
IndiaPix/IndiaPicture via Getty Images

Reflecting India

An intersectional and longitudinal analysis of popular scripted television from 2018 to 2022

Geena Davis Institute on Gender in Media
If she can see it, she can be it.™
Google
USC Viterbi School of Engineering
IAA International Advertising Association

Table of Contents

Purpose of Study

The Geena Davis Institute on Gender in Media partnered with University of Southern California's Signal Analysis and Interpretation Laboratory (which contributed insights and analysis), Google Research (which provided the machine learning technology), and the International Advertising Association India Chapter (which provided subject matter advisement) to undertake this longitudinal representation study of the most popular scripted TV series in India over the past five years, across five different languages.

The goal of this report is to analyze the evolution of representation of gender, age, skin tone, and their intersections in scripted television in India. The novelty of this study is in leveraging artificial intelligence advances in computer vision and natural language processing to analyze media at scale, and in inferring insights that are less tenable to harness manually. This process provides a powerful technology-enabled lens to understand media representation, and these insights can be used to encourage more equitable representation in media.

Entertainment media, such as scripted television, profoundly shapes the minds of viewers as well as social narratives. It influences how we perceive the world, how we see ourselves holding place in society, what we should value, what we should respect, what careers we may pursue, who gets to be the hero, and more.

This study is an expansion of our previous study *#SeeItBeIt: What Families Are Seeing on TV*, in which we analyzed 12 years of representation in popular TV in the U.S.[1] To expand our understanding of the global media environment, we now apply this lens to the Indian TV industry.

Home to over 1.3 billion people, the Indian subcontinent includes a number of ethnicities, languages, cultures, and religions. India also has a vibrant entertainment-media ecosystem, which offers content dubbed and adapted across languages and regions—and television plays a key part. For example, TV-owning households make up two-thirds of the nearly 300 million Indian households, and this level of penetration has nearly doubled from 2001 to 2020.[2] Therefore, studying the representation and portrayals of characters in these stories is important to understanding how these stories are told, and in understanding their societal impact. For example, a study of the portrayal of women in Indian TV soap operas reports that despite governmental policies and industry efforts, there has been little improvement in the portrayal of women.[3] For this reason, there is a critical need for supporting such analyses both at scale and in nuanced, objective ways to gain insights about media narratives. Computational approaches that can process and map constructs of human representation—such as perceived gender, age, and other human attributes—can facilitate such analyses.

Given the profound influence television has on people's daily lives, this endeavor represents a critical undertaking in identifying representation inequalities and offers potential solutions to the gaps that persist.

This study is pioneering in four meaningful ways:

1. **Intersectionality of human-centric signals:** With Google Research's machine learning innovations, we were able to infer human-centric signals at scale, including perceived gender expression, perceived age, and perceived skin tone. We also extrapolated many intersectional patterns across these signals.

2. **Scaled Analysis:** Leveraging this technology, we were able to process a large volume of data: 430 hours of video viewing time, about 15 million frames, and a total of about 38 million face tracks. If done manually, processing this data would have taken a significantly longer time and cost much, much more.

3. **Longitudinal Trends:** With the speed of automation, we were able to extend the study across five languages (Hindi, Bengali, Tamil, Telugu, and Kannada) and a time span of five years, giving us historical trends across 5 Indian languages.

4. **Language Analysis:** We leveraged Google Research's natural language processing innovations, which are novel to this report. These allowed us to understand character portrayal patterns. For this analysis, we used large language models to study names mentioned in dialogue and explored future avenues of its application.

Data insights about representation in entertainment media can help eliminate unconscious bias that reinforces harmful stereotypes and unfavorable behavior toward different groups, such as gender, skin tone, body image, and age. Our sincere hope is that this report will be a tool to enhance and support efforts already being made to improve equitability across media.

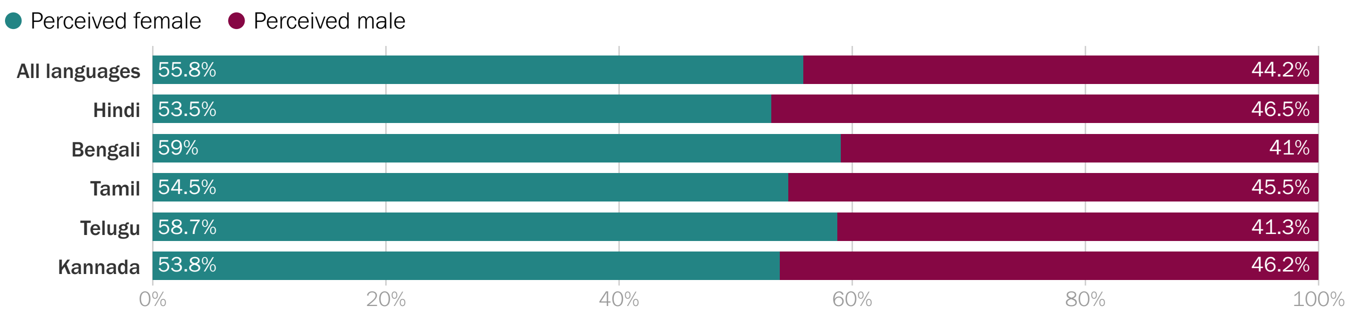
Executive Summary

For this study, we leveraged machine learning innovations in computer vision (CV) and natural language processing (NLP) to analyze representation of perceived gender expression, perceived age, perceived skin tone, screen time, and language usage. Using these technologies, we inferred representation of human-centric signals at scale, inferring longitudinal trends, cross-language patterns, and intersectional insights. We analyzed the most popular scripted television series in India over five years, from 2018 to 2022, according to the Broadcast Audience Research Council India, and across five languages: Hindi, Bengali, Tamil, Kannada, and Telugu. We also applied NLP methods on the transcribed dialogue to analyze the frequency of person-name mentions and whether the mentioned names are perceived to be more masculine or more feminine. Altogether, we analyzed 1,188 episodes, amounting to 430 hours of footage, 15 million frames processed, 38 million faces, and 1.8 million words.

The main findings are as follows:

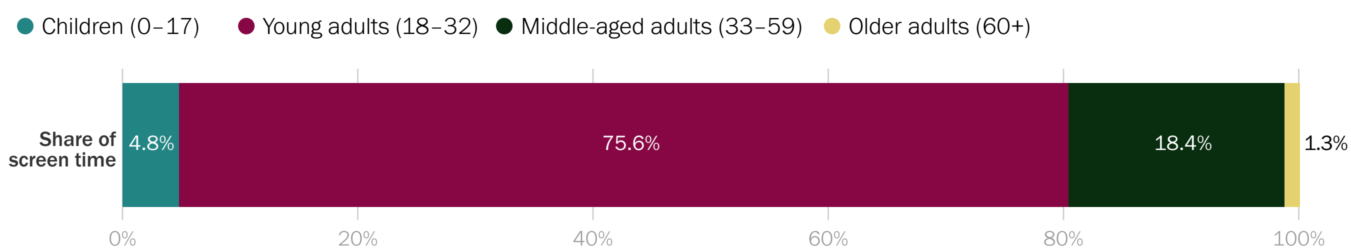
1. **Across all years and all languages, the share of screen time for female characters is higher than for male characters (55.8% compared with 44.2%).** The longitudinal analysis suggests these proportions are steady from 2018 to 2022. Across all years, female characters have the highest share of screen time in Bengali- (59.0%) and Telugu-language (58.7%) series.

Share of screen time for male and female characters



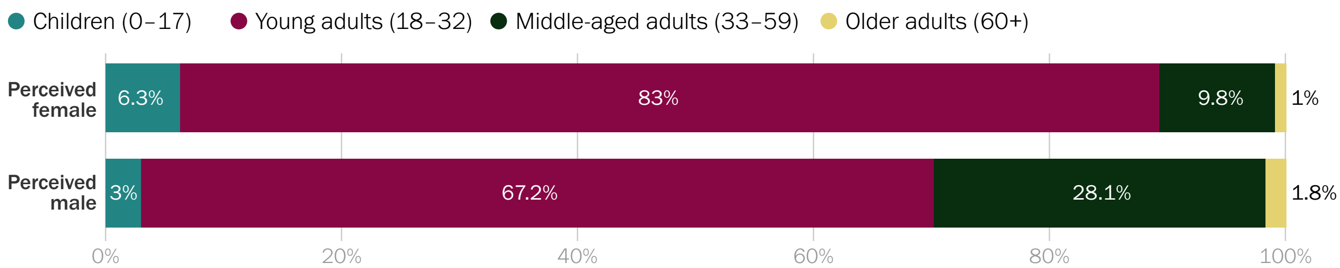
2. **Across all years and all languages, the vast majority of characters shown on screen are young adults (ages 18 to 33).** Young adults account for 75.6% of all characters perceived on screen, while those ages 60 and older are given only 1.3% of all screen time.

