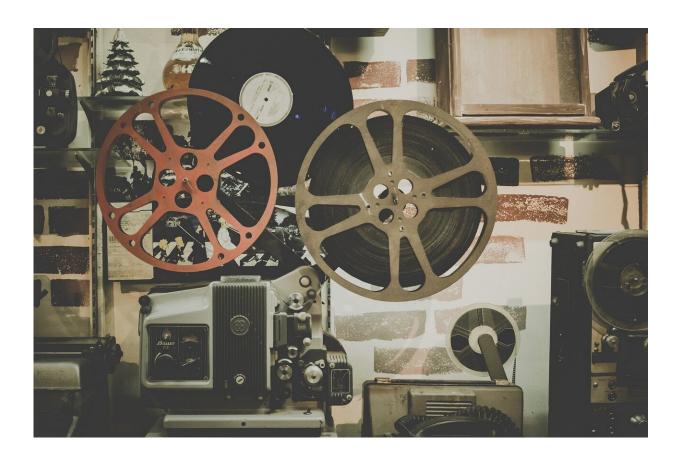# AI identifies social bias trends in Bollywood, Hollywood movies

February 25 2021

Credit: Pixabay/CC0 Public Domain

Babies whose births were depicted in Bollywood films from the 1950s and 60s were more often than not boys; in today's films, boy and girl newborns are about evenly split. In the 50s and 60s, dowries were

socially acceptable; today, not so much. And Bollywood's conception of beauty has remained consistent through the years: beautiful women have fair skin.

Fans and critics of Bollywood—the popular name for a $2.1 billion film industry centered in Mumbai, India—might have some inkling of all this, particularly as movies often reflect changes in the culture. But these insights came via an automated computer analysis designed by Carnegie Mellon University computer scientists.

The researchers, led by Kunal Khadilkar and Ashiqur R. KhudaBukhsh of CMU's Language Technologies Institute (LTI), gathered 100 Bollywood movies from each of the past seven decades along with 100 of the top-grossing Hollywood moves from the same periods. They then used statistical language models to analyze subtitles of those 1,400 films for gender and social biases, looking for such factors as what words are closely associated with each other.

"Most cultural studies of movies might consider five or 10 movies," said Khadilkar, a master's student in LTI. "Our method can look at 2,000 movies in a matter of days."

It's a method that enables people to study cultural issues with much more precision, said Tom Mitchell, Founders University Professor in the School of Computer Science and a co-author of the study.

"We're talking about statistical, automated analysis of movies at scale and across time," Mitchell said. "It gives us a finer probe for understanding the cultural themes implicit in these films." The same natural language processing tools might be used to rapidly analyze hundreds or thousands of books, magazine articles, radio transcripts or social media posts, he added.

For instance, the researchers assessed beauty conventions in movies by using a so-called cloze test. Essentially, it's a fill-in-the-blank exercise: "A beautiful woman should have BLANK skin." A language model normally would predict "soft" as the answer, they noted. But when the model was trained with the Bollywood subtitles, the consistent prediction became "fair." The same thing happened when Hollywood subtitles were used, though the bias was less pronounced.

To assess the prevalence of male characters, the researchers used a metric called Male Pronoun Ratio (MPR), which compares the occurrence of male pronouns such as "he" and "him" with the total occurrences of male and female pronouns. From 1950 through today, the MPR for Bollywood and Hollywood movies ranged from roughly 60 to 65 MPR. By contrast, the MPR for a selection of Google Books dropped from near 75 in the 1950s to parity, about 50, in the 2020s.

Dowries—monetary or property gifts from a bride's family to the groom's—were common in India before they were outlawed in the early 1960s. Looking at words associated with dowry over the years, the researchers found such words as "loan," "debt" and "jewelry" in Bollywood films of the 50s, which suggested compliance. By the 1970s, other words, such as "consent" and "responsibility," began to appear. Finally, in the 2000s, the words most closely associated with dowry—including "trouble," "divorce" and "refused"—indicate noncompliance or its consequences.

"All of these things we kind of knew," said KhudaBukhsh, an LTI project scientist, "but now we have numbers to quantify them. And we can also see the progress over the last 70 years as these biases have been reduced."

A research paper by Khadilkar, KhudaBukhsh and Mitchell was presented at the Association for the Advancement of Artificial

Intelligence virtual conference earlier this month.

Provided by Carnegie Mellon University