Towards Real Diversity and Gender Equality in Artificial Intelligence

Advancement Report

November 2023

GPAI | THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

This report was developed by Experts and Specialists involved in the Global Partnership on Artificial Intelligence's project "Towards Real Diversity and Gender Equality in Artificial Intelligence: Evidence-Based Promising Practices and Recommendations". The report reflects the personal opinions of the GPAI Experts and External Experts involved and does not necessarily reflect the views of the Experts' organisations, GPAI, or GPAI Members. GPAI is a separate entity from the OECD and accordingly, the opinions expressed and arguments employed therein do not reflect the views of the OECD or its Members.

Acknowledgements

This report was developed in the context of the project *"Towards Real Diversity and Gender Equality in AI: Evidence-Based Promising Practices and Recommendations*," with the steering of the Project Co-Leads and the guidance of the Project Advisory Group, supported by the GPAI Responsible AI Working Group. The GPAI Responsible AI Working Group agreed to declassify this report and make it publicly available.

Co-Leads:
Paola Ricaurte Quijano*, Tecnológico de Monterrey
Benjamin Prud'homme‡, Mila - Quebec Artificial Intelligence Institute

The report was prepared by: **Niobe Haitas‡**, Mila, with the contributions of the project co-leads **Paola Ricaurte Quijano***, Tecnológico de Monterrey and **Benjamin Prud'homme†**, Mila, **Wanda Muñoz‡**, Feminist AI Research Network, and **Stephanie King**, CEIMIA, as well as research partners **Shamira Ahmed‡**, Data Economy Policy Hub, **Leslie Evelin Salgado Arzuaga‡**, University of Calgary, **Shazade Jameson‡**, Tilburg University, **Florian Lebret‡**, Université Laval, **Ana Gabriela Ayala Núñez‡**, Tecnológico de Monterrey, **Ivanna Martínez Polo‡**, Tecnológico de Monterrey, **Gargi Sharma‡**, CLIMA Fund; **Selene Yang‡**, Global f<a+i>r network & Geochicas and **Razieh Shirzadkhani‡**, Mila.

We want to thank the members of the Project Advisory Group for their support throughout the process: **Lucia Velasco***, School of Transnational Governance, European University Institute; **Karine Gentelet†**, Université du Québec en Outaouais (UQO) & International Observatory on the Societal Impacts of AI and Digital Technology (OBVIA), **Nicole Kaniki†**, Senomi Solutions Inc, **Ricardo Baeza-Yates***, Institute for Experiential AI of Northeastern University; **Karen de Brouwer Vásquez†**, Community of Practitioners and Experts in Project Management for Results in Development in Latin America and the Caribbean (COPLAC); **Kudakwashe Dandajena***, African Institute for Mathematical Sciences (AIMS) and University of the Western Cape; **Golnoosh Farnadi†**, McGill University and Mila; **Ivana Feldfeber‡**, DataGénero; **Alison Gillwald‡**, Research ICT Africa; **Gloria Guerrero‡**, ILDA; **Toshiya Jitsuzumi***, Chuo University; **Ching-Yi Liu***, National Taiwan University; **Catherine Régis***, Université de Montréal and Mila; **Michael Running Wolf†**, Indigenous AI and Mila; **Juliana Sakai***, Transparência Brasil; **Prateek Sibal***, UNESCO; **Elissa Strome†**, Canadian Institute for Advanced Research (CIFAR); **Jaco du Toit****, UNESCO;Shaz **Eliane Ubalijoro†**, CIFOR-ICRAF, **Nicole Osayande†**; Mila, **Jessica Fjeld†**, Harvard Law School; **Zinnya Del Villar‡**, Data Pop Alliance, **Inese Podgaiska***; Association of Nordic Engineers and **Anupama Srikonda†**. We also want to express our appreciation to the multiple organizations and individuals who have contributed and continue to contribute to the project, such as: **Research ICT Africa**, **Data Pop Alliance** and **Derechos Digitales** for conducting the regional consultations.

GPAI wants to acknowledge the tireless efforts of colleagues at the International Centre of Expertise in Montréal on Artificial Intelligence (CEIMIA) and GPAI's Responsible AI Working Group. We are grateful, in particular, for the support of **Camille Seguin**, **Laëtitia Vu** and **Stephanie King** from CEIMIA, and for the dedication of the Working Group Co-Chairs **Raja Chatila***, Sorbonne University and **Catherine Régis***, Université de Montréal and Mila.

* Expert
** Observer
† Invited Specialist
‡ Contracted Parties by the CofEs to contribute to projects

Citation

GPAI 2023. *Towards Real Diversity and Gender Equality in Artificial Intelligence: Advancement Report*, November 2023, Global Partnership on AI.

Table of Contents

Table of Figures

Executive Summary

This is an Advancement Report for the Global Partnership on Artificial Intelligence (GPAI) project "Towards Real Diversity and Gender Equality in Artificial Intelligence: Evidence-Based Promising Practices and Recommendations." It describes, at a high level, the strategy, approach, and progress of the project thus far in its efforts to provide governments and other stakeholders of the artificial intelligence (AI) ecosystem with recommendations, tools, and promising practices to integrate Diversity and Gender Equality (DGE) considerations into the AI life cycle and related policy-making.

The report starts with an overview of the human rights perspective, which serves as the framework upon which this project is building. By acknowledging domains where AI systems can pose risks and harms to global populations, and further, where they pose disproportionate risks and harms to women and other marginalized populations due to a lack of consideration for these groups throughout the AI life cycle, the need to address such inequalities becomes clear.

Stakeholders of the AI ecosystem (e.g., governments, academia, industry, and civil society) currently lack practical, evidence-based guidance on what measures to take to address DGE in ways that comply with human rights, can be evaluated, and can demonstrate a tangible, sustainable impact (UN General Assembly 2015). Taking meaningful, measurable action is necessary. According to the United Nations (1979) and to good practices in advancing DGE considerations, responding to today's challenges necessitates adopting a twin-track approach, which includes actions and indicators in two areas:

1. **Targeted interventions to help persons from marginalized groups** (and their representative organizations) **to know their rights and increase their capacities.** These interventions must exist to improve knowledge and networks, while increasing capacities to equally access rights, resources, and opportunities, and participate in decision-making.

2. **Mainstreaming gender and diversity**: A strategy to make DGE a cross-cutting issue in all initiatives. All institutions – companies, academia, and generally, all systems – must **take concrete action to ensure that systemic barriers that perpetuate the exclusion of women and marginalized groups are eliminated**. This may include, for instance, funding and action plans to incorporate anti-racist, feminist and enabling initiatives and procedures.

As the UN Human Rights Council (2015) affirms, "*inclusion is not only about including those who are traditionally excluded but must also be about dismantling the many forms of discrimination that contribute to the persistent marginalization of groups on the basis of arbitrary distinctions, such as their age, their gender or the color of their skin.*"

Achieving equality requires a shift from mere equal treatment to a focus on effective action on DGE, including resources, monitoring, and reporting. This shift also means that existing power structures must be challenged, including the distribution of power and resources. Failure to do so risks perpetuating and amplifying the current culture and values that dominate our world order, particularly within the technology industry, to the detriment of diversity and equality, and thus of human rights.

"Tokenistic" strategies are insufficient to address the harms caused or amplified by AI systems, and the lack of DGE in AI ecosystems. Research (UNESCO, 2020; IMF, 2018; OECD, 2015; ILO, 2009) consistently demonstrates that increasing DGE enhances the quality, usability and effectiveness of AI products, improves the achievement of their intended outcomes, and significantly boosts economic opportunities and national GDPs. These outcomes are important to everyone.

To achieve these aims, the project is undertaking a number of tangible actions to assess the current state of the intersection of DGE and AI, including a literature review; regional consultations with diverse stakeholders (civil society, academia, government and industry), focusing on individuals and organizations who identify as and represent marginalized groups; collecting, mapping and analyzing existing initiatives globally, to provide examples of promising practices and resources to integrate DGE in AI; and an environmental scan of DGE efforts in the AI ecosystem. This report provides an overview of the structure, methodologies, and current status of each of the above-mentioned activities. It also reflects on lessons learned to date, and lists the next steps to be taken.

Throughout its activities, the project places particular emphasis on intersectionality, encompassing identities and characteristics such as gender, race, ethnicity, disability, among others, and their intersections. Given the global scope of the project and the need to integrate numerous perspectives, it is essential to pursue meaningful engagement with marginalized groups, to be able to best capture their perspectives on issues specific to their communities and AI.

