AEQUITAS

D2.1 Bias Report

Co-funded by
the European Union

AEQUITAS

About the project

AEQUITAS – Preparation of the CSOs and public healthcare sector to address gender and racial biases in AI is an European Union funded project that strengthens the capacity of civil society organizations and healthcare professionals to recognize and address discriminatory biases in biomedical artificial intelligence. By combining training, awareness campaigns, and policy development, AEQUITAS promotes the fair and ethical use of AI in healthcare. The project will establish a European network of CSOs and hospitals, develop a Database of AI biases, and produce an AI Regulatory Model with policy recommendations shared with policymakers, legal professionals, and healthcare providers.

The project is coordinated by the Center for the Study of Democracy (Bulgaria) and implemented by a consortium of ten partners. These include Centre Diotima (Greece), Universitätsklinikum Köln (Germany), Innovation Hive – KypSelí Kainotomias (Greece), C.I.P. Citizens in Power (Cyprus), Women‘s Issues Information Center (Lithuania), Lobby Europeo de Mujeres en España (Spain), TIA Formazione Internazionale APS (Italy), and Health Citizens – European ResearchInstitute (Portugal).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Call/Topic	CERV-2024-CHAR-LITI-CHARTE
Project Title	Preparation of the CSOs and public healthcare sector to address gender and racial biases that might arise from the wide usage of AI in order to protect and promote fundamental rights
Project Acronym	AEQUITAS
Grant Agreement No.	101215009
Project Website	https://aequitasproject.eu/

Document Properties

Deliverable No.	D 2.1
Work Package	WP2 Mapping of gender and racial biases in biomedical AI and database and guidelines development following the principles of the EU Charter for the protection of fundamental rights
Author/s	UKK, CSD

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Contributor/s	All partners
Reviewed	
Approved	
Name	Bias Report
Version	4.0
Date	26/2/26
Dissemination Level	

History of Changes

Version	**Date**	**Comments**	**Main Authors**
1.0	18/12/25	Initial version uploaded to the drive	Maria Christoforaki
2.0	5/2/26	Final version open for comments	Maria Christoforaki
3.0	12/2/26	Integration of comments	Maria Christoforaki
4.0	26/2/26	Integration of Translations	Maria Christoforaki

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Table of contents

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Introduction

Part of the AEQUITAS project is to create a database about gender and racial biases in Artificial Intelligence (AI) medical applications, specifically focused on three diseases: cardiovascular diseases, diabetes, and depression.

To complete this task, the consortium partners first need to collect a variety of sources on the biases listed above. The University Hospital of Cologne (Universitätsklinikum Köln, UKK), as the task leader and domain expert, organised the information collection activity and provided the mapping template, which the partners used to map the sources, ensuring that the relevant information could be easily transferred to the database.

This report presents the theoretical and scientific foundations that guided the selection of the collection activity and mapping template accompanied by case studies showcasing the different kinds of bias; the policy implications that biomedical AI induced biases impact the rights protected by the EU Charter of Fundamental Rights; a description of the data collection activity; the mapping template; a list of sources collected by the AEQUITAS partners; and other supporting material. The rest of the report is structured as follows:

First, we present the theoretical background regarding our work in an introduction on the AI Medical applications and the concept of bias in computer systems and medicine. We begin by focusing on Medicine, first presenting how race and gender bias manifest in medical care, and second, how moral issues arising in the practice of medicine and biomedical research are addressed by Bioethics, offering a brief introduction to the four Bioethics principles (Autonomy, Non-Maleficence, Beneficence, and Justice).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Subsequently, we turn to the AI domain, presenting the types of bias that can be observed in AI systems as they manifest in the Machine Learning and Artificial Intelligence (ML/AI) pipeline. Each type of bias is accompanied by examples and a case study of gender and racial biases and their impact at a societal level regarding the three focus diseases of the AEQUITAS project (cardiovascular, diabetes, and depression), drawn from sources collected by the AEQUITAS partners after the completion of T2.2. When this was not possible because the collected material did not clearly demonstrate the specific type of AI bias under consideration, an alternative case was presented from another medical domain that was easily generalisable to the AEQUITAS target diseases. The case study descriptions, along with the collected sources, draw on additional scientific resources as needed to support them.

Finally, in the Policy implications section, it is demonstrated how the various types of AI biases impact the fundamental rights protected by the EU Charter, most notably the precepts of human dignity, equality before the law, non-discrimination, as well as the right to integrity of the person, the right to healthcare, data protection, and the right to an effective remedy, concluding in the safeguards that can be put in place in conformity assessments, post-market monitoring, and public-sector procurement.

The report closes with the References and the following Appendices:

Appendix 1: Source Collection and Mapping Method, which contains the mapping template and describes the collection, mapping, and information assessment process conducted during the T2.1 and T2.2 tasks.

Appendix 2: Contains supporting material for the T2.1 and T2.2 tasks, i.e., slides from partner meetings describing the process, presented by UKK.

Appendix 3: The list of sources collected by the AEQUITAS partners.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

AI Medical Applications

The rise of AI applications in recent years has heavily impacted Medicine, including digitised data acquisition, machine learning, and computing infrastructure (Yu et al., 2018). Especially the introduction of deep learning algorithms in areas such as computer vision and natural language processing has revolutionised computer applications in radiology, pathology, cardiology, diabetology, psychiatry, oncology, etc., (Esteva et al., 2019; Koteluk et al., 2021; Rajpurkar et al., 2022; Gou et al., 2024). The World Health Organisation (WHO) lists the following application domains of AI systems in health care: diagnosis and prediction-based diagnosis, clinical care, research and drug development, health systems management and planning, public health and public health surveillance, health promotion, disease prevention, prediction-based surveillance, emergency preparedness, and outbreak response (World Health Organization, 2021).

However, the advent of AI applications in medicine comes with a set of challenges, such as implementation challenges, including model trust and data limitations, accountability issues, which include regulatory challenges and proper responsibility attribution, and ensuring fairness via ethical data use, equitable distribution of benefits and bias detection and mitigation (Rajpurkar et al., 2022).

The AEQUITAS project focuses on gender and race bias cases in cardiovascular diseases, diabetes, and depression. AI medical applications support cardiovascular care through clinical decision support, telemedicine, risk assessment, customised therapy, predictive analytics, and remote monitoring (Bernstein et al., 2025; Naskar et al., 2025), improve diabetes management (including patient monitoring and self-management), diagnosis, treatment, and prevention (Contreras & Vehi, 2018; Khalifa & Albadawy, 2024; Naskar et al., 2025; Sheng et al., 2024). Regarding depression, they are involved in screening, diagnosis, and treatment (Alhuwaydi, 2024) with a special focus on detection and screening with the use of Large Language Models (LLMs) (Cao

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

et al., 2025; Kumari et al., 2025; Mao et al., 2023; Wang et al., 2025). In all of the above areas, there are bias challenges, for example, regarding cardiovascular diseases, diabetes, and depression, see (van Assen et al. 2024), (Cronjé et al. 2023), (Dang et al. 2024), respectively.

Challenges such as bias, whether in Medicine or AI, are addressed through a combination of Bioethics and AI Ethics. In the following section, we provide a brief overview of these two types of applied ethics domains that served as the theoretical and scientific foundation for developing the Bias Mapping template.

Ethics and Bias in Medicine and AI

Bioethics and Bias in Medicine

Bias in medicine is well documented: see, for example, (Hammond et al. 2021) regarding cognitive bias, which consists of systematic errors in thinking due to human processing limitations or inappropriate mental models, and (FitzGerald and Hurst 2017) for implicit bias involving associations outside conscious awareness that lead to a negative evaluation of a person on the basis of irrelevant characteristics such as race or gender.

Racial bias in Medicine is well studied in the USA case, for example, where it is documented that African Americans, as well as those in other minority groups, receive fewer procedures and poorer-quality medical care, by obtaining less aggressive treatment, experiencing lower rates of surgical treatment, and receive fewer referrals to specialists than white individuals (Bowser, 2001; Williams & Wyatt, 2015).

Gender bias can be attributed to gender blindness and stereotyped preconceptions about men and women (Hamberg, 2008), added to a generalised lack of knowledge about the functioning of the female human body and its biological differences from the male body. For example, critically ill women 50 years and older were less likely than

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

critically ill men to be admitted to an intensive care unit (ICU) (Bierman, 2007), and even male mouse models are overall more represented than female models in basic, preclinical, and surgical biomedical research (Yoon et al., 2014).

It is also important to note that LGBT+ individuals experience discrimination in terms of access to healthcare and are subject to stereotypes that do not affect the heterosexual population. These social and cultural factors perpetuate discrimination and have an impact on health. For example, a study in the USA, based on data from the 2013–14 National Health Interview Survey (NHIS), found that LGB adults reported higher levels of poor health, functional limitations, severe psychological distress and difficulties affording healthcare compared to their heterosexual counterparts. These inequities are driven by minority group stress and multifaceted societal marginalisation (Liu et al., 2023).

On the other hand, medicine as a discipline is held to high ethical standards from ancient times to the present (Baker & McCullough, 2008). For centuries, there has been a social expectation that a physician will follow ethical rules of professional responsibility set by the standards of their profession, as manifested via professional norms ranging from the Hippocratic Oath from 400 BCE (Miles, 2005), to the Declarations of Geneva and Helsinki (Tröhler, 2008). As pointed out by (Vevaina et al., 1993), physicians are held responsible to conform to the ethical code of their profession by the investment that society makes in their education (monetary and the use of its members as learning material throughout the physician's training and career), and the virtual monopoly that their profession is granted through licensing.

Biomedical ethics (or bioethics) is a domain of practical (or applied) ethics that addresses the moral issues arising in the practice of medicine and biomedical research (Vevaina et al., 1993). Central to biomedical ethics are the four principles as defined by Beauchamp and Childress (Beauchamp & Childress, 2019):

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

1. Autonomy: respecting the decision-making capacities of autonomous persons. Two general conditions are essential for autonomy: liberty, manifested as independence from controlling influences, and agency, namely, the capacity for intentional action.

2. Nonmaleficence: avoiding the causation of harm.

3. Beneficence: taking positive steps to help others, specifically, preventing evil or harm, removing evil or harm, and promoting good.

4. Justice: distributing benefits, risks, and costs fairly. Justice is interpreted as fair, equitable, and appropriate treatment for individuals and groups, given the many disparities in health care and research based on race, ethnicity, gender, and social status.

AI Ethics and Bias

The introduction of AI and the rapid development of AI applications have raised a variety of ethical issues (Christoforaki & Beyan, 2022), bias and discrimination featuring prominently among them.

Thus, AI ethics was developed as a domain of practical (or applied) ethics that comprises “a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies” (Leslie, 2019, p. 3).

AI ethics draws from both bioethics (the four principles presented above) and human rights discourse, the latter including, a.o., entitlement to equal freedom and dignity under the law, the protection of civil, political, and social rights, the universal recognition of personhood, and the right to free and unencumbered participation in the life of the community (Leslie, 2019).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

The four bioethics principles amended with Explicability are rendered for AI in (Floridi et al., 2018) as follows:

1. Autonomy, as the power for humans to decide whether to decide, and containing the risk of delegating too much to machines.

2. Non-maleficence, as preventing harms arising either from the intent of humans or the not-predicted behaviour of machines.

3. Beneficence, as promoting well-being, preserving dignity, and sustaining the planet.

4. Justice, as preventing and eliminating already existing unfair discriminations, as well as new harms, and ensuring the equal distribution of AI benefits.

5. Explicability, defined as understanding and holding accountable the decision-making processes of AI.

Thus, AI Ethics has also converged on a set of principles based on the four classic principles of medical ethics, as well as other approaches, summarised in (Christoforaki & Beyan, 2022). However, as noted in (Mittelstadt, 2019), compared to medicine, AI development lacks: (1) common aims and fiduciary duties, (2) professional history and norms, (3) proven methods to translate principles into practice, and (4) robust legal and professional accountability mechanisms; this undermines the success of the principled approach. Naturally, there is also a complex regulatory landscape governing the development and use of AI in the EU, including anti-discrimination laws, a subject, however, out of the scope of this report.

Regarding human rights, according to a 2018 report funded by the Council of Europe (Committee of experts on internet intermediaries (MSI-NET), 2018), human rights that are particularly affected by algorithms and automated data processing techniques include:

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- Free trial and due process
- Privacy and data protection
- Freedom of expression
- Effective remedy
- Freedom of assembly and association
- Prohibition of discrimination
- Social rights and access to public services
- The right to free elections

Biased algorithms are mentioned explicitly as possible discrimination factors against societal groups based on age, sexual orientation, race, gender, or socio-economic standing (Committee of experts on internet intermediaries (MSI-NET), 2018, p. 27). Additionally, the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law specifically mentions that the member states “shall adopt or maintain measures with a view to ensuring that activities within the lifecycle of artificial intelligence systems respect equality, including gender equality, and the prohibition of discrimination, as provided under applicable international and domestic law”, (Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, 2024, p. 4).

Civil society organisations (CSOs) as stakeholders in the health care ecosystem (Vayena et al., 2018) can play a significant role in identifying and addressing AI bias and AI governance in general via advocacy for ethical AI development, holding stakeholders accountable, educating the public, representing marginalised communities, shaping policy and regulatory frameworks, and fostering collaboration between governments, tech companies, and the public (Korir, 2024).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Within this theoretical framework, a variety of technical solutions are explicitly developed to address bias. In the following section, we present a classification of AI-induced biases that served as the basis for our mapping template, focusing on their impact on gender and racial discrimination. Human mental biases (Hofmann, 2023), for example, cognitive biases, such as confirmation or availability bias, although highly impactful in Medicine, are considered out of scope for the current project.

Bias in AI systems

Bias in computer systems is defined in (Friedman & Nissenbaum, 1996, p. 332) as a term " [referring] to computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favour of others. A system discriminates unfairly if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate."

According to (Friedman & Nissenbaum, 1996), bias in computer systems can be distinguished into three categories: pre-existing bias, technical bias, and emergent bias. In the following subsection, we examine each kind of bias and illustrate it with case studies, as manifested in the scientific literature.

Pre-existing Bias

Pre-existing bias stems from biases in social institutions, practices, and attitudes that already exist and are independent and usually present prior to the creation of the system. This kind of bias is incorporated into the system either consciously or unconsciously, sometimes even when the system creators are trying to avoid it.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Case study: Diagnosing Cardiovascular Diseases in Women

Cardiovascular diseases (CVDs) have been commonly perceived as “men’s disease”, and this has contributed to underdiagnosis and treatment for women. As shown in (Al Hamid et al., 2024), a systematic review on the issue, CVDs were less reported among women who either showed milder symptoms than men or had their symptoms misdiagnosed as gastrointestinal or anxiety-related symptoms; thus, women were offered fewer diagnostic tests, medicines, and were referred to cardiologists and/or hospitalisation less often. Furthermore, if hospitalised, women were less likely to receive a coronary intervention. Hence, women had their risk factors under-considered by physicians, especially by male physicians. Given the fact that women remain underrepresented in the field of cardiology (Fatunde et al., 2025), it may be concluded that women are less likely to receive proper health care due to already existing biases.

AI systems are trained using data collected by existing practices, so an AI CVD diagnosis system will incorporate this bias, creating discrimination against women, irrespective of any choices during the technical implementation.

Technical Bias

Technical bias arises from technical constraints or considerations, particularly when system creators attempt to make human constructs amenable to computers, such as quantifying the qualitative, discretising the continuous, or formalising the nonformal. Additionally, decontextualising algorithms from the environments in which they operate might cause them to fail to treat all groups fairly under all significant conditions.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Case study: Predictive Accuracy of Stroke Risk Prediction Models Across Black and White populations

(Hong et al., 2023) performed a retrospective study of predictive accuracy of stroke risk comparing existing stroke-specific risk prediction models and novel machine learning techniques involving, among other criteria, the race of the patients. All algorithms exhibited worse discrimination in Black individuals than in White individuals. This situation, according to the authors, may be attributed to risk factors not captured in the data, such as insurance type, language barriers, and other factors resulting from differential access to health care services, i.e., the data are decontextualised from the socioeconomic environment in which they were produced. At the same time, all the above-mentioned risk factors are constructs difficult to represent in a form amenable to computers. To all the above, we might also add that state-of-the-art AI algorithms are by nature opaque regarding the features they select to achieve high accuracy (Knight, 2017), thus rendering even their creators unable to explain how they work, and thus control whether any of the above-mentioned socioeconomic factors are really taken into consideration in the AI system's internal workings.