Share of screen time for children, young adult, adult, and older characters



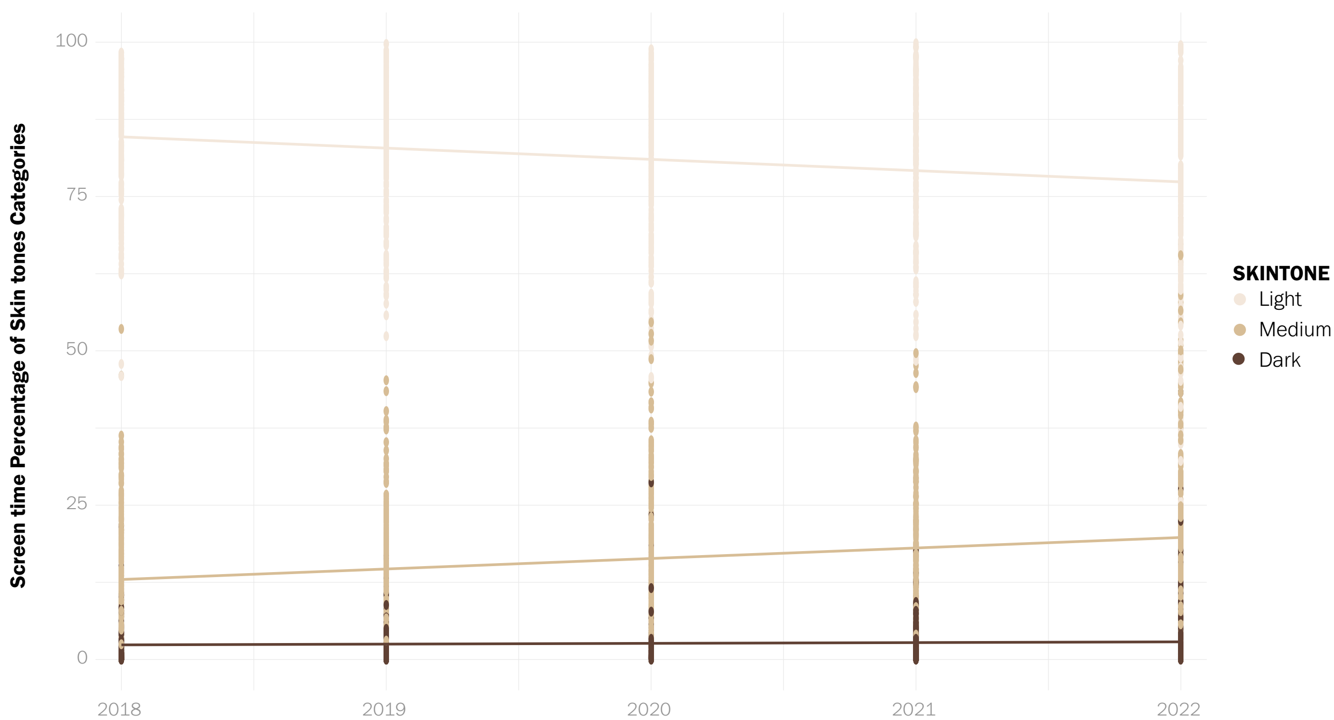
3. **Older characters are rarely seen; there is more age diversity among male characters than female characters**. Girls (ages 0 to 17) and young women (18 to 32) receive 89.3% of female screen time, while boys (0 to 17) and young men (18 to 32) account for 70.2% of male screen time. For middle-aged (33 to 59) and older (60 and older) characters, more screen time is given to male characters (67.6% and 54.9%, respectively). This suggests that older women are less visible on screen than older men.

Share of male and female characters from different age groups



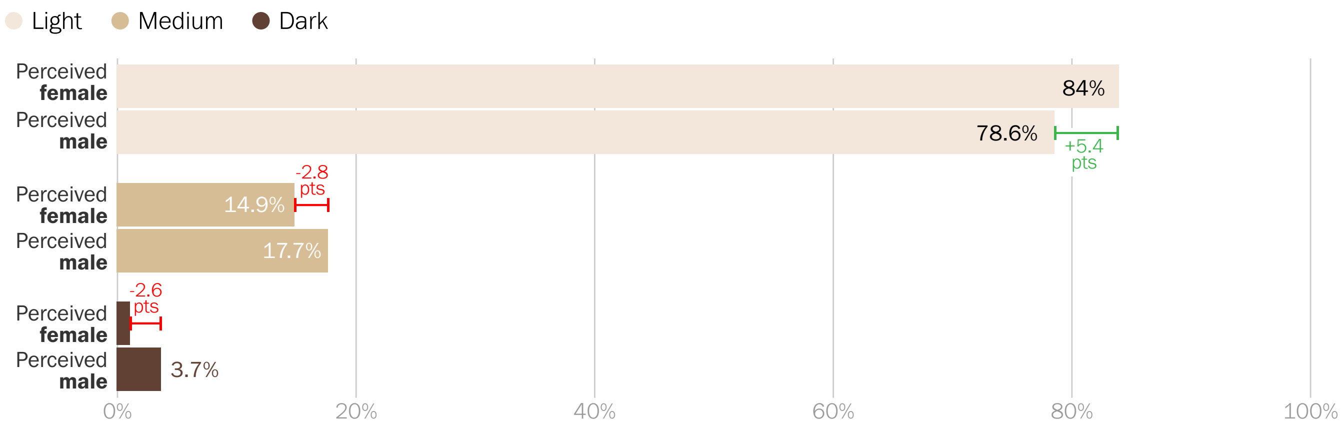
4. **Characters with a light skin tone are eight times more common on screen than characters with medium or dark skin tones; but the percentage of screen time given to characters with a medium tone has increased slightly over time.** Across all years and regions, applying the Monk Skin Tone scale (MST) (which categorizes skin tones across a range from 1, the lightest, to 10, the darkest), we find that 81.2% of all screen time is given to characters who have a light skin tone (1 to 3 MST); only 2.2% of screen time is given to characters who have a dark skin tone (8 to 10 MST). From 2018 to 2022, the share of screen time for characters with a light skin decreased by about 8 percentage points, while the share of screen time for characters with a medium skin tone increased about 7 percentage points.

Share of screen time for light, medium, and dark skin tones, 2018 to 2022



5. **When on screen, female characters have lighter skin tones than male characters do.** Female characters on screen are more likely to have a light skin tone than male characters are (84.0% compared with 78.6%). This may suggest that female characters are being held to Eurocentric beauty standards more rigorously than male characters are. Characters with dark or medium skin tones are given the most screen time when they are older adults — for characters 60 and older, those with a dark skin tone occupy 3.9% of screen time and characters with a medium skin tone occupy 21.9% of screen time (compared with screen time for all characters under 18, where characters with a dark skin tone occupy 1.6% of screen time and characters with a medium skin tone occupy 6.7% of screen time).

Light, medium, and dark skin-tone representation among male and female characters

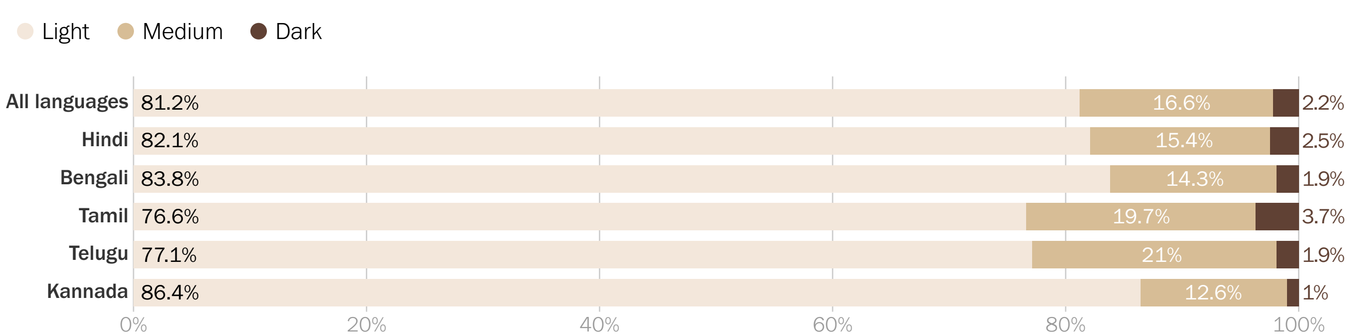


Note: Positive differences represent more screen time for women than for men of the same MST category. Negative differences represent more screen time given to men than women of the same MST category.

Looking at skin tone by age and gender, we find that most characters overall are young adults (ages 18 to 32) with a light skin tone — 70.0% of female characters and 52.9% of male characters. For male characters, the next-most dominant group was middle-aged adults (33 to 59) with a light skin tone (21.4%). This shows a clear preference for featuring younger, lighter-skinned characters, especially for female characters, whereas male characters are allotted slightly more age and skin-tone diversity.

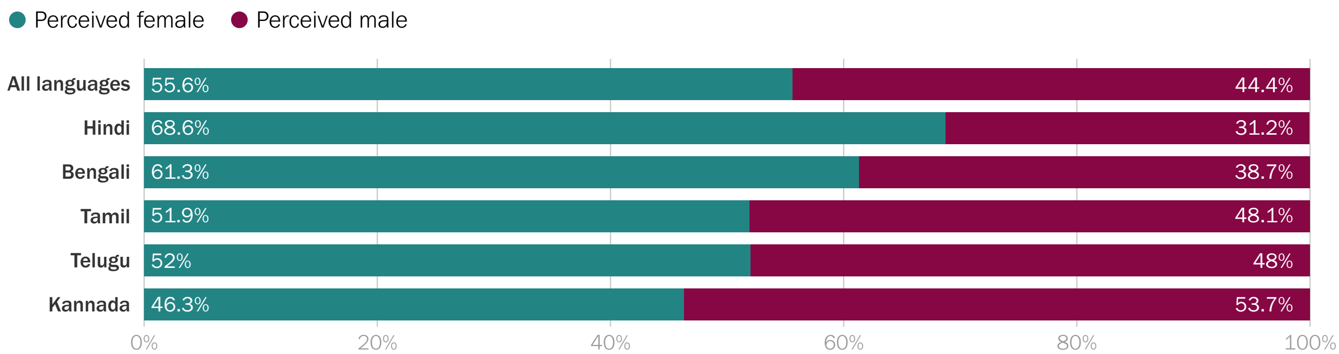
6. **Characters in Tamil- and Telugu-language series have the most skin-tone diversity of the languages analyzed.** In Tamil- and Telugu-language shows, characters with medium or dark skin tones occupy more screen time than they do in shows featuring other languages (23.5% for Tamil and 22.9% for Telugu, compared with 17.9% for Hindi, 16.2% for Bengali, and 13.6% for Kannada). From 2018 to 2022, along with series in Hindi, series in Tamil and Telugu showed a significant decrease in screen time given to characters with a light skin tone, and increased screen time given to characters with a medium skin tone.