The project team hopes that this advancement report can serve as an initial exploration of these crucial issues, building on existing international dialogues among stakeholders of the global AI ecosystem surrounding the responsibilities of all parties to respect and honour human rights.

Diversity and Gender Equality in AI: A Human Rights Perspective

AI offers a wide range of possibilities to enhance the well-being of different groups and contribute to the Sustainable Development Goals (Vinuesa et al., 2020). However, its use can deepen economic, knowledge, gender, and cultural divides (UNESCO, 2021). Indeed, while AI can enhance efficiency and productivity, it also raises profound ethical concerns, given its potential to embed and amplify biases, contribute to environmental and ecological degradation, exacerbate existing inequities, further harm marginalized groups, and undermine human dignity, through prejudice, discrimination, and stereotyping (Amrute et al., 2022; UNESCO, 2023; A/HRC/44/57).

Today, AI is still generally designed, developed, monitored, and evaluated without systematic DGE approaches. This has negative consequences: it precludes AI from achieving its potential for social good, and increases harm to already marginalized groups. The 2020 Report by GPAI's Responsible AI Working Group and The Future Society found that AI initiatives (including those that self-identify as "AI for SDG" or "ethical AI") do not sufficiently include *"the Global South and marginalized communities, such as people with disabilities, indigenous groups, the LGBTI+ community, persons living below the poverty line and migrants. There are of course notable and recent exceptions, such as the A+ Alliance for Inclusive Algorithms. This lack of diversity risks undermining the effectiveness and credibility of Responsible AI initiatives as well as their ability to scale. Importantly, it risks perpetuating existing inequalities and biases and misinforming policy priorities. This is particularly problematic for cross-regional collaborations*" (GPAI, 2020).

International laws and frameworks, such as the 2030 Agenda for SDGs (A/RES/70/1), recognize women's and diverse populations' right to participate in all decision-making processes. While there has been some progress, these populations remain largely excluded from AI as active agents and decision-makers (Arora et al., 2023), even though ethical issues raised by AI systems are impacting critical areas such as governance, social interactions, democracy, rule of law, environment, data protection, gender equality, and protection of human rights:

- **Employment:** AI systems are narrowing employment opportunities (Barbieri et al., 2021) and making it difficult to enforce anti-discrimination laws in the context of disability (Whittaker et al., 2019);
- **Housing:** AI systems are undermining housing equality (So et al. 2022);
- **Health care:** AI systems are used to monitor and censor women, putting reproductive rights at risk (Peña & Varon, 2019b);
- **Misinformation, disinformation, and hate speech:** AI is directly linked to the increase in misinformation, disinformation and hate speech online (A/HRC/44/57). Certain AI tools can now create convincing, realistic-looking explicit content or pornographic deep fakes (Hunter, 2023). Such instances add to the harassment faced online, especially by women, and highlight the insufficiency of current laws in protecting women online (Kelleher, 2023);
- **Migration:** AI systems have been used to further punitive border policies, preventing already vulnerable people from seeking asylum and exposing them to the risk of refoulement (McGregor & Molnar, 2023; Dumbrava, 2021; Lehtonen & Aalto, 2017; Parks & Caplan, 2017). AI is also being used to track facial expressions or recognise emotions at security checkpoints to decide whether or not an individual is a threat (Podoletz, 2023);

- **Weapons:** Increased autonomy in the critical functions of weapons systems can lead to life-and-death decisions being delegated to AI systems, raising humanitarian, legal, ethical and security concerns (UNSG & ICRC, 2023). Autonomous weapons systems could affect marginalized groups disproportionately, including persons with disabilities (A/HRC/49/52);

- **Social services:** AI systems trained on biased data sets are allocating fewer resources and less support to persons with disabilities (A/HRC/49/52). Public-private ventures are harming girls of Indigenous and/or immigrant backgrounds and people living in poverty (Balmaceda et al., 2023);

- **Reinforcing fabricated categorizations:** Facial recognition systems have higher error rates for dark-skinned and female faces, and impose and entrench stereotypes of what different genders are supposed to look like (Buolamwini & Gebru, 2018; Ciston, 2019; Costanza-Chock, 2018). Furthermore, data categorization, in general, has been conducted in a top-down manner and has reinforced classifications created not by the data subjects, but by those with a vested interest in wielding power over them (Masiero & Das, 2019);

- **Carceral system:** AI systems are used to engage in mass surveillance of public spaces (Perez-Desrosiers, 2021). Moreover, using AI in the criminal justice system is entrenching stereotypes and undermining the presumption of innocence and fair trials (Benjamin, 2016; Tahir, 2019; Angwin et al., 2016; Jansen, 2018).

These examples reveal a pattern where AI systems pose disproportionate risks and harms to women and marginalized groups. Because these groups have not been considered, there are missed opportunities for AI to be more impactful. Many recent initiatives have highlighted both the positive impact and risks of AI for gender equality (UNESCO, OECD, & IADB, 2022), but **the AI ecosystem lacks practical, evidence-based guidance on what measures should be taken to address DGE** in ways that comply with human rights, can be evaluated, and can demonstrate impact.

To address the inequality reflected in AI ecosystems, a systemic approach based on human rights principles is required (Prabhakaran et al., 2022). This approach acknowledges that historically structured inequalities lie at the heart of marginalization and exclusion. By confronting these uncomfortable realities at the individual, community, institutional and systemic levels, we pave the way for effective means to address them and create opportunities for social transformation. This in turn fosters meaningful inclusion, inclusive growth, sustainable development and more peaceful societies, all of which are objectives of the 2030 Agenda for SDGs to which GPAI is committed. Justice, human rights and equality must occupy a central – rather than peripheral – role in the design, development, deployment and use of AI: to increase its effectiveness, to promote trust in AI, and because it is a human right for everyone, including marginalized groups, to participate in shaping our future societies, which includes and requires benefiting from advances in technology.

Project Objectives

The project's overall objective is to contribute to ensuring that AI ecosystems have the appropriate tools, frameworks, and resources to incorporate effective DGE strategies throughout the AI life cycle, and to demonstrate their impact with indicators in accordance with human rights, the OECD AI Principles, and the UNESCO Recommendation on the Ethics of AI. The project emphasizes intersectionality, considering factors such as race, ethnicity and disability, among others. Specifically, this report provides an update on how the project's activities seek to respond to existing

gaps and contribute to ensuring a just, responsible, and inclusive AI life cycle. The project's three goals are to:

1. frame the intersection of AI with DGE issues as a matter of human rights within the global context, taking a holistic approach based on an intersectional perspective;

2. document and represent the voices of marginalized groups, including showcasing diverse promising practices and resources; and

3. propose practical and actionable recommendations for moving forward.

The project's final report will present challenges and recommendations in an empirically grounded and accessible manner, based on dialogues with representatives of marginalized groups, promising practices, and existing resources both from the AI ecosystem and from initiatives to advance DGE in other sectors.

Finding Common Ground

While United Nations organizations (e.g., UNWOMEN, and UNICEF 2017) have adopted working definitions of gender equality and diversity-related terms, there is currently no globally shared understanding of definitions in this domain – a key challenge encountered by the project. Stakeholders sometimes use the terms *gender* and *diversity* interchangeably, or one word may predominate over the other. To facilitate a common understanding of these issues, the project is developing a Glossary to provide definitions of the key concepts that guide its research. As part of this work, the project will consider existing conceptual frameworks established by the Human Rights conventions and intergovernmental organizations, including defining gender, gender equality, gender equity, gender discrimination, racial discrimination, diversity, intersectionality, and disability.

A Systemic Approach

To advance the project's work, it is crucial to understand the problems associated with the lack of DGE in AI as a result of social inequities and power imbalances. This problem is systemic and structural, and as such, technical fixes are insufficient. The task at hand therefore encompasses more than rectifying biased data sets or algorithms to make them representative or accurate. Addressing inequality in AI ecosystems systemically and structurally, following a Human Rights framework, involves a critical examination of:

- Current human resources, social, education, and other practices that represent social, economic, attitudinal, physical, communication and other accessibility barriers for marginalized groups;

- The relevance, effectiveness, and appropriateness of AI-based interventions, particularly in well-resourced countries;

- Foundational policies and their underpinning social decisions, and the inclusiveness of AI policy-making decisions;

- Decision-making related specifically to the use of public resources for AI-powered solutions in order to achieve specific policy outcomes;

- The process by which human rights guide AI developments, and are monitored, at every stage of the AI life cycle; and

- The existence of sufficient systems in societies to ensure accountability, guarantees of non-repetition, and reparation when damage is caused by the use of AI.

