BROOKINGS

Report

Determining systematic differences in human graders for machine learning-based automated hiring

Mike H. M. Teodorescu, Nailya Ordabayeva, Marios Kokkodis, Abhishek Unnam, and Varun Aggarwal
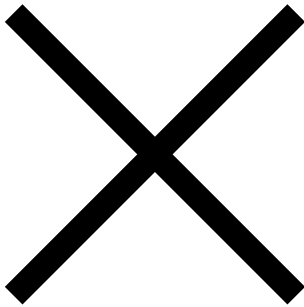Thursday, June 23, 2022

Editor's Note:

Abstract

Firms routinely utilize natural language processing combined with other machine learning (ML) tools to assess prospective employees through automated resume classification based on pre-codified skill databases. The rush to automation can however backfire by encoding unintentional bias against groups of candidates. We run two experiments with human evaluators from two different countries to determine how cultural differences may affect hiring decisions. We use hiring materials provided by an international skill testing firm which runs hiring assessments for Fortune 500 companies. The company conducts a video-based interview assessment using machine learning, which grades job applicants automatically based on verbal and visual cues. Our study has three objectives: to compare the automatic assessments of the video interviews to assessments of the same interviews by human graders in order to assess how they differ; to examine which characteristics of human graders may lead to systematic differences in their assessments; and to propose a method to correct human evaluations using automation. We find that systematic differences can exist across human graders and that some of these differences can be accounted for by an ML tool if measured at the time of training.
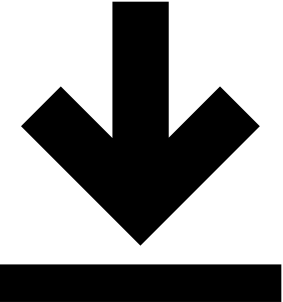
Download the full working paper here.

Get updates on economics from Brookings

Enter Email [Enter Email] Subscribe

No thanks, just download the file.