Data4Good: Designing for Diversity and Development

Data4Good: Designing for Diversity and Development

Luigi De Russis
Politecnico di Torino
Torino, Italy
luigi.derussis@polito.it

Neha Kumar
Georgia Institute of Technology
Atlanta, USA
neha.kumar@gatech.edu

Akhil Mathur
Nokia Bell Labs
Cambridge, UK
akhil.mathur@nokia-bell-labs.com

ABSTRACT

We are witnessing unprecedented datafication of the society we live in, alongside rapid advances in the fields of Artificial Intelligence and Machine Learning. However, emergent data-driven applications are systematically discriminating against many diverse populations. A major driver of the bias are the data, which typically align with predominantly Western definitions and lack representation from multilingually diverse and resource-constrained regions across the world. Therefore, data-driven approaches can benefit from integration of a more human-centred orientation before being used to inform the design, deployment, and evaluation of technologies in various contexts. This workshop seeks to advance these and similar conversations, by inviting researchers and practitioners in interdisciplinary domains to engage in conversation around how appropriate human-centred design can contribute to addressing data-related challenges among marginalised and under-represented/underserved groups.

CCS CONCEPTS

• **Human-centered computing** → *Human computer interaction (HCI)*; **Collaborative and social computing**.

KEYWORDS

AI for social good, data literacies, multilingual/multicultural contexts, diversity, interdisciplinary computing

ACM Reference Format:
Luigi De Russis, Neha Kumar, and Akhil Mathur. 2020. Data4Good: Designing for Diversity and Development. In *International Conference on Advanced Visual Interfaces (AVI '20), September 28-October 2, 2020, Salerno, Italy*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Even as computing makes transformative advancements in data-driven sub-disciplines such as Artificial Intelligence (AI) and Machine Learning (ML), and we witness unprecedented datafication of the society we live in, there is also growing awareness that emergent applications are systematically discriminating against many different populations.

A major driver of the bias are *data*, or the lack thereof. As long as the data are drawn using methods that align with a Western, universalist approach, gender, ethnic, racial, and cultural biases are likely to persist. Zou and Schiebinger [2] have noted that over 45% of ImageNet [1] data, which computer vision research draws from significantly, comes from the United States, where only 4% of the world's population resides. India and China together contribute only 3% of ImageNet data, while representing 36% of the world's population. The result is that a photograph of a traditional US bride is annotated accurately, while a photograph of a North Indian bride is recognised as "performance art" and "costume".

Given this lack of representation in data, problems related to the robustness and representativeness of data infrastructures become more pressing. For data to be meaningful, they must be collected, stored, understood, analysed, and visualised, all from a holistic and contextually appropriate perspective. There may be challenges encountered in each of these stages; these challenges are exacerbated when we consider the cultural, technological, and/or infrastructural specificities of multilingually diverse and resource-constrained regions across the world. This is true for parts of the Global North as well as the Global South.

In many application domains such as global health, education, gender equality, agriculture, and others, the data burden is borne by workers from socio-culturally and economically diverse backgrounds. Low digital expertise and different vantage points can mean that these workers lack the kind of data literacies required of them by their employers. Data-driven approaches can benefit from integration of a more human-centred orientation before being used to inform the design, deployment, and evaluation of technologies in various less-served contexts. These are some of the important conversations that our workshop seeks to advance.

We invite researchers and practitioners in **interdisciplinary domains** intersecting Human-Computer Interaction (HCI), AI, ML, design, and/or global development to engage in dialog on the topics above. A key priority of our workshop is to have submissions from an *intellectually diverse* and *global group* of participants to further discussions on how appropriate human-centred design can contribute to addressing data-related challenges among marginalised and under-represented/underserved groups around the world. In particular, we solicit participation across more and less technical researchers in HCI who are motivated to address the topics below.

2 TOPICS OF INTEREST

Topics at the workshop must consider the integration of human-centred design in data-driven approaches, touching upon under-resourced contexts across the Global South and North. They include, but are not limited to:

Interfaces and Visualisation

AVI '20, September 28-October 2, 2020, Salerno, Italy

ACM ISBN 978-1-4503-xxxx-x/20/09…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

- Novel interfaces for deriving qualitative/quantitative insights from data
- Interfaces to support data literacy in multilingual contexts
- Information visualisation tools/techniques for data literacy
- Data literacy for end-users in under-resourced contexts

Data Infrastructures for Social Good

- Data collection and field research
- Data quality
- Data sharing
- Privacy and transparency in data analytics

Human-centred design of data-driven approaches

- Interfaces for explainable AI
- User-centred evaluations, techniques, and methods for AI and social good
- Study of public concerns with AI-based technologies

Data work in specific application areas, such as public/global health, education, agriculture, gender equality, refugee resettlement, and more.

3 EVENT FORMAT

The format of the workshop, summarised in Table 1, will alternate between panels (where each panel will consist of a group of short presentations by participants who made workshop submissions, arranged according to themes) and breakout sessions that will focus on design challenges for ideating around solutions for particular contexts (again, dependent on submissions). These contexts will be chosen based on the themes that emerge from the submissions.

Table 1: Workshop organisation

Time	Activity
2:30-3	Introductions
3-3:45	Panel I
3:45:4:15	Breakout Session I
4:15-4:30	Break
4:30-5:15	Panel II
5:15-5:45	Breakout Session II
5:45-6	Summaries and Closing Remarks

3.1 Target Audiences

The workshop will be open for participation to a global and intellectually diverse audience of researchers and practitioners. We foresee interest from industrial labs (e.g., Google AI) and startups working on developing technology solutions in development contexts, as well as academic researchers working on designing user interfaces and experiences for low-literacy users in the Global South (such as from the HCI4D, ML4D, and AI for Social Good focus areas).

We hope, in particular, to attract participation from researchers who may be interested to explore technology interventions to support new migrants coming into the EU, as they will likely face synergistic challenges in their research. In all, we expect to have 25-30 participants attend.

4 ORGANISERS

Luigi De Russis is an Assistant Professor (RTDb) at Politecnico di Torino, Italy, where he mainly focuses on Human-Computer Interaction and, more specifically, investigates how to overcome challenges in interaction between humans and computers, especially in contexts where the interaction happens in complex settings (e.g., as within Connected and Smart Environments).

Neha Kumar is an Assistant Professor at the Georgia Institute of Technology, USA, where she conducts research at the intersection of human-centred computing and global development, with a focus on global health and community informatics.

Akhil Mathur is a Principal Research Scientist at Nokia Bell Labs in Cambridge, United Kingdom where he does research on Mobile Sensing and Deep Learning for Edge Devices. His current focus is on developing techniques to improve the robustness and execution latency of deep learning algorithms on edge devices.

REFERENCES

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. https://doi.org/10.1109/cvpr.2009.5206848

[2] James Zou and Londa Schiebinger. 2018. AI can be sexist and racist — it's time to make it fair. *Nature* 559, 7714 (jul 2018), 324–326. https://doi.org/10.1038/d41586-018-05707-8