Diversify with Intersectionally Fairer AI

Project 101107969

Deliverable number: D1.1
Deliverable title: Project Handbook

Document ID: DIVERSIFAIR_D1.1-v0.3
Dissemination level:
Main Author(s): Lizette Maljaars, Steven Vethman
Issue date: February 29, 2024

Co-funded by the European Union

EACEA.A – Erasmus+, EU Solidarity Corps A.2 – Skills and Innovation

Disclaimer and acknowledgment

Co-funded by the European Union

This project has received funding from the European Education and Culture Executive Agency (EACEA) in the framework of Erasmus+, EU Solidarity Corps A.2 – Skills and Innovation under grant agreement 101107969.

Disclaimer

DIVERSIFAIR consortium

TNO innovation for life

UCD DUBLIN

eticas

WOMEN 4CYBER
EUROPEAN CYBER SECURITY ORGANISATION

SciencesPo

WOMEN IN AI

Turing College

CorTexter

Document information

Document change log

Version	Date	Comments
v0.1	14-09-2023	First draft of the document
v0.2	05-10-2023	Revised version based on the comments of the MT members (see Sections 2.1 and 2.2 for names)
V0.3	29-02-2024	Minor adjustment related to

Document distribution log

Version	Date	Distributed to
v0.1	14-09-2023	Reviewers
v0.2	05-10-2023	GA
V0.3	29-02-2024	GA

Verification and approval

	Name	Date
Verification of final draft by WP leader	Lizette Maljaars	29-02-2024
Check before upload by coordinator	Lizette Maljaars	29-02-2024

Executive summary

DIVERSIFAIR will develop a pioneering EU wide approach to deal with intersectional fairness in AI based on use cases, data, and models that originate or have been deployed in Europe. It will use innovative tools and structure them practically in two main knowledge hubs, containing user-friendly educational content in 2 main areas of fairer AI-Equality4Tech and Tech4Equality, accompanied by transversal use cases connecting across all project activities.

Equality4Tech encompasses:

1. Sector-specific educational kits and Fair AI Scrum methodology for data scientists and AI practitioners
2. Non-tech stakeholders AI BIAS awareness training.

Tech4Equality detects sectional & intersectional bias from a technological point of view. It develops a Recruitment scenario on intersectional bias mitigation and detection to experiment and bring practical knowledge to AI experts and tech-related professionals.

Eight consortium partners from 6 EU countries – academic/ research institutions create open-source reports, scientific articles, educational materials & lecture series on AI Intersectional fairness and Ethical auditing. DIVERSIFAIR builds a deeper understanding of harms and discriminatory impact of AI-facilitated outcomes on people's lives. It prepares the ground for effectively aligning AI systems' development with EU values and robust regulation in the field of FAIR AI. VET partner consolidates and disseminates an eco-friendly, fully digitalized AI Intersectional bias educational program to beginner and seasoned AI professionals. The NGO partners craft and deliver AI Bias awareness training for social workers, HRs, policy-builders, and other non-tech stakeholders. DIVERSIFAIR incorporates voices of underprivileged groups. AI Bias training is freely accessible for them.

More than 500 educated tech and non-tech participants, 6 education courses, 5 AI audits with sectorial case studies and large dissemination campaign to labour market players and EU policy-making institutions.

Definitions and abbreviations

AI	Artificial Intelligence
ALTAI	Assessment List of Trustworthy Artificial Intelligence
CA	Consortium Agreement
CEO	Chief Executive Officer
MT	Management Team
EC	European Commission
EU	European Union
DoA	Description of Action (a.k.a. DoW)
DoW	Description of Work (a.k.a. DoA)
DPO	Data Protection Officer
GA	General Assembly (by context differentiated from Grant Agreement)
GA	Grant Agreement (by context differentiated from General Assembly)
HR	Human Resources
LGBTQ	Lesbian, Gay, Bisexual, Transgender, Queer or Questioning
NGO	Non-Governmental Organization
QPR	Quarterly Progress Report
SOTA	State-of-the-Art
UWV	Dutch Unemployment Agency
VET	Vocational Education Training
WP	Work Package

Table of Contents

1. Introduction

The development of the Artificial Intelligence (AI) sector internationally has created a set of needs in terms of training and education, two main needs of which this project responds to. The first is overcoming the scarcity of AI experts on algorithmic fairness who possess the right skills when it comes to identifying and addressing bias and discrimination in AI. The second is effectively incorporating an intersectional approach to understanding and addressing bias, fairness, discrimination, and equality in AI to prevent the emergence of invisible minority groups that would suffer systematic AI-driven disadvantage.

The project partners aspire to shed light on the pressing need to mitigate the potential of AI applications to produce discriminatory impact by focusing on principles such as algorithmic fairness or transparency. Recognising that existing work on bias and fairness is still primarily done in a techno-centric manner, DIVERSIFAIR will bring together technical and non-technical expertise in an interdisciplinary manner through two complementary approaches: Equality4Tech and Tech4Equality.1 By doing this, DIVERSIFAIR will boost innovation putting the focus on digital skills as they are increasingly important in all job profiles across the entire labour market.