Light, medium, and dark skin-tone representation among all characters, by language



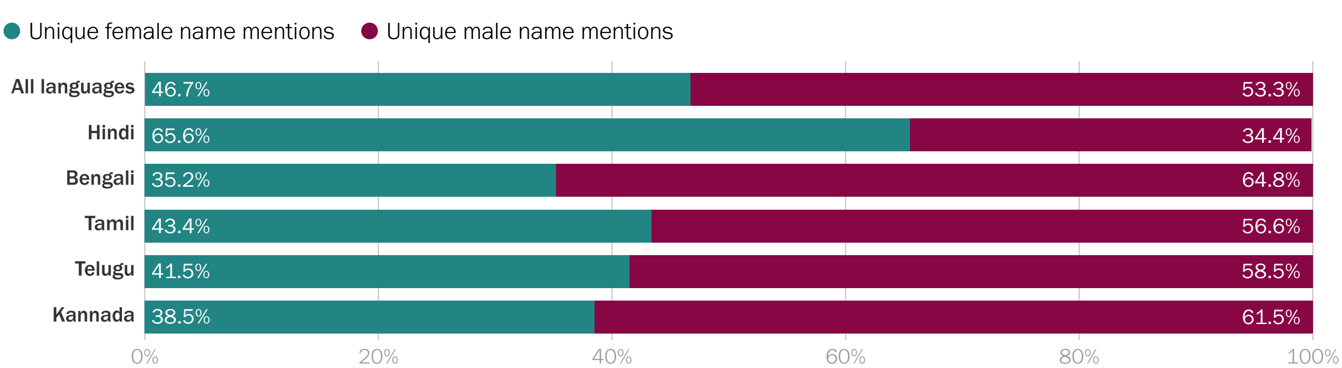
7. **Across all years and all languages, the percentage of female names mentioned is higher than that for male names mentioned (55.6% compared with 44.4%).** Across all years, female names have the highest proportion of mentions in Hindi (68.8%) and Bengali (61.3%) series. In Kannada, there is a reverse trend, with male names having a slightly higher percentage of mentions (53.7%). Across all years and languages, the proportion of male and female names mentioned is 55.6% and 44.4%, respectively, and follows the distribution of male and female screen time (55.8% and 44.2%). For Bengali-language series, female names mentioned have decreased roughly 15 percentage points and male names mentioned have increased roughly 15 percentage points. Male names mentioned have also increased roughly 32 percentage points in Hindi TV series. The distribution of gendered names has remained steady for the other languages.

Mentions of male and female character names



8. **Across all years and all languages, the percentage of unique female names mentioned is lower than that for male names mentioned (46.7% compared with 53.3%).** Although the mention of female names occurs at a higher frequency than male names mentioned (similar to screen time), there are fewer *unique* female names being mentioned than unique male names. In other words, while female characters overall are seen on screen more than male characters, the mention of unique female names in dialogue is less than that of unique male names. Across all years, female characters have the lowest proportion of unique name mentions in Bengali (35.2%), even though they occupy the majority of total name mentions (61.3%). In Hindi series, the proportion of unique female names (65.6%) remains higher than that of unique male names (34.4%), although dropping slightly relative to total female names (68.8%). Tamil, Telugu, and Kannada series have lower proportion of unique female character mentions and experience a drop relative to proportion of total female character mentions.

Unique mentions of male and female names



Data and Methodology

The data for this analysis is the most popular scripted broadcast television series in India from 2018-2022 in 5 languages (Bengali, Hindi, Kannada, Tamil, and Telugu), according to Broadcast Audience Research Council (BARC) India metrics. For each year, we look at the 10 most popular series in each of the 5 languages. The sample includes the five latest episodes from each series during the year in which it was ranked most popular, where available, for a total of 1,188 episodes.[4]

The Geena Davis Institute on Gender in Media, in collaboration with IAA India Chapter, curated the broadcast TV dataset based on BARC India ranking metrics, described above. As the next step, the Signal Analysis and Interpretation Laboratory at the University of Southern California's Viterbi School of Engineering and the Institute, in partnership with Google Research, applied the Representation ML Pipeline (RP) to perform visual and language analysis on the dataset to derive raw analytics of character presence and screen time.

The Representation ML Pipeline (RP) was designed to run at scale to process large volumes of data and uses machine learning models to infer visual and language signals. The RP derives visual signals perceived from face images: perceived age, perceived gender expression, and perceived skin tone. The RP aggregates these signals and their intersections over face tracks across the videos. The current study does not attempt to measure the attributes of the actors cast to the attributes of their roles (e.g., an older actor playing the role of a younger character).

The language analysis component of the RP leverages natural language processing capabilities and the multilingual nature of large language models to extract names mentioned in dialogues and to infer perceived gender associated with the name. The pipeline aggregates this data at a language level to allow comparisons across languages. We recognize that gender is multifaceted; using algorithmic estimates in the RP of perceived gender expression and gendered names is one way to identify gender inclusion. For more information about the RP, please see the Appendix.

Using the output measures from the RP vision and language analysis, the Geena Davis Institute on Gender in Media and the Signal Analysis and Interpretation Laboratory at the University of Southern California's Viterbi School of Engineering collaborated on conducting statistical analyses and insights. We report statistical estimates of "screen time" from a linear mixed model. Linear mixed models are appropriate for examining longitudinal data like this, where some series are in the dataset over multiple years because of their longstanding popularity. This model accounts for data dependencies in the observations. The data presented in this report include analysis of the raw data, and insights based on the statistical analysis of the raw data.

Findings

The analyses presented below are done at scale using machine learning to infer “visual appearance” and “spoken language usage,” and therefore provide high-level quantitative findings of screen time and language usage for popular scripted television in India. The findings cannot provide a deeply nuanced understanding of how a character is portrayed (e.g., if a character in a leadership role versus a subservient role). In other words, while these findings are important for understanding who is seen on screen, they do not account for potentially problematic portrayals, such as reinforcement of harmful stereotypes, the lack of agency, limited narrative prominence, or subjugation; these are areas of ongoing research in computational understanding of media portrayals.

From 2018-2022, we found that women occupy more than half of screen time in Indian television, but there is room for improvement in representation of older adults, people with a medium or dark skin tone, and in intersectional diversity among women, especially. While these overall findings are true across languages, they do vary in degrees. For example, popular series in Tamil see older adults and those with a medium or dark skin tone occupy more screen time than series in the other languages (Hindi, Bengali, Telugu, Kannada). In contrast, the share of screen time occupied by women, older people, and people with a medium or dark skin tone was below-average for series in Kannada and Hindi. Table 1 presents an overview of our gender, age and skin tone findings across all years, for series in the 5 languages analyzed.

TABLE 1

Ranking by language across perceived gender, age, and skin tone

PERCEIVED GENDER EXPRESSION (RANK BY MOST WOMEN ON SCREEN)	PERCEIVED AGE (RANK BY MOST 60 AND OLDER CHARACTER ON SCREEN)	PERCEIVED SKIN TONE (RANK BY MOST CHARACTERS WITH MEDIUM AND DARK SKIN TONE)
Bengali (59.0%)	Tamil (2.1%)	Tamil (23.4%)
Telugu (58.7%)	Bengali (2.0%)	Telugu (22.9%)
Tamil (54.5%)	Kannada (0.9%)	Hindi (17.9%)
Kannada (53.8%)	Hindi (0.8%)	Bengali (16.2%)
Hindi (53.5%)	Telugu (0.7%)	Kannada (13.6%)

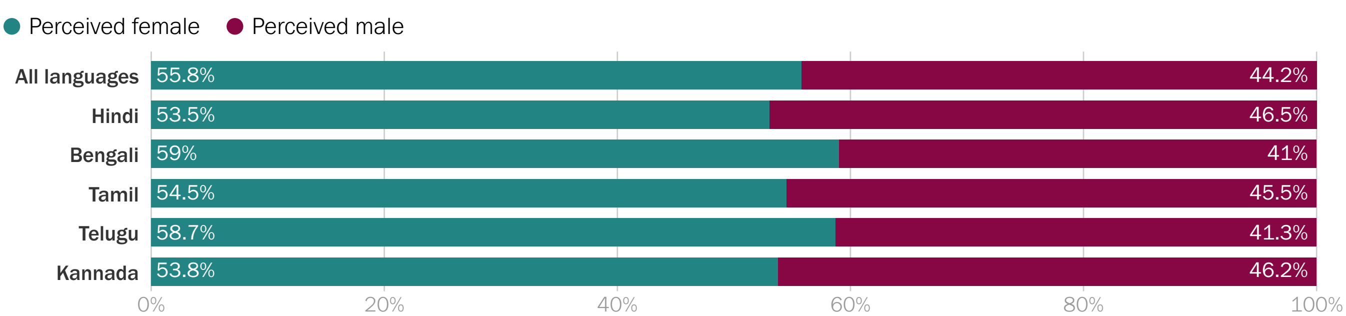
Note: The bottom three in each column are below the overall average for all Indian television. The top two in each column are above the overall average for all Indian television.

GENDER REPRESENTATION ON SCREEN

How has women's representation on TV in India evolved over the last 5 years? In our analysis of the most popular scripted series in India across 5 languages from 2018 to 2022, we find that **women occupy more screen time than men**, accounting for 55.8% (compared to 44.2% of screen time given to men). In contrast, our analysis of popular television programming in the US found that female characters occupy 38.2% of all screen time (on average from 2010 to 2021).[5] **Series in Bengali have the greatest share of screen time for women,** with women occupying 59.0% of all screen time, **followed closely by series in Telugu**, where women occupy 58.7% of all screen time. Since most families in India (98%) are a 'One-Unit, One-TV' household and women spend a lot of time co-viewing with their families,[6] it is possible that women's TV preferences were favored in this dataset and they may prefer shows with more female representation.

CHART 1

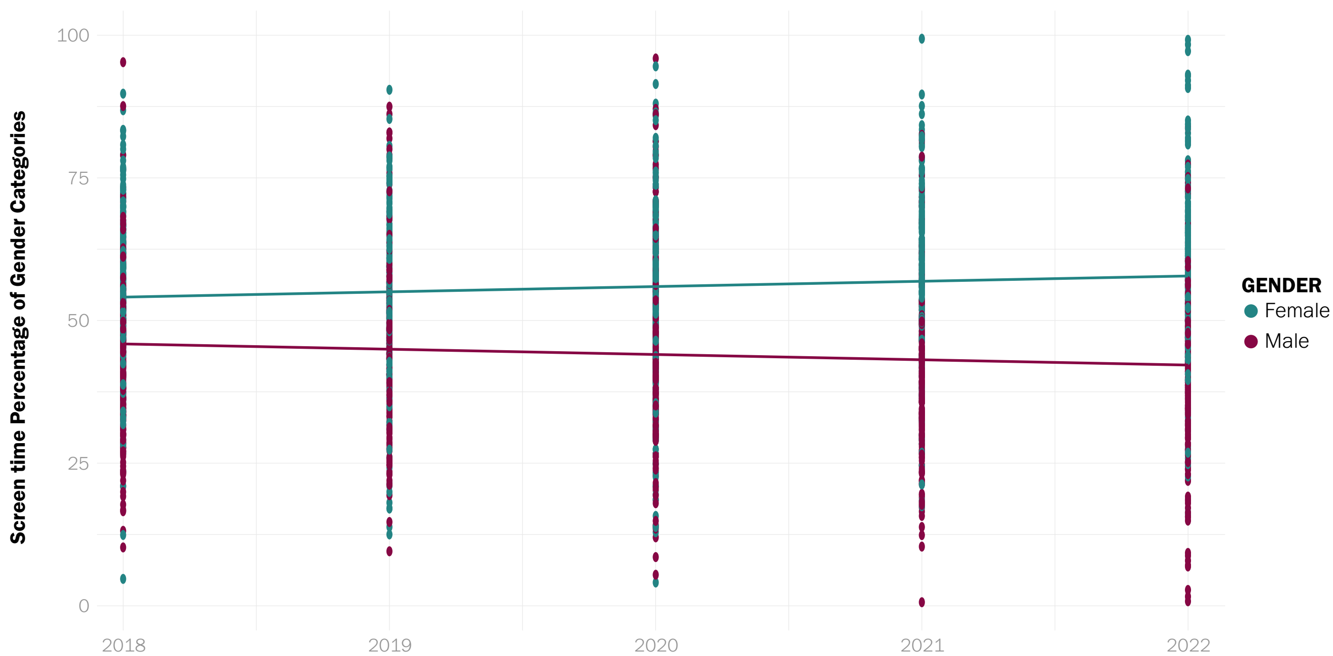
Share of screen time for female and male characters



While over the time frame screen time slightly decreases for men and slightly increases for women, the longitudinal analysis suggests these changes are not statistically significant and the proportions remain relatively steady from 2018 to 2022. Gendered screen time does not vary significantly for any of the languages analyzed, over this time frame.

