

This is a forum for perspectives on designing for communities marginalized by economics, social status, infrastructure, or policies. It will discuss design methods, theoretical and conceptual contributions, and methodological engagements for underserved communities. — **Nithya Sambasivan, Editor**

Toward Responsible AI for the Next Billion Users

Nithya Sambasivan and Jess Holbrook, Google

Recent advances in computing power, increases in the quantity and quality of data, and algorithmic breakthroughs have led to a rise in machine intelligence, creating opportunities that simply did not exist a few decades ago. From stock trading to detecting disease-vector progression to personalized music recommendations, AI is starting to be integrated into diverse domains of human life. And as countries like India, Brazil, and Nigeria experience massive growth online, AI technologies are increasingly intersecting with new user groups, applications, datasets, and regulations.

Much of AI's path has been shaped by its originating contexts in Western nations. As AI touches the fundamental underpinnings of the technological universe, it is incumbent upon us in the HCI community to ask the *who* questions. We call upon the community to challenge implicit assumptions and biases, and to integrate various global communities into the discourse and development of AI. The ground realities of growing Internet penetration, novel applications, multiple languages, low-end devices, services with global reach, cultural norms, and more require globally relevant AI models and products. While AI is still emerging in the Global South, engaging and analyzing today will help us create an inclusive AI in the near future. An intimate understanding of user practices, value systems, and

implications for various communities worldwide is essential to creating human-centric AI that is meaningful and ethical for all.

In this article, we present research provocations for AI for the next billion users, to spur a conversation on the implicit beliefs, biases, and issues that may be normalized in AI. As much of AI's functioning is still not well understood or fully developed, we believe these areas for research are crucial to shaping inclusive AI as it becomes more complex, powerful, and present in daily life. We bring our perspectives as HCI and social scientists who work closely with AI researchers. We have started to address some of these areas in our research and invite further exploration from the research community.

BE ROBUST TO MULTILINGUALISM AND DIVERSE LITERACIES

As the profiles of people coming online change over the next few years, new

Insights

- Implicit beliefs, biases, and issues from Western contexts may be normalized in AI, making it less globally responsible.
- We need to employ a global lens in our core AI assumptions, whether it is the training data, model performance, or explainability of systems.
- It's important to ask the difficult questions on positive and negative impacts early on, rather than introducing repairs and post-hoc fixes.

languages and diverse literacies will come into the fray. It is estimated that 40 percent of Nigerians and 29 percent of Indians were non-literate in 2015 [1]. Many of the current assumptions of AI systems may need rethinking, such as what constitutes user models, meaningful interactions, training data, and signals to improve AI models. Consider the possibility that lower literacy may shape their technology use to specific online activities, such as visual browsing or memorized sequences. A corresponding consideration is whether lower-literacy users may be served low-quality or impersonal outputs by current AI models, as the models may use features based on majority literate users. For example, if a user chooses three videos in a row based on the relevance of video thumbnails or simply the order of presentation (versus the actual user goals), the models can form a kind of feedback loop where more results based on the spurious behavioral signals are presented and watched because the user is unaware of any alternatives. In addition, the formulation of queries and requests needs to be designed to include low-literate users, as they face difficulty with abstraction (an ability derived from formal education), shown in research by Thies et al. [2].

Many countries in the Global South are officially or informally multilingual, with a combination of official, native, and trade languages. Multilingualism and vernacular languages present interesting new challenges for NLP and AI. As voice interfaces surge in popularity, we



are moving past English, a dominant language on the Web, to less-indexed vernacular languages and multilingualism. Code mixing and transliteration of native languages, such as Akan-English, Taglish, or Hinglish are employed in everyday online interactions due to the difficulty of local-language input and comfort with language switching in daily life. Personal assistants and interfaces need to consider language switching and local languages—as researchers of Project Melange show, Indians switch to various languages depending on emotion and context, which is a key insight for personal AI interfaces. Large digital corpora of languages tend to be in dominant languages, like Hindi, but languages spoken by smaller or indigenous communities like Odia, Kurdish, or Twi find little representation online, posing new challenges.

PERSONALIZE EXPERIENCES WITH THE SHARED IDENTITY MODEL

AI personalization models take user behaviors, preferences, feedback, and characteristics to create personalized experiences. However, prevailing identity paradigms like “one user, one account” are regularly challenged in the case of shared device use, which may occur due to social and economic norms. Features today include personal variables like heartbeat data, financial transactions, and video watch time; however, when personal features are used to make inferences for multiple user scenarios, the recommendations could be irrelevant, offensive, or even adversarial. For example, in our research we have found that women’s phones tend to be shared or monitored, and complex privacy practices are used when devices travel across hands; but users

can be implicated when content is revealed accidentally [3].

New identity models like phone-number-based account creation present personalization challenges with multiple accounts, throwaway SIMs, and recovery. Infrastructural realities like poor or intermittent connectivity and low-end devices lead to questions on decentralization, AI training and inference at the edge, and the updating of personalization models.

EXPLAINABILITY AND TRANSPARENCY SHOULD MAKE SENSE FOR NEW TECHNOLOGY USERS

Explainability of the inputs, decision making, and outputs of AI systems is crucial in many use cases to avoid blind spots in data, find bias, and help users understand the model’s decisions. It is important to recognize that new technology users may face technology-



PROACTIVELY ADDRESS FAIRNESS IN DATA SKews

As Internet access is still emerging in the Global South, there can be substantial differences in the demographics of who can and cannot come online. In environments where most of the audience and authors are from majority user groups, online curation biases toward them—whether it's curation by a human, a social network, or an algorithm. Optimizing for the total user base will inherently disenfranchise any minority group. Machine learning, which is increasingly used in feeds, searches, and recommendations, typically uses large annotated datasets to train its algorithms. But when specific user groups are underrepresented, training data does not come in equal numbers from them.

Take the case of women—two-thirds of countries around the world have more men than women online [4]; in India, only 29 percent of women are online [5]. If the amount of training data from women is not equal to that from men, the resulting curation bias will produce algorithms that favor men. Men's attributes may therefore be overrepresented relative to their real-world prevalence, and their content preferences—assumed based on time spent, up votes, down votes, searches, or shares—can influence content recommendations such as image detection, recommendations, or suggested actions. Although certain content may exist, recommendation engines might not suggest it because the group's interests have been down-weighted by algorithms that optimize for total popularity. Semantic mismatches between how platforms and users describe content also add to the challenge of finding relevant content.

Boosting the training data through targeted crowdsourcing, partnerships with community organizations, training the models to account for bias, and reinterpreting queries through the lens of how the community might search could all lead to reduced bias. At the same time, the specific privacy needs and concerns of each of these groups must be taken into account and respected throughout any data-collection process.



familiarity and data-literacy realities, raising important research questions about explainability across differing literacy levels. As new technology users are poised to be at the receiving end of new AI systems for microloans, healthcare, and education, explaining the decision making should help provide recourse to users and force developers to create unbiased, less discriminatory models, especially when the stakes are high.

Transparency and control are other equally important principles for new technology users. Knowing what types of personal data are gathered, how that data is used, and how to control the system's functioning are essential

elements for people to feel in control of the AI systems they use. Without them, people can create elaborate and often inaccurate models of how an AI system works and subsequent workarounds. For example, Shaina of Kanpur, India, reported how when she watches a risqué video, she finds it difficult to control the platform's recommendations; instead, she "searches for five to six other videos on different topics" to stop the site from recommending more risqué videos. While Shaina understands that her inputs affect the outputs, how can she have a better understanding and control over the system's recommendations, instead of resorting to complex workarounds?

CHECK THE POTENTIAL SOCIETAL ABUSE OF AI

As AI becomes more sophisticated in its ability to not only detect and recognize but also manipulate entities, it increases the risk of causing serious damage to various underserved communities. An important research area lies in proactively understanding the potential algorithmic manipulation of any personally identifiable or attributable content to cause harm to individuals. Deepfakes, fake images, and videos powered by deep learning present new challenges with manufacturing fake news, malicious content, and pornographic content. Such synthesis techniques can cause serious harm to non-privileged groups through wider circulation on social networks, with broad social repurcussions. Take the case of Indian journalist Rana Ayyub, who discusses how she has been constantly harassed through deepfake pornographic videos made of her, causing massive reputation damage [6]. Such image-manipulation incidents travel virally and further impact online expression; in our research on gender equity, we found that 61 percent of women across seven countries, including Brazil, India, and Indonesia, proactively uploaded profile photos using non-face images like flowers, animals, landscapes, and group photos to avoid personal-image manipulation, based on incidents they had heard about in the news.

Safety issues are complex and require the joint involvement of technology policy, law and order, and institutional change for any lasting change. Safeguards against bad actors and anti-abuse management are essential in the formative design principles of systems. Technology moderation and takedown policies should grow to encompass various cultural contexts. We should practice inclusive and participatory design and always consider all the stakeholders of the system; even if we leave out 5 percent of stakeholders and the technology works well for 95 percent of cases, there is the potential for unintended consequences.

UNDERSTAND AI POLICY IMPLICATIONS GLOBALLY

A growing number of AI researchers are building laudable applications for social good domains in healthcare, agriculture, social justice, and more. At the same time, we should not lose sight of how current AI trends and policies on automation and digitization affect societies all over the world. Difficult, polarizing questions are being raised about the impact of automation on jobs, skills, and wages. Most Global South economies are heavy on the informal and outsourcing sectors, such as call centers, data entry, and low-level factory jobs. Entire industries are vulnerable to job displacement. In our ongoing research on the future of work in vocational sectors in India, we find that most technicians have little to no awareness of the future-of-work discourse or development. Skill reinvention of economically disadvantaged workers is key to resilience and job readiness in the future. Policy interventions for a jobless future like universal basic income are being proposed in the Global North, but such proposals need to take into consideration the realities of low incomes, corruption, and very large populations in other regional contexts.

Governments in the Global South are increasingly pushing for the digitization of nation states to eliminate middlemen, for social welfare decisions, and to increase resource-allocation efficiency. Algorithms and data are now being used for human welfare in the Global South, for example, vaccine deployments and food rations. Citizenry are increasingly defined by digitization, such as the Aadhar program. In a context where large groups of people are already below the poverty line and have fragile access to social welfare, errors and biases in automated decision making can be serious (e.g., misallocating food rations). Audits, interface evaluations, and public user studies could bring the concerns to the fore.

A CALL TO ACTION

It is our responsibility as the HCI research community to influence the ways in which AI is perceived, adopted, and normalized globally. There are

groups in industry and academia pushing this conversation, including the People + AI Research team (PAIR) at Google (you can see our thinking on a human-centered approach to AI at <https://design.google/library/ai/>), the AI Now Institute, and the Algorithmic Fairness and Opacity Working Group.

Our call to action is to employ a global lens to our core AI assumptions, whether it is the training data, model performance, explainability of systems, or deciding which human needs to address with AI in the first place. Only then can we truly make AI work for humanity.

ENDNOTES

1. CIA World Factbook, 2015; <https://www.cia.gov/library/publications/the-world-factbook/fields/2103.html>
2. Thies, I.M. User interface design for low-literate and novice users: Past, present and future. *Foundations and Trends® in Human–Computer Interaction* 8, 1 (2015), 1–72.
3. Sambasivan, N., Checkley, G., Batool, A., Gaytán-Lugo, L.S., Matthews, T., Consolvo, S., and Churchill, E. “Privacy is not for me, it’s for those rich women”: Performative privacy practices on mobile phones by women in South Asia. *Proc. of SOUPS 2018*. USENIX Association, 2018.
4. ITU Facts and Figures, 2017; <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf>
5. UNICEF State of the World’s Children, 2017; https://www.unicef.org/publications/files/SOWC_2017_ENG_WEB.pdf.
6. Ayyub, R. In India, journalists face slut-shaming and rape threats. *New York Times*. May 22, 2018; <https://www.nytimes.com/2018/05/22/opinion/india-journalists-slut-shaming-rape.html>

 **Nithya Sambasivan** is a UX researcher in Google AI. She co-leads research on building human-centered AI in emerging markets with Jess Holbrook. She has a Ph.D. from UC Irvine and an M.S. in HCI from Georgia Tech. Her research has won top awards at HCI and ICTD conferences.

→ nithyasamba@google.com

 **Jess Holbrook** is a UX manager and UX researcher in Google AI. He and his team take a human-centered and technology-inspired approach to building AI-powered products like Google Clips, Lens, and AIY. He co-leads the People + AI Research (PAIR) group, to provide accessible AI to help people solve meaningful problems themselves. He has a Ph.D. in psychology from the University of Oregon.

→ jessh@google.com