A major limitation in existing approaches to algorithmic bias and fairness is that they focus on one of the characteristics protected under anti-discrimination law, e.g., ethnicity, race, gender, sexual orientation, disability, age, religion, or belief, which is then used to translate principles of fairness and equality in the context of application of AI. Such algorithmic fairness approaches, however, omit to consider social identities or individual positions that emerge at the intersection of above-mentioned protected characteristics. In other words - they omit cases where the individual identifies with or belongs to multiple protected groups – e.g., a woman in the AI or Cyber sector who is also of African American origin and a single mother. Hence, current fairness approaches are not suited to addressing intersectional discrimination and disadvantage that arise at the crossroads of multiple social memberships protected under anti-discrimination law. For example, presumably accurate image-recognition tools display significant errors when tested on dark-skinned men (in contrast to white men) and perform even poorer when tested on images of dark-skinned women. Next to that, a failure to factor in the intersection of race and gender results in invisibility loops. Hence, the system's accuracy is compromised and facilitates the digital erasure of the affected groups, excluding their interests from the model's calculations and outcome.

The intersectional lens in conceptualizing fairness in AI also goes beyond testing performance across attribute-based groupings, and provides a framework for critically evaluating the broader design context and core objectives of AI systems.

1.1 DIVERSIFAIR as social inclusion promoter

To overcome the major gap described above, project partners set the general objective of the DIVERSIFAIR project - to contribute to strengthening Europe’s innovation capacity and resilience by boosting innovation through cooperation and flow of knowledge among higher education, vocational education, and training in a broad socioeconomic environment, including research. They will make it possible for 36 months through upskilling their digital knowledge, boosting innovative sustainable ideas and solutions, and fostering cross sectoral partnerships. As a result, the project will contribute to deconstructing society’s invisibility loops and help replacing one-dimensional with intersectional fairness approaches to AI. It will ultimately contribute to a fairer AI founded on the EU principles of equality and inclusion. The consortium, consisting of two higher education and one VET Institutions,

one research organizations (including for applied research), two business companies (inclusion innovation start-up and technology & innovation company) as well as two NGOs related to gender equality and AI will achieve social awareness, create international policy-making influence in six EU member states in Northern, Western, South, and Central Europe, and train tech and non-tech experts in futureproof AI skills.

Our experts achieve the inclusion, rehabilitation, and wellness promotion of excluded groups by spanning out a 3-pillar approach:

1. Craft a body of academic and practical knowledge on intersectional AI bias: State-Of-The-Art (SOTA) definitions, typology of harms, standards and principles, requirements for intersectional fairness in AI, best practices and real-life examples in AI
2. Develop educational AI tools and processes:
    a. Equality4Tech for AI auditing and bias detection - 'Fair AI Scrum' methodology.
    b. Tech4Equality for bias mitigation and prevention – we will use it for training tech experts to produce algorithmic transparency in anti-bias intersectional AI technologies.
3. Develop AI bias awareness educational programs for the following target groups:
    a. AI professionals (e.g. Data Scientists, Machine learning engineers) – > 'Fair AI Scrum' & Intersectionality bias in AI courses;
    b. Non-technical AI users/ implementers within particular sectors -> General educational courses on sectional- and intersectional bias, and potential social, ethical, and legal challenges around AI technologies and bias.
    c. Human Resources (HR) Experts, Employment agencies and social workers –> Bias awareness course – Minimizing bias and discrimination at the workplace and on the job market via disseminating for free the Bias Language detection application by the partner CorTexter in 5 EU countries and 6 organizations - 1 in Germany, 1 in France, 1 in Belgium, and 3 in The Netherlands (one of which the Dutch Unemployment agency - UWV).

Lastly, whereas most of the groundwork gets done in the SHERPA3 and SATORI4 projects on AI ethics, DIVERSIFAIR expands on that base by developing a specific knowledge line around intersectional fairness, which will be disseminated via Europe`s open access platform AI-on-Demand.5

1.2 Horizontal priorities addressed

In harmony with the EC`s Assessment List for Trustworthy Artificial Intelligence (ALTAI)6 for self-assessment from 2020 and the EU Coordinated Plan on AI, reviewed in 20217, DIVERSIFAIR contributes to and is fully in line with the following horizontal priorities of the Erasmus+ Program:

Inclusion and diversity

The project creates preconditions for promoting equal opportunities and access, inclusion, diversity, and fairness across all sectors. DIVERSIFAIR develops educational AI tools and processes and puts mechanisms and resources at a disposal through open sources. When designing the project's concept and activities, the consortium used an inclusive approach, making the results of the project accessible to diverse participants. The partnership envisages granting access to the AI Intersectional bias educational program for women or representatives of other underrepresented and underprivileged groups and by doing this promotes equality and non-discrimination. The consortium partners' Women4Cyber and Women in AI will play an active role here.

Digital Transformation and Social Resilience

Development Integrating the expertise of VET, academic, private, and public stakeholders in fully digitalized educational programs on intersectional bias, DIVERSIFAIR guarantees the much-needed ongoing and congruent digital transformation and resiliency of different stakeholders in a post-COVID-19 world. Offering innovative multidisciplinary AI education for AI experts, HRs, social workers, as well as Equality Bodies and Data Protection Agencies across the EU and beyond, DIVERSIFAIR will contribute to tackling AI-driven intersectional bias and discrimination for a more cohesive European society. It will promote resilience and excellence through quality, inclusive and flexible VET and establish a new lifelong learning culture emphasizing the relevance of C-VET and digitalization.