CHART 2

Share of screen time for female and male characters, 2018 to 2022

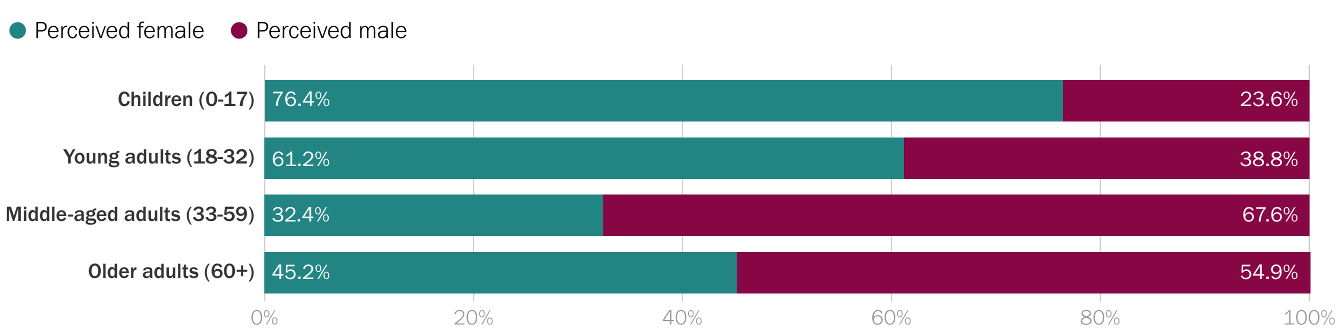


Gender at the Intersection of Age

How does the share of screen time for older women compare to older men and younger women? When looking at the intersection of gender and age, a telling pattern emerges. Older characters, especially older women, are rarely seen. Among middle-aged (33 to 59) and older adults (60+) onscreen, more than half of screen time is given to male characters (67.6% to men 33 to 59 and 54.9% to men 60 and older). On the flip side, among young characters 0 to 17, young girls occupy 3 times the screen time of young boys (76.4% compared to 23.6%). Among young adult characters (18 to 32), women occupy almost twice as much screen time as men (61.2% compared to 38.8%). In short, women outnumber men among younger characters (under 33), but men outnumber women among older characters (33 and older).

CHART 3

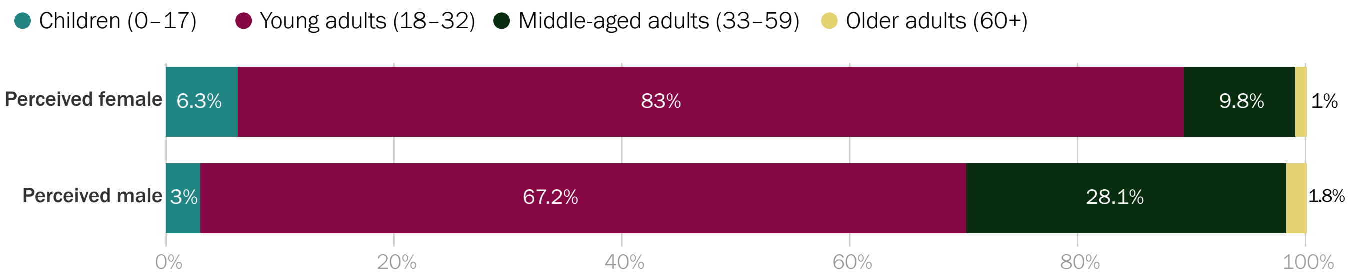
Share of screen time given to different age groups by gender



Regardless of gender, we find that youth is favored on television. Overall, girls (0 to 17) and young women (18 to 32) receive 89.3% of female screen time, while boys (0 to 17) and young men (18 to 32) account for 70.2% of male screen time. Only 1.0% of female characters' screen time and 1.8% of male characters' screen time is given to people 60 and older. This phenomenon is mirrored across the globe. According to a similar study in the U.S., just 1.2% of female screen time in 2021 and 5.5% of male screen time is given to characters over 60.[7]

CHART 4

Share of screen time given to female and male characters from different age groups



When looking at differences across languages, **the finding that older women are rarely seen remains true, especially in Kannada, Hindi, and Telugu series** (where women 60 and older receive less than 1.0% of all female screen time). **Women 60 and over are given the most screen time in series in Tamil and Bengali**; here, women are given 1.5% and 1.4% of female screen time, respectively.

When looking at age differences for men across languages, we also see that **men are more diverse with respect to age** and **that older men (60 and older) receive more screen time than older women (1.8% compared to 1.0%)**. Men 60 and older receive the most screen time in Bengali and Tamil series (3.3% and 2.9%, respectively).

CHART 5

Share of female and male screen time for different age groups, by language

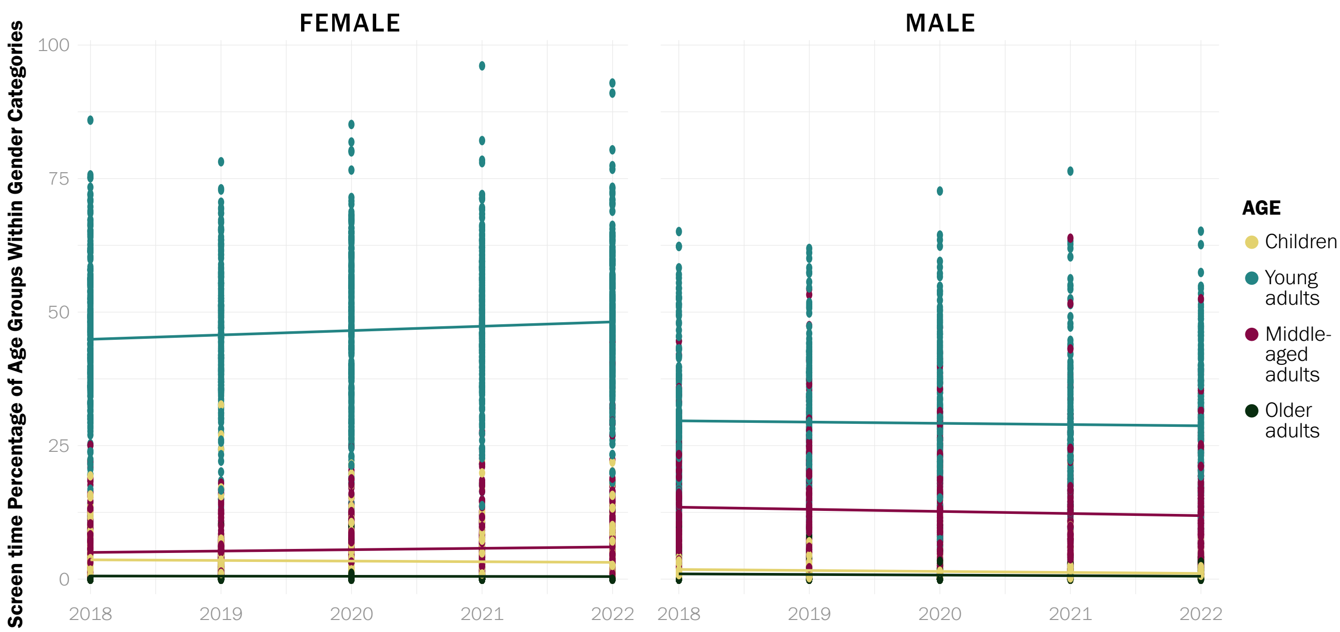
Perceived female / Perceived male

	ALL LANGUAGES		HINDI		BENGALI		TAMIL		TELUGU		KANNADA	
	Perceived female	Perceived male	Perceived female	Perceived male	Perceived female	Perceived male	Perceived female	Perceived male	Perceived female	Perceived male	Perceived female	Perceived male
Children (0-17)	6.3%	3.0%	6.1%	3.1%	5.4%	2.6%	6.5%	3.6%	6.8%	2.0%	6.5%	3.6%
Young adults (18-32)	83.0%	67.2%	83.5%	75.8%	83.1%	65.6%	79.8%	59.7%	85.7%	69.2%	82.9%	64.7%
Middle-aged adults (33-59)	9.8%	28.1%	9.7%	20.4%	10.1%	28.8%	12.2%	30.8%	7.1%	27.8%	9.7%	30.8%
Older adults (60+)	1.0%	1.8%	0.7%	0.8%	1.4%	3.3%	1.5%	2.9%	0.4%	1.0%	0.9%	0.9%

Over time, young women (18 to 32), and girls and boys (0 to 17) have seen significant changes in their share of screen time. Young women's screen time has increased roughly 2.4 points over the 5 years analyzed. Girls' and boys' screen time has decreased roughly 0.6 and 0.9 points, respectively. Older characters' screen time has been steady.