Emergent Bias

Emergent bias manifests in a context of use with real users, typically after a design is completed, as a result of changing societal knowledge that cannot be, or is not, incorporated into the system design, or a population with different knowledge or cultural values than assumed in the design.

Case study: Dataset shifts

A dataset shift is a mismatch between the distributions of the training and test datasets during algorithm development and may lead to disparate performance at the subgroup level (Chen et al., 2023).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

In skin cancer detection, for example, many imaging datasets used to train AI algorithms to detect skin cancer are sourced from countries with fair-skinned populations (Guo et al., 2021), thus underrepresenting certain demographic groups. AI algorithms trained with these datasets underperform when applied in countries with a more diverse population, discriminating against dark-skinned individuals. Datasets are difficult and expensive to collect, annotate, and validate, making it necessary for AI systems developed in low- and middle-income countries to rely on publicly available datasets that may not reflect their population distribution, resulting in a mismatch between the source and target populations. The same may also occur in high-income countries, for example, due to population shifts from increased immigration, or in variations in self-reported race. As noted in (Chen et al., 2023), “since it is now accepted that race is a social construct and that there is greater genetic variability within a particular race than there is between races” […] “the medical community has begun to realise that the taxonomies of the past do not adequately represent the groups of people that they purport to” and “they can obscure culture, history, socioeconomic status and other confounders of fairness.”

Kinds of bias specific to the ML/AL pipeline

While the above is valid for all computer systems, AI applications have more specific requirements, so we needed a more fine-grained taxonomy. Consequently, we decided to follow the bias classification presented in (Suresh & Guttag, 2020), as it identifies the types of bias at each ML/AI pipeline step, as illustrated in Figure 1.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

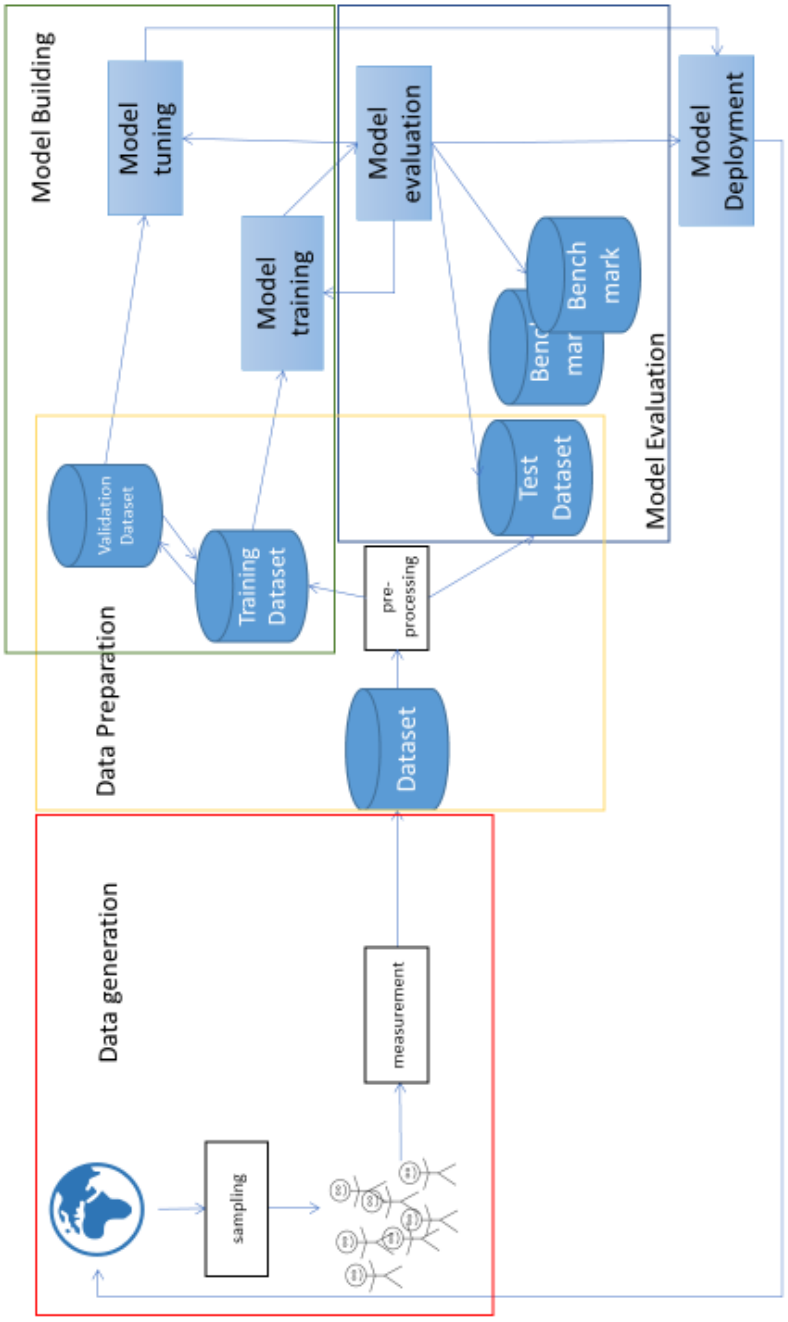


Figure 1 ML/AI Pipeline. Image adapted from (Suresh & Guttag, 2020)

A typical ML/AI pipeline can be described as follows:

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- **Data generation**. The creation of an ML/AI system starts with data generation. This involves first collecting and preparing data to compile a dataset for the AI system. Existing data in the world must be collected by identifying a target population sample. The next step is to define and measure features relevant to the application to be implemented, and/or annotate data with appropriate labels. This is an expensive and lengthy process, so more often than not, AI practitioners use existing datasets (either public or purchased).

- **Data Preparation**. At this stage, the dataset is split into three sets, namely: the *training dataset* - the actual dataset used to train the model; the *validation dataset*, a sample of data used to provide an evaluation of a model fit on the training dataset while tuning model hyperparameters (model parameters that cannot be learned from the data, e.g. the number of layers and neurons in a neural network model). At this phase, the data may need to be preprocessed (e.g., cleaned, normalized); and the *test dataset*, the part of the data used to evaluate the final model providing a gold standard once a model is completely trained.

- **Model building**. At this phase, the model is trained on the training data and fine-tuned by adjusting hyperparameters on the validation dataset.

- **Model Evaluation**. The trained model is evaluated using the test dataset and sometimes benchmark datasets, which are independently compiled datasets used to demonstrate model robustness and/or enable comparison with other methods.

- **Model Deployment**. Application of the model in a real-world setting. This may lead to changes depending on the results and can also create a feedback loop to the beginning of the pipeline.

Taking into consideration the phases of the ML/AI pipeline described above, we adopt the bias classification of (Suresh & Guttag, 2020). Specifically, they identify the following bias categories: historical, representation bias, measurement, aggregation,

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

learning, evaluation, and deployment bias. In the following subsections, we define the above-listed biases and offer case studies from the sources collected for the project.

Historical Bias

Historical bias corresponds to Pre-existing bias as defined by (Friedman & Nissenbaum, 1996), which incorporates already existing prejudices and stereotypes in the data. An example can be seen in (Calderone, 1990), which examines whether the frequency of pain and sedative medication administered to postoperative coronary artery bypass graft (CABG) patients differs according to patient gender and age. The result revealed that male patients and patients 61 years or younger were administered pain medication significantly more frequently than female patients and patients 62 years old and over, who instead were administered sedative medication significantly more frequently. The Case study about Predictive Accuracy of Stroke Risk Prediction Models Across Black and White populations shocases that in the Pre-existing Bias subsection; however, we are going to present another case study showcasing historical bias regarding the use of AI in mental health.

Case study: Artificial Intelligence in mental health and the biases of language-based models

(Straw & Callison-Burch, 2020) present a systematic literature review of uses of NLP in mental health, with the aim of identifying how these biases may widen health inequalities. AI models that use NLP to profile mental health collect large datasets of expressive language, typically acquired from social media, online forums, blogs, and chat rooms. However, these data are already influenced by an individual's personal background and social context.

Specifically, regarding gender and language, there is an extensive bibliography (on the English language) summarised in (Pennebaker et al., 2003), which reveals differences in women's and men's use of words. For example, women use less assertive speech,

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

manifesting in greater politeness, less swearing, more intensifiers (e.g., really, so), and more hedges (i.e., qualifiers or uncertain words such as sort of, perhaps, or maybe). Men, on the other hand, were described as directive, precise, and also less emotional in their language use, which is characterised by references to quantity, judgmental adjectives (e.g., good, dumb), elliptical sentences ("Great picture."), and "I" references. As the authors remark, these differences are consistent with a sociological framework of gender differences but can also be attributed to alternative explanations, such as women's greater social engagement.

Regarding mental health, men and women write suicide notes expressing suicidal distress differently; women internalise negative emotions, while men express increasing anger (Straw & Callison-Burch, 2020). An AI system that screens for mental health issues for one gender may be inappropriate for another (and this is considering gender in a binary context, which excludes a large part of the population).

Representation bias

Representation bias occurs when the development sample underrepresents some part of the population during the data collection phase. This may arise in the following ways: when defining the target population, if it does not reflect the use population; when defining the target population, if it contains underrepresented groups; when sampling from the target population, if the sampling method is limited or uneven. Representation bias results in poor generalisation for a subset of the user population. A typical example of representational bias concerns skin cancer detection, since many imaging datasets underrepresent certain demographic groups, causing machine learning models to train on images of primarily fair-skinned individuals (Guo et al., 2021). Taking into consideration the target diseases of the AEQUITAS project, we present a case study on representation bias regarding race in type 2 diabetes.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Case study: Assessing racial bias in type 2 diabetes risk prediction algorithms

According to (Cronjé et al., 2023), regarding the USA population, despite their comparatively lower risk, non-Hispanic White groups remain overrepresented in the diabetes risk prediction literature. In a different review on Ethnoracial Equity in Artificial Intelligence for Diabetes Management, in the reviewed articles that reported race, the average distribution was 69.5% White, 17.1% Black, and 3.7% Asian, while only 2 articles reported inclusion of Native American participants (Pham et al., 2021).

It is well documented that the inequities in diabetes outcomes are largely driven by complex, interrelated social determinants of health, including access to healthy food, quality healthcare, insurance status, educational barriers, and differential rates of technology adoption. These outcomes include higher rates of complications and worse glycemic control among minoritised and low-income populations (Alipour & Alipour, 2025).

As a result, an AI system trained on existing datasets would generalise poorly, leading to biased predictive models that may favour individuals of certain racial groups, for example, in preventive action.

Measurement bias

Measurement bias occurs when choosing, collecting, or computing features and labels to use in a prediction problem, especially when using a proxy (an approximation of a construct that is not directly encoded or observable). An example can be found in a study by (Obermeyer et al. 2019), where health costs were used as a proxy to predict and rank which patients would benefit most from extra care, resulting in race discrimination. However, health costs are a poor proxy for health needs because black patients, facing disproportionate levels of poverty, often spend less on health care than whites. Because of this bias, the algorithm falsely concluded that blacks were

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

healthier than equally sick white patients, thus triaging them as a lower priority patient when accessing health care services.

Other sources of measurement bias can occur when the measurement method varies across groups, for example, when two groups are monitored for the same behavior, but one of them is monitored more stringently or frequently than the other. Similarly, measurement accuracy can vary across groups, which in medical applications may lead to systematically higher rates of misdiagnosis or underdiagnosis in certain groups. For example, physicians are more likely to underestimate the pain of black patients relative to nonblack patients due to false beliefs about biological differences between blacks and whites, leading to black patients being less likely to be given pain medications and, if given, they receive lower quantities (Hoffman et al., 2016).

Case study: Racial and Ethnic Differences in the Association Between Mean Glucose and Hemoglobin A1c

The A1C test measures the average amount of glucose (sugar) in blood and is used to detect prediabetes or help diagnose Type 2 diabetes. However, A1C is only an indirect measure and not causally linked to health outcomes, since there are numerous ways that the relationship between direct measures of glycemia (the concentration of glucose in the blood) and A1C can be directly altered. There is even substantial variation in the glycemia–A1C relationship across individuals and even within individuals over time. Furthermore, studies have reported significantly higher hemoglobin A1c (A1C) in African American patients than in White patients with the same mean glucose (Karter et al., 2023).

If an AI system designed to diagnose diabetes is trained to use A1C test results as a proxy for glycemia without accounting for other factors, such as patient race, this may lead to premature diabetes diagnoses and inappropriate treatment, resulting in biased health care quality and health inequities. However, as noted in (Alipour & Alipour, 2025), a systematic review of biases that could affect the equity of AI/ML models in

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

diabetes (including measurement bias), while the studies reviewed explicitly mention that measurement bias may propagate through AI models if not corrected, none of them accounted for such biases during model development, explicitly mitigated them or reported correcting for differences in measurement accuracy.

Aggregation bias

Aggregation bias results when a one-size-fits-all model is used for a dataset that includes diverse groups of people or things.

We can consider the example of mapping input data (e.g., a person's income) to labels that describe them (e.g., low, middle, high) being assumed to be consistent across subsets of the data. In reality, a person's background or culture can change what those numbers actually mean. For example, a "high" income in a small rural town or low- or middle-income country might mean something very different than it does in a major city or a high-income country.

Case study: Digital health tools for the passive monitoring of depression

The use of digital tools to measure physiological and behavioural variables for the passive monitoring of depression is addressed by (De Angel et al., 2022), a systematic review on the issue. The reviewed articles examined associations between depression and objective behavioural data obtained from smartphone and wearable device sensors. These data were mapped into features used by the AI models to make predictions, corresponding to sleep, physical activity, circadian rhythm, sociability, location, and phone use.

However, the authors underline the heterogeneity that arises from the diversity of methods used to create these features. For example, the feature "sleep quality" can be defined by measuring counts of awakenings, the total number of minutes awake, or the proportion of awake vs. asleep in a sleep session, while we must also take into

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

consideration the differences in how sensors in different devices describe an event as "sleep". As all the above differentiations are not considered and are collectively lumped as "sleep quality", and since a dataset might be sourced from people or groups with different backgrounds, cultures, or norms, this feature may have a different meaning for each of these groups or individuals.

Aggregating such data into a single feature may result in a system that does not fit any group, or that privileges the dominant population if there is also representation bias. For example, there is evidence that there are sex differences in sleep between men and women, while the latter are often underrepresented in sleep research. Additionally, other factors not usually taken into consideration for sleep patterns and disorders are not distinguishing gender as a social construct from biological sex, and not considering intersectional identities defined by age, race, and socioeconomic class (Lok et al., 2024).

Learning bias

Learning bias arises when modelling choices amplify performance disparities across different examples in the data. An example concerns differential privacy, a mechanism used in AI systems that ensures that, by examining a system's output, it is not possible to determine whether a specific individual's data were included in the original dataset. Differential privacy is used in healthcare datasets to protect sensitive patient information, for example, in the case of rare diseases, where each patient's case is more or less unique in a limited area covered by a hospital, so even if the data are anonymised, it is not very difficult to deduce the person's identity. However, it has been shown that differential privacy reduces the influence of underrepresented data on the model; thus, if the AI system is biased to begin with, applying a privacy-enhancing measure exacerbates this bias even more (Bagdasaryan & Shmatikov, 2019).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Case Study: Differential privacy and health disparities

In September 2018, the US Census Bureau announced that they would implement differential privacy on data products derived from 2020 census data. However, (Santos-Lozada et al., 2020) researched the way the implementation of differential privacy can alter knowledge about health disparities in mortality, especially for racial or ethnic minorities in small areas and less urban settings. Their results suggested that differential privacy will more strongly affect mortality rate estimates for non-Hispanic blacks and Hispanics than estimates for non-Hispanic whites.