The project proposal is fully in line with the following objectives of the Alliances for Education and Enterprises call:

Fostering new, innovative, and multidisciplinary approaches to teaching and learning: innovation in education design and delivery, teaching methods, assessment techniques, learning environments, and new future skills development. The project will upskill students across the spectrum from high school via university to CEOs, HRs, legal representatives, and policymakers which guarantees a cross-functional change with EU-wide synergic success. The course for data professionals is fully online, self-paced and incorporates active peer-to-peer learning.

Fostering corporate social responsibility – using AI-based solutions to drive digital and real-life equity and social inclusion for women and girls and other groups at risk of discrimination, such as senior job seekers, ethnic minorities, people with disabilities, and the LGBTQ community. The Inclusion HR software partner CorTexter will promote project deliverables dissemination and positive influence on corporate and social domains, like the Dutch and Belgian Social services institutions. Equality bodies and data protection agencies, as enforcers of equality and data protection law in Europe, play a key role in tackling algorithmic discrimination and will therefore represent a key target group for the educational materials developed by DIVERSIFAIR. Thus, the project promotes social inclusion in a centralized manner by enabling management in public and private sectors to gain awareness of the risk AI algorithms used at their company entail upon the inclusion of specific employees or potential employee groups. Ultimately, this will move the needle towards enhanced inclusion of the so-called “invisible” groups, equal involvement in the recruitment and selection process, and fair distribution of higher-paying jobs, re-training, and upskilling opportunities.

Facilitating the flow and co-creation of knowledge between higher education and vocational education and training, research, the public sector, and the business sector. The inevitable advantage that DIVERSIFAIR brings comes from the versatile consortium, consciously involved in cutting-edge educational innovation. It is brought about by higher education & VET institutions, NGOs, and scientific and innovation companies.

Building and supporting effective and efficient higher education and vocational education and training systems, which are connected and inclusive and contribute to innovation. The material for the AI professionals’ course will be open-source and TURING will serve as a practical example of how it can be utilised by vocational schools with a modern approach (online, self-paced, practical-project based, peer-to-peer supported studies). By demonstrating that this coursework brings additional

value to data scientists, the result will encourage other vocational schools to start focusing more on the AI fairness.

1.3 Purpose of the document

The purpose of this Project Handbook is to describe the procedures and processes that are generic for all Work Packages and that enable smooth cooperation.

1.4 Document structure

Chapter 2 describes the management structures, including the nominees for the various boards. Chapter 3 is dedicated to specific quality management procedures, including communication structures and tools, the peer reviewing process for high quality deliverables, as well as risk management and other quality assurance means. In Chapter 4, the technical infrastructure for communication and collaboration is presented. Chapter 5 outlines the specific ethical guidelines that the project is following. In Chapter 6 the consortium's strategy towards openness is described and relates to open source in terms of software as well as open access in terms of publications and other project results.

1.5 Deviations from the original Description of Action (DoA)

In WP 5, Woman in AI and Women 4 Cyber have agreed on a redistribution of budget between them, pending approval from the grant authorities.

1.6 Description of work related to deliverable as given in DoA

The Project handbook contains operational procedures and processes for the project regarding information sharing, meetings, quality assurance, risk management and ethics. After being submitted in M3, the Project handbook remains a dynamic document. E.g, if the data management plan needs to be updated, that results in a new version of the Project handbook.

1.7 Time deviations from the original DoA

As the project was kicked-off three months after the formal start of the consortium according to the Grant Agreement, this deliverable was submitted 1 months later than anticipated.

1.8 Content deviations from the original DoA

There are no deviations from the DoA; the content of this deliverable is in line with the plan.

2. Management structure

Both the Grant Agreement (GA) and the Consortium Agreement (CA) specify a number of bodies for the management of the project. Though the GA and CA, being legal documents

that can be found on SharePoint 'Grant Agreement'[1] and 'Consortium Agreement', take precedence over this handbook, the following sections specify the operational view of these bodies.

2.1 Work package (WP)

The work package (WP) is the building block of the project. The WP leader

1. organizes the WP,
2. prepares and chairs WP meetings,
3. organizes the production of the results of the WP,
4. represents the WP in the Management Team (MT).

Current WP leaders are shown in Table 2-1.

Table 2-1: current WP leaders

WP	WP name	Current WP leader
1	Project management and coordination	Lizette Maljaars (TNO)
2	Developing concepts of Intersectional fairness, IS bias for AI Auditing	Susan Leavy (UCD)
3	Sectorial case studies and development of SOTA typology on tooling and methodologies for the detection/ mitigation of intersectional bias in AI	Gemma Galdon-Clavel (ETICAS)
4	Implementation – Delivery of intersectionally Fair AI Tech Educational course	Rūta Gumbelevičiūtė (TURING College)
5	Communication & Impact maximisation	Marisa Tschopp a.i. (Women In AI)

2.2 Management Team (MT)

The Management Team consists of:

- The WP leaders of all (active) WPs (Table 2-1)
- The scientific lead of the project (Table 2-2)
- The consortium manager (Table 2-2)

Table 2-2: Additional MT members

Role	Person
Scientific lead	Ilina Georgieva and Steven Vethman (TNO)
Consortium manager	Lizette Maljaars (TNO)

The consortium manager organizes and chairs the MT meetings. The MT manages the coordination between the WPs. The MT has a mandate from the General Assembly (GA) for all day-to-day management.

The GA members and task managers, even if not MT member, are welcome at the MT meetings.

[1] Access to links referring to SharePoint and/or Teams sites have restricted access for consortium members only.

2.3 General Assembly (GA)

The GA consists of one representative of each partner. The current GA-members are listed in Table 2-3. The members of the GA are referred to as 'partner manager'.

The GA takes all decisions that affect the direction of the project. The GA members are addressed for any issue, technical or administrative, concerning that partner.

Table 2-3: Current partner managers

Nr	Partner	Partner manager
1	TNO	Lizette Maljaars
2	University College Dublin	Susan Leavy
3	Sciences PO	Raphaële Xenidis
4	ETICAS	Matteo Mastracci
5	TURING College	Rūta Gumbelevičiūtė
6	Women4Cyber	Saskia Brugman a.i.
7	Women In AI	Hanan Salam a.i.
8	CorTexter	Miguel van Bodegom

2.4 Advisory boards

DIVERSIFAIR will not have any advisory boards from outside the project.

3. Quality procedures and Code of Conduct

3.1 Internal communication structures and procedures

3.1.1 Team members joining or leaving the consortium

The partner managers will inform the MT and GA if one of their team members leaves the project, or if they plan to introduce a new team member. The Consortium management will arrange that leaving team members will be excluded from communication and access to documentation. The Consortium management will grant new team members access to collaboration tools and infrastructure (see Section 4), and will include them in communication.

3.1.2 GA meetings

GA meetings will be scheduled once every nine months. If important decisions need to be taken at GA level, then an ad-hoc meeting can be scheduled. The agenda will be distributed at least three weeks before the meeting (10 days in case of an ad-hoc meeting). All members of the GA or MT can enlist agenda items for the GA meeting, agenda items that require a decision must be identified as such. No minutes are taken at the GA meetings, but decisions and actions of the GA are listed on sharepoint. The consortium management shares the decisions and actions with the GA and MT members via e-mail, when needed, they may additionally be discussed in the first MT meeting after the GA meeting.

3.1.3 MT meetings

Every month[2], the MT has a conference call. The main purpose of these meetings is the alignment of work between the WPs. The MT will work with a recurring agenda containing the following topics:

- Progress (incl. milestones & deliverables)
- Risks and issues
- Budget and timelines
- GA decisions and actions

MT and/or GA members can enlist additional agenda items up to a week before the next meeting. MT members will prepare for the meeting by updating the Progress tracker for MT meetings-file on ShareP at least two days in advance. This progress tracker serves to maximize the efficiency of the meetings as progress is clear before the meeting, so the meeting can focus only on items that need discussion or a decision. The actual agenda of each meeting will be stored on SharePoint. Decisions and action points of the MT meetings will be listed on SharePoint and are communicated to all MT and GA members by the consortium management via e-mail. In addition, decisions, action points and a list of the actual participants will be added to the agenda of the respective MT meeting within two working days after the meeting. This allows GA members to react, e.g. if decisions are taken in an MT meeting and a GA member considers that decision to require GA endorsement.

3.1.4 WP and task meetings

For meetings within the WP, the WP leaders have full freedom to arrange them as they wish. The only constraint will be the travel budget of the partners. If a partner is not participating fully in the WP or task, and there is a risk of that partner becoming a 'defaulting partner', as defined in the CA, then the following steps will be taken.

- The manager of the task/WP will have a private discussion with the partner. The result will be recorded in an e-mail, sent in Cc to the consortium manager. In the unlikely case the WP leading partner is not fully participating, any partner in the WP can signal this to the consortium manager, initiating the next step immediately.
- If this fails to produce the desired behavior or if a WP leader is not participating fully in the WP, the consortium manager will have a private discussion with the partner. The result will be recorded in an e-mail, sent in Cc to the General Assembly.
- If this fails to produce the desired behavior, the GA starts the 'defaulting partner procedure' as defined in the CA.

3.2 External communication structures & procedures

3.2.1 Communication, dissemination and visibility

Communication, dissemination and visibility is a core activity of the DIVERSIFAIR project (WP5) and will take place as detailed out on page 61 of Section 4.2 of the Technical Description (Part B) in the Grant Agreement.

[2] Partners agreed to start with a monthly meeting although quarterly meetings may be preferred to minimize meeting load. To balance meeting load with consortium needs (i.e. meeting deliverables within time, budget and quality requirements), the frequency may be reviewed and adjusted on a quarterly basis.

3.2.2 Communication to the European Commission

All communication with the European Commission (EC), and in particular with the project officer (PO), will be coordinated by the consortium management as defined in Table 3-1. Where needed, the Scientific lead will support the consortium management in this.

Table 3-1: Consortium management

Role	Person
Consortium manager	Lizette Maljaars (TNO)
Consortium management support	Catelijne Rauch (TNO)

3.3 Quality of (non-)deliverables and peer review

Reviews are the key elements in the quality assurance of a project like DIVERSIFAIR. For the review process, there is a distinction between review of deliverables and the review of other material.

3.3.1 Deliverables

Deliverables can be planned well, since a global description of the content, the submission date and the partners working on it are defined in the DoA. The review process will be done in three stages:

- Structure or scope review
- Content review
- GA check

Two independent reviewers are appointed by the MT for each deliverable, and in principle both perform the structure/scope and the content review. Reviewers are considered independent when they are not the authors of the deliverable. Of course, others are free to review too, but the appointed reviewers take on the quality assurance responsibility for the deliverable.

The deliverable editors are responsible for timely submission for Structure or scope review and Content review to the reviewers, with cc to the consortium manager. The consortium manager will submit the reviewed deliverable to the GA for the GA check. The consortium management will uploaded the fully reviewed deliverables to the Participant Portal – Continuous Reporting for submission to the granting authority. The timeline for deliverables is depicted in Figure 3-1.

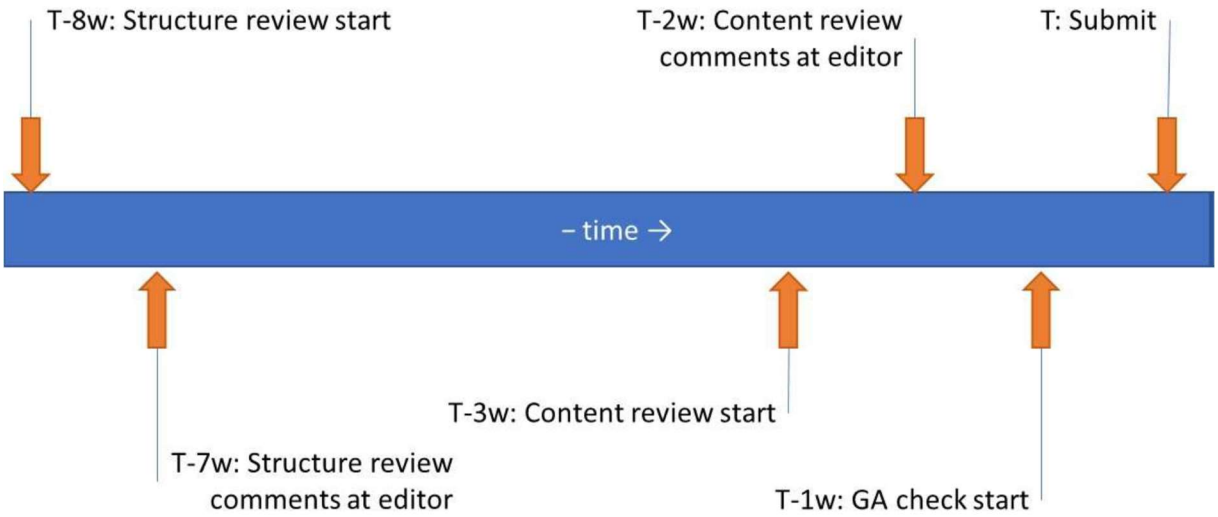

Figure 3-1: Timeline for deliverable review.

3.3.2 Structure or scope review

The input for the structure review is the structure description of the deliverable. The structure description consists of at least two levels in the table of contents, chapters and sections. At section level there is:

- a 5-line description of the content
- the responsible partner/person for generating the content
- the expected number of pages as indicator for the level of detail

The structure review starts as soon as the structure description is available, but not later than 8 weeks before the submission date of the deliverable. Reviewer comments are to be submitted to the deliverable editor 7 weeks before the submission date.

3.3.3 Content review

The input to the content review is the full deliverable text; only supporting parts – references, list of abbreviations and annexes – might still need completion. The content review starts at the latest 3 weeks before the submission date. Review comments are submitted to the deliverable editor 2 weeks before the submission date. In general, the content review contains four main attention areas.

DoA coverage

- Is the scope and the content of the deliverable consistent with the intention of the deliverable as stated in the DoA?
- In case of deviations, are they fully and plausibly motivated?
- Are the relations to other DIVERSIFAIR activities/deliverables clear?

Target audience

- Is the target audience clear?
- In case of multiple target groups, is it clear what parts of the deliverable are intended for each audience?
- Are the management summary, introduction and conclusions/recommendations at the level, and using the language, of the target audience? Note: The detailed content might be too

detailed for all target groups, but the sections mentioned here should be fit for all target groups.
- Are the conclusions fully backed by the preceding material and are recommendations actionable?

Language and structure
- Is the language used proper English? Signal use of national variants, and in case of doubt consult a native English speaker.
- Is the text well-structured, e.g. using lists and tables where appropriate?
- Do chapters have a local introduction/purpose and local conclusions/recommendations?
- Are illustrations and diagrams used to support the text where appropriate? If taken from external sources, is the attribution correct/complete?
- Are relevant references to literature included?

Technical content
- For the editor/WP leader to judge during the review process

3.3.4 GA check

The GA members receive the deliverable at least one week before the submission date. They check that the deliverable does not disclose commercially sensitive information of their organization. If the deliverable contains material from non-partners that is made available via their organization, the GA member checks that the deliverable respects the confidentiality agreements made by their organization with the non-partners. Note: the GA check is not a classical review. It is an 'emergency brake' if confidential material is about to be disclosed that was not noted by authors and reviewers.

3.3.5 Non-deliverables

For non-deliverables, such as publications and dissemination material, the procedure for deliverables will be used where applicable and with a timeline that fits the material. In all cases the MT is required to be informed via the WP leader about the intention to publish DIVERSIFAIR material as early as possible, with a minimum of 4 weeks. The MT will decide on the review procedure for that case. This is enabled by WP leaders signalling planned academic publications or conference contributions to the Scientific lead and signalling nonacademic work to the WP lead. Since there are many types of material, this handbook cannot provide details for all cases. We distinguish the following broad categories of material.

- Dissemination material (e.g. flyer, website, leaflets, popular science publications). The default reviewer is the consortium manager, supported by one or more partner managers.
- Scientific publication or conference presentation. The default reviewer is the scientific lead, supported by one or more partner managers.

3.4 Risk management

In the GA, the results of an initial risk assessment are listed. This is considered the initial risk register. When a partner or WP leader identifies the following situations, it should be communicated with the consortium management as soon as possible.

- an existing risk becoming an issue
- a substantial change of a risk (e.g. if the expected probability or impact of the risk changes)
- a new risk or issue

At the latest during the closest upcoming MT, this risk or issue and potential measures will be discussed, to avoid a risk becoming an issue, and added to the risk register. At the subsequent MT, it will be checked whether the risk containing measures were/are sufficient, or if the risk has become an issue. The risk register will be reviewed and updated in the MT on an approximately quarterly basis.

3.5 Project templates

The DIVERSIFAIR project intends to use a consistent 'project style'. This is implemented by providing templates for the deliverables, the presentations and dissemination materials. This document, for example, is based on the structure of the project handbook[3] of the uCARe consortium, led by TNO. More project style templates can be produced by WP5 when needed. All available project style templates can be found on SharePoint in the folder 'Templates'.

4. Tools and collaboration infrastructure

4.1 Access to collaboration tools

Partner managers should inform Project management support in case an existing member leaves, or a new member joins. For new members, please provide a name, e-mail address and a mobile telephone number (able to receive a text message (SMS) for the 2-stage authentication) to facilitate creating a partner account. This partner account will provide access to the DIVERSIFAIR SharePoint (Section 4.2) and the Teams site (Section 4.3). All project members have to provide their contact details in the project member list. If a project member leaves the project, the partner account will be deactivated, and the member will be removed from the Teams channels.

4.2 Document sharing

A key element in a project like DIVERSIFAIR is collecting, sharing and analyzing information, and the collaborative production of reports on the results. For both purposes a SharePoint environment has been created. Within this SharePoint environment directories are available to store work documents for each WP, and to store deliverables. Furthermore, lists will be maintained for project members and external contact persons; see Figure 4-1.

[3] uCARe-D1.6-v1.0, Guiding document for pollutant reducing operations and maintenance of NRMM, PTW and HD, S. Hausberger, 05-11-2019

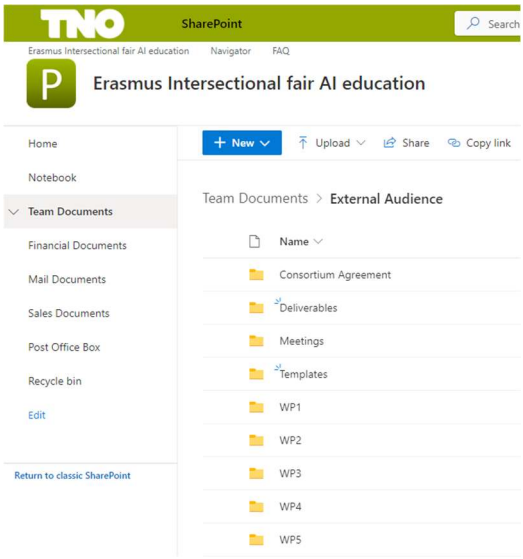

Figure 4-1: Snapshot of the SharePoint environment of DIVERSIFAIR.

4.3 Day to day information exchange

Day to day information exchange may be based on e-mail, telephone and via the chat function of the Teams channels listed in Table 4-1. All project members will be granted access to the Teams site, as well as the general channel. In addition, each work package has its own channel. These Teams channels serve to exchange messages and links to the documents that are stored on SharePoint, rather than the documents themselves, to avoid creating parallel versions.

Please do not include a document larger than 50 kB in an e-mail. Instead, store the document on SharePoint, and send the link via e-mail.

Table 4-1: Teams channels related to the DIVERSIFAIR project.

Name	Purpose
P060.53869-Erasmus Intersectional fair AI education-02	Teams site for DIVERSIFAIR that hosts all channels
General channel	Messages relevant for all project members
WP1 – project management	Messages only relevant to project members involved in the respective WP
WP2 – Concepts of intersectional fairness	
WP3 – Case studies and SOTA	
WP4 – Implementation	
WP5 – Communication and impact maximization	

4.4 Online meetings

Online meetings will be organized via Teams.

4.5 Quarterly Progress Reports (QPR)

To make sure that effort and expected results are balanced well, each partner should report their effort to the MT on a quarterly basis. To this end, QPR will be used, which is an Excel based tool where each partner reports the person months spent in the recently closed quarter for each WP. Figure 4-2 shows a part of the QPR Excel sheet for the first 3 months. To minimize the effort for each partner manager,

only the person months per WP are monitored. Direct costs will only be reported in the EU review after 18 and 36 months.

partner	TNO	Cortexter	Eticas	NUID UCD	Sciences PO	Turing	Women4Cyber	Women in AI
WP1								
WP2								
WP3								
WP4								
WP5								

Figure 4-2: Fragment of the QPR.

The consortium management will consolidate all partner inputs. The MT will monitor whether the reported effort matches the expected output of each partner.

QPR timeline:

- The partner managers receive a request for QPR reporting on the first working day of the month after closing a quarter.
- The partner manager reports the effort at the latest on the 15th of the month after closing a quarter.
- The consolidated QPR report is available at the latest on the 22nd of the month after closing a quarter and will be on the agenda of the first MT after the 22nd.

5. Ethical guidelines

Ethics is an integral part of responsible research, from the conceptual phase to the publication of research results. The DIVERSIFAIR consortium considers adhering to ethical guidelines as a continuous effort and is clearly committed to not only inventorize relevant ethical aspects upfront, but also proactively identify and mitigate potential ethical issues that may arise during the course of the project. Therefore, the first version of this project handbook covers the ethical guidelines based on a first assessment of ethical aspects that the DIVERSIFAIR consortium will take into account. This will be periodically reviewed by and discussed with external experts. If additional ethical aspects come up during the course of the project that require additions to or modifications of the ethical guidelines, the project handbook will be updated accordingly.

The initial guidelines capture the following:

- DIVERSIFAIR will adhere to the European Code of Conduct for Research Integrity from All European Academies, which is recognized by the European Commission as the reference document for all EU-funded research projects[4].
- The DIVERSIFAIR consortium is aware that it (as any consortium) does not represent all (intersections of) groups experiencing unfairness. The consortium will always put an effort to communicate about the potential known harms of our results, whilst taking an active stance

[4] ALLEA (2023) The European Code of Conduct for Research Integrity – Revised Edition 2023. Berlin. DOI 10.26356/ECOC; https://allea.org/code-of-conduct/#:~:text=The%20European%20Code%20of%20Conduct%20for%20Research%20Integrity%20serves%20the,and%20for%20all%20research%20settings.

to understand harms of those demographic memberships that are not represented in this consortium.

- The DIVERSIFAIR consortium will be open and transparent in their communication, internally as well as externally (e.g. to fairly select students that will obtain a voucher to participate in courses).
- DIVERSIFAIR will fairly select participants for training courses.
- Although a goal is to detect bias in AI algorithms, DIVERSIFAIR will non-judgmentally address organizations that allow the consortium to audit their algorithms, thus focusing on opportunities to increase fairness, i.e. the positive side of detecting and mitigating bias.

5.1 Data protection and privacy

First of all, it should be noted that the participation in Algorithmic Audits (WP3) hardly generates any threat to the privacy of the participating organization. Likewise, participation of students to the educational program (WP4) is not expected to generate a significant threat to their privacy. However, to also account for exceptional cases with a small risk, DIVERSIFAIR will take appropriate measures. During any data collection process data protection issues involved with handling of personal data will be addressed by the following strategies:

- Volunteers to be enrolled will be exhaustively informed, so that they are able to autonomously decide whether they give their consent to participate or not. The purposes of the research, the procedures as well as the handling of their data (protection, storage) will be explained. For online interviews these explanations will be part of the initial briefing of interviewees, for face-to-face interventions informed consent (see below) shall be agreed and signed by both, the study participants as well as the respective research partner.
- The data exploitation will be in line with the GDPR and the respective national data protection acts. Since data privacy is only under threat when data are traced back to individuals – they may become identifiable and the data may be abused – we will anonymise all data. Furthermore, where identification data is not required by the research task at hand, those data shall not be recorded, following the privacy by design principle.
- For the Algorithm Audits, the participating organization, e.g. a Civil Society organization, might act as data gathering organization and provide only the anonymised data to DIVERSIFAIR. The data gathered through questionnaires, interviews, observational studies at the workplace, focus groups, workshops and other possible data gathering methods during this research will be anonymised and therefore, the data cannot be traced back to the individual.
- Data will be stored only in anonymous forms so the identities of the participants will only be known by the research partners involved. If raw data need to be shared among the consortium partners, this will only be done after a confidentiality agreement (template to be aligned upfront) has been signed.
- Reports based on interviews, focus group and other data gathering methods will be based on aggregated information and comprise anonymous quotations respectively.

The strategies above will be elaborated per audit/educational program and require the approval of the Data Protection Officer (DPO) of:

- The data collecting organisation, either a consortium partner and/or an external organization participating in an Algorithm Audit or educational program
- TNO if the data collecting organisation has no DPO.

In addition, the consent of the DPO may be required in case:

- A consortium partner acts as DIVERSIFAIR contact person for Algorithm Audits or educational programme.

The DPOs of the partners, including their contact data, are listed in Table 6-1.

Table 6-1: Contact data of Data Protection Officers.

Organization	DPO Name	E-mail	Phone
TNO	Remy van den Boom	Remy.vandenboom@tno.nl	+31615083470
University College Dublin	n/a		
IEPP France	n/a		
ETICAS	n/a		
TURING College	n/a		
Women4Cyber	n/a	n/a	n/a
Women In AI	n/a		
CorTexter	n/a		

5.2 Communication strategy

Participants to Algorithm Audits and/or educational program will be made aware of the potential benefits and identified risks of participating in the project at all times. The main means of communicating benefits and risks to the individual is by means of informed consent (template to be aligned upfront). Prior to consent, each individual participant in any of the studies in DIVERSIFAIR will be clearly informed of its goals, its possible adverse events, and the possibility to refuse to enter or to retract at any time with no consequences. This will be done through a project information sheet or the informed consent form and it will be reinforced verbally. In order to make sure that participants are able to recall what they agree upon when signing, the informed consent forms will be provided in the native language of the participants. In addition, the consortium partners will make sure that the informed consent forms are written in a language suitable for the target group(s). The DIVERSIFAIR contact person for Algorithmic Audits or educational program is responsible for the communication, either directly to the participants or indirectly via the external organization participating in an Algorithmic Audit.

6. Open access and open research data

The DIVERSIFAIR project firmly believes in openness to be a major factor for innovation. There are many examples of how open innovation and open source are successful models, especially in domains where many different stakeholders are required to bring about effective change. Openness has many facets. The most important ones for the DIVERSIFAIR consortium are:

- **Open project collaboration**. All partners are committed to developing (working for) relationships with external partners for mutual benefit. Making contacts with similar projects and establishing collaboration with potential organizations for Algorithmic Audits is considered beneficial for all. Open collaboration in DIVERSIFAIR is understood in a transdisciplinary way, opening research processes to the wider public and in particular allowing stakeholders to build upon the results of WP3, and access the educational program.

However, internal project communications and non-deliverable documentation will not be made publicly available and/or accessible.

- **Open source technology**. Code developed within the DIVERSIFAIR consortium is proprietary, as outlined in the Grant Agreement and/or Consortium Agreement.
- **Open access to scientific results**. From a scientific perspective, the consortium clearly favours open access to its scientific output, and (research) papers as well as methodologies will be made publicly available.
- **Open access to research data**. The general policy of the DIVERSIFAIR project is to apply "open by default" to its research data, with exceptions being made based on privacy, competitiveness, and ethical rules on anonymity as described above (section 5) are thus highly relevant and need to be agreed with each of the participants to Algorithmic Audits and the educational program. Moreover, the consortium will proactively check whether their research is accessible for target audience feedback, and create manuals for access.

The open access strategy will be detailed in the following sections.

6.1 Open access strategy for publications

In line with the EC policy initiative on open access[5], which refers to the practice of granting free online access to research articles, the project is committed to follow a publication strategy considering a mix of both 'Green open access' (immediate or delayed open access that is provided through self-archiving) and 'Gold open access' (immediate open access that is provided by a publisher) as far as possible.

All deliverables (reports, software, data, media, other) labelled as "public" will be made accessible via the DIVERSIFAIR website (insert link when available). Here we will store deliverables with permanent identifiers.

Where appropriate, the results will also be published via ResearchGate (https://www.researchgate.net/), preferably via the accounts of scientists that already have a track record in this domain (i.e. no DIVERSIFAIR account). All outcomes of the project labelled as "public" will be distributed under specific free/open license, where the authors retain the authors' rights but the users can redistribute the content freely. In particular, the supporting material Algorithmic Audits such as leaflets, instruction video, etc., will be free for exploitation to facilitate replication and/or scaling up of these audits.

6.2 Data management plan (DMP)

This is a first version of the DMP for DIVERSIFAIR, which provides an analysis of the main aspects to be followed by the project's data management policy. The DMP evolves in the course of the project and will be updated accordingly as research data is collected.

The DMP is particularly relevant to the the Algorithmic Audits and monitoring the impact of the educational program.

It is expected that the project will produce the following open research data:

WP1:

- None besides the indicated deliverables.

WP2:

- None besides the indicated deliverables.

WP3:

[5] Research and innovation (europa.eu)

- All indicated deliverables
- Results of Algorithmic Audits

WP4:

- All indicated deliverables
- Results from satisfaction questionnaires
- Results of workshop to assess the bias detection solution

WP5:

- All indicated deliverables
- Results from surveys/stakeholder interviews

7. Conclusions and recommendations

This handbook describes the main procedures of the DIVERSIFAIR project to operate successfully and effectively in order to achieve high quality project results following a responsible research and innovation approach. Open access, ethics, and engagement of societal actors are amongst the key elements. While this handbook is provided in the form of a report and deliverable it is a living document and may be challenged and updated by the consortium during the project. The processes described here are implemented in the daily operations of the consortium, and most of the elements are separately available on SharePoint. The management reports will include updates on any crucial changes in the handbook.