This critical examination involves undertaking work along two paths: (1) **improving the capacities of marginalized groups** related to AI and the knowledge of human rights approaches, and how to participate in AI-related forums; and (2) **examining the policies, practices and procedures** of AI stakeholders **to eliminate existing barriers to equal participation of women and marginalized groups**. Despite countless efforts to explain the issue and the evidence of the harms caused, there has been little advancement to date. This project endeavours to tackle root causes; provide a common ground on the issues and a conceptual understanding of the terminology; and offer actionable recommendations, resources, and tools for stakeholders of the AI ecosystem to increase DGE throughout the AI life cycle, and in all AI policies and governance frameworks.

Towards Real Diversity and Gender Equality in AI: A GPAI Endeavour

With the aim of collecting a wide range of perspectives and voices from around the world on evidence-based promising practices and recommendations, the project seeks to move beyond mere identification of problems (the "what"), to explore implementation strategies, frameworks, and tools (the "how") to integrate DGE approaches in AI. Moreover, the final project outputs will focus on identifying the necessary conditions for governments and other stakeholders to effectively implement these solutions, tools, and frameworks. There has been significant progress across all project activities to date, as outlined in the following sections. However, all project components remain in a continuous state of development, with the final delivery of outputs set for 2024.

Research Design

The project's approach is founded primarily on a participatory, qualitative, and desk-based research methodology, with significant emphasis placed on the engagement with the stakeholders consulted, including our Project Advisory Group (PAG). The project's activities comprise five key components:

1. **Literature Review:** Aiming to identify key themes, theories, methodologies, and gaps in the literature on the topic of integrating DGE into the AI life cycle. This review covers existing practices and critiques of the current discourse around, and practices related to, DGE in AI.

2. **Regional Consultations:** Engaging localized expertise from five regions worldwide: Latin America and the Caribbean, Sub-Saharan Africa, Middle East and North Africa, North America and Europe, and Asia and the Pacific. Stakeholders included representatives from academia, civil society, industry, and government. A participatory qualitative data collection methodology was developed based on an interview guide and an iterative approach to explore emerging themes in collaboration with delivery partners. Consultations are ongoing.

3. **Community Perspectives:** Outreach conducted by sharing a first version of the final report with civil society organizations and persons self-identifying as members of marginalized groups for additional feedback. These individuals and organizations also responded to a set of questions on the integration of DGE in AI.

4. **Promising Practices and Resources:** A selection of existing initiatives or solutions aiming to integrate DGE in AI, and a more elaborate analysis through use cases. Such initiatives include technical, capacity building, policy, and community engagement processes.

5. **Environmental Scan:** A mapping of existing initiatives that aim to include DGE in the AI life cycle, complete with labelling and annotation.

To validate the project's methodological steps, a Project Advisory Group was formed for regular consultation. The members of this group include GPAI Experts and External Experts/Specialists, in order to benefit from their expertise in topics such as diversity, equality, and issues specific to Indigenous communities. Members of this group include representatives from organizations such as UNESCO, Université de Montréal, Université du Québec en Outaouais, Research ICT Africa, Mila, CIFAR, Center for International Forestry Research and World Agroforestry (CIFOR-ICRAF), Transparência Brasil, Chuo University, Princeton University, Senomi Solutions, University of the Western Cape, Universitat Pompeu Fabra, Iniciativa Latinoamericana por los Datos Abiertos (ILDA), COPLAC, DataGénero, and Indigenous AI. The project is grateful for the support of the Project Advisory Group thus far, and is keen to continue to expand this group of individuals as needed to ensure diverse and robust representation, especially from the groups and intersections that this project seeks to explore as part of its objectives.

Literature Review

The scope of the literature review is twofold, seeking to collect (1) principles that have been established by top-down governance institutions; and (2) specific solutions that address concerns named by communities experiencing (or at risk of experiencing) the harms of AI development, particularly harms that result from the oppressive structures that underpin various facets of society (e.g., social, economic, political, epistemological, spiritual, health, technological relations, etc.). The research question posed was, "What are the principal socio-technical challenges, trends, gaps, voids, solutions (social and technical) that address Equity, Diversity and Inclusion in AI, in particular addressing diverse and marginalized communities of the world, and countries of the Global Majority?" The review focuses on the (in)equities in the AI life cycle highlighted by international organizations, such as United Nations bodies and the European Union, academic researchers, and most importantly, civil society organizations that support the interests of those minoritized by the status quo. By centring the perspectives of those who have been silenced or minoritized, or those not considered as part of the development of AI, the literature review also seeks to provide a brief glimpse into what equitable AI governance and technological relationships can look like.

Regional Consultations

Design

To conduct regional consultations, the project collaborated with regional experts (delivery partners) across five regions: sub-Saharan Africa, the Middle East and North Africa (MENA), Latin America and the Caribbean, North America and Europe, and Asia and the Pacific. Delivery partners included Data Pop Alliance, Derechos Digitales, and Research ICT Africa, as well as a number of distributed Research Associates. Delivery partners were selected based on their availability and capacity to conduct the consultations in a short period, as well as their existing networks. Delivery partners mapped key stakeholders and organizations for each region, streamlining the organization of the regional consultations.

To ensure consistent data collection for later comparison and aggregation, we devised a methodological concept note and a set of tools for the delivery partners. These resources were translated into French, Spanish, and Portuguese, where relevant, and included:

- **Methodological Concept Note:** Outlining the overarching principles within which the project required these consultations to be held.

- **Interview Guides:** A list of questions adapted to different stakeholders (academia, civil society, government, and industry) and targeted groups (e.g., Indigenous communities). The guide aimed to explore gaps and opportunities in the field of DGE in AI. Interview guides were used for individual interviews, and modified iteratively based on initial findings, allowing for validation of those findings during subsequent round-table discussions. Questions were grouped into the following themes:

    - Understanding needs and expectations;
    - Existing resources, programs and initiatives;
    - Overcoming barriers and challenges; and
    - Proposing ideas and solutions.

- **Sociodemographic Questionnaire (Optional):** Questions related to respondents' diverse characteristics, such as gender identity, sexual orientation, background, religion, and ancestral heritage. Participants' responses contributed to our efforts to promote inclusivity and cultural awareness, and enhanced our ability to track DGE progress and gain deeper insights into effective DGE strategies. The development of the questionnaire was the result of in-depth discussions on terms relating to race and ethnic backgrounds with various experts in the field, including Equity Diversity and Inclusion specialist Nicole Kaniki.

- **Informed Consent Form:** Outlining the project's objectives, procedures, potential benefits, and associated risks, as well as steps undertaken to ensure data privacy and confidentiality. The participants could choose whether or not to consent to audio recording, and whether they preferred their contributions to remain anonymous.

- **Glossary:** Shared with participants in advance of the consultations, with the aim of creating a common understanding on AI, gender, and diversity terms;

- **Data Collection Database Templates (Online):** A Microsoft Excel template for data collection to facilitate anonymous reporting and standardized sharing of participant information, and a Microsoft Excel template to map organizations working in the field of AI and DGE, as well as representatives of marginalized organizations (per region);

- **Reporting Template:** With various sections setting out expectations for reporting on the regional consultations, based on Quality Standards for Reporting Qualitative Research (O'Brien et al., 2014);

- **Trint and Deepl Software:** Access to Trint software for the automatic transcription of audio files and access to Deepl software for the automatic translation of transcripts;

- **Recommendations for Data Privacy and Confidentiality:** Recommendations to delivery partners to ensure participants' privacy was appropriately and adequately respected.

Implementation

Participants consisted of a diverse array of stakeholders representing academia, civil society, industry and government. Their selection was designed to ensure a diversity of perspectives, and also depended on their availability and interest in participating in the consultation. Regional consultations took various forms, encompassing individual interviews, round-table discussions, and written contributions, with the overarching objective of capturing a wide spectrum of perspectives

and voices. Aiming for an inclusive approach, the project consulted with Indigenous Peoples in North and Latin America, and will continue to reach out to more groups and representatives of marginalized communities.