These findings were supported by (Kurz et al., 2022), who show that applying differential privacy to the same data can result in a misrepresentation of Medicaid participation rates among already marginalised racial and ethnic groups. Specifically, these rates for certain combinations of county, race, and ethnicity differed between the differential privacy data results and the original data, sometimes exceeding 10%. Additionally, non-Hispanic White individuals were the only ethnic and racial subgroup for which the differential privacy algorithm accurately captured Medicaid participation rates. This finding may have important implications for health policy, as Census data are used to plan government programs, allocate resources, and evaluate and track policies.

Evaluation bias

Evaluation bias occurs when the benchmark data used for a particular task do not represent the use population. Benchmarks are standardised datasets used to measure a model's quality, enabling quantitative comparison of models. Subsequently, there is a risk of encouraging the development and deployment of models that perform well only on the subset of the data represented in the benchmark. Thus, discrimination against vulnerable subgroups or individuals can occur if the benchmark is subject to historical, representational, or measurement bias.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

In health care, the reasons for underrepresentation of specific populations in datasets can be either because individuals or groups are absent from datasets (for example, pregnant women, due to ethical constraints) or because people are incorrectly or inappropriately categorised into groups (for example, categories of “mixed ethnicity” or “other”). The root causes for this may include social and technical or legal/ethical reasons, such as structural barriers to receiving healthcare, technical obstacles to the capture or digitisation of relevant health data, individual and structural limitations regarding consent for data sharing, and legal or ethical restrictions on data sharing preventing data accessibility, among others (Arora et al., 2023). The result is that AI systems calibrated to such benchmarks may underperform when applied to individuals from an underrepresented group. However, it is important to note that benchmark validity is a more generic issue and is not limited to bias (Brooks, 2025).

Case study: Skin image datasets

Skin imaging datasets underrepresent certain demographic groups, as most images in these datasets come from populations in North America or Europe and predominantly depict fair-skinned individuals (Guo et al., 2021). Because of the high cost and difficulty of constructing these datasets, apart from training models, they can also be used as benchmarks.

The case study that illustrates emergent bias, i.e., the skin cancer image datasets used to train prediction models, is an example of an inappropriate benchmark when the user population comes from underrepresented groups (Guo et al., 2021). A similar case, although not related to AI, shows the generality of the problem, which had to do with pulse oximeters (devices that measure blood oxygen saturation, used, for example, in cases of heart attack or failure), which are shown to work more accurately in light-pigmented skin (Sjoding et al., 2020).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Representation, measurement, aggregation, learning, and evaluation bias can be mapped to technical bias defined by (Friedman & Nissenbaum, 1996).

Deployment bias

Deployment bias arises when there is a mismatch between the problem a model is intended to solve and the way it is actually used, which can cause harm, especially when combined with cognitive biases such as confirmation and automation biases. Deployment bias is the same as emergent bias defined by (Friedman & Nissenbaum, 1996).

Case study: Domain shift

The case of data shift is documented in the emergent bias subsection about skin cancer detection. Additionally, we can define the case of domain shift, which occurs when a system is implemented, has passed regulatory authorisation, and is deployed in clinical practice, but is applied to a different patient cohort from that for which it was trained. As an example, a system can be developed for a hospital in a high-income country and be deployed in a low- or middle-income country without taking into consideration factors such as the sociodemographic characteristics of patients, or whether the patients have the same overall risk level compared to the ones included in the training data (Vokinger et al., 2021).

Policy implications

The evidence mapped in Deliverable D2.1 demonstrates that gender and racial biases in biomedical AI are not incidental or isolated technical flaws, but systemic risks emerging across the full lifecycle of AI systems used in healthcare. In cardiovascular disease, depression, and diabetes, bias arises from historically skewed clinical datasets,

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

unequal diagnostic practices, proxy variables that encode structural inequalities, and deployment contexts that unevenly distribute both benefits and harms. These findings confirm that biomedical AI directly engages multiple rights and principles protected by the EU Charter of Fundamental Rights, most notably the precepts of human dignity, equality before the law, and non-discrimination, as well as the right to integrity of the person, the right to healthcare, data protection, and the right to an effective remedy.

Against this background, EU and national policy frameworks governing AI in healthcare must treat bias mitigation not as a voluntary ethical add-on but as a binding component of lawful, rights-compliant AI deployment. European and national regulatory efforts concerning AI in healthcare should be seen as situated within the wider fundamental rights framework governing AI (see Novossiolova, 2025; Novossiolova et al., 2025; Kasapi, 2025). The EU AI Act provides a necessary regulatory backbone by classifying most biomedical AI as high-risk systems, yet its effectiveness in practice will depend on how fundamental-rights safeguards are operationalised in conformity assessments, post-market monitoring, and public-sector procurement.

First, guarantees for meaningful human oversight must be strengthened and specified for biomedical AI systems throughout their lifecycle. Clinical AI tools used for diagnosis, risk stratification, screening, or treatment support should in no case function as de facto autonomous decision-makers. Human oversight must include not only the possibility of override by healthcare professionals, but also clear institutional responsibility for understanding system limitations, known bias risks, and subgroup performance gaps. In line with the Charter's protection of human dignity and integrity, healthcare professionals should be trained and institutionally supported to critically interrogate AI outputs rather than defer to them. This requires embedding AI literacy, bias awareness, and fundamental-rights training into medical education and continuous professional development.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Transparency obligations should be interpreted expansively in healthcare contexts. Patients and healthcare users must be informed whenever AI systems are used in clinical decision-making that affects them, including in screening, prioritisation, or risk scoring. Where AI-generated outputs inform public healthcare services, these outputs should be clearly identifiable as such, and accompanied by accessible explanations of their role, limitations, and known bias risks. Individuals should also be informed when their personal data are used for AI training, testing, or continuous learning, particularly where sensitive health data are involved. These transparency measures are essential to uphold the Charter rights to data protection and effective remedy, and to enable individuals to meaningfully challenge decisions that may adversely affect them.

Second, fundamental-rights impact assessment must become a routine, enforceable requirement for biomedical AI systems, extending beyond pre-market checks to continuous evaluation during deployment. The empirical evidence in D2.1 shows that many bias harms only become visible once AI systems interact with real populations and clinical workflows, particularly through intersectional effects involving gender, race, age, and socio-economic status. Rights-based impact assessments, such as those inspired by the Council of Europe's HUDERIA methodology (Methodology for the Risk and Impact Assessment of Artificial Intelligence Systems From the Point of View of Human Rights, Democracy and the Rule of Law), should therefore be mandatory for high-risk medical AI, explicitly examining differential performance and outcomes across protected groups. These assessments must involve meaningful stakeholder participation, including civil society organisations, patient representatives, and equality bodies, in order to surface harms that may be invisible from a purely technical or clinical perspective.

Periodic audits of biomedical AI systems should be required to verify continued compliance with fundamental-rights standards, with particular attention to bias drift, dataset shifts, and changes in clinical use over time. Where audits reveal persistent or unmitigable discriminatory effects, there must be clear legal and institutional pathways

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

to restrict, suspend, or terminate the use of the system. The right to healthcare cannot justify the continued deployment of AI tools that systematically disadvantage certain groups, even if aggregate performance metrics appear favourable.

Third, EU and national authorities must address the risk of misuse and secondary harms associated with biomedical AI. This includes cybersecurity vulnerabilities that could compromise system integrity or enable malicious manipulation of clinical outputs, as well as the repurposing of health AI for surveillance, profiling, or exclusionary practices. Biomedical AI systems should be subject to regular security assessments and robust incident-reporting obligations, with clear accountability mechanisms in cases where biased or compromised systems lead to rights violations. Liability frameworks should ensure that responsibility cannot be deflected solely onto individual clinicians when harms are structurally embedded in AI design or deployment decisions.

Fourth, the promotion of ethical and responsible practices must be embedded across the entire biomedical AI value chain. Developers should be required to proactively address bias risks through representative data collection, careful target and proxy selection, subgroup-specific validation, and transparent reporting of performance across gender and racial groups. Importantly, the evidence reviewed in D2.1 shows that “fairness through unawareness” and purely technical debiasing strategies are often insufficient in healthcare settings. Regulatory guidance and standards should therefore move beyond abstract fairness metrics and require developers to demonstrate clinically meaningful equity outcomes, assessed in relation to real healthcare pathways and access patterns.

Public procurement and funding policies play a crucial role in shaping developer incentives. Healthcare authorities and public hospitals should integrate fundamental rights and bias criteria into procurement decisions for AI systems, favouring solutions that demonstrate robust, transparent, and independently verified bias mitigation

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

practices. EU funding instruments, including future research and innovation programmes, should continue to prioritise projects that combine technical innovation with rights-based governance, stakeholder engagement, and capacity building, in line with the AEQUITAS model.

Finally, strengthening societal resilience to biased biomedical AI requires sustained investment in public awareness, civil society engagement, and cross-sector collaboration. Individuals must be empowered to understand their rights in AI-mediated healthcare and the mechanisms available to protect them. Civil society organisations, equality bodies, and patient groups should be recognised as essential actors in monitoring AI impacts, supporting affected individuals, and informing policy development. Cooperation between governments, healthcare providers, researchers, industry, and civil society is necessary to ensure that the benefits of biomedical AI are shared equitably and do not reinforce existing health inequalities.

Taken together, the findings of Deliverable D2.1 support a clear policy conclusion: biomedical AI can only be considered trustworthy and legitimate in the EU when its design, deployment, and governance are firmly anchored in the protection of fundamental rights. The EU AI Act, interpreted through the lens of the EU Charter of Fundamental Rights and operationalised via concrete oversight, impact assessment, and accountability mechanisms, provides a critical opportunity to ensure that innovation in healthcare advances equity rather than reproducing historical patterns of discrimination.

References

Al Hamid, A., Beckett, R., Wilson, M., Jalal, Z., Cheema, E., Al-Jumeily Obe, D., Coombs, T., Ralebitso-Senior, K., & Assi, S. (2024). Gender Bias in Diagnosis, Prevention,

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

and Treatment of Cardiovascular Diseases: A Systematic Review. *Cureus*, *16*(2), e54264. https://doi.org/10.7759/cureus.54264

Alhuwaydi, A. M. (2024). Exploring the Role of Artificial Intelligence in Mental Healthcare: Current Trends and Future Directions – A Narrative Review for a Comprehensive Insight. *Risk Management and Healthcare Policy*, *17*, 1339–1348. https://doi.org/10.2147/RMHP.S461562

Alipour, M., & Alipour, A. (2025). Algorithmic Bias in AI-Based Diabetes Care: Systematic Review of Model Performance, Equity Reporting, and Physiological Label Bias. *InfoScience Trends*, *2*(5), 33–46. https://doi.org/10.61186/ist.202502.05.04

Arora, A., Alderman, J. E., Palmer, J., Ganapathi, S., Laws, E., McCradden, M. D., Oakden-Rayner, L., Pfohl, S. R., Ghassemi, M., McKay, F., Treanor, D., Rostamzadeh, N., Mateen, B., Gath, J., Adebajo, A. O., Kuku, S., Matin, R., Heller, K., Sapey, E., … Liu, X. (2023). The value of standards for health datasets in artificial intelligence-based applications. *Nature Medicine*, *29*(11), 2929–2938. https://doi.org/10.1038/s41591-023-02608-w

Bagdasaryan, E., & Shmatikov, V. (2019). *Differential Privacy Has Disparate Impact on Model Accuracy* (arXiv:1905.12101). arXiv. https://doi.org/10.48550/arXiv.1905.12101

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Baker, R. B., & McCullough, L. B. (Eds). (2008). *The Cambridge World History of Medical Ethics*. Cambridge University Press. https://doi.org/10.1017/CHOL9780521888790

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (8th ed).

Bernstein, B. S., Streather, S., & O'Gallagher, K. (2025). The Emerging Role of Artificial Intelligence in Heart Failure. *Future Cardiology*, *21*(10), 795–801. https://doi.org/10.1080/14796678.2025.2523155

Bierman, A. S. (2007). Sex matters: Gender disparities in quality and outcomes of care. *CMAJ : Canadian Medical Association Journal*, *177*(12), 1520–1521. https://doi.org/10.1503/cmaj.071541

Bowser, R. (2001). Racial bias in medical treatment. *Dick. L. Rev.*, *105*(3), 365.

Brooks, M. (2025). Is your AI benchmark lying to you? *Nature*, *644*(8075), 294–296. https://doi.org/10.1038/d41586-025-02462-5

Calderone, K. L. (1990). The influence of gender on the frequency of pain and sedative medication administered to postoperative patients. *Sex Roles*, *23*(11), 713–725. https://doi.org/10.1007/BF00289259

Cao, Y., Dai, J., Wang, Z., Zhang, Y., Shen, X., Liu, Y., & Tian, Y. (2025). Machine Learning Approaches for Depression Detection on Social Media: A Systematic Review of

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Biases and Methodological Challenges. *Journal of Behavioral Data Science*, *5*(1), 67–102. https://doi.org/10.35566/jbds/caoyc

Chen, R. J., Wang, J. J., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Lu, M. Y., Sahai, S., & Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering*, *7*(6), Article 6. https://doi.org/10.1038/s41551-023-01056-8

Christoforaki, M., & Beyan, O. (2022). AI Ethics—A Bird's Eye View. *AI Ethics—A Bird's Eye View*, *12*(9), Article 9. https://doi.org/10.3390/app12094130

Committee of experts on internet intermediaries (MSI-NET). (2018). *Algorithms and human rights—Study on the human rights dimensions of automated data processing techniques and possible regulatory implications* [Study]. Council of Europe. https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5

Contreras, I., & Vehi, J. (2018). Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. *Journal of Medical Internet Research*, *20*(5), e10775. https://doi.org/10.2196/10775

Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Council of Europe Treaty Series—No. 225 (2024). https://rm.coe.int/1680afae3c

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Cronjé, H. T., Katsiferis, A., Elsenburg, L. K., Andersen, T. O., Rod, N. H., Nguyen, T.-L., & Varga, T. V. (2023). Assessing racial bias in type 2 diabetes risk prediction algorithms. *PLOS Global Public Health*, *3*(5), e0001556. https://doi.org/10.1371/journal.pgph.0001556

Dang, V. N., Cascarano, A., Mulder, R. H., Cecil, C., Zuluaga, M. A., Hernández-González, J., & Lekadir, K. (2024). Fairness and bias correction in machine learning for depression prediction across four study populations. *Scientific Reports*, *14*(1), 7848. https://doi.org/10.1038/s41598-024-58427-7

De Angel, V., Lewis, S., White, K., Oetzmann, C., Leightley, D., Oprea, E., Lavelle, G., Matcham, F., Pace, A., Mohr, D. C., Dobson, R., & Hotopf, M. (2022). Digital health tools for the passive monitoring of depression: A systematic review of methods. *Npj Digital Medicine*, *5*(1), 3. https://doi.org/10.1038/s41746-021-00548-8

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), Article 1. https://doi.org/10.1038/s41591-018-0316-z

Fatunde, O. A., Grant, J. K., Lara-Breitinger, K., Kizzee, O. P., Savic, J., LeMond, L., & Hayes, S. N. (2025). Gender Disparities in Cardiology. *JACC: Advances*, *4*(4), 101642. https://doi.org/10.1016/j.jacadv.2025.101642

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

FitzGerald, C., & Hurst, S. (2017). Implicit bias in healthcare professionals: A systematic review. *BMC Medical Ethics*, *18*(1), 19. https://doi.org/10.1186/s12910-017-0179-8

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347. https://doi.org/10.1145/230538.230561

Gou, F., Liu, J., Xiao, C., & Wu, J. (2024). Research on Artificial-Intelligence-Assisted Medicine: A Survey on Medical Artificial Intelligence. *Diagnostics*, *14*(14), 1472. https://doi.org/10.3390/diagnostics14141472

Guo, L. N., Lee, M. S., Kassamali, B., Mita, C., & Nambudiri, V. E. (2021). Bias in, bias out: Underreporting and underrepresentation of diverse skin types in machine learning research for skin cancer detection-A scoping review. *Journal of the American Academy of Dermatology*, S0190-9622(21)02086-7. https://doi.org/10.1016/j.jaad.2021.06.884