CHART 6

Screen time for female and male characters by age cohort, 2018 to 2022

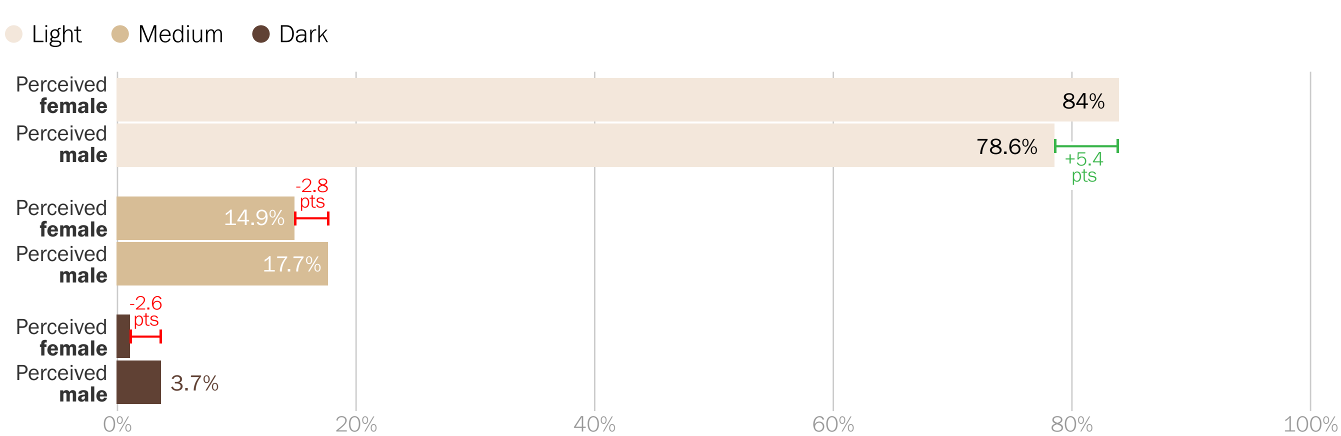


Gender at the Intersection of Skin Tone

How does female characters' screen time differ by skin tone? Even though the concepts of beauty and skin tone diversity in India have evolved,[8] **we see lighter skin tone is favored**, which is true also for films in Bollywood[9] and advertising in India.[10] Female characters are more likely than male characters to have a light skin tone and less likely to have a medium or dark skin tone. Also, when female characters are on screen, they are more likely to have a light skin tone than when male characters are on screen (84.0% compared to 78.6%). Of women's screen time, 16.0% is given to women with a medium or dark skin tone. In comparison, men with a medium or dark skin tone occupy 21.4% of men's screen time. **This may suggest that women are being held to more Eurocentric beauty standards than men** and that male characters are more diverse with respect to skin tone.

CHART 7

Light, medium, and dark skin tone representation among female and male characters



*Positive differences represent more screen time given to women of the respective skin tone. Negative differences represent more screen time given to men of the respective skin tone.

Across all languages, **women with a light skin tone receive more screen time than women with a medium or dark skin tone, especially in Kannada, Bengali, and Hindi**. Women with a light skin tone receive 90.1% of female screen time in Kannada, 86.2% of female screen time in Bengali, and 85.5% of female screen time in Hindi. Series originating in languages in the south (Telugu and Tamil), but also Hindi, give the most screen time to women with a medium or dark skin tone (21.4%, 19.9%, and 14.5%, respectively).

Men on screen are more diverse with respect to skin tone across all languages. Men with a medium or dark skin tone occupy the greatest share of screen time in series in languages spoken primarily in the south, such as Tamil (27.8% of male screen time) and Telugu (24.6% of male screen time), but also Hindi (21.0% of male screen time). Series in Kannada and Bengali show the highest share of men with a light skin tone (82.2% and 80.7%, respectively).

TABLE 2

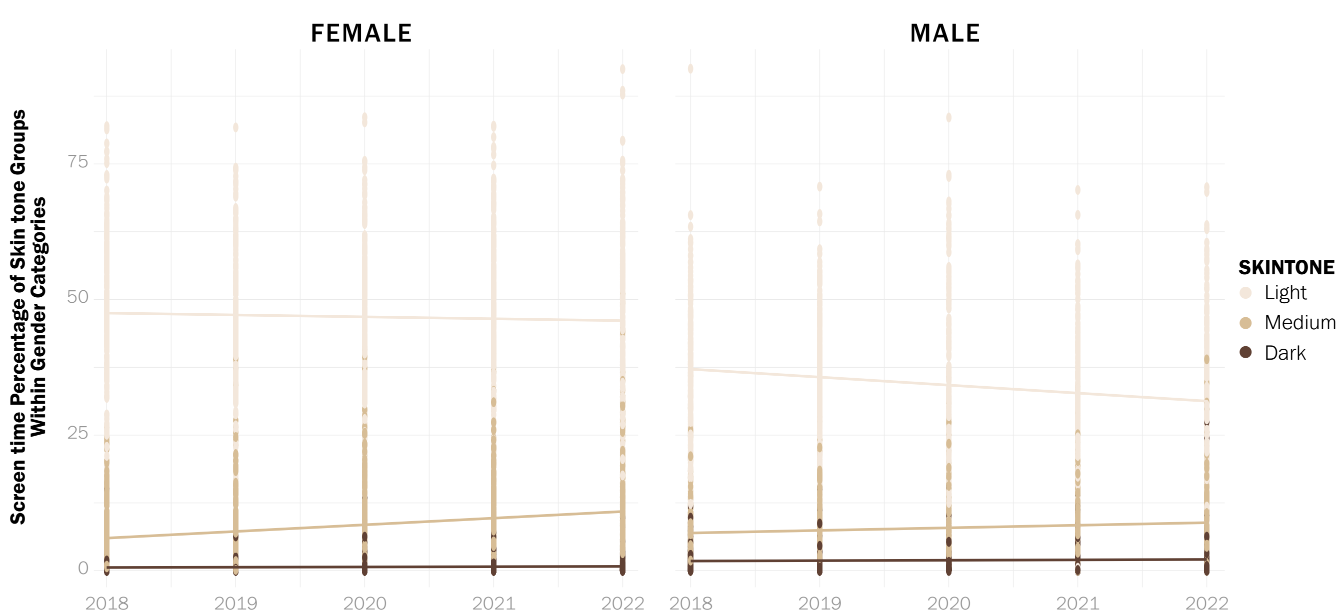
Light, medium, and dark skin tone representation among female and male characters, by language

	LIGHT SKIN TONE		MEDIUM SKIN TONE		DARK SKIN TONE	
	PERCEIVED FEMALE	PERCEIVED MALE	PERCEIVED FEMALE	PERCEIVED MALE	PERCEIVED FEMALE	PERCEIVED MALE
All Languages	84.0%	78.6%	14.9%	17.7%	1.1%	3.7%
Hindi	85.5%	79.1%	12.8%	17.6%	1.7%	3.4%
Bengali	86.2%	80.7%	12.6%	16.4%	1.2%	2.9%
Tamil	80.1%	72.2%	18.3%	21.2%	1.6%	6.6%
Telugu	78.6%	75.4%	20.6%	21.2%	0.8%	3.4%
Kannada	90.1%	82.2%	9.5%	15.9%	0.4%	1.9%

Over time, **we see a slight trend toward more diverse skin tones both for women and men on screen**. From 2018 to 2022, the share of screen time has decreased for women and men with a light skin tone (3.0 and 5.1 points, respectively), but has increased for women and men with a medium skin tone (4.9 and 2.3 points, respectively). Screen time for men and women with a dark skin tone has remained relatively stable over the past 5 years.

CHART 8

Screen time for male and female characters by skin tone, 2018-2022



Gender at the Intersections of Age and Skin Tone

When looking at skin tone by age and gender, **there is a preference for featuring younger and lighter-skinned characters, especially for female characters**. The majority of female and male characters overall are young adults (18 to 32) with a light skin tone — 70.0% of female characters and 52.9% of male characters. For female characters, the next most dominant group was young women (18 to 32) with a medium skin tone (12.1%). For male characters, the next most dominant group was middle-aged men (33 to 59) with a light skin tone (21.4%). **Older women and men with a dark skin tone are nearly erased from the television screen** — 0.0% of female characters and 0.1% of male characters.

TABLE 3

Female and male screen time by age and skin tone

	LIGHT SKIN TONE		MEDIUM SKIN TONE		DARK SKIN TONE	
	PERCEIVED FEMALE	PERCEIVED MALE	PERCEIVED FEMALE	PERCEIVED MALE	PERCEIVED FEMALE	PERCEIVED MALE
Children (0-17)	5.4%	2.6%	0.8%	0.4%	0.1%	0.1%
Young adults (18-32)	70.0%	52.9%	12.1%	11.8%	1.0%	2.4%
Middle-aged adults (33-59)	8.2%	21.4%	1.5%	5.7%	0.1%	1.1%
Older adults (60+)	0.6%	1.2%	0.4%	0.5%	0.0%	0.1%

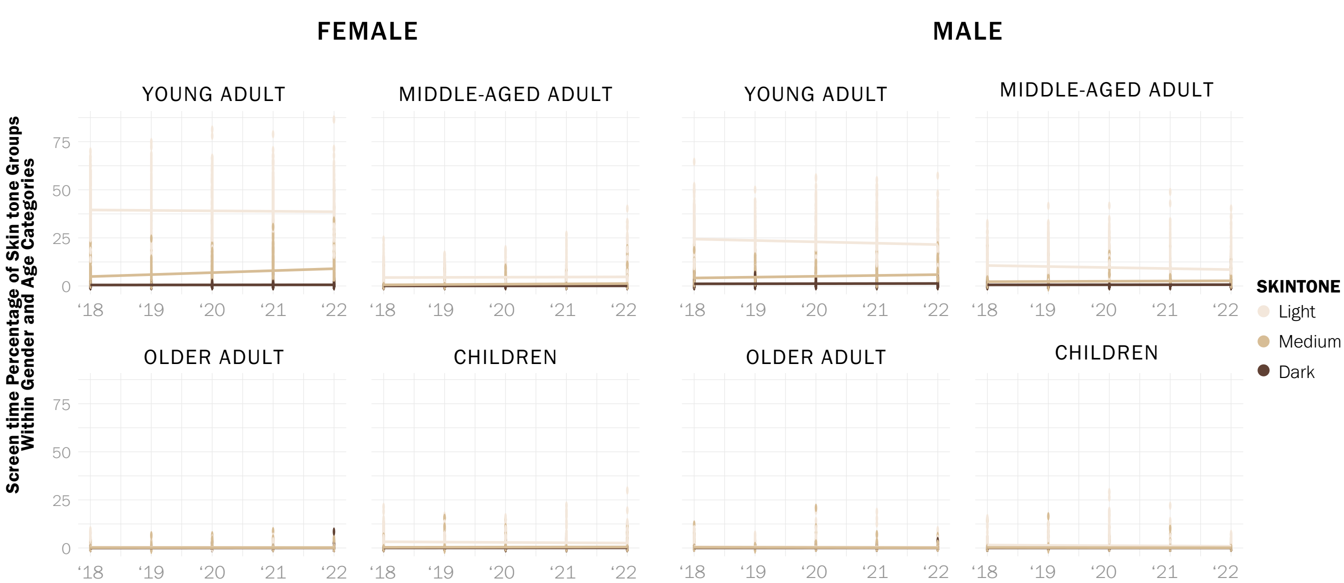
Despite this, some of the following findings may suggest a trend toward more diversity across gender, age, and skin tone:

- Over time, the share of screen time awarded to children (0 to 17) with a light skin tone has decreased, for girls (1.0 points) and boys (0.7 points).
- The share of screen time given to young adults (18 to 32) with a medium skin tone has increased, for women (4.1 points) and men (1.8 points).
- The share of screen time for young men (18 to 32) and middle-aged men (33 to 59) with a light skin tone has decreased (2.3 points and 1.6 points, respectively).