The interviews and/or round tables were conducted via teleconference platforms (e.g., Zoom or Big Blue Button). Interviews were one-on-one discussions with the designated researcher, lasting approximately one hour, whereas round tables involved a group discussion led by a facilitator to foster an open and interactive environment and encourage respectful and inclusive discussions. If participants consented, the sessions were recorded, and/or notes taken. Following the interview and/or round table, the recordings were transcribed. Additionally, anonymous socio-demographic information was collected from participants, to better understand the diverse perspectives contributing to this study.

The insights and discussions generated during the interviews and/or round tables were captured as qualitative data, and analyzed thematically to identify emerging themes, convergence and divergence of opinions, consensus among participants, promising practices, lessons learned, and knowledge sharing. Following the interview and/or round table, participants had the opportunity to review and validate whether their contributions had been accurately captured and represented in the draft report, providing additional feedback as and where needed.

For the regional consultations, the project reached out to more than 284 people, consulting with 180 people from a total of 46 countries. More specifically, there was representation from ten countries in Africa (Kenya, Nigeria, Zimbabwe, Cameroon, Mauritius, Uganda, South Africa, Tunisia, Swaziland, and Burkina Faso); nine countries in Latin America (Brazil, Ecuador, Colombia, Argentina, Paraguay, Mexico, Uruguay, Bolivia, and Chile); nine countries in North America and Europe (Belgium, France, Germany, Portugal, Romania, United Kingdom, Hungary, USA, and Canada); nine countries in the MENA region (Egypt, Israel, Iran, Lebanon, Mauritania, Morocco, Saudi Arabia, Tunisia, and the United Arab Emirates); and ten countries in the Asia-Pacific region (Australia, India, Japan, Nepal, Philippines, Singapore, South Korea, Thailand, Vietnam, and Malaysia). Statistics on the regional consultations can be found in Appendix 1.

A note on Indigenous Perspectives

Interviews with 16 Indigenous people from Latin America and North America were also conducted to understand their unique perspectives and shed light on the intersection of cultural values, identity, and the rapid evolution of AI. Interview guides were prepared, addressing the following topics:

- Cultural background and heritage;
- Personal experience with technology and AI;
- Representation, harm and community values;
- Governance and sovereignty;
- Public policy; and
- Technology and projects.

Six representatives of Indigenous Communities of North America (three from Canada, three from the USA) were interviewed. Four participants represented academia (three students, one professional) and two represented legal/policy/government. Additionally, ten interviews were conducted with people from indigenous communities in Southeast Mexico. The interviewees were speakers and/or revitalizers of the Maya, Tzotzil, Tzeltal, Chol, Mixteco and Zoque Ayapaneco languages.

Perspectives from Community Organizations

In addition to regional consultations, the project aimed to capture the perspectives of civil society organizations that represent marginalized groups (such as Indigenous Peoples, people with disabilities, gender and sex dissidences, immigrants, and refugees, among others). The project reached out to 19 civil society organizations representing such groups to ask them to share their perspectives on the intersection of AI and their community, which eight organizations and individuals agreed to do. The project offered compensation for reviewing a draft version of the outputs and sharing feedback with a specific focus on practical strategies, frameworks, and tools that facilitate the inclusion of their perspectives into the AI life cycle. The project also sought their responses to the following three questions:

1. What are your insights into the relationship between [*your marginalized community*] and AI in the context of addressing issues related to gender equality, diversity and AI?

2. What suggestions do you have for (a) governments, (b) industry players, (c) civil society organizations, and (d) academia to enhance their efforts in promoting diversity and gender equality within the field of AI?

3. In your view, does the project output accurately reflect your perspective on [*your marginalized community*] and AI? If not, how can we enhance its alignment with your viewpoint?

Promising Practices and Resources

The project has been working on identifying and highlighting effective, positive approaches, strategies, and methods regarding the integration of DGE in the AI life cycle. The vision for the project is that these initiatives can serve as inspirations for others to either adopt, adapt to their specific contexts, or replicate. For the selection of Promising Practices and Resources, an online search was conducted in July and August 2023, using the research terms *projects*, *AI*, *equality*, *inclusion*, and *gender*. There were 51 cases included in the initial results that addressed DGE in AI from various angles, from technical bias mitigation to developing policies for inclusive technologies. The project defined the following criteria to help short-list these cases:

A. diversity and inclusion as core to project goals,
B. transparent and accessible documentation,
C. providing an actionable, potentially replicable, approach, and
D. representing regional diversity.

To evaluate each practice against the defined criteria, an instrument/evaluation matrix was designed to curate a selection of 25 promising practices and resources, according to a feminist and socio-constructivist lens (MacKenzie & Wajcman 1999). It should be noted that defining a standardized methodology to evaluate impact is especially challenging where the goal is to move beyond the traditional definition of success in terms of numerical specifications.

Use Cases: Seeing Promising Practices and Resources in Action

The project will conduct further narrative explorations of the curated shortlist of 8 to 10 “Promising Practices and Resources,” resulting in detailed use cases of the ‘Promising Practices and Resources’ in action. The project aims for a selection that is diverse across regions, types of projects, and stakeholders, and will classify the use cases according to the type of stakeholder, region of origin, and nature of the contribution. The objective is to create space for a diversity of perspectives to be showcased, while maintaining a focus on actionable approaches. To delve more

deeply into the experience and analyses of our use cases, further interviews will be conducted with those users and/or impacted stakeholders from the short-listed use cases.

Environmental Scan

The project is also continuing to map existing DGE initiatives applicable to the AI life cycle, classifying each under parameters such as economic contexts, business functions, designs, model types, application areas, uses pertaining to people and planet, stakeholders, DGE topics, type of tools, etc. While the identification and mapping work is ongoing (and the classification categories are not yet finalized), promising initiatives such as toolkits, papers, guidelines, documents, and online resources are being identified; the current database includes over 400 resources. The findings from this effort should be shared in a repository format, for ease of accessibility and dissemination.

Learnings and Challenges for Inclusion

Conducting a project on sensitive topics such as DGE can present challenges. To date, several challenges have been faced when completing project activities. These are summarized below to share learnings from the project with the broader DGE and ethical AI ecosystems. The project also seeks to learn from these challenges, and remain aware of their potential impact on its final findings, conclusions, and recommendations.

Lack of Participation in Consultations

The project faced challenges in ensuring participation in the consultations, particularly in certain regions. In many cases, efforts to reach out to specific groups went unanswered. In the case of the MENA region, topics that could be deemed controversial or taboo, such as those related to sexual orientation and gender expression, affected the number of interviews, although many organizations and representatives of LGBTQIA+ communities were contacted. Many countries have attitudes, levels of acceptance, and laws that can result in limited acceptance and protection for these individuals. Ensuring participation was also challenging due to differences in time zones, which made scheduling difficult. Finally, participation by large proportions of some populations was hindered by language barriers, due to the vast number of languages spoken in particular regions, combined with low English proficiency.

Lack of Access to Technological Resources

Consultations to date were held virtually. Consequently, the project has only consulted with people who have access to these types of resources, thereby excluding people who are digitally marginalized. This is an important caveat for most regions where digital exclusion is widespread.

Unequal Representation of Stakeholder Groups

To date, the project has had more representation in consultations from civil society and academic representatives than from representatives of government or private industry.

Overrepresentation of Experts and Highly Educated Individuals

Despite the project's original intention to engage with both experts and non-experts, the interview guides were designed in a way that rendered them less accessible to non-specialized (specifically in DGE and AI) audiences. Additionally, as the socio-demographic questionnaire showed, most participants had a high level of education: approximately 84% of respondents had either a Master's

or Doctoral degree; see Figure 5 in Appendix 1. This group of participants therefore does not provide adequate representation of the voices of less formally educated or marginalized groups.

Lack of Alignment on DGE and AI Topics Prior to Consultations

Although the project intended to brief and prepare consultation participants on AI and DGE topics prior to their participation in the consultations, efforts were limited due to time constraints. The project was therefore unable to evaluate whether these briefing efforts were sufficient or adequate. As such, it is possible that participants may not have felt fully prepared to participate in the discussion, particularly on issues outside of their own domains of expertise.

Looking Forward

Following these initial stages and learnings, the immediate next steps will be to:

A. Strengthen the methodology with specific actions to include voices of those groups who are currently underrepresented in the work, which will include further mapping and contacting of additional organizations representing marginalized groups; and

B. Identify additional promising practices and use cases, and describe in detail those that have already been identified.