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Hamberg, K. (2008). Gender Bias in Medicine. *Women's Health*, *4*(3), 237–243. https://doi.org/10.2217/17455057.4.3.237

Hammond, M. E. H., Stehlik, J., Drakos, S. G., & Kfoury, A. G. (2021). Bias in Medicine. *JACC: Basic to Translational Science*, *6*(1), 78–85. https://doi.org/10.1016/j.jacbts.2020.07.012

Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences*, *113*(16), 4296–4301. https://doi.org/10.1073/pnas.1516047113

Hofmann, B. (2023). Biases in bioethics: A narrative review. *BMC Medical Ethics*, *24*(1), 17. https://doi.org/10.1186/s12910-023-00894-0

Hong, C., Pencina, M. J., Wojdyla, D. M., Hall, J. L., Judd, S. E., Cary, M., Engelhard, M. M., Berchuck, S., Xian, Y., D'Agostino, R., Sr, Howard, G., Kissela, B., & Henao, R. (2023). Predictive Accuracy of Stroke Risk Prediction Models Across Black and White Race, Sex, and Age Groups. *JAMA*, *329*(4), 306–317. https://doi.org/10.1001/jama.2022.24683

Karter, A. J., Parker, M. M., Moffet, H. H., & Gilliam, L. K. (2023). Racial and Ethnic Differences in the Association Between Mean Glucose and Hemoglobin A1c. *Diabetes Technology & Therapeutics*, *25*(10), 697–704. https://doi.org/10.1089/dia.2023.0153

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Kasapi, Z., Markov, D., Novossiolova, T., Laanpere, L., Rünne, E., Pastukhov, O., & Iwańska, K. (2025). *Fundamental rights implications of AI use: Background study*. REFRAIME Initiative, Center for the Study of Democracy. https://refraime.csd.eu/wp-content/uploads/fundamental-rights-implications-of-ai-use_background-study.pdf

Khalifa, M., & Albadawy, M. (2024). Artificial intelligence for diabetes: Enhancing prevention, diagnosis, and effective management. *Computer Methods and Programs in Biomedicine Update*, *5*, 100141. https://doi.org/10.1016/j.cmpbup.2024.100141

Knight, W. (2017, April 11). The Dark Secret at the Heart of AI. *MIT Technology Review*. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/

Korir, K. (2024, October 4). The Role of Civil Society in AI Governance. *Medium*. https://medium.com/@kiplangatkorir/the-role-of-civil-society-in-ai-governance-30138dcc916c

Koteluk, O., Wartecki, A., Mazurek, S., Kołodziejczak, I., & Mackiewicz, A. (2021). How Do Machines Learn? Artificial Intelligence as a New Era in Medicine. *Journal of Personalized Medicine*, *11*(1), Article 1. https://doi.org/10.3390/jpm11010032

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Kumari, M., Singh, G., & Pande, S. D. (2025). A Survey of Current Progress in Depression Detection Using Deep Learning and Machine Learning. *Biomedical Materials & Devices*. https://doi.org/10.1007/s44174-025-00301-9

Kurz, C. F., König, A. N., Emmert-Fees, K. M. F., & Allen, L. D. (2022). The effect of differential privacy on Medicaid participation among racial and ethnic minority groups. *Health Services Research*, *57*(S2), 207–213. https://doi.org/10.1111/1475-6773.14000

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. Zenodo. https://doi.org/10.5281/ZENODO.3240529

Liu, M., Sandhu, S., Reisner, S. L., Gonzales, G., & Keuroghlian, A. S. (2023). Health Status and Health Care Access Among Lesbian, Gay, and Bisexual Adults in the US, 2013 to 2018. *JAMA Internal Medicine*, *183*(4), 380–383. https://doi.org/10.1001/jamainternmed.2022.6523

Lok, R., Qian, J., & Chellappa, S. L. (2024). Sex differences in sleep, circadian rhythms, and metabolism: Implications for precision medicine. *Sleep Medicine Reviews*, *75*, 101926. https://doi.org/10.1016/j.smrv.2024.101926

Mao, K., Wu, Y., & Chen, J. (2023). A systematic review on automated clinical depression diagnosis. *Npj Mental Health Research*, *2*(1), 20. https://doi.org/10.1038/s44184-023-00040-z

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Miles, S. H. (2005). *The Hippocratic Oath and the Ethics of Medicine*. Oxford University Press, USA.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Naskar, S., Sharma, S., Kuotsu, K., Halder, S., Pal, G., Saha, S., Mondal, S., Biswas, U. K., Jana, M., & Bhattacharjee, S. (2025). The biomedical applications of artificial intelligence: An overview of decades of research. *Journal of Drug Targeting*, *33*(5), 717–748. https://doi.org/10.1080/1061186X.2024.2448711

Novossiolova, T. (2025). *Human-Centric AI Innovation and Use: Role of the Charter of Fundamental Rights of the EU*. REFRAIME Initiative. https://reframe.csd.eu/wp-content/uploads/reframe-policy-brief_en.pdf

Novossiolova, T., Iwańska, K., Pastukhov, O., Markov, D., Skoric, V., Laanpere, L., Rünne, E., & Vitoratos, S. (2025). *Development of AI systems and fundamental rights: Good practices for awareness, assessment, and accountability (AIM Framework)*. Center for the Study of Democracy. https://reframe.csd.eu/wp-content/uploads/reframe-framework-ai-systems_en.pdf

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, *54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Pham, Q., Gamble, A., Hearn, J., & Cafazzo, J. A. (2021). The Need for Ethnoracial Equity in Artificial Intelligence for Diabetes Management: Review and Recommendations. *Journal of Medical Internet Research*, *23*(2), e22320. https://doi.org/10.2196/22320

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), Article 1. https://doi.org/10.1038/s41591-021-01614-0

Santos-Lozada, A. R., Howard, J. T., & Verdery, A. M. (2020). How differential privacy will affect our understanding of health disparities in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(24), 13405–13412. https://doi.org/10.1073/pnas.2003714117

Sheng, B., Pushpanathan, K., Guan, Z., Lim, Q. H., Lim, Z. W., Yew, S. M. E., Goh, J. H. L., Bee, Y. M., Sabanayagam, C., Sevdalis, N., Lim, C. C., Lim, C. T., Shaw, J., Jia, W., Ekinci, E. I., Simó, R., Lim, L.-L., Li, H., & Tham, Y.-C. (2024). Artificial intelligence for diabetes care: Current and future prospects. *The Lancet Diabetes & Endocrinology*, *12*(8), 569–595. https://doi.org/10.1016/S2213-8587(24)00154-2

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E., & Valley, T. S. (2020). Racial Bias in Pulse Oximetry Measurement. *New England Journal of Medicine*, *383*(25), 2477–2478. https://doi.org/10.1056/NEJMc2029240

Straw, I., & Callison-Burch, C. (2020). Artificial Intelligence in mental health and the biases of language based models. *PLOS ONE*, *15*(12), e0240376. https://doi.org/10.1371/journal.pone.0240376

Suresh, H., & Guttag, J. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO '21*, 1–9. https://doi.org/10.1145/3465416.3483305

Tröhler, U. (2008). The Historical Development of International Codes of Ethics for Human Subjects Research. In R. B. Baker & L. B. McCullough (Eds), *The Cambridge World History of Medical Ethics* (1st edn, pp. 566–575). Cambridge University Press. https://doi.org/10.1017/CHOL9780521888790.052

van Assen, M., Beecy, A., Gershon, G., Newsome, J., Trivedi, H., & Gichoya, J. (2024). Implications of Bias in Artificial Intelligence: Considerations for Cardiovascular Imaging. *Current Atherosclerosis Reports*, *26*(4), 91–102. https://doi.org/10.1007/s11883-024-01190-x

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Vayena, E., Dzenowagis, J., Brownstein, J. S., & Sheikh, A. (2018). Policy implications of big data in the health sector. *Bulletin of the World Health Organization*, *96*(1), 66–68. https://doi.org/10.2471/BLT.17.197426

Vevaina, J. R., Nora, L. M., & Bone, R. C. (1993). Issues in biomedical ethics. *Disease-a-Month*, *39*(12), 874–925.

Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S. (2021). Mitigating bias in machine learning for medicine. *Communications Medicine*, *1*(1), 25. https://doi.org/10.1038/s43856-021-00028-w

Wang, L., Wang, C., Li, C., Murai, T., Bai, Y., Song, Z., Zhang, S., Zhang, Q., Huang, Y., Bi, X., & Jiang, J. (2025). AI-assisted multi-modal information for the screening of depression: A systematic review and meta-analysis. *Npj Digital Medicine*, *8*(1), 523. https://doi.org/10.1038/s41746-025-01933-3

Williams, D. R., & Wyatt, R. (2015). Racial Bias in Health Care and Health: Challenges and Opportunities. *JAMA*, *314*(6), 555. https://doi.org/10.1001/jama.2015.9260

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization. https://www.who.int/publications/i/item/9789240029200

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Yoon, D. Y., Mansukhani, N. A., Stubbs, V. C., Helenowski, I. B., Woodruff, T. K., & Kibbe, M. R. (2014). Sex bias exists in basic science and translational surgical research. *Surgery*, *156*(3), 508–516. https://doi.org/10.1016/j.surg.2014.07.001

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, *2*(10), 719–731. https://doi.org/10.1038/s41551-018-0305-z

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Appendix 1. Source Collection and Mapping Method

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Collection of sources for the database and the Mapping template

Collection of sources: process, method and tools

At the AEQUITAS Kick-off meeting on 25/9/25 in Larissa, Greece, UKK presented the WP2 Task 2.1 and 2.2 overview, as well as an introduction to the concept of bias in medicine and AI. The slides for this presentation can be found in the Appendix 2: Meeting Slides. As per the contractual agreement, the partners were asked to collect sources regarding race and/or gender bias in AI systems, which, when mapped to a template, can be used to populate the bias database produced by AEQUITAS. Since the partners belong to a variety of organisations (academic institutions, companies, and civil society organisations), each can contribute according to their expertise, providing a coverage of the domain via multiple points of view.

A spreadsheet was deemed the most efficient solution for information gathering, as a well-known tool to all project partners. Google Sheets was the preferred option, since a. Google solutions are used for general project documentation and communication, and b. All the partners were familiar with and had access to them. Additionally, the gathered material did not contain any sensitive data, so there was no need for a more sophisticated tool to ensure privacy and data integrity.

An initial template was supplied by UKK in the form of a Google Sheet, which served as a guide for the required information. The template was based on the project proposal desiderata and the relevant bibliography as summarily presented in the previous section.

The template is available in Table 1 of the Information Gathering and Mapping templates section. The partners were also supplied with detailed instructions and definitions where needed, as well as dropdown lists to enable them to use a consistent terminology when possible, with the capability to record more than one option for each source (e.g., sources that covered more than one disease). The instructions given

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

to the partners can be found in Table 3 of the Instructions for information gathering section.

A further meeting via Zoom was held on 10/10/25 to clarify the source collection process and get feedback from the AEQUITAS partners. Then, the information gathering phase was continued till mid-November 2025.

The following information was collected for each source:

- **Case added by**: the abbreviation of the name of each partner as designated in the project proposal.
- **Source**: full citation of the source in an acknowledged academic style (preferably APA), in the case of scholarly publications, or title and URL in other cases. The following sources for gender and race biases in medical AI applications were suggested:
    - **Scholarly publications**: academic papers and preprints that could be found using Google Scholar, PubMed, Scopus, Web of Science, or generative AI applications as search tools (verified to avoid fabricated, factually inaccurate answers generated by LLMs, commonly known as “hallucinations.”).
    - **Legal cases** that the partners might have access to, especially in their own mother tongue. The data should not contain sensitive information, effectively limiting the contributions on cases that were made public.
    - **Blogs and news** about gender/race bias, mainly as a starting point for more reliable information, such as scholarly publications.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- **Stakeholder Groups** (e.g., Patient Groups) and Civil Society Organisations that the AEQUITAS partners have access to. This could provide us with reports, registered incidents, or other similar material.
- **Repositories of AI bias cases**, for example, the **AI Incident Database** (https://incidentdatabase.ai/), which indexes the collective history of harms or near harms realized by the deployment of AI systems.
- **Any other source of information** they may have access to.

- **Disease name**: the three diseases relevant to the project proposal. A dropdown list with options Cardiovascular, Diabetes, Depression, and Not Sure/Other. For each disease, a definition according to the WHO was also provided. The Not Sure/Other option was provided to accommodate cases where the disease was not clear (e.g., mental health).
- **Type of source**: dropdown list with the kinds of sources as presented above, namely: scholarly publication, legal case, webpage, repository, and other.
- **Application scope**: records the intended use of the AI system referenced in the source. A dropdown list with the following options was provided: Customised therapy, Diagnosis, Disease prevention, Disease self-management, Drug development, Health promotion, Imaging, Management and planning, Prediction, Prediction-based surveillance, Public health, Risk assessment, Surveillance, Treatment response, Other.
- **Type of bias**: dropdown list of bias types as presented in (Suresh & Guttag, 2021) and described in Bias in AI systems above, plus the "Other" option. Detailed definitions -according to (Suresh & Guttag, 2021)- and examples were provided for each bias type, so that it would be easy for the partners to identify the bias described in each source.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- **Short description**: Free text field, for the partners to provide some information about the source content.
- **Impact on**: Free text field to fill with either gender or race, or be more precise.
- **Comment**: Free text field to record issues or doubts regarding the source
- **Coverage**: EU, US, Global, etc.
- **Language:** The language the source was written in, e.g., English, Greek, Spanish, etc.

Year of Publishing: No sources before 2000 were accepted, as the AI domain is evolving rapidly. The information-gathering process took place from the beginning of October to mid-November 2025.

Information Gathering Assessment

By the end of the information-gathering period, the AEQUITAS partners had collected 114 sources, mostly scholarly publications but also master's theses, reports from a variety of organisations, and webpages (blogs and news pages).

Some entries were duplicates of the same source, so there are 100 unique sources. The geographical coverage of most sources was global (especially the review and introductory scholarly publications), followed by the USA and then the EU countries and the UK.

A Zoom meeting was held on 13/11/25 with InnoHive and CSD as the partners responsible for the WP2 tasks, to discuss the findings of the Information gathering assessment. The slides from this meeting are in the Slides of the Information gathering assessment meeting section.

The scholarly publications can be classified into the following:

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- Specific research reporting specific experiments or AI systems. May also include bias detection and/or mitigation strategies. Typically, these papers include information about the dataset(s) and model(s) used, the types of biases measured, the fairness and accuracy measures used, the mitigation strategies applied, and recommendations.

- Reviews of papers regarding specific diseases and/or types of bias. They may refer to race or gender or both, but they can also include other kinds of discrimination, for example, ageism or classism. May also include guidelines or frameworks for detecting or mitigating bias, lists of bias types, fairness measures, and mitigation strategies.

- Introductory papers on AI, Bias, Medical bias. They present the landscape of AI applications for one or more of the AEQUITAS focus diseases, explaining how these applications work and summarising the ethical and, sometimes, legal challenges (including bias and discrimination). They may also include lists of types of biases (either AI or cognitive), fairness measures, and mitigation strategies, as well as recommendations, guidelines, checklists, and frameworks.

- Papers that are tangential to our purposes. For example, they may refer to medical/cognitive biases, not AI; generic forms of diseases we are interested in, e.g., mental health, with a passing mention to depression; other diseases, not the ones we are interested in; and Medical products.

The non-scholarly sources refer to one or more of the following:

- Review of current status or specific instances of discrimination due to some kind of AI bias (e.g., news pages and blogs).

- Bias definitions and explanations, usually with examples.

- Legislation against discrimination by AI systems.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- Socioeconomic cause of (gender, race, or other) bias.
- They use social science methods, such as interviews or focus groups, and not technical methods to assess bias in medical AI systems.
- Recommendations, frameworks, guidelines (e.g., WHO) for bias detection or mitigation

One more Zoom meeting was held on 28/11/25 with the WP2 partners, with a specific focus on database implementation and collecting feedback, which was incorporated into the development of the mapping template.