Other cohorts' share of screen time has remained relatively stable over the past 5 years.

CHART 9

Share of screen time for male and female characters by age and skin tone, 2018 to 2022



AGE REPRESENTATION ON SCREEN

Which age cohorts are more visible on screen? Television in India is viewed by multi-generational groups of friends and families, who come together to watch.[11] As such, the TV screen can be a place to bring people together from different genders, ages, spoken language(s), and appearances. Furthermore, TV serves as a window into popular Indian culture and is a source of inspiration for younger audiences.

The next segment spotlights screen time across different age cohorts. Overall, young adults enjoy the highest screen time on scripted television across languages, whereas the oldest age cohort is nearly invisible. Skin tone diversity on screen also varies with age, with older adults having the highest share of characters with a medium or dark skin tone. In contrast, children and young adults are more likely to have a light skin tone compared to other age cohorts.

The pattern across all languages shows that young adults (18 to 32) occupy the most screen time while older adults (60 and older) occupy the least. Young adults (18 to 32) account for three-fourths (75.6%) of all screen time, while those 60 and older occupy only 1.3%. By language, older adults (60 and older) receive the most screen time in Tamil (2.1%) and Bengali (2.0%) series. Young adults (18 to 32) receive the highest share of screen time in series originating in Telugu, Hindi, and Bengali, occupying more than three-fourths of all screen time from each language.

TABLE 4

Share of screen time for children and teen, young adult, adult, and older characters, by language

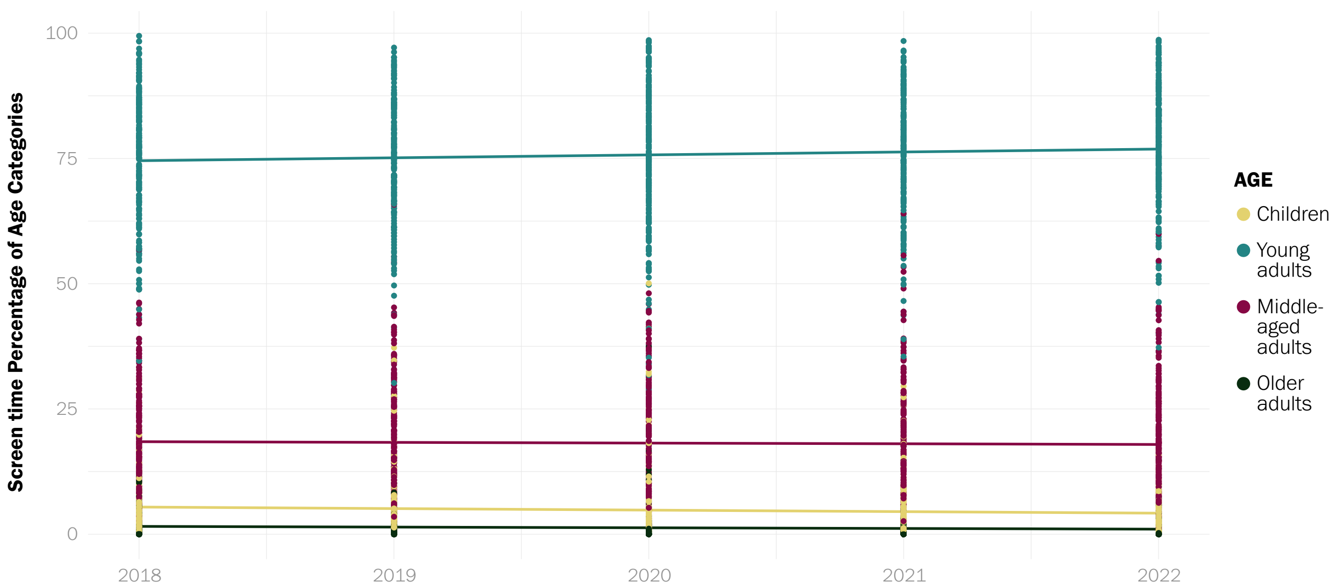
	ALL LANGUAGES	HINDI	BENGALI	TAMIL	TELUGU	KANNADA
Children (0-17)	4.8%	4.6%	4.2%	5.2%	4.8%	5.1%
Young adults (18-32)	75.6%	76.5%	76.2%	70.6%	78.8%	73.6%
Middle-aged adults (33-59)	18.4%	16.0%	17.6%	22.1%	15.7%	20.5%
Older adults (60+)	1.3%	0.8%	2.0%	2.1%	0.7%	0.9%

From 2018 to 2022, children's (0 to 17) screen time has significantly decreased (1.5 points) and the share of young adults' screen time has significantly increased (2.2 points). The share of screen time given to middle-aged and older adults has remained relatively stable over this time frame. This could be because children characters on long-running series, which are popular in India, have aged into young adult characters, or due to the evolution of regulations that limit children's participation in the labor force, such as the entertainment industry. However, this trend does not translate into increased screen time for middle-aged or older adults, as their screen time remains relatively the same from 2018 to 2022.

When comparing languages, in Tamil and Telugu series there is a decline in screen time for children (0 to 17) and an increase in screen time for young adults (18 to 32). Bengali television also shows a significant increase of 5.1 points in young adults' share of screen time from 2018 to 2022. Screen time for middle-aged and older adults has remained stable across all languages.

CHART 10

Share of screen time given to different age cohorts, 2018 to 2022



Age at the Intersection of Skin Tone

Does skin tone diversity increase with age? Yes. Characters 60 and over have the greatest skin tone diversity, compared to the other age cohorts. Of characters 60 and older, 3 in 4 have a light skin tone, compared to 79.1% of middle-aged, 82.2% of young adults, and 91.7% of children on screen. Children (0 to 17) have the least skin tone diversity — just 6.7% have a medium skin tone and 1.6% have a dark skin tone. Across all age cohorts, characters with a light skin tone have the majority of screen time in all languages.

CHART 11

Share of screen time for different age cohorts with different skin tone

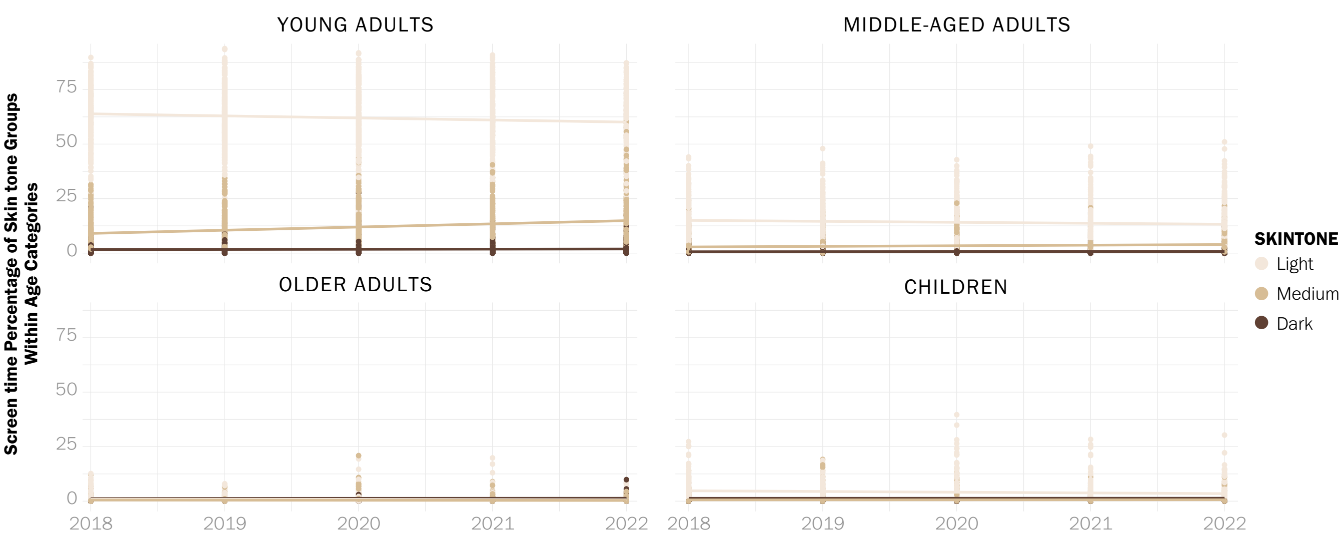
	LIGHT	MEDIUM	DARK
Children (0-17)	91.7%	6.7%	1.6%
Young adults (18-32)	82.2%	15.7%	2.1%
Middle-aged adults (33-59)	79.1%	17.9%	3.1%
Older adults (60+)	74.3%	21.9%	3.9%

Series in Tamil and Telugu have the greatest representation of characters with a medium or dark skin tone, yielding the highest share of screen time given to children, young adults, and middle-aged adults with a medium or dark skin tone. Older adults with a dark and medium skin tone receive the highest screen time in Tamil and Hindi television.

From 2018 to 2022, the share of screen time for characters with a light skin tone has decreased across all age cohorts except for older adults (60 and older), where the share has remained relatively stable. For example, the share of screen time for children (0 to 17) with a light skin tone has decreased 1.7 points. The share of screen time occupied by young adults (18 to 33) with a medium skin tone has increased over time (6.0 points). Other age cohorts at the intersection of skin tones are stable over time. **This suggests a trend toward more skin tone diversity for children and young and middle-aged adults.**

CHART 12

Share of screen time of different age cohorts and different skin tones, 2018 to 2022




simon2579/E+ via Getty Images

SKIN TONE REPRESENTATION ON SCREEN

How diverse is skin tone representation on screen in popular scripted TV in India? Regardless of the region or language, **Indians with lighter skin tones are favored in Indian television,** suggesting Eurocentric beauty standards are reinforced by popular scripted TV across the country. Applying the Monk Skin Tone scale (MST) (which ranges from 1 to 10), we grouped skin tones into three categories: light skin tone (1 to 3 MST); medium skin tone (4 to 7 MST); and dark skin tone (8 to 10 MST). For more information on the Monk Skin Tone scale (MST) and our grouping of skin tones, see the Appendix.

Across all years and regions, we find that 81.2% of all screen time is given to characters that have a light skin tone. Only 2.2% of screen time is given to characters that have a dark skin tone.

We expected to see some skin tone variation by language, because there are skin tone differences between the regions in which these languages are spoken. Characters in series in Tamil and Telugu have the most skin tone diversity of the languages analyzed. In Tamil and Telugu, characters with a medium or dark skin tone occupy more screen time than they do in the other languages (23.5% for Tamil and 22.9% for Telugu, compared to 17.9% for Hindi, 16.2% for Bengali, and 13.6% for Kannada). **However, across all languages more screen time is given to characters with a light skin tone than any other skin tone**.

TABLE 5

Light, medium, and dark skin tone representation among all characters, by language

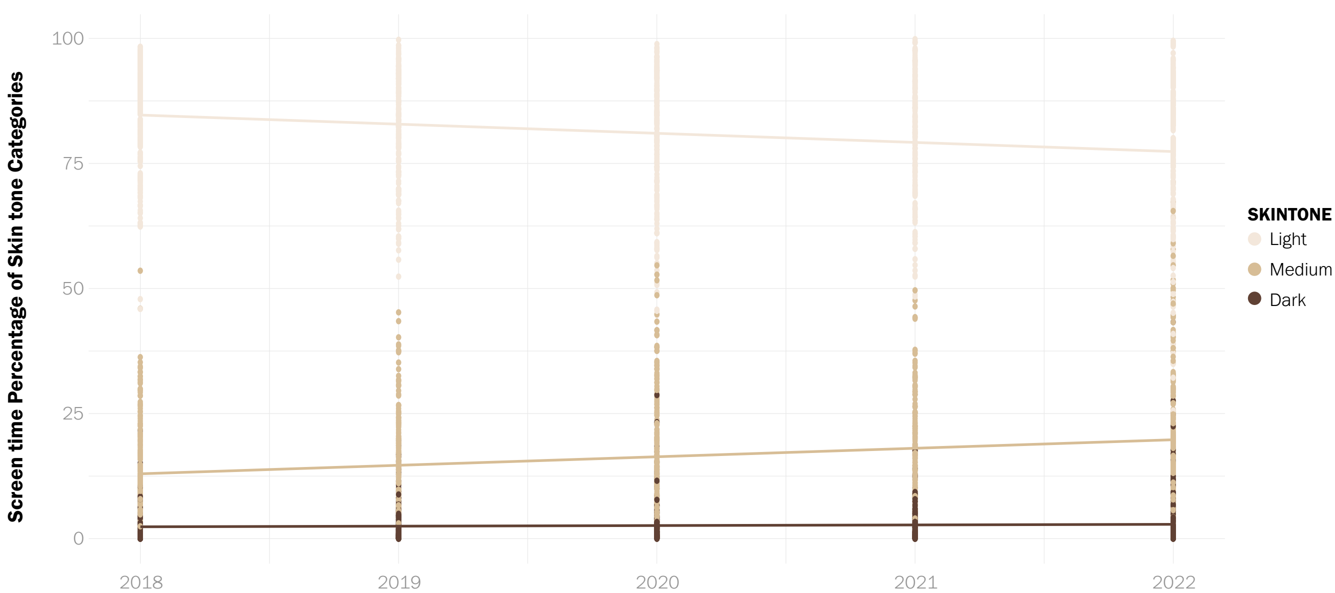
	ALL LANGUAGES	HINDI	BENGALI	TAMIL	TELUGU	KANNADA
Light skin tone	81.2%	82.1%	83.8%	76.6%	77.1%	86.4%
Medium skin tone	16.6%	15.4%	14.3%	19.7%	21.0%	12.6%
Dark skin tone	2.2%	2.5%	1.9%	3.7%	1.9%	1.0%

While most characters shown on screen have a light skin tone, the percentage of screen time given to characters with a light skin tone has decreased slightly over time. From 2018 to 2022, the share of screen time for characters with a light skin decreased by about 8 points while the share of screen time for characters with a medium skin tone has increased about 7 points. **However, representation of characters with dark skin tone was largely steady over this time frame.** Our previous study that looked at popular TV in the US found a decline in characters with light skin tone from 2010 to 2021, from 81% of screen time to 55% of screen time.[12]

From 2018 to 2022, series in Hindi, Tamil and Telugu showed a significant decrease in screen time given to characters with a light skin tone (6.4 points, 18.1 points, and 15.8 points, respectively) and a significant increase in screen time given to characters with a medium skin tone (roughly 6.7 points, 14.4 points, and 16.1 points, respectively). We also found that the share of screen time occupied by characters with a dark skin tone has increased significantly (3.7 points) in series originating in Tamil. The share of screen time occupied by characters with other skin tones and in other languages not reported has remained relatively stable over time.

CHART 13

Share of screen time for light, medium, and dark skin tone, 2018 to 2022



A COMPARISON: POPULAR SCRIPTED TV IN THE U.S. AND INDIA

In our U.S. study published in 2022, the most popular scripted television dataset included a large portion of police procedurals, medical dramas, and half-hour comedies, while the Indian scripted television dataset includes mostly mythological and drama serials.[13] Because of these general differences, as well as population differences, especially with respect to age and skin tone, it is expected that these regions will differ in representation of gender, age and skin tone on screen. Table 6 presents an overview comparing gender, age and skin tone on screen for India (2018 to 2022) and the U.S. (2010 to 2021).

TABLE 6

A comparison of findings for screen time in the U.S. and Indian popular scripted television

	POPULAR SCRIPTED TV IN INDIA (2018-2022)	POPULAR SCRIPTED TV IN THE U.S. (2010-2021)
Perceived female	55.8%	38.2%
Perceived light skin tone	81.2%	78.3%
Perceived older adult (60+)	1.3%	3.4%

LANGUAGE ANALYSIS

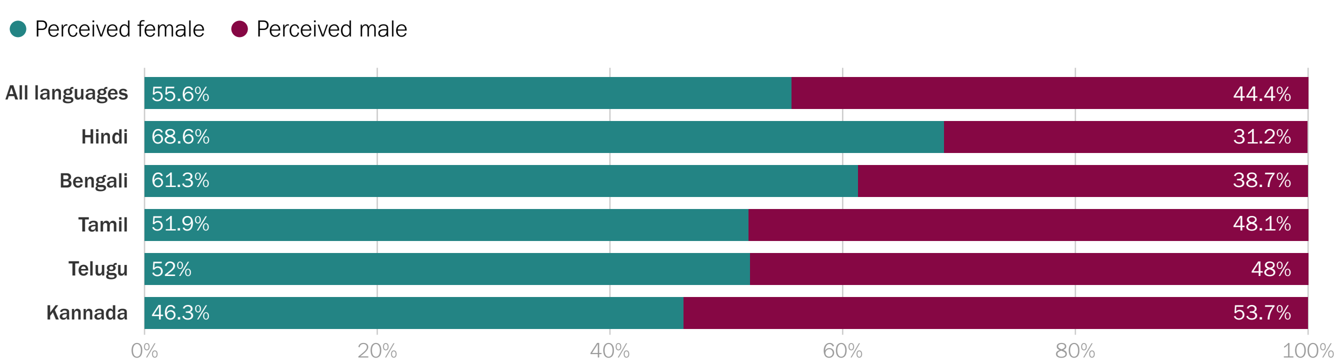
By using the RP's natural language processing capabilities, we analyzed **how often women were spoken** about by extracting all instances of person-name mentions in the dialogue and inferring whether the name is more likely to be feminine or masculine. Second, we were able to analyze **how often the extracted names were unique names.** In other words, removing names that were repeated across a series or episode.

Gender and Name Mentions

How often are women spoken about by name on screen? Across all years and all languages, the share of names mentioned identified as belonging to women is higher than that of men (55.6% compared with 44.4%) and follows the distribution of female and male screen time (55.8% compared with 44.2%). Across all years, women's names have the highest proportion of mentions in Hindi (68.8%) and Bengali (61.3%) series. In Kannada, there is a reverse trend, with men's names having a slightly higher percentage of mentions (53.7% of all names).

CHART 14

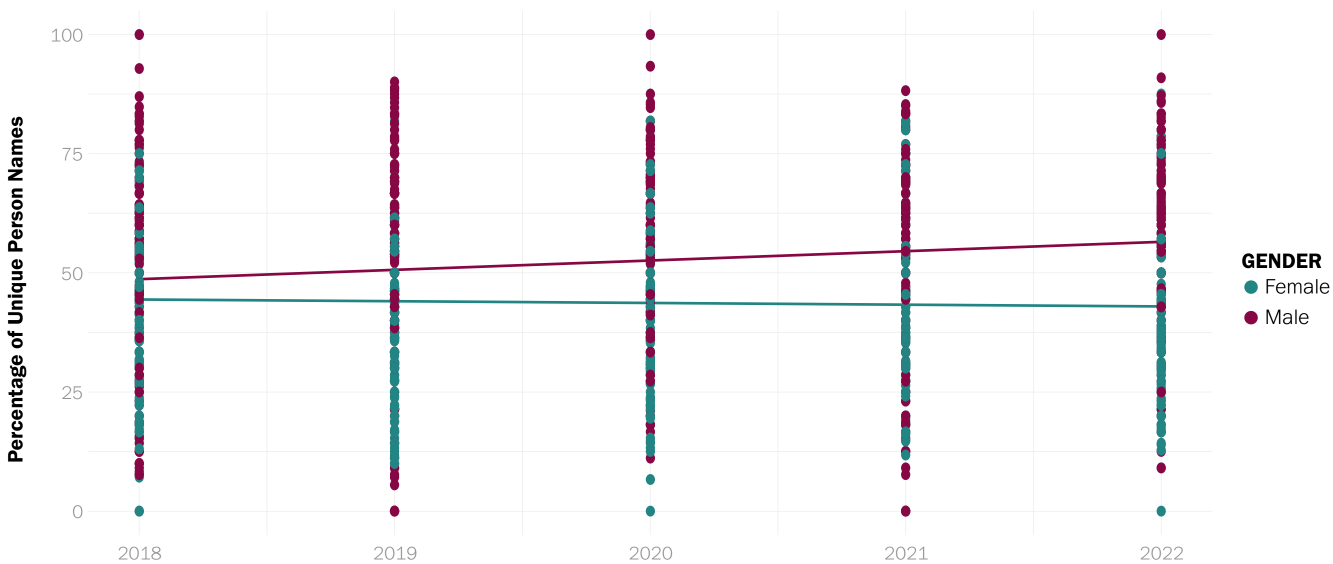
Mentions of female and male identified names



Across languages, the share of female and male identified names mentioned remained relatively stable over time. Female names mentioned have decreased roughly 15 points over time in Bengali TV series, and male character mentions have increased roughly 15 points over time. Male names mentioned have also increased in Hindi TV series by 32 points. The distribution of gendered names has remained steady for the other languages.