These additional actions will allow the project to strengthen its coherence between the subject matter (equality) and its methodology, and enable it to provide more precise, specific recommendations to the AI ecosystem. Further specific actions will include:

- Updating the project's governance model to form an Advisory Committee with representation of marginalized groups to participate in designing and drafting the final report;
- Conducting a mapping exercise of international and regional representative organizations of marginalized groups, which, if they consent, will be included in the final report.
- Reaching out to additional organizations and people representing marginalized communities for consultations, such as Black in AI and Māori groups from New Zealand, people on the move (e.g.,migrants, refugees, displaced persons, asylum seekers, etc.), LGBTQIA+ individuals, neurodiverse persons, and persons with disabilities;
- Fine-tuning project recommendations by type of stakeholder;
- Preparing specific recommendations to include, and eliminate barriers to the inclusion of, specific groups in the AI ecosystem;
- Develop the platform and/or repository for dissemination of the findings of the environmental scan; and
- Prepare a repository of "Promising Practices and Resources" classified by topic, language, and stakeholder type.

The GPAI project 'Towards Real Diversity and Gender Equality in Artificial Intelligence: Evidence-Based Promising Practices and Recommendations' welcomes inputs from the broader community to help build upon its existing endeavours, towards delivery of final outputs in 2024. The project hopes the work conducted as part of this effort will help GPAI Members, and the global AI community more broadly, to adopt, adapt, and replicate impactful DGE initiatives into all respective contexts— creating meaningful, measurable action in advancing diversity and gender equality in AI.

References

[1] Amrute S., Singh R., Guzman R.L. (2022) *A primer on AI from the Majority World*. Data and Society. Available from: https://datasociety.net/wp-content/uploads/2022/09/09142022_AIMW_Primer_eng_final.pdf

[2] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016) *Machine bias*. ProPublica. Available from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[3] Arora, A., Barrett, M., Lee, E., Oborn, E., & Prince, K. (2023) *Risk and the future of AI: Algorithmic bias, data colonialism, and marginalization*. Information and Organization, 33(3), 100478. https://doi.org/10.1016/j.infoandorg.2023.100478

[4] Balmaceda, T., Pedace, C., & Schleider, T. (2023) *What Artificial Intelligence is hiding: Microsoft and vulnerable girls in northern Argentina.* Transnational Institute. https://www.tni.org/files/2023-04/What%20Artificial%20Intelligence%20is%20hiding.pdf

[5] Barbieri, D., Caisl, J., Lanfredi, G., Linkeviciute, J., Mollard, B., Ochmann, J., Peciukonis, V., Reingarde, J., & Kullman, M. (2021) *Artificial intelligence, platform work and gender equality.* Available from*:* https://eige.europa.eu/sites/default/files/documents/artificial_intelligence_platform_work_and_gender_equality.pdf

[6] Benjamin, R. (2016) *Catching our breath: Critical race STS and the carceral imagination.* Engaging Science, Technology, and Society, 2, 145-156.

[7] Buolamwini, J., & Gebru, T. (2018, January). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. In *Conference on Fairness, Accountability and Transparency* (pp. 77-91).

[8] Ciston, S. (2019) *Intersectional AI is essential: Polyvocal, multimodal, experimental methods to save artificial intelligence*. Journal of Science and Technology of the Arts, 11(2), 3-8.

[9] Costanza-Chock, S. (2018) *Design justice, AI, and escape from the matrix of domination*. Journal of Design and Science, 3(5).

[10] Dumbrava, C. (2021) *Artificial intelligence at EU borders – Overview of applications and key issues*. EESC: European Economic and Social Committee. Belgium. Available from: https://policycommons.net/artifacts/3178234/artificial-intelligence-at-eu-borders/3976753/. CID: 20.500.12592/13jr74.

[11] GPAI, The Future Society and CEIMIA (2020) *Areas for Future Action in the Responsible AI Ecosystem*. Available from: https://gpai.ai/projects/responsible-ai/areas-for-future-action-in-responsible-ai.pdf

[12] Hunter, T. (2023) *AI porn is easy to make now. For women, that's a nightmare*. Washington Post. Available from: https://www.washingtonpost.com/technology/2023/02/13/ai-porn-deepfakes-women-consent/

[13] ILO (2009) *The price of exclusion: The economic consequences of excluding people with disabilities from the world of work*. Prepared by Sebastian Buckup. Available from: https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---ifp_skills/documents/publication/wcms_119305.pdf

[14] IMF Staff Discussion Note. Prepared by Ostry, J.D., Alvarez, J., Espinoza, R., & Papageorgiou, C. (2018) *Economic Gains from gender inclusion: New mechanisms, new evidence*. Available from: https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2018/10/09/Economic-Gains-From-Gender-Inclusion-New-Mechanisms-New-Evidence-45543

[15] Jansen, F. (2018) *Data driven policing in the context of Europe.* Data Justice Lab. Available from: https://www.datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-Data-Driven-Policing-EU.pdf
[16] Kelleher, K. (2023) *Revenge porn and deep fake technology: The latest iteration of online abuse*. Dome. Available from: https://sites.bu.edu/dome/2023/08/10/revenge-porn-and-deep-fake-technology-the-latest-iteration-of-online-abuse/
[17] Lehtonen, P., & Aalto, P. (2017) *Smart and secure borders through automated border control systems in the EU? The views of political stakeholders in the Member States.* European Security, 26(2), 207-225.
[18] MacKenzie, D., & Wajcman, J. (1999) *The social shaping of technology* (2nd ed.). Maidenhead, UK: Open University Press.
[19] Masiero, S., & Das, S. (2019) *Datafying anti-poverty programmes: Implications for data justice*. Information, Communication & Society, 22(7), 916-933.
[20] McGregor, M., & Molnar, P. (2023) *Digital border governance: A human rights based approach.* University of Essex & UN High Commissioner for Human Rights. Available from: https://www.ohchr.org/sites/default/files/2023-09/Digital-Border-Governance-A-Human-Rights-Based-Approach.pdf
[21] O'Brien, B.C., Harris, I.B., Beckman, T.J., Reed, D.A., & Cook, D.A. (2014) *Standards for reporting qualitative research: A synthesis of recommendations*. Academic Medicine, 89(9), 1245-1251. doi: 10.1097/ACM.0000000000000388. PMID: 24979285.
[22] OECD Forum (2015) *Why a push for gender equality makes sound economic sense*. Available from: https://www.oecd.org/social/push-gender-equality-economic-sense.htm
[23] Parks, L., & Kaplan, C. (2017) *Life in the age of drone warfare*. Durham: Duke University Press
[24] Peña, P., & Varon, J. (2019) *Consent to our data bodies: Lessons from feminist theories to enforce data protection.* Coding Rights. Available from: https://codingrights.org/docs/ConsentToOurDataBodies.pdf
[25] Perez-Desrosiers, D. (2021) AI application in surveillance for public safety: Adverse risks for contemporary societies. In Keskin, T., & Kiggins, R.D. (Eds.), *Towards an International Political Economy of Artificial Intelligence*. Cham, Switzerland: Springer
[26] Podoletz, L. (2023) *We have to talk about emotional AI and crime*. AI and Society, 38, 1067-1082
[27] Prabhakaran, V., Mitchell, M., Gebru, T., & Gabriel, I. (2022) *A human rights-based approach to responsible AI*. arXiv:2210.02667
[28] So, W., Lohia, P., Pimplikar, R., Hosoi, A. E., & D'Ignazio, C. (2022) *Beyond fairness: Reparative algorithms to address historical injustices of housing discrimination in the US*. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency* (pp. 988-1004)
[29] Tahir, M. (2019) *Violence work and the police order*. Public Culture, 31(3), 409-418
[30] UNESCO (2020) *Artificial intelligence and gender equality: Key findings of UNESCO's global dialogue*. Available from: file:///Users/niobehaitas/Downloads/374174eng%20(1).pdf
[31] UNESCO (2021) *Recommendation on the ethics of artificial intelligence.* SHS/BIO/REC-AIETHICS/2021. Available from https://unesdoc.unesco.org/ark:/48223/pf0000380455
[32] UNESCO (2023) *Recommendation on the ethics of artificial iIntelligence: Key facts.* SHS/2023/PI/H/1. Available from https://unesdoc.unesco.org/ark:/48223/pf0000385082

[33] UNESCO, OECD, & IDB (2022) *The effects of AI on the working lives of women*. Available from: https://publications.iadb.org/en/effects-ai-working-lives-women
[34] UN General Assembly (21 October 2015) *Transforming our world: The 2030 Agenda for Sustainable Development*. Available from: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/291/89/PDF/N1529189.pdf?OpenElement
[35] UN Human Rights Council (1979) *Convention on the elimination of all forms of discrimination against women*, Article 1. Available from: https://www.ohchr.org/sites/default/files/cedaw.pdf
[36] UN Human Rights Council (2015) *Empowerment, inclusion, equality: Accelerating sustainable development with human rights*. Pamphlet, available from: https://www.ohchr.org/sites/default/files/Documents/Issues/MDGs/Post2015/EIEPamphlet.pdf
[37] UN Human Rights Council (18 June 2020) *Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*, A/HRC/44/57. Available from: undocs.org/en/A/HRC/44/57
[38] UN Human Rights Council (28 December 2021) *Report of the Special Rapporteur on the rights of persons with disabilities,* A/HRC/49/52. Available from: undocs.org/en/A/HRC/49/52
[39] UNICEF (2017) *Gender equality: Glossary of Terms and Concepts.* Available from: https://www.unicef.org/rosa/media/1761/file/Genderglossarytermsandconcepts.pdf
[40] UN Secretary General (2023) *Note to correspondents: Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for states to establish new prohibitions and restrictions on autonomous weapon systems.* Available from: https://www.un.org/sg/en/content/sg/note-correspondents/2023-10-05/note-correspondents-joint-call-the-united-nations-secretary-general-and-the-president-of-the-international-committee-of-the-red-cross-for-states-establish-new#:~:text=In%20a%20joint%20appeal%20today,weapon%20systems%2C%20to%20protect%20humanity
[41] UNWOMEN *GenderTerm*: *UN Women online resources on the use of gender-inclusive language*. Available from: https://www.unwomen.org/en/digital-library/genderterm?gclid=Cj0KCQjwtJKqBhCaARIsAN_yS_kt__8_L15clXhyrUaWn6risoT2hRwVm-xbd84XKV6BvYGDLvy7JgsaAgX6EALw_wcB
[42] Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Fuso Nerini, F. (2020) *The role of artificial intelligence in achieving the Sustainable Development Goals*. *Nature Communications*, 11(1), 1-10
[43] Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, L., Mills, M., ... & West, S. M. (2019) *Disability, bias, and AI.* AI Now Institute. Available from: https://ainowinstitute.org/wp-content/uploads/2023/04/disabilitybiasai-2019.pdf.

Appendix 1: Statistics from Regional Consultations

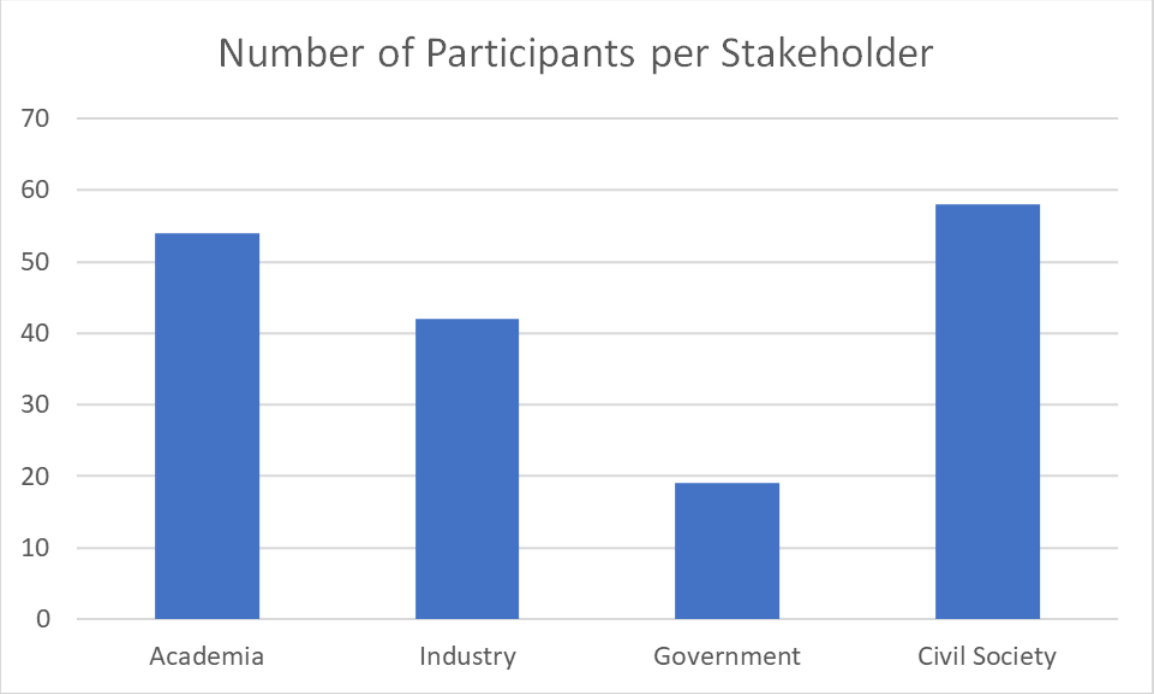

Figure 1: Number of Consultation Participants per Stakeholder Type

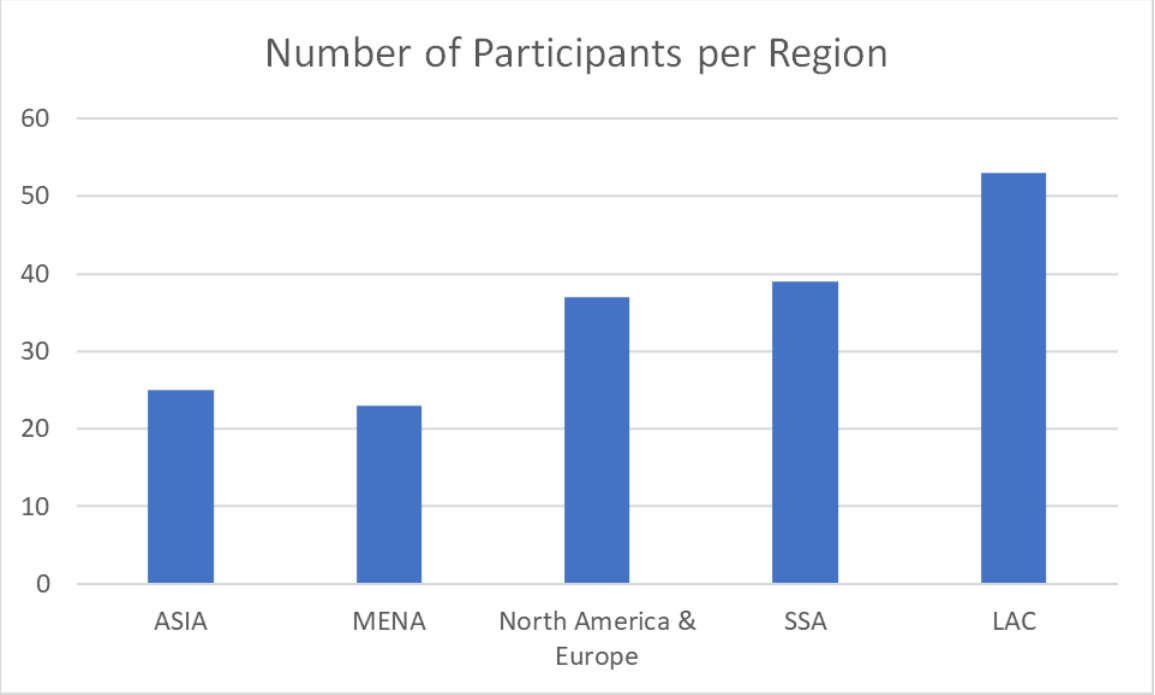

Figure 2: Number of Consultation Participants per Region

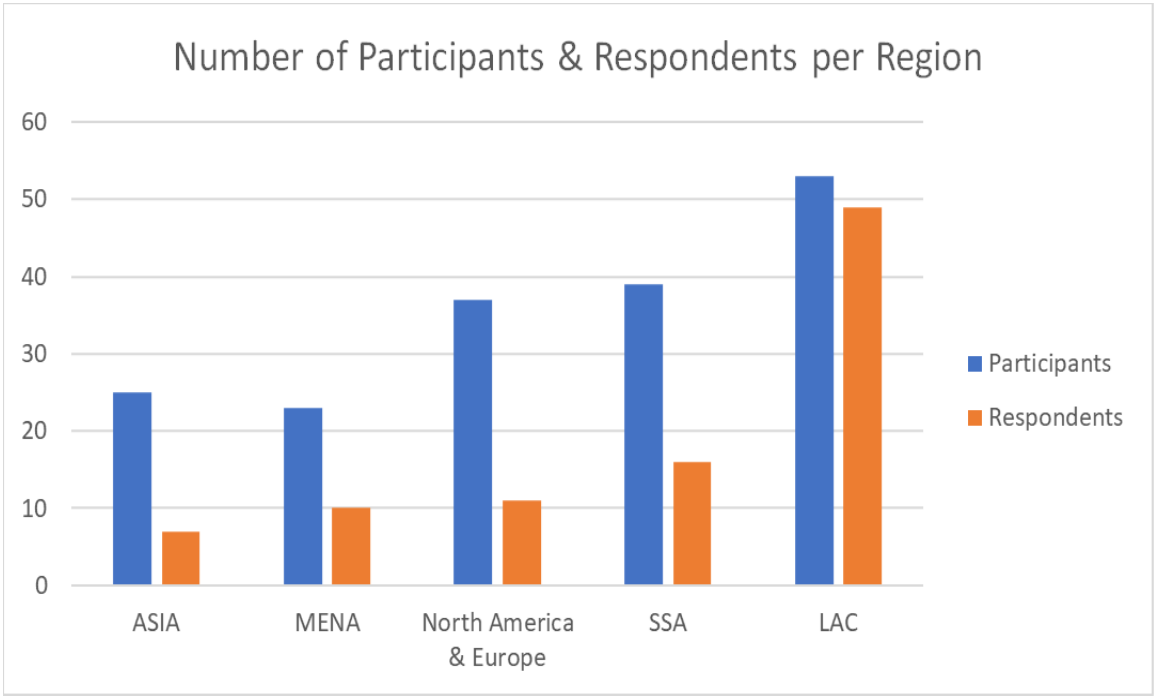

Figure 3: Number of Consultation Participants (Blue) and Socio-Demographic Questionnaire Respondents (Orange)

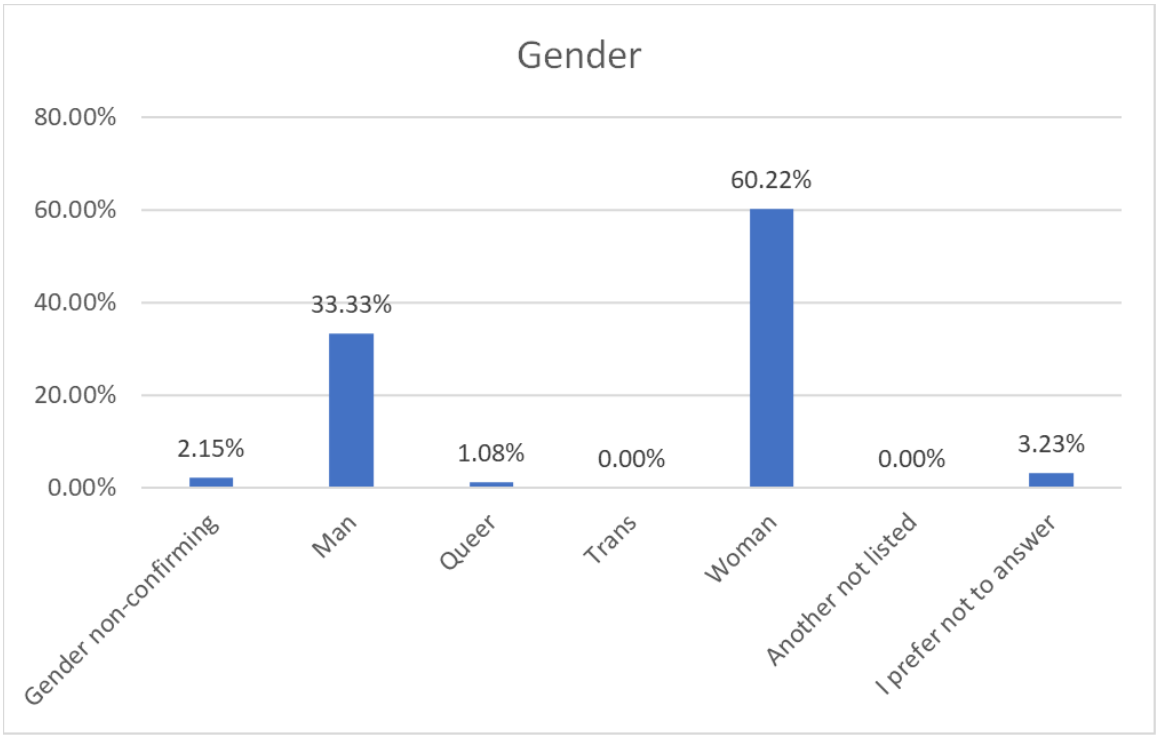

Figure 4: Gender Breakdown: Percentages among Questionnaire Respondents

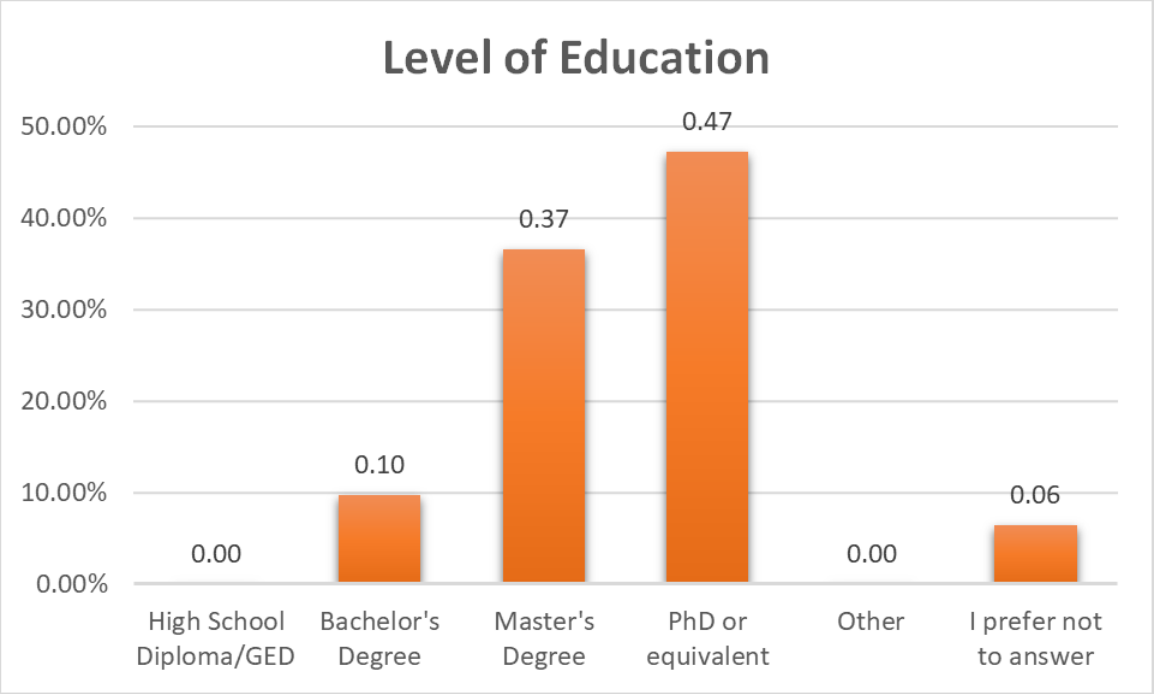

Figure 5: Level of Education: Percentages among Questionnaire Respondents

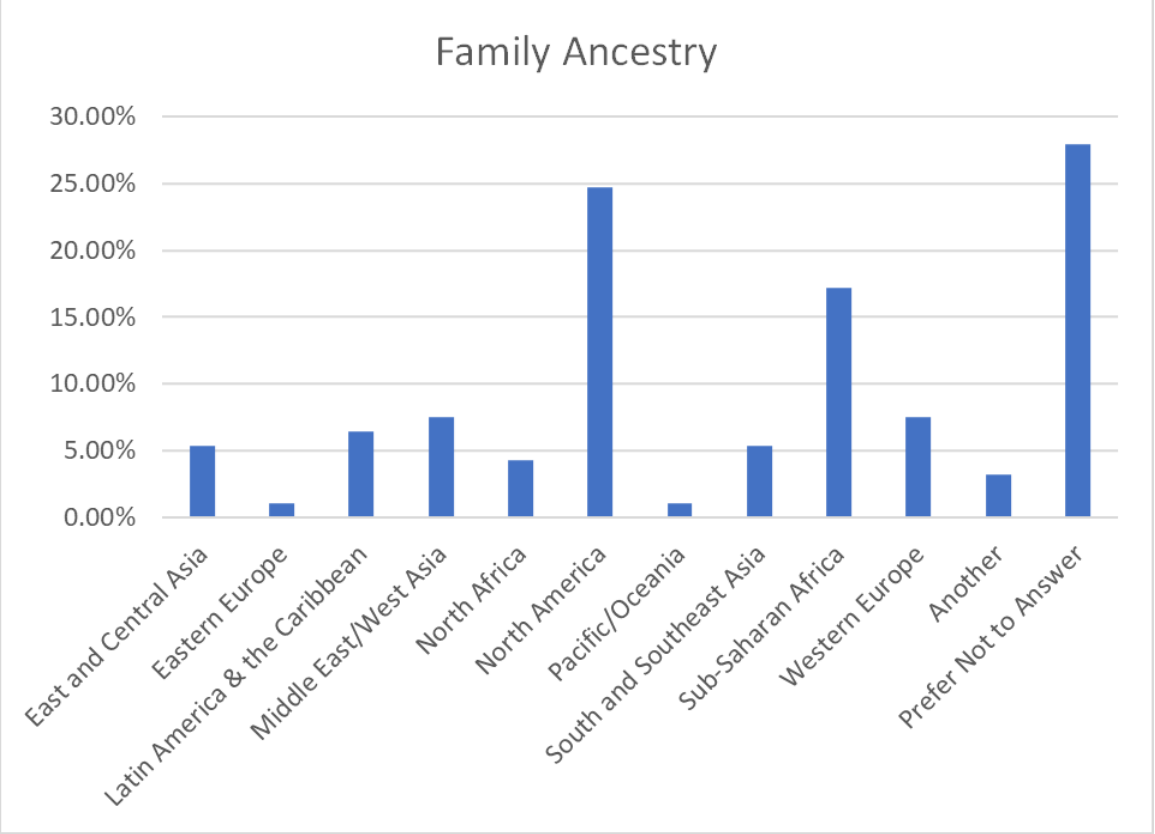

Figure 6: Family Ancestry: Percentages among Questionnaire Respondents

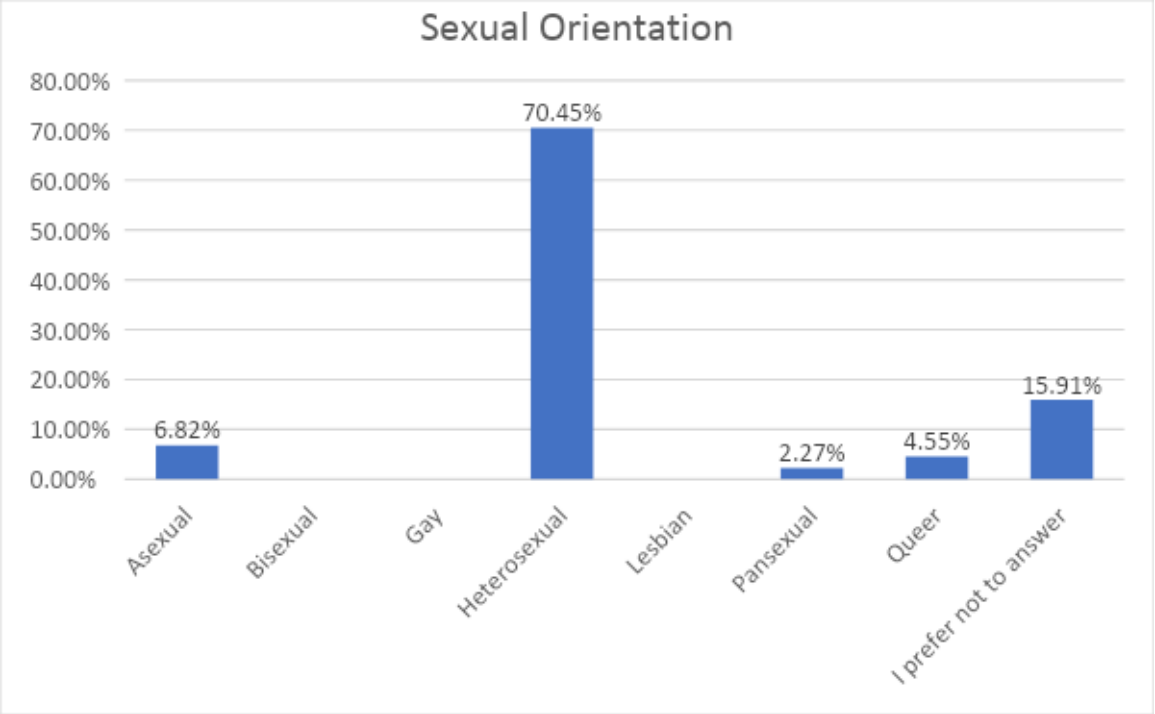

Figure 7: Sexual Orientation: Percentages among Questionnaire Respondents

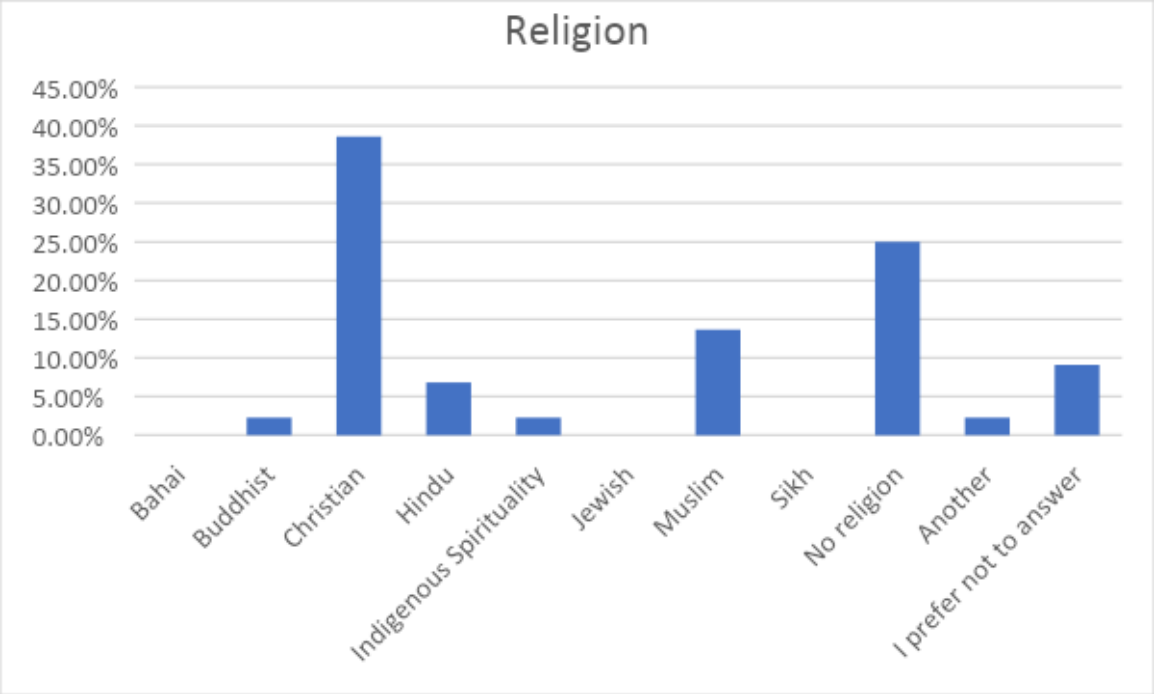

Figure 8: Religion: Percentages among Questionnaire Respondents