Mapping template

After the information-gathering and source assessment phases were complete, a mapping template was developed. The template was based on the initial source-gathering template (see Appendix: Table 1 Information Gathering template), which was expanded to take into consideration the results of the source assessment presented in the previous section.

Specifically, to the initial fields of the sources gathering template (Case Added By, Source, Disease, Type of source, Application Scope, Type of Bias, Short description, Impact on, Comment, Coverage, Language, and Year of Publishing), the following were added:

- AI domain: the specific AI domain the source is about, for example, LLMs, Image processing, structured data (e.g., EHR), biosignals, etc. Medical AI applications can span a variety of AI subdomains depending on the data they manipulate or the techniques they use to make predictions. This field would help the end user categorise and retrieve the information they are looking for more easily.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- Dataset: This field records the specific dataset(s) that were used to train and evaluate the system. Specific kinds of bias (e.g., representation, measurement, and evaluation bias) are linked to the dataset(s), so it is useful for the user to have this piece of information, especially in the case of public datasets.

- AI Model: Records the model(s) used for the predictions. May provide information on correlations of specific models with kinds of bias or application domains. For example, LLMs can be correlated to gender bias.

- Mitigation technique: List of the mitigation techniques used or proposed in the source. While bias cannot be avoided, it can be diagnosed and mitigated. Quite a lot of sources present these techniques in the context of medical AI applications.

- Fairness measures: List of fairness measures used or suggested in the source. There are various approaches to measuring fairness (e.g., demographic parity, proportional parity, equalized odds, predictive rate parity, false positive rate parity, etc.). An algorithm may employ a variety of constraints to ensure that one or more definitions of fairness are satisfied.

- Accuracy Measures: List of the accuracy measures used in the source. Accuracy is how a model's performance is tested and measured before its release. It relates to how a model is evaluated (e.g., it may lead to evaluation bias depending on the benchmarks used). Additionally, specific AI models (e.g., LLMs) are known to produce inaccurate results, potentially leading to deployment bias.

- Model evaluation: The way the model was evaluated. Some sources use formal evaluation methods, such as benchmarks, while others use human evaluators, depending on the application. For example, the responses of an LLM used to diagnose depression might be evaluated by human experts, while a system that

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

uses images to predict cardiovascular disease might be tested against a benchmark. Both of them might lead to different kinds of bias.

- Definitions of: Some sources contain definitions of kinds of bias or fairness measures. This is usually the case with informative sources or reviews. It is useful to be able to retrieve this kind of information to provide the user with an educational resource.

- Recommendations, guidelines, or frameworks: Some of the sources contain lists of prescriptive rules for the avoidance or mitigation of bias and/or ensuring fair, robust, and responsible AI in Medicine. Users should be able to retrieve this kind of information, which is equally essential to all AEQUITAS target groups.

- Organisation issuing the guidelines or recommendations: Records which organisation issued the recommendations or guidelines (e.g., WHO).

- Informative or review: describes the kind of source, when it does not refer to a specific case or experiment. A helpful piece of information for users who wish to gain an overall view of the bias issues in the target diseases and would like to retrieve only sources that can provide it.

The full mapping template is presented in the Appendix Table 2 Mapping template.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Information Gathering and Mapping templates

Table 1 Information Gathering template

Case Added By	
Source	
Disease	
Type of source	
Application Scope	
Type of Bias	
Short description	
Impact on	
Comment	
Coverage (EU, US, Global)	
Language	
Year of Publishing (no sources before 2000)	

Table 2 Mapping template

Case Added By	
Source	
Disease	
Type of source	
Application Scope	
Type of Bias	
Short description	
Impact on	
Comment	
Coverage (EU, US, Global)	
Language	
Year of Publishing (no sources before 2000)	
AI domain	
Dataset	
AI Model	
Mitigation technique	
Fairness measures	
Accuracy Measures	
Model evaluation	
Contains definitions of	
Contains recommendations, guidelines, or frameworks	
Organisation issuing the guidelines or recommendations	
informative or review	

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Instructions for information gathering

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Table 3 Instructions for the information gathering task

Source	
Type of source	**Definition and examples**
Scholarly publication	Academic papers, preprints, technical reports usually found using Google Scholar, PubMed, Scopus, Web of Science, arXiv or other academic repositories
Legal case	Documents relating to legal proceedings
Webpage	Webpage or news about biomedical AI biases. These resources are not always reliable, but can be used to find more reliable sources like the ones above.
Stakeholder Group	Cases that have been reported to/by a stakeholder group, e.g. patient associations
Repository	There are a variety of repositories about AI incidents, some related to bias as well. An example is e.g., AI Incident Database https: //incidentdatabase.ai/
Other	Other kinds of sources not covered above. In case you use this option, please describe the source type in the Comment cell

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Disease	
Name	**Definition**
Diabetes	Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Insulin is a hormone that regulates blood glucose. Hyperglycaemia, also called raised blood glucose or raised blood sugar, is a common effect of uncontrolled diabetes and, over time, leads to serious damage to many of the body's systems, especially the nerves and blood vessels. More Info: https://www. who.int/news-room/fact-sheets/detail/diabetes
Depression	Depressive disorder (also known as depression) is a common mental disorder. It involves a depressed mood or loss of pleasure or interest in activities for long periods of time. More info: https://www.who.int/news-room/fact-sheets/detail/depression
Cardiovascular	Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels. They include: coronary heart disease – a disease of the blood vessels supplying the heart muscle; cerebrovascular disease – a disease of the blood vessels supplying the brain; peripheral arterial disease – a disease of blood vessels supplying the arms and legs; rheumatic heart disease – damage to the heart muscle and heart valves from rheumatic fever, caused by streptococcal bacteria; congenital heart disease – birth defects that affect the normal development and functioning of the heart caused by malformations of the heart structure from birth; and deep vein thrombosis and pulmonary embolism – blood clots in the leg veins, which can dislodge and move to the heart and lungs. More info: https://www.who. int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
Not Sure/Other	This option is used when you think the disease is relevant, but you do not feel confident using any of the above categories. In that case, please elaborate in the Comment cell.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Application Scope
Customized therapy
Diagnosis
Disease prevention
Disease self-management
Drug development
Health promotion
Imaging
Management and planning
Prediction
Prediction-based surveillance
Public health
Risk assessment
Surveillance
Treatment response
Other

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

BIAS		
Type of Bias	**Definitions according to (Suresh and Guttag 2021)**	**Example**
Historical Bias	Historical bias arises even if data is perfectly measured and sampled, if the world as it is or was leads to a model that produces harmful outcomes. Such a system, even if it reflects the world accurately, can still inflict harm on a population. Considerations of historical bias often involve evaluating the representational harm (such as reinforcing a stereotype) to a particular group.	Pain assessment according to gender and age, word embeddings reflect real-world biases about women and ethnic minorities, and an embedding model trained on data from a particular decade reflects the biases of that time
Representation bias	Representation bias occurs when the development sample underrepresents some part of the population, and subsequently fails to generalize well for a subset of the use population.	In skin cancer detection, many imaging datasets underrepresent certain demographic groups, causing machine learning models to train on images of primarily fair-skinned individuals leaving minorities behind.
Measurement bias	Measurement bias occurs when choosing, collecting, or computing features and labels to use in a prediction problem. Typically, a feature or label is a proxy (a concrete measurement) chosen to approximate some construct (an idea or concept) that is not directly encoded or observable. Proxies become problematic when they are poor reflections of the target construct and/or are generated differently across groups, which can occur when: (1) The proxy is an oversimplification of a more complex construct (2) The method of measurement varies across groups (3) The accuracy of measurement varies across groups	Using Health costs as a proxy for health needs (poor reflection of the target construct)
Aggregation bias	Aggregation bias arises when a one-size-fits-all model is used for data in which there are underlying groups or types of examples that should be considered differently. Underlying aggregation bias is an assumption that the mapping from inputs to labels is consistent across subsets of the data. In reality, this is often not the case. A particular dataset might represent people or	The use of a sole HbA1c point to diagnose Diabetes in all ethnic populations while its values are higher in Blacks, Asians, and Latinos without diabetes when compared to White persons

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

	groups with different backgrounds, cultures or norms, and a given variable can mean something quite different across them. Aggregation bias can lead to a model that is not optimal for any group, or a model that is fit to the dominant population (e.g., if there is also representation bias).	
Learning bias	Learning bias arises when modeling choices amplify performance disparities across different examples in the data. For example, an important modeling choice is the objective function that an ML algorithm learns to optimize during training. Typically, these functions encode some measure of accuracy on the task (e.g., crossentropy loss for classification problems or mean squared error for regression problems). However, issues can arise when prioritizing one objective (e.g., overall accuracy) damages another (e.g., disparate impact)	differential privacy (i.e., preventing them from inadvertently revealing excessive identifying information about the training examples during use) while improving privacy, reduces the influence of underrepresented data on the model, and subsequently leads to a model with worse performance on that data (as compared to a model without differentially private training).
Evaluation bias	Evaluation bias occurs when the benchmark data used for a particular task does not represent the use population. A model is optimized on its training data, but its quality is often measured on benchmarks. A misrepresentative benchmark encourages the development and deployment of models that perform well only on the subset of the data represented by the benchmark data. Evaluation bias ultimately arises because of a desire to quantitatively compare models against each other. Applying different models to a set of external datasets attempts to serve this purpose, but is often extended to make general statements about how good a model is.	Skin cancer detection benchmarks are usually composed of images of fair skinned individuals

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Deployment bias	Deployment bias arises when there is a mismatch between the problem a model is intended to solve and the way in which it is actually used. This often occurs when a system is built and evaluated as if it were fully autonomous, while in reality, it operates in a complicated sociotechnical system moderated by institutional structures and human decision-makers.	The COMPAS system is a model intended to predict a person's likelihood of committing a future crime. In practice, however, these tools may be used in "off-label" ways, such as to help determine the length of a sentence.
Other	In case of bias not described adequately by the above. If you use this option, please elaborate in the Comment cell. There are a lot of kinds of biases not covered in the above taxonomy, so we can expand it if necessary.	

Impact on	**Comment**
This is a free text field. You may just write: gender or race or be more precise	This is a free text field. You may include remarks or elaborate if you use the Other option in the previous fields

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Appendix 2: Meeting Slides

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

UKK Slides of the Kick-Off Meeting

UNIVERSITY OF COLOGNE

BI-K Institute for Biomedical Informatics University of Cologne

WP2: MAPPING OF GENDER AND RACIAL BIASES IN BIOMEDICAL AI AND DATABASE AND GUIDELINES DEVELOPMENT FOLLOWING THE PRINCIPLES OF THE EU CHARTER FOR THE PROTECTION OF FUNDAMENTAL RIGHTS

Maria Christoforaki

Aequitas Kick-Off Meeting

25.9.2025

Institute for Biomedical Informatics / Medical Faculty / Maria Christoforaki

01

WP DESCRIPTION

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Duration and Objectives

- Duration: M1-M10
- Objectives:
- To organise a methodical, evidence-based mapping of gender and racial biases in biomedical AI;
- To compile the acquired results and present them in the form of a consolidated report with policy recommendations, testing the principles for future policies;
- To develop the AEQUITAS database following principles of emerging Knowledge Graph paradigms to allow for increased interoperability and connectedness with other research data and information infrastructures;
- To develop an AI regulatory model to be utilized by healthcare staff and CSOs

Tasks

Task	Task Name	Description	Participants
T2.1	Mapping of the gender and racial biases in biomedical AI	• Make the preparations for the mapping activity: the division of tasks among partners, the approach of the target group, as well as the approach that partners will follow in order to achieve the extraction of quality and valuable data and results. • Develop the necessary templates for the mapping, which will entail the following activities: ◦ Categories of biomedical AI (e.g. Predictive diagnosis, drug development, treatment planning) ◦ Sources of bias (data representation, development of algorithms, clinical applications) ◦ Gender and racial dimensions of AI and impact analysis • The partners will examine the biases that are presented in the data collection and algorithm development.	• UOC • ALL
T2.2	Compilation of results and development of the Report	After the partners have conducted the mapping of the gender and racial biases in biomedical AI related to cardiovascular disease, depression and diabetes, they will proceed with the compilation of results UOC, as the partner with expertise in research in biomedical AI, will be responsible for the compilation of results and the development of the Report. After it is finalized, the project partners will continue with the translation of the Report in their national languages (EL/DE/BG/LT/ES/IT/PT) . Development of the Bias Report M6	• UOC • ALL
T2.3	Development of the AEQUITAS database	Provide feedback via questionnaires	• InnoHive • ALL
T2.4	Development of the AI regulatory model	Project Partners will develop the model, translate it and provide feedback by filling out structured questionnaires	• CSD • ALL

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Milestones

Milestone No (continuous numbering not linked to WP)	Milestone Name	Work Package No	Lead Beneficiary	Description	Due Date (month number)	Means of Verification
MS3	Development of the Bias Report	2	UOC	The Report will contain the necessary information and data that was gathered through the mapping activity on the different gender and racial biases distinguished at different levels of AI development and application.	M4->M6	Research on data regarding the diseases that concern the project (cardiovascular, depression and diabetes).
MS4	Technical development of the database	2	InnoHive	The project partners will develop the database that will entail the gender and racial biases that have been reported in biomedical AI. This database will act as a guiding tool for healthcare staff and civil society organizations in order to help minimize these societal biases that could lead to misdiagnosis and mistreatment.	M8	Structured questionnaires by the project partners and the NSAGs.
MS5	Structure of the Model	2	CSD	The partners will take into account the findings of the mapping activity in order to develop the Model following the principles protected by the EU Charter of fundamental rights. The Model will encompass biomedical AI rules of conduct in correlation to the protection of the EU fundamental rights.	M10	Collection of feedback from the partners and the NSAGs.

UNIVERSITY OF COLOGNE Institute for Biomedical Informatics / Medical Faculty / Maria Christoforaki 25.9.2025 5

Deliverables

Deliverable No (continuous numbering linked to WP)	Deliverable Name	Work Package No	Lead Beneficiary	Type	Dissemination Level	Due Date (month number)	Description (including format and language)
D2.1	Biases Report	2	UOC	R	PU	M4->M6	Electronic, EN/EL/DE/BG/LT/ES/IT/PT, approx. 100-120 pages
D2.2	AEQUITAS Database	2	InnoHive	DATA	PU	M8	Database, electronic, available in English
D2.3	AEQUITAS AI Regulatory Model	2	CSD	OTHER	PU	M10	Electronic, EN/EL/DE/BG/LT/ES/IT/PT, approx. 50 pages

UNIVERSITY OF COLOGNE Institute for Biomedical Informatics / Medical Faculty / Maria Christoforaki 25.9.2025 6

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

02

THE USE OF ML/AI IN MEDICINE

AI Applications for Health

- Health care
  - Diagnosis and prediction-based diagnosis
  - Clinical care
- In health research and drug development
- In health systems management and planning
- In public health and public health surveillance
  - Health promotion
  - Disease prevention
  - Surveillance
    - prediction-based surveillance
    - emergency preparedness
  - Outbreak response

(World Health Organization 2021)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

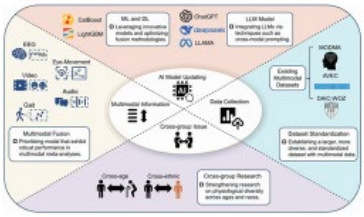
Co-funded by the European Union

AEQUITAS

Biomedical AI related to cardiovascular disease, depression and diabetes

Schematic representation of applications of ai in cardiovascular disease

Schematic representation of applications of ai in diabetes



Outlook for an AI multi-modal physiological and behavioural information integration for the screening of depression.

Images from (Naskar et al. 2025)

Image from (Wang et al. 2025)

03

BIAS IN AI/ML

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

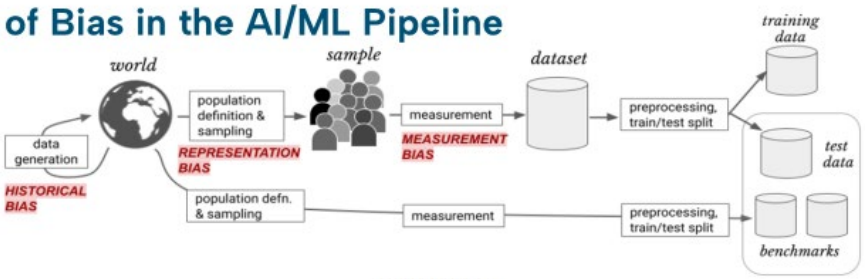
AEQUITAS

What is bias

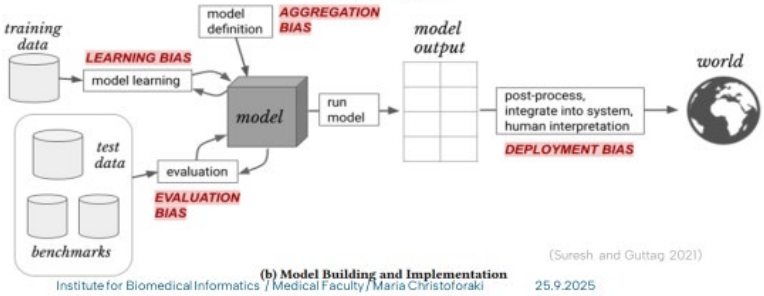
- We use the term bias to refer to computer systems that **systematically** and **unfairly** discriminate against certain individuals or groups of individuals in favor of others. A system discriminates **unfairly** if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate
- Two points follow:
  - First, unfair discrimination alone does not give rise to bias unless it occurs systematically
  - Second, systematic discrimination does not establish bias unless it is joined with an unfair outcome

(Friedman and Nissenbaum 1996)

Kinds of Bias in the AI/ML Pipeline



(a) Data Generation



(Suresh and Guttag 2021)

(b) Model Building and Implementation

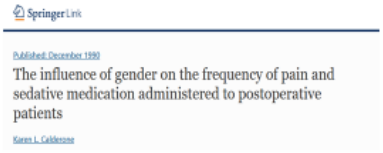
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Historical Bias

- Incorporating already existing prejudices and stereotypes in the data
- Example: Pain assessment
- Male patients were administered pain medication significantly more frequently than female patients
- Female patients were administered sedative medication significantly more frequently than male patients
- Patients 61 years or younger received pain medication significantly more frequently than those patients 62 years and over


SpringerLink

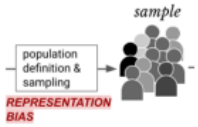
Published: December 1990

The influence of gender on the frequency of pain and sedative medication administered to postoperative patients

Karen L. Calderone

https://link.springer.com/article/10.1007/BF00289259

Representation Bias



- When defining the target population, if it does not reflect the use population
- When defining the target population, if it contains underrepresented groups
- When sampling from the target population, if the sampling method is limited or uneven

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Representation Bias Example: Skin Cancer Detection

- Many imaging datasets underrepresent certain demographic groups, causing machine learning models to train on images of primarily fair-skinned individuals

Source of training images	No. of studies†
Most common countries (>3 studies)	
United States	16
Italy	11
Austria	10
Greece	8
United Kingdom	7
Germany	5
South Korea	5
Australia	4
France	4
Netherlands	4
Image repositories	
ISIC (predominantly United States, Austria, Australia, and Spain)	46
Interactive Atlas of Dermoscopy (Italy and Austria)	18
PH² (Portugal)	11
HAM10000* (Austria and Australia)	6
MED-NODE (Netherlands)	5
Dermofit Image Library (United Kingdom)	2

HAM10000, Human Against Machine with 10000 training images; *ISIC*, International Skin Imaging Collaboration.
*HAM10000 is also included in the International Skin Imaging Collaboration archive of images.
†Studies were counted multiple times if training images were derived from multiple countries or image sources.

Table from (Guo et al. 2021)

Measurement Bias

- Occurs when choosing, collecting, or computing features and labels used in a prediction problem as proxies for some construct that is not directly encoded or observable
  - The proxy is an oversimplification of a more complex construct.
- The method of measurement varies across groups
- The accuracy of measurement varies across groups

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Example:
Racial bias in population health management algorithm

- Optum health services company algorithm used health costs to predict and rank which patients would benefit most from extra care
- Health costs as a proxy for health needs is biased
  - black patients are less affluent and
  - spend less on health care than whites
- The algorithm concluded that blacks were healthier than equally sick white patients

Number of chronic illnesses versus algorithm-predicted risk by race

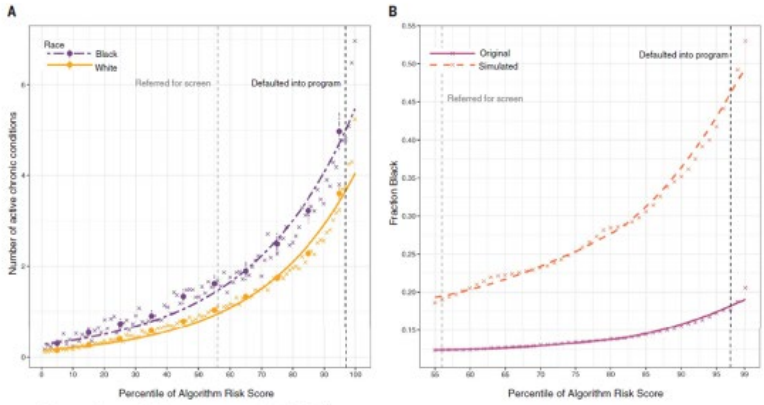


Figure from (Obermeyer, et al. 2019)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Costs versus algorithm–predicted risk and costs versus health by race

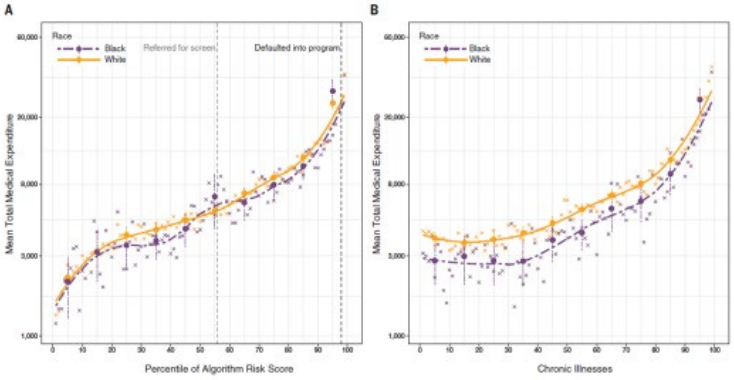
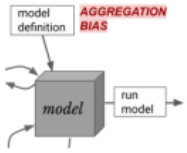


Figure from (Obermeyer, et al. 2019)

Aggregation Bias



- Underlying aggregation bias is an assumption that the mapping from inputs to labels is consistent across subsets of the data
- This is may vary if the dataset represents people or groups with different backgrounds, cultures or norms
- Example:
  - HbA1c has been considered the reference test for the assessment of glycaemic control in individuals with diabetes mellitus (DM)
  - HbA1c values are higher in Blacks, Asians, and Latinos without DM when compared to White persons
  - Might have an impact on the use of a single HbA1c point to diagnose DM in all ethnic populations.

Effect of ethnicity on HbA1c levels in individuals without diabetes: Systematic review and meta-analysis

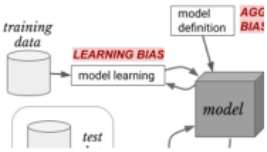
Gabriela Cavagnolli, Ana Laura Pimentel, Priscila Aparecida Correa Freitas, Jorge Luiz Gross, Joíza Lins Camargo

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

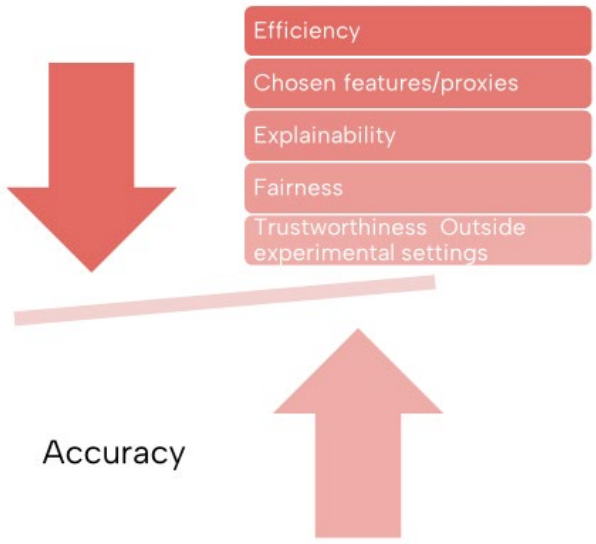
Co-funded by the European Union

AEQUITAS

Learning Bias



- Arises when modelling choices amplify performance disparities across different examples in the data
- Model design choices are made to maximize test-set accuracy by means of an objective function that an algorithm learns to optimize during training
- Problems arise when we prioritize one objective, such as test-set accuracy, over others
- Example: Differential privacy reduces the influence of underrepresented data on the model

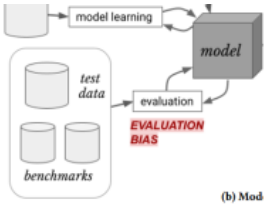


Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Evaluation Bias



(b) Mod

- Occurs when the benchmark data used for a particular task does not represent the use population
- A model is optimized on its training data, but its quality is often measured on benchmarks
- We use benchmarks so we can compare one model to another, i.e., to identify the respective strengths and weaknesses of a given methodology in contrast with others
- This evaluation and comparison is done by regarding their ability to learn patterns in 'benchmark' datasets that have been applied as 'standards'

Example: Skin Cancer Detection

- Many imaging datasets underrepresent certain demographic groups, causing machine learning models to train on images of primarily fair-skinned individuals

Source of training images	No. of studies[†]
Most common countries (>3 studies)	
United States	16
Italy	11
Austria	10
Greece	8
United Kingdom	7
Germany	5
South Korea	5
Australia	4
France	4
Netherlands	4
Image repositories	
ISIC (predominantly United States, Austria, Australia, and Spain)	46
Interactive Atlas of Dermoscopy (Italy and Austria)	18
PH^2 (Portugal)	11
HAM10000* (Austria and Australia)	6
MED-NODE (Netherlands)	5
Dermofit Image Library (United Kingdom)	2

HAM10000, Human Against Machine with 10000 training images; *ISIC*, International Skin Imaging Collaboration.
*HAM10000 is also included in the International Skin Imaging Collaboration archive of images.
[†]Studies were counted multiple times if training images were derived from multiple countries or image sources.

Table from (Guo et al. 2021)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Deployment Bias

- Arises when there is a mismatch between the problem a model is intended to solve and the way in which it is actually used
- Can cause harm especially when combined with cognitive biases like
  - confirmation bias
  - automation bias

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

04

IMPACT OF BIAS

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Gender Bias Impact Example: Clinical Trials

- In Clinical trials, due to technical and bioethical considerations, such as
  - attempt to reduce the impact of the estrous cycle in experimental studies
  - protective policies for women of childbearing age in clinical research
- Result:
  - Some of the treatments that currently exist for several diseases are not adequately evaluated in women who are likely to be underrepresented in clinical trials
  - Women typically report more adverse event reactions compared with men

(Cirillo et al., 2020)

05

NEXT STEPS

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Timeline for UOC Coordinated activities

	Oct 2025	Nov 2025	Dec 2025	Jan 2026	15 Feb 2026–28 Feb 2026
UOC	Bias Mapping template			Compiling the results	Translation in EL/DE/BG/LT/ES/IT/PT
Partners	Phase I: gathering information on the gender and racial biases	Phase II: assessment and gathering information on the gender and racial biases	Mapping of the gender and racial biases in biomedical AI related cardiovascular disease, depression and diabetes according to the template		

Potential Sources of Cases and Info Gathering File

- Scholarly publications (Google Scholar, PubMed, Scopus, Web of Science)
- Legal cases
- Blogs and news as a starting point for more reliable information
- Stakeholder Groups (e.g., Patient Groups) and Civil Society Organisations
- Repositories, e.g. **AI Incident Database** https://incidentdatabase.ai/
- Other…
- Information Gathering file: https://docs.google.com/spreadsheets/d/1ik5q3gnF5H-n7-gb5MaXKUV38HgEFLVgMneUYBab89U/edit?usp=sharing
- @All partners: please send @maria.christoforaki@uk-koeln.de one contact person for WP2, or put the name in the respective cell of **Instructions and Definitions!Contact Person(s)** Google Sheet above

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

References

- Cavagnolli, Gabriela, Ana Laura Pimentel, Priscila Aparecida Correa Freitas, Jorge Luiz Gross, and Joiza Lins Camargo. 'Effect of Ethnicity on HbAlc Levels in Individuals without Diabetes: Systematic Review and Meta-Analysis'. *PLOS ONE* 12, no. 2 (2017): e0171315. https://doi.org/10.1371/journal.pone.0171315.
- Cirillo, Davide, Silvina Catuara-Solarz, Czuee Morey, et al. 'Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare'. *Npj Digital Medicine* 3, no. 1 (2020): 81. https://doi.org/10.1038/s41746-020-0288-5.
- Friedman, Batya, and Helen Nissenbaum. 'Bias in Computer Systems'. ACM Transactions on Information Systems 14, no. 3 (1 July 1996): 330–47. https://doi.org/10.1145/230538.230561.
- Guo, Lisa N., Michelle S. Lee, Bina Kassamali, Carol Mita, and Vinod E. Nambudiri. 'Bias in, Bias out: Underreporting and Underrepresentation of Diverse Skin Types in Machine Learning Research for Skin Cancer Detection-A Scoping Review'. *Journal of the American Academy of Dermatology*, 10 July 2021, S0190-9622(21)02086-7. https://doi.org/10.1016/j.jaad.2021.06.884.
- Naskar, Sweet, Suraj Sharma, Ketousetuo Kuotsu, et al. 'The Biomedical Applications of Artificial Intelligence: An Overview of Decades of Research'. *Journal of Drug Targeting* 33, no. 5 (2025): 717–48. https://doi.org/10.1080/1061186X.2024.2448711.
- Suresh, Harini, and John Guttag. 'A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle'. In Equity and Access in Algorithms, Mechanisms, and Optimization, 1–9. EAAMO '21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3465416.3483305.
- Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations'. Science 366, no. 6464 (25 October 2019): 447–53. https://doi.org/10.1126/science.aax2342.
- Wang, Luyao, Chenhan Wang, Chenyang Li, et al. 'AI-Assisted Multi-Modal Information for the Screening of Depression: A Systematic Review and Meta-Analysis'. *Npj Digital Medicine* 8, no. 1 (2025): 523. https://doi.org/10.1038/s41746-025-01933-3.
- World Health Organization. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. World Health Organization, 2021. https://www.who.int/publications/i/item/9789240029200.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Slides of the Information gathering assessment meeting

AEQUITAS

Funded by the European Union

Information Gathering Assessment

Maria Christoforaki
WP2 Meeting 13.11.25

AEQUITAS

Funded by the European Union

Table of contents

1. Description of Sources
2. Way to go
3. Next Steps

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Commission-EU. Neither the European Union nor the European Commission can be held responsible for them Project code: 101191047 — 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Source Description

- Assessment of material gathered till 31.10
- Mostly Scholarly Publications
  - Journal, Conference, and preprints
  - A few duplicates
- Master's thesis
- Reports
  - Variety of organizations, e.g., WHO
- Webpages
  - Blogs and news

Content of Scholarly Publications

- Specific research reporting
- Reviews
- Guidelines
- Introductory papers on AI, Bias, Medical bias
- Some refer to
  - medical/cognitive biases, not AI
  - Generic form of diseases we are interested in, e.g., mental health, with a passing mention to depression
  - Other diseases, not the ones we are interested in
  - Medical products

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

• Specific research reporting

- Dataset
- Model
- Bias and fairness measurements
- Accuracy measurements
- Mitigation
- Recommendations

• Other material content

- Review of current status (useful for snowballing)
- Bias definition
- Legislation
- Socioeconomic cause of bias
- Social sciences methods (interviews, focus groups)
- Recommendations, frameworks, guidelines

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

• What are we interested in mapping?

1. What is already in the information gathering template
2. Any of the features presented previously
3. Are we going to use standards or ontologies?
4. What would be interesting to our intended users?
5. Should we expand/narrow our material scope?

• Next Steps

1. Create a mapping template
2. Map the sources
3. Write the deliverable draft
4. Collect feedback, create a second version, and translate

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Appendix 3: Collected sources

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

No	Source
1	A. Skvortsova, S. H. Meeuwis, R. C. Vos, H. M. M. Vos, H. van Middendorp, D. S. Veldhuijzen, A. W. M. Evers (2023), Implicit gender bias in the diagnosis and treatment of type 2 diabetes: A randomized online study, https://onlinelibrary.wiley.com/doi/10.1111/dme.15087
2	Abd-Alrazaq, A., AlSaad, R., Shuweihdi, F., Ahmed, A., Aziz, S., & Sheikh, J. (2023). Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. Npj Digital Medicine, 6(1), 1–16. https://doi.org/10.1038/s41746-023-00828-5
3	Achtari, M., Salihu, A., Muller, O., Abbé, E., Clair, C., Schwarz, J., & Fournier, S. (2024). Gender Bias in AI's Perception of Cardiovascular Risk. Journal of Medical Internet Research, 26:e54242. https://www.jmir.org/2024/1/e54242
4	Adedinsewo, Demilade A., Amy W. Pollak, Sabrina D. Phillips, et al. 'Cardiovascular Disease Screening in Women: Leveraging Artificial Intelligence and Digital Tools'. Circulation Research 130, no. 4 (2022): 673–90. https://doi.org/10.1161/CIRCRESAHA.121.319876.
5	al Hamid, A., Beckett, R., Wilson, M., Jalal, Z., Cheema, E., Al-Jumeily OBE, D., Coombs, T., Ralebitso-Senior, K., & Assi, S. (2024). Gender Bias in Diagnosis, Prevention, and Treatment of Cardiovascular Diseases: A Systematic Review. Cureus. https://doi.org/10.7759/cureus.54264
6	Alday, E.A.P, Rad, A.B, Reyna, M.A, Sadr, N, Gu, A, Li, Q., Dumitru, M., Xue, J., Albert, D., Sameni, R., Clifford, G.D. Age, sex and race bias in automated arrhythmia detectors, Journal of Electrocardiology, 2022 Sep-Oct:74:5-9. DOI: https://doi.org/10.1016/j.jelectrocard.2024.01.006
7	Alipour, M., & Alipour, A. (2025). Algorithmic Bias in AI-Based Diabetes Care: Systematic Review of Model Performance, Equity Reporting, and Physiological Label Bias. *InfoScience Trends, 2*(5), 33–46. https://doi.org/10.61186/ist.202502.05.04
8	Amaya-Santos, S., Jiménez-Pernett, J., & Bermudez-Tamayo, C. (2024). Health for whom? Intersectionality and biases in the use of artificial intelligence in clinical diagnosis. *Anales Del Sistema Sanitario de Navarra, 47*(2). https://doi.org/10.23938/ASSN.1077
9	American Heart Association. (2024, March 12). Race, Racism, and Risk Prediction for Cardiovascular Disease. https://newsroom.heart.org/news/race-racism-and-risk-prediction-for-cardiovascular-disease
10	Amponsah D, Thamman R, Brandt E, James C, Spector-Bagdady K, Yong CM. Artificial Intelligence to Promote Racial and Ethnic Cardiovascular Health Equity. Curr Cardiovasc Risk Rep. 2024 Nov;18(11):153-162. doi: 10.1007/s12170-024-00745-6. Epub 2024 Aug 20. PMID: 40144330; PMCID: PMC11938301, https://pmc.ncbi.nlm.nih.gov/articles/PMC11938301/

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

11	Barnty, B., Joseph, O. & Ok, E. (2025). Bias and Fairness in AI-Based Mental Health Models. *Ladoke Akintola University of Technology.* https://www.researchgate.net/publication/389214235_Bias_and_Fairness_in_AI-Based_Mental_Health_Models
12	Bhimavarapu, U. (2025). Bias in AI-Driven Diabetes Prediction Models: Challenges, Impacts, and Mitigation Strategies. In Artificial Intelligence (AI) in Healthcare: A Comprehensive Overview (pp. 195–214). IGI Global Scientific Publishing. https://www.igi-global.com/chapter/bias-in-ai-driven-diabetes-prediction-models/376753
13	Boothroyd, G. (2025, March 24). Minding the Gaps : Neuroethics, AI, and Depression. Non-profit Quarterly. https://nonprofitquarterly.org/minding-the-gaps-neuroethics-ai-and-depression/
14	Bouguettaya, A., Stuart, E. M., & Aboujaoude, E. (2025). Racial bias in AI-mediated psychiatric diagnosis and treatment: a qualitative comparison of four large language models. *Npj Digital Medicine, 8*(1), 332. https://doi.org/10.1038/s41746-025-01746-4
15	Burlina, P., Joshi, N., Paul, W., Pacheco, K.D., Bressler, N.M. Addressing Artificial Intelligence Bias in Retinal Diagnostics. Translational Vision Science & Technology February 2021, Vol.10, 13. Addressing Artificial Intelligence Bias in Retinal Diagnostics. https://doi.org/10.1167/tvst.10.2.13
16	Buslón, N., Cirillo, D., Rios, O., & Perera del Rosario, S. (2025). Exploring Gender Bias in AI for Personalized Medicine: Focus Group Study With Trans Community Members. *Journal of Medical Internet Research, 27*, e72325–e72325. https://doi.org/10.2196/72325
17	Cau, R., Pisu, F., Suri, J. S., & Saba, L. (2024). Addressing hidden risks: Systematic review of artificial intelligence biases across racial and ethnic groups in cardiovascular diseases. European Journal of Radiology, 111867. https://doi.org/10.1016/j.ejrad.2024.111867
18	Cavagnolli, Gabriela, Ana Laura Pimentel, Priscila Aparecida Correa Freitas, Jorge Luiz Gross, and Joíza Lins Camargo. 'Effect of Ethnicity on HbA1c Levels in Individuals without Diabetes: Systematic Review and Meta-Analysis'. PLOS ONE 12, no. 2 (2017): e0171315. https://doi.org/10.1371/journal.pone.0171315
19	Center, C.-S. M. (2025, June 30). Cedars-Sinai Study Shows Racial Bias in AI-Generated Treatment Regimens for Psychiatric Patients. Cedars-Sinai Study Shows Racial Bias in AI-Generated Treatment Regimens for Psychiatric Patients ; Cedars-Sinai Medical Center. https://www.cedars-sinai.org/newsroom/cedars-sinai-study-shows-racial-bias-in-ai-generated-treatment-regimens-for-psychiatric-patients/

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

20	Chin, M. H., Afsar-Manesh, N., Bierman, A. S., Chang, C., Colón-Rodríguez, C. J., Dullabh, P., Duran, D. G., Fair, M., Hernandez-Boussard, T., Hightower, M., Jain, A., Jordan, W. B., Konya, S., Moore, R. H., Moore, T. T., Rodriguez, R., Shaheen, G., Snyder, L. P., Srinivasan, M., … Ohno-Machado, L. (2023). Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care. JAMA Network Open, 6(12), e2345050. https://doi.org/10.1001/jamanetworkopen.2023.45050
21	Chrissos, D. (2023). Will Artificial Intelligence and ChatGPT Replace the Clinical Doctor? Hellenic Journal of Cardiology / ΕΚΕ Magazine, 64(4), 268-276. Athens: Hellenic Society of Cardiology. https://www.hcs.gr/wp-content/uploads/2024/03/11.-%CE%9C%CF%80%CE%BF%CF%81%CE%BF%CF%8D%CE%BD-%CE%B7-%CE%A4%CE%B5%CF%87%CE%BD%CE%B7%CF%84%CE%AE-%CE%9D%CE%BF%CE%B7%CE%BC%CE%BF%CF%83%CF%8D%CE%BD%CE%B7-%CE%BA%CE%B1%CE%B9-%CF%84%CE%BF-ChatGPT-%CE%BD%CE%B1-%CF%85%CF%80%CE%BF%CE%BA%CE%B1%CF%84%CE%B1%CF%83%CF%84%CE%AE%CF%83%CE%BF%CF%85%CE%BD-%CF%84%CE%BF%CE%BD-%CE%BA%CE%BB%CE%B9%CE%BD%CE%B9%CE%BA%CF%8C-%CE%B9%CE%B1%CF%84%CF%81%CF%8C-1.pdf
22	Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Rementeria, M. J., Chadha, A. S., & Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. Npj Digital Medicine, 3(1), 81. https://doi.org/10.1038/s41746-020-0288-5
23	Cronjé HT, Katsiferis A, Elsenburg LK, Andersen TO, Rod NH, Nguyen T-L, et al. (2023) Assessing racial bias in type 2 diabetes risk prediction algorithms. PLOS Glob Public Health 3(5): e0001556. https://doi.org/10.1371/journal.pgph.0001556
24	Cross, J. L., Choma, M. A., & Onofrey, J. A. (2024). Bias in medical AI: Implications for clinical decision-making. PLOS digital health, 3(11), e0000651. https://doi.org/10.1371/journal.pdig.0000651
25	Cruz, A. (2020). Fairness-Aware Hyperparameter Optimization. Repositorio-aberto.up.pt. https://repositorio-aberto.up.pt/handle/10216/128959
26	De Angel V., Lewis S., White K., Oetzmann C., Leightley D., Oprea E., Lavelle G., Matcham F., Pace A., Mohr D., Dobson R., Hotopf M. (2022). Digital health tools for the passive monitoring of depression: a systematic review of methods. DOI: 10.1038/s41746-021-00548-8
27	Department of health and social care, Independent report: Equity in medical devices: independent review - summary report, Published 11 March 2024, https://www.gov.uk/government/publications/equity-in-medical-devices-independent-review-final-report/equity-in-medical-devices-independent-review-summary-report

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

28	Duffy, G., Clarke, S.L., Christensen, M., He, B., Yuan, N., Cheng, S., & Ouyang, D. (2022). Deep Learning Discovery of Demographic Biomarkers in Echocardiography https://arxiv.org/abs/2207.06421
29	ECO. (2024, September 18). Artificial intelligence: a revolution at the service of health. https://eco.sapo.pt/2024/09/18/inteligencia-artificial-uma-revolucao-ao-servico-da-saude/
30	EPRS | European Parliamentary Research Service, Scientific Foresight Unit (STOA), PE 729.512 (June 2022). Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts, https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU%282022%29729512#:~:text=4,and%20ethical%20and%20societal%20impacts
31	Food and Drug Administration. (2025, January 06). FDA proposes updated recommendations to help improve performance of pulse oximeters across skin tones. https://www.fda.gov/news-events/press-announcements/fda-proposes-updated-recommendations-help-improve-performance-pulse-oximeters-across-skin-tones
32	Gardiner, H. & Mutebi, N. (2025, January 31). AI and mental healthcare: ethical and regulatory considerations. (POSTnote 738). UK Parliament Post. https://researchbriefings.files.parliament.uk/documents/POST-PN-0738/POST-PN-0738.pdf
33	Gierschmann, L. (2024). The Role of Gender: Gender Fairness in the Detection of Depression Symptoms on Social Media. Studenttheses.uu.nl. https://studenttheses.uu.nl/handle/20.500.12932/47734
34	Hong, C., Pencina, M. J., Wojdyla, D. M., Hall, J. L., Judd, S. E., Cary, M., Engelhard, M. M., Berchuck, S., Xian, Y., D'Agostino, R., Howard, G., Kissela, B., & Henao, R. (2023). Predictive Accuracy of Stroke Risk Prediction Models Across Black and White Race, Sex, and Age Groups. JAMA, 329(4), 306. https://doi.org/10.1001/jama.2022.24683
35	https://www.bhf.org.uk/what-we-do/news-from-the-bhf/news-archive/2024/july/ai-reduces-underdiagnosis-of-common-heart-failure-in-black-patients-new-research-finds , original paper: Wu, J., Biswas, D., Brown, S., Ryan, M., Bernstein, B. S., Tam To, B., ... & O'Gallagher, K. (2025). Artificial intelligence methods to detect heart failure with preserved ejection fraction within electronic health records: an equitable disease detection model. European Heart Journal-Digital Health, ztaf107. https://academic.oup.com/ehjdh/advance-article/doi/10.1093/ehjdh/ztaf107/8256371
36	Hussain, S. A., Bresnahan, M., & Zhuang, J. (2025). The bias algorithm: how AI in healthcare exacerbates ethnic and racial disparities – a scoping review. *Ethnicity & Health, 30*(2), 197–214. https://doi.org/10.1080/13557858.2024.2422848

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

37	Ive, J., Bondaronek, P., Yadav, V., Santel, D., Glauser, T., Cheng, T., Strawn, J. R., Agasthya, G., Tschida, J., Choo, S., Chandrashekar, M., Kapadia, A. J., & Pestian, J. (2024). A Data-Centric Approach to Detecting and Mitigating Demographic Bias in Pediatric Mental Health Text: A Case Study in Anxiety Detection. ArXiv.org. https://arxiv.org/abs/2501.00129
38	Jerlyn Q.H. Ho, Andree Hartanto, Andrew Koh, Nadyanna M. Majeed, Gender biases within Artificial Intelligence and ChatGPT: Evidence, Sources of Biases and Solutions, Computers in Human Behavior: Artificial Humans, Volume 4, 2025, 100145, ISSN 2949-8821, https://www.sciencedirect.com/science/article/pii/S2949882125000295?via%3Dihub
39	Junias, O., Kini, P., & Chaspari, T. (2025). Assessing Algorithmic Bias in Language-Based Depression Detection: A Comparison of DNN and LLM Approaches. ArXiv.org. https://arxiv.org/abs/2509.25795
40	Kamulegeya, L. H., Okello, M., Bwanika, J. M., Musinguzi, D., Lubega, W., Rusoke, D., Nassiwa, F., & Börve, A. (2019). Using artificial intelligence on dermatology conditions in Uganda: A case for diversity in training data sets for machine learning. https://doi.org/10.1101/826057
41	Kaur, D., Hughes, J. W., Rogers, A. J., Kang, G., Narayan, S. M., Ashley, E. A., & Perez, M. v. (2024). Race, Sex, and Age Disparities in the Performance of ECG Deep Learning Models Predicting Heart Failure. Circulation: Heart Failure, 17(1). https://doi.org/10.1161/CIRCHEARTFAILURE.123.010879
42	Kim, J.-W., Yoon, H., Oh, W., Jung, D., Yoon, S.-H., Kim, D.-J., Lee, D.-H., Lee, S.-Y., & Yang, C.-M. (2025). Domain Adversarial Training for Mitigating Gender Bias in Speech-based Mental Health Detection. https://arxiv.org/abs/2505.03359
43	Langlais, É. L., Dupont, J., Martin, A., & Nguyen, P. (2022). Novel artificial intelligence applications in cardiology: Current landscape, limitations, and the road to real-world applications. Journal of Cardiovascular Translational Research, 16(3), 513–525. https://doi.org/10.1007/s12265-022-10260-x
44	Lee, T., Puyol-Antón, E., Ruijsink, B., Shi, M., & King, A. P. (2022). *A Systematic Study of Race and Sex Bias in CNN-Based Cardiac MR Segmentation* (pp. 233–244). https://doi.org/10.1007/978-3-031-23443-9_22
45	Li, F., Wu, P., Ong, H. H., Peterson, J. F., Wei, W.-Q., & Zhao, J. (2023). Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction. Journal of Biomedical Informatics, 138, 104294. https://doi.org/10.1016/j.jbi.2023.104294
46	Li, Z., Keel, S., Liu, C., He, Y., Meng, W., Scheetz, J., Lee, P. Y., Shaw, J., Ting, D., Wong, T. Y., Taylor, H., Chang, R., & He, M. (2018). An Automated Grading System for Detection of Vision-Threatening Referable Diabetic Retinopathy on the Basis of Color Fundus Photographs. Diabetes Care, 41(12), 2509–2516. https://doi.org/10.2337/dc18-0147

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

47	Margaux Achtari, Adil Salihu, Olivier Muller, Emmanuel Abbé, Carole Clair, Joëlle Schwarz , Stephane Fournier, J Med Internet Res. (2024), Gender Bias in AI's Perception of Cardiovascular Risk, https://pmc.ncbi.nlm.nih.gov/articles/PMC11538872/
48	Matos, J. (2023). Research Frameworks towards Health Equity. Repositorio-aberto.up.pt. https://repositorio-aberto.up.pt/handle/10216/150994
49	Medicines&Healthcare Products Regulatory Agency, Digital Mental Health Technology - Regulation and Evaluation for Safe and Effective Products: Device characterisation, regulatory qualification and classification - Version 1.2, https://assets.publishing.service.gov.uk/media/6866572fadfe29730ea3a9d5/MHRA_guidance_on_DMHT_-_Device_characterisation_regulatory_qualification_and_classification.pdf
50	Mihan, A., Pandey, A., & van Spall, H. G. (2024). Mitigating the risk of artificial intelligence bias in cardiovascular care. The Lancet Digital Health, 6(10), e749–e754. https://doi.org/10.1016/S2589-7500(24)00155-9
51	Mihan, A., Pandey, A., & van Spall, H. G. C. (2024). Artificial intelligence bias in the prediction and detection of cardiovascular disease. Npj Cardiovascular Health, 1(1), 31. https://doi.org/10.1038/s44325-024-00031-9
52	Miller, R. J., Zhang, Y., Ahmed, H., Lee, T., & Patel, S. (2022). Mitigating bias in deep learning for diagnosis of coronary artery disease from myocardial perfusion SPECT images. European Journal of Nuclear Medicine and Molecular Imaging, 50(2), 387–397. https://doi.org/10.1007/s00259-022-05972-w
53	Mohd Hamdan, M. D. H., Ab Jabal, M. F., Abdul Rahman, S., & Kapi, A. Y. (2025). Fairness And Bias Mitigation in AI Models for Diabetes Diagnosis: A Comparative Evaluation of Algorithmic Approaches. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 17(3), 27–34. https://doi.org/10.54554/jtec.2025.17.03.004
54	Mooghali, M., Stroud, A. M., Yoo, D. W., Barry, B. A., Grimshaw, A. A., Ross, J. S., Zhu, X., & Miller, J. E. (2024). Trustworthy and ethical AI-enabled cardiovascular care: a rapid review. BMC Medical Informatics and Decision Making, 24(1), 247. https://doi.org/10.1186/s12911-024-02653-6
55	Mosteiro, P., Prieto, A., Solares, C., García, M., & Fernández, Á. (2022). Bias discovery in machine learning models for mental health. Information, 13(5), 237. https://doi.org/10.3390/info13050237
56	Muzammil, M.A, Javid, S, Afridi, A.K., Siddineni, R., Shahabi, M., Haseeb, M., Fariha, F.N.U., Kumar, S., Zaveri, S., Nashwan, A.J. Artificial intelligence-enhanced electrocardiography for accurate diagnosis and management of cardiovascular diseases. Journal of Electrocardiology Volume 83, March–April 2024, Pages 30-40. https://doi.org/10.1016/j.jelectrocard.2024.01.006

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

57	Nancy Lapid, (March 28, 2024). AI fails to detect depression signs in social media posts by Black Americans, study finds. https://www.reuters.com/business/healthcare-pharmaceuticals/ai-fails-detect-depression-signs-social-media-posts-by-black-americans-study-2024-03-28/
58	Naskar, S., Sharma, S., Kuotsu, K., Halder, S., Pal, G., Saha, S., Mondal, S., Biswas, U. K., Jana, M., & Bhattacharjee, S. (2025). The biomedical applications of artificial intelligence: an overview of decades of research. Journal of Drug Targeting, 33(5), 717–748. https://doi.org/10.1080/1061186X.2024.2448711
59	Norori, N., Hu, Q., Aellen, F.M., Faraci, F.D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. Patterns, 2(10), 100347. https://www.cell.com/patterns/fulltext/S2666-3899(21)00202-6?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2666389921002026%3Fshowall%3Dtrue
60	Noseworthy, P. A., Attia, Z. I., Brewer, L. C., Hayes, S. N., Yao, X., Kapa, S., Friedman, P. A., & Lopez-Jimenez, F. (2020). Assessing and Mitigating Bias in Medical Artificial Intelligence. Circulation: Arrhythmia and Electrophysiology, 13(3). https://doi.org/10.1161/circep.119.007988
61	Olczak, V. (2024). Gender equality and artificial intelligence : navigating the EU policy frameworks for a feminist future. Gender five plus. https://genderfiveplus.org/wp-content/uploads/2024/11/Report-Victoire-AI-1.pdf
62	Omar, M., Soffer, S., Agbareia, R., Bragazzi, N. L., Apakama, D. U., Horowitz, C. R., Charney, A. W., Freeman, R., Kummer, B., Glicksberg, B. S., Nadkarni, G. N., & Klang, E. (2025). Sociodemographic biases in medical decision making by large language models. Nature Medicine, 31(6), 1873–1881. https://doi.org/10.1038/s41591-025-03626-6
63	Otokiti, A. U., Shih, H., & Williams, K. S. (2025). Gender and racial bias unveiled: clinical artificial intelligence (AI) and machine learning (ML) algorithms are fanning the flames of inequity. Oxford Open Digital Health, 3. https://doi.org/10.1093/oodh/oqaf027
64	Paixão Cansado, Marta, (2024), Inteligência Artificial no setor da saúde: desafios jurídicos e regulação, No 187, GEE Papers, Gabinete de Estratégia e Estudos, Ministério da Economia, https://EconPapers.repec.org/RePEc:mde:wpaper:187
65	Palmer, K. & Lee McFarling, U. (2024, September 3). Doctors use problematic race-based algorithms to guide care every day. Why are they so hard to change. STAT10. https://www.statnews.com/2024/09/03/embedded-bias-investigation-health-equity-clinical-algorithms/

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

66	Patel, D., Chetarajupalli, C., Khan, S., Khan, S., Patel, T., Joshua, S., & Millis, R.M. (2025). A Narrative Review on Ethical Considerations and Challenges in AI-Driven Cardiology. Annals of Medicine & Surgery, 87, 4152–4164 https://journals.lww.com/annals-of-medicine-and-surgery/fulltext/2025/07000/a_narrative_review_on_ethical_considerations_and.26.aspx
67	Pham, Q., Wiljer, D., Cafazzo, J., & et al. (2021). The need for ethnoracial equity in artificial intelligence for diabetes management: Review and recommendations. Journal of Medical Internet Research, 23(2), e22320. https://doi.org/10.2196/22320
68	Pias, T. S., Su, Y., Tang, X., Wang, H., Faghani, S., & Yao, D. (2025). Enhancing Fairness and Accuracy in Diagnosing Type 2 Diabetes in Young Adult Population. IEEE Journal of Biomedical and Health Informatics, 1–10. https://doi.org/10.1109/jbhi.2025.3616312
69	Published in National Health Law Program by Cassandra LaRose and Elizabeth Edwards, https://healthlaw.org/1557-final-rule-protects-against-bias-in-health-care-algorithms/?utm_source=chatgpt.com
70	Puyol-Antón, E., Ruijsink, B., Piechnik, S. K., Neubauer, S., Petersen, S. E., Razavi, R., & King, A. P. (2021). *Fairness in Cardiac MR Image Analysis: An Investigation of Bias Due to Data Imbalance in Deep Learning Based Segmentation* (pp. 413–423). https://doi.org/10.1007/978-3-030-87199-4_39
71	Rai, S., Stade, E. C., Giorgi, S., Fodeh, S. J., Ungar, L. H., & Guntuku, S. C. (2024). Key language markers of depression on social media depend on race. Proceedings of the National Academy of Sciences, 121(14), e2319837121. https://doi.org/10.1073/pnas.2319837121
72	Ratwani, R. M., Sutton, K., & Galarraga, J. E. (2024). Addressing AI Algorithmic Bias in Health Care. *JAMA, 332*(13), 1051. https://doi.org/10.1001/jama.2024.13486
73	Ratwani, R. M., Sutton, K., & Galarraga, J. E. (2024). Addressing AI Algorithmic Bias in Health Care. JAMA, 332(13), 1051–1052. https://doi.org/10.1001/jama.2024.13486
74	Ross, C. & Herman, B. (2023, November 14). *UnitedHealth faces class action lawsuit over algorithmic care denials in Medicare Advantage plans.* STAT10. UnitedHealth sued over use of algorithm in Medicare Advantage plans
75	Sarma, A. D., & Devi, M. (2025). Artificial intelligence in diabetes management: transformative potential, challenges, and opportunities in healthcare. *Hormones, 24*(2), 307–322. https://doi.org/10.1007/s42000-025-00644-4
76	Sasseville, M., Ouellet, S., Rhéaume, C., Sahlia, M., Couture, V., Després, P., Paquette, J.-S., Darmon, D., Bergeron, F., & Gagnon, M.-P. (2025). Bias Mitigation in Primary Health Care Artificial Intelligence Models: Scoping

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

	Review. Journal of Medical Internet Research https://www.jmir.org/2025/1/e60269
77	Scheetz J., Koca D., McGuinness M., Holloway E., Tan Z., Zhu Z., O'Day R., Sandhu S., MacIsaac R., Gilfillan C., Angus Turner A., Stuart Keel S., He M.(2021). Real-world artificial intelligence-based opportunistic screening for diabetic retinopathy in endocrinology and indigenous healthcare settings in Australia. https://www.nature.com/articles/s41598-021-94178-5
78	Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nature Medicine, 27(12), 2176–2182. https://doi.org/10.1038/s41591-021-01595-0
79	Sheng, B., Pushpanathan, K., Guan, Z., Lim, Q. H., Lim, Z. W., Yew, S. M. E., Goh, J. H. L., Bee, Y. M., Sabanayagam, C., Sevdalis, N., Lim, C. C., Lim, C. T., Shaw, J., Jia, W., Ekinci, E. I., Simó, R., Lim, L.-L., Li, H., & Tham, Y.-C. (2024). Artificial intelligence for diabetes care: Current and future prospects. The Lancet Diabetes & Endocrinology, 12(8), 569–595. https://doi.org/10.1016/s2213-8587(24)00154-2
80	Sinha, Chaitali & Schryer-Roy, Anne-Marie. (2018). Digital health, gender and health equity: invisible imperatives. Journal of public health (Oxford, England). https://pmc.ncbi.nlm.nih.gov/articles/PMC6294032/
81	Sogancioglu, G., Mosteiro, P., Salah, A.A., Scheepers, F., Kaya,H. Fairness in AI-Based Mental Health: Clinician Perspectives and Bias Mitigation Proceedings of the Seventh AAAI/ACM Conference on AI, Ethics, and Society (AIES 2024) https://research-portal.uu.nl/ws/files/243475617/31732-Article_Text-35796-1-2-20241016.pdf
82	Starre Vartan, (2019). Racial Bias Found in a Major Health Care Risk Algorithm, https://www.scientificamerican.com/article/racial-bias-found-in-a-major-health-care-risk-algorithm/
83	Straw I, Callison-Burch C. Artificial Intelligence in mental health and the biases of language based models. PLoS One. 2020 Dec 17;15(12):e0240376. doi: 10.1371/journal.pone.0240376. PMID: 33332380; PMCID: PMC7745984.
84	Sue Haupt, Bronwyn Graham, Jane Hirst (2024). Can AI fight sex and gender bias in healthcare? https://www.unsw.edu.au/newsroom/news/2024/10/can-ai-fight-sex-and-gender-bias-in-healthcare-
85	Tat, E., Bhatt, D.L., & Rabbat, M.G. (2020). Addressing bias: artificial intelligence in cardiovascular medicine. The Lancet Digital Health, 2(12), e635–e636.doi: 10.1016/S2589-7500(20)30249-1 https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30249-1/fulltext

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

86	The Lancet Digital Health. (2024). Promises and challenges of digital tools in cardiovascular care. The Lancet Digital Health, 6(10), e673. https://doi.org/10.1016/S2589-7500(24)00194-8
87	Timmons A., Tutul A., Avramidis K., Duong J., Carta K., Walters S., Jumonville G., Carrasco A., Freitag G., Romero D., Ahle M., Comer J., Narayanan S., Khurd I., Chaspari T. (2025). Developing personalized algorithms for sensing mental health symptoms in daily life. https://www.nature.com/articles/s44184-025-00147-5?fromPaywallRec=false
88	Tina Wall (2025), Head of MediaNew AI algorithm uses mammograms to accurately predict cardiovascular risk in women. https://www.georgeinstitute.org/news-and-media/news/new-ai-algorithm-uses-mammograms-to-accurately-predict-cardiovascular-risk-in-women
89	Travis Zack, Eric Lehman, Mirac Suzgun, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health. 2024 Jan;6(1):e12-e22. doi: 10.1016/S2589-7500(23)00225-X.
90	Tytgat, N., Marques, P., Gho, J., Stefański, B., & Cornelissen, F. (2023). Racial disparities in continuous glucose monitoring-based 60-min blood glucose predictions in people with type 1 diabetes. PLOS Digital Health, 2(9): e0000918. https://journals.plos.org/digitalhealth/article?id=10.1371%2Fjournal.pdig.0000918&utm
91	van Assen M., Beecy A., Gershon G., Newsome J., Trivedi H., Gichoya J. (2024). Implications of Bias in Artificial Intelligence: Considerations for Cardiovascular Imaging. DOI: 10.1007/s11883-024-01190-x
92	van Houten, H. (2020, November 16). *For fair and equal healthcare, we need fair and bias-free AI.* Philips. https://www.philips.com/a-w/about/news/archive/blogs/innovation-matters/2020/20201116-for-fair-and-equal-healthcare-we-need-fair-and-bias-free-ai.html
93	Vien Ngoc Dang, Cascarano, A., Mulder, R. H., Cecil, C., Zuluaga, M. A., Jerónimo Hernández-González, & Karim Lekadir. (2024). Fairness and bias correction in machine learning for depression prediction across four study populations. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-58427-7
94	Wang, S.C.Y.. et al. (2024). AI-based diabetes care: risk prediction models and bias concerns. NPJ Digital Medicine https://www.nature.com/articles/s41746-024-01034-7?utm
95	World Health Organization. (2024). *Ethics and governance of artificial intelligence for health : guidance on large multi-modal models.* https://iris.who.int/server/api/core/bitstreams/e9e62c65-6045-481e-bd04-20e206bc5039/content

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

96	Wu, G., Tewari, S., Wong, A., Chung, E., Chim, I., Hoang, B., Mansubi, N., Shams, A., del Buono, M., Srinivasan, S., Paliath-Pathiyal, H., & Khan, O. (2025). Chatbots and Diabetes: Is There Gender Bias? *Journal of Patient Experience, 12*. https://doi.org/10.1177/23743735251380954
97	Yang M., El-Attar A., Chaspari T. (2024). Deconstructing demographic bias in speech-based machine learning models for digital health. Frontiers in Digital Health. http://DOI: 10.3389/fdgth.2024.1351637
98	Yousufi, M., Damaševičius, R., & Maskeliūnas, R. (2024). Multimodal Fusion of EEG and Audio Spectrogram for Major Depressive Disorder Recognition Using Modified DenseNet121. *Brain Sciences, 14*(10), 1018. https://doi.org/10.3390/brainsci14101018
99	Zainab Al-Zanbouri, Sharma, G., & Raza, S. (2024). Equity in Healthcare: Analyzing Disparities in Machine Learning Predictions of Diabetic Patient Readmissions. 660–669. https://doi.org/10.1109/ichi61247.2024.00105
100	Zhang, A., Yuksekgonul, M., Guild, J., Zou, J., & Wu, J. C. (2023). ChatGPT Exhibits Gender and Racial Biases in Acute Coronary Syndrome Management. ArXiv.org. https://doi.org/10.48550/arXiv.2311.14703

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Translations

In the following sections, we provide the translation of the main report (i.e., without the appendices and the bibliography table) in the following partner languages in the order stated in the GA: EL/DE/BG/LT/ES/IT/PT.

The translations were provided by the partners according their respective native languages.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. Project code: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI