 16, 2023. <https://www.nora.ai/news-and-events/news/30-women-in-norway-changing-the-field-of-ai.html>
28. "Role Models in AI: Rebekah Agwunobi | by AI4ALL Team | AI4ALL." 2018. Medium. <https://medium.com/ai4allorg/role-models-in-ai-rebekah-agwunobi-f3bc438580da>
29. "The tag game | Activities." n.d. Scouts. Accessed June 16, 2023. <https://www.scouts.org.uk/activities/the-tag-game/>

30. "Describe and draw | Activities." n.d. Scouts. Accessed June 16, 2023. <https://www.scouts.org.uk/activities/draw-the-description/>
31. Katherine Johnson, https://en.wikipedia.org/wiki/Katherine_Johnson
32. "Inspirational women: Hedy Lamarr - Code Week." 2022. EU Code Week blog. <https://blog.codeweek.eu/inspirational-women-hedy-lamarr/>
33. UNESCO Institute for Statistics, Glossary, <http://uis.unesco.org/en/glossary>
34. College of the Environment, University of Washington, Diversity, Equity and Inclusion Glossary ,<https://environment.uw.edu/about/diversity-equity-inclusion/tools-and-additional-resources/glossary-dei-concepts/>
35. Harvard Human Resources, Glossary of Diversity, Inclusion and Belonging Terms, https://edib.harvard.edu/files/dib/files/dib_glossary.pdf

Image Credits

Front, back covers – Own design, using image designed by [shuvojit11 / Freepik](#)

Pages 6, 14 – Own design, using image designed by [rawpixel.com / Freepik](#)

Page 28 – Image designed by [rawpixel.com / Freepik](#)

Page 34 – Image "[Katherine Johnson Receives Presidential Medal of Freedom \(NHQ201511240003\)](#)" by [NASA HQ PHOTO](#) licensed under [CC BY-NC-ND 2.0](#)

Page 38 – Image "[Hedy Lamarr](#)" by [ReznorH](#) is marked with [Public Domain Mark 1.0](#).

Pages 43 to 61 – Screenshots taken from the website [learningml.org](#)

Page 57 – Section image designed [by Freepik](#)

Page 63 – Own source

Fostering Artificial Intelligence at Schools

Inclusion guide for (non-formal) educators in the
context of Artificial Intelligence education



Co-funded by the
Erasmus+ Programme
of the European Union



Universidad
Rey Juan Carlos

Teatro Circo de Braga
EM, S.A.