CHART 15

Mentions of male and female names overall, 2018 to 2022



Across all years and all languages, the percentage of *unique* female names mentioned is lower than male names mentioned (46.7% compared with 53.3%). Unique mentions are new names said in the dialogue. This could indicate that although women occupy more screen time and more female names are spoken about than male names, there are fewer individual female names mentioned than male names. In other words, while female characters overall are seen on screen more than male characters, they are uniquely the subject of dialogue less often than male characters. It could also indicate that named female characters are receiving more attention and development than named male characters, as they are being spoken about more frequently in the dialogue.

Across all years, female names have the lowest proportion of unique mentions in Bengali (35.2%), even though they occupy the majority of total mentions (61.3%) in that language. In Hindi series, the proportion of unique female names (65.6%) is higher than that of male names (34.4%), but less than total female names mentioned (68.8%). Tamil, Telugu, and Kannada series have lower proportions of unique female names mentioned than total female names mentioned.

Over time, the percentage of unique male names has increased significantly — 7.6 points from 2018 to 2022. The percentage of unique female names has remained relatively stable. By language, the percentage of unique male names has increased for Hindi series — 28.3 points from 2018 to 2022. Unique female and male names remained relatively stable for other languages.

Deepak Sethi/E+ via Getty Images

Recommendations

Based on these findings, the Institute presents the following recommendations to diversify inclusion in popular TV in India with respect to gender, age, and skin tone.

1. **Continue to show an equal share of women and men on screen but cast more diverse women with respect to age and skin tone.** Since women were found to be shown as younger and to have lighter skin tone compared to men on screen, increase the intersectional diversity of women onscreen by featuring more middle-aged women, older women, and women with a medium or dark skin tone.

2. **Show more middle-aged and older adults onscreen.** Children and young adults are well represented on screen in popular scripted Indian television. Bring more age diversity to screen by writing narratives that feature older adults, especially older women, such as stories about intergenerational families, female leaders, or matriarchs.

3. **Increase skin tone diversity on screen.** Cast characters with more diverse skin tones in prominent roles. By increasing screen time given to those with a medium or dark skin tone, popular scripted Indian television series will more accurately represent the diversity of skin tones and people of India.

Appendix

THE REPRESENTATION ML PIPELINE (RP) METHODOLOGY

The Representation ML Pipeline (RP) was developed by Google Research's MUSE team (Media Understanding for Social Exploration). It provides AI-enabled technology, across computer vision (CV) and natural language processing (NLP) modalities, for measuring and understanding the presence and portrayals of people in media content at scale. The RP uses machine learning models to infer human-centric signals perceived from faces: perceived age, perceived gender expression, and perceived skin tone, and whether people on screen are speaking. As well, the RP provides NLP capabilities via the use of LLMs (large language models) to understand language usage patterns.

The first step of the pipeline runs a model that localizes the appearance of faces in a video at the rate of 10 frames per second. This approach comes with some limitations despite continued development and training. In particular, the accuracy of the system decreases for situations of low light or poor image quality. Since most of the content we analyzed is produced in high quality, we believe this limitation has minimal effect on our findings.

Once the faces in frame have been localized, the model uses the output features from the bounding box and facial landmarks to compute a cropped region of the image around each face. We then run these "face thumbnail" images through several classification models to produce the attributes that we wish to measure. Note that this approach means that any face that is not detected in the first stage of the pipeline cannot be classified and therefore will not contribute to our overall statistics.

Share of screen time for a group (e.g., over 60 characters with light skin tone) in each video is based on the fraction of detected faces associated with different possible attribute combinations. For example, to estimate the share of screen time of middle-aged female characters in a video, we divide the number of feminine faces estimated to be between the ages of 33 and 60 by the total number of detected faces in the video. Normalizing by the total number of faces in a video allows for a comparison of these summary estimates across the videos in the study sample.

Details of the specific classifiers and summary groups used in this study follow.

Perceived Gender Expression

We recognize that gender is not a simple binary attribute and that one's gender identity may not match one's gender expression. For the purposes of this study, we use a model that classifies perceived gender expression as either predominantly feminine or predominantly masculine based on face thumbnails. Because we are interested in how characters depicted on screen are perceived by the audience, we use the algorithmic estimates of perceived gender expression to identify female and male characters. We do not analyze representation of people whose gender expression falls outside of this binary representation, and we do not attempt to infer the gender identity of the actors cast in the roles. We note that norms around gender expression vary across cultures and that no single aspect of a person's appearance suffices to determine their gender expression. A limitation of this study is that the model does not consider aspects of a person's appearance outside of the cropped face thumbnail.

Perceived Age

The age attribute is computed by a multiclass classifier that produces an estimated probability distribution over the age range zero through 100 years, in intervals of 0.5 years. We estimate the age of a face as the mean of this probability distribution for the purposes of statistical analysis. For the purposes of this study, age is grouped into four groups: children (under 18), young adults (over 18, under 33), middle-aged adults (over 33, under 60), and older adults (over 60).

Skin Tone

The RP uses a multiclass classifier model to measure skin tone from the face using a 10-point scale referred to as the Monk Skin Tone (MST) Scale developed by Harvard professor Dr. Ellis Monk. For the purposes of this study the 10-point scale has been further grouped into three coarse categories that we refer to as "light," "medium," and "dark" skin tones.

Language Analysis

The language analysis is performed on automatic speech recognition (ASR) based transcription of the video content, as we don't have access to the actual TV series scripts. The ASR transcription provides, in essence, subtitles for the TV series which would be less precise than having the screenplays/scripts which detail who the speaker is, among other factors.

Our approach is to prompt a large language model (LLM) to extract person names from the transcript, along with the name's corresponding gender association. This approach gives us results across all five languages: Hindi, Bengali, Kannada, Tamil and Telugu. We also validated a representative sample of the LLM output through human annotators.

ENDNOTES

1. Baruah, Sabyasachee, Digbalay Bose, Meredith Conroy, Shrikanth S. Narayanan, Susanna Ricco, Komal Singh, Krishna Somandepalli. 2022. “#SeeItBeIt: What Families Are Seeing on TV.” The Geena Davis Institute on Gender in Media.
2. BARC India. 2020. “TV Universe Estimates 2020.” Broadcast Audience Research Council India. https://www.barcindia.co.in/whitepaper/barc-india-tv-universe-estimates-2020.pdf.
3. Sahu, Sudhansubala. (2018). “Revisiting television in India: Mapping the portrayal of women in soap operas.” Sociological Bulletin, 67(2), no. 2 (2018): 204-219. https://www.jstor.org/stable/48566220.
4. When a series did not newly air episodes in the year in which it was ranked as one of the top 10 most popular series, the five most recent episodes available were selected. When the five latest episodes of a series were not available, we selected up to five of the latest available episodes from that respective year.
5. Baruah, Sabyasachee, Digbalay Bose, Meredith Conroy, Shrikanth S. Narayanan, Susanna Ricco, Komal Singh, Krishna Somandepalli. 2022. “#SeeItBeIt: What Families Are Seeing on TV.” The Geena Davis Institute on Gender in Media.
6. BARC India. 2018. “Impact of Co-viewing on TV Viewership.” Broadcast Audience Research Council India. https://www.barcindia.co.in/whitepaper/impact-of-co-viewing-on-tv-viewership.pdf.
7. Baruah, Sabyasachee, Digbalay Bose, Meredith Conroy, Shrikanth S. Narayanan, Susanna Ricco, Komal Singh, Krishna Somandepalli. 2022. “#SeeItBeIt: What Families Are Seeing on TV.” The Geena Davis Institute on Gender in Media.
8. Yasir, Sameer. & Jeffrey Gettleman. (2020, June 28). “India Debates Skin-Tone Bias as Beauty Companies Alter Ads.” The New York Times. https://www.nytimes.com/2020/06/28/world/asia/india-skin-color-unilever.html?smid=url-share.
9. Peters, Rebecca. (2021, February 24). “Colorism, Castism, and Gentrification in Bollywood.” The Jugaad Project. www.thejugaadproject.pub/home/colorism-bollywood.
10. GDIGM and UNICEF. (2021, April). “Gender Bias and Inclusion in Advertising in India.” The Geena Davis Institute on Gender in Media. https://seejane.org/research-informs-empowers/gender-bias-and-inclusion-in-advertising-in-india/.
11. BARC India. 2018. “Impact of Co-viewing on TV Viewership.” Broadcast Audience Research Council India. https://www.barcindia.co.in/whitepaper/impact-of-co-viewing-on-tv-viewership.pdf.
12. Baruah, Sabyasachee, Digbalay Bose, Meredith Conroy, Shrikanth S. Narayanan, Susanna Ricco, Komal Singh, Krishna Somandepalli. 2022. “#SeeItBeIt: What Families Are Seeing on TV.” The Geena Davis Institute on Gender in Media.
13. Baruah, Sabyasachee, Digbalay Bose, Meredith Conroy, Shrikanth S. Narayanan, Susanna Ricco, Komal Singh, Krishna Somandepalli. 2022. “#SeeItBeIt: What Families Are Seeing on TