

**Title:** Panel Discussion on <AI and marginalized languages>

**Abstract:** During the past year or so, we have seen rapidly growing interest and excitement in large-scale language models and their applications to various domains beyond traditional problems in natural language processing and machine learning. These large-scale language models are not anymore just a proof of concept but have been productized and are being served as actual products to millions of users all over the world. Although it is indeed an exciting development, these language models have been trained on a large corpus that may not be representative of all the languages in the world and may focus disproportionately on already-better-served languages, such as English and European languages. This raises both questions and concerns about the potential for these language models to exacerbate the issue of digital divide as well as inequality and inequity in information access. As a main venue of publication as well as discussion in the field of artificial intelligence, ICML is thus hosting a panel discussion to discuss the issue at the intersection of artificial intelligence, with particular focus on recent large-scale language models, and marginalized languages, in 2023.

**Date/Time:** July 26 2023 (Wednesday) at 4-5.30pm HST

**Panelists:**

- **Steven Bird** (<https://language-lab.cdu.edu.au/>, <http://www.stevenbird.net/>) is an extremely well-respected NLP researcher who's been dedicating himself full time on documenting and revitalizing indigenous languages in Australia. He is a professor at Charles Darwin University.
- **Hady Elsahar** (<https://www.hadyelsahar.io/>) has been working on NLP for marginalized languages in Africa ever since graduate school. He is an active member of Masakhane, which aims to create NLP systems for African languages from scratch by creating new corpora as well as systems. He is a Research Scientist at Meta AI.
- **Keoni Kealoha Mahelona** (<https://www.linkedin.com/in/kmahelona/>) has worked on building NLP systems for indigenous languages in Pacific as a CTO at Te Hiku Media (<https://tehiku.nz/>).
- **Ji Hyung Moon** (<https://www.linkedin.com/in/mjihyung/>) was the lead author of KLUE (<https://arxiv.org/abs/2105.09680>) which created a new set of benchmark suites for evaluating Korean NLP systems, all from scratch. She is a co-founder and CTO of Softly AI.
- **Shruti Rijhwani** (<https://shrutirij.github.io/#about-me>) is one of the organizers of AmericasNLP workshop [http://turing.iimas.unam.mx/americasnlp/2023\\_workshop.html](http://turing.iimas.unam.mx/americasnlp/2023_workshop.html) that focuses on NLP for indigenous languages in the Americas (both North and South). She has worked on NLP for endangered languages over the past few years. She is a Research Scientist at Google.

**ICML Page:** <https://icml.cc/virtual/2023/panel/25588>

**Submit your questions** at <https://forms.gle/5yFG4vE4mEGUj3g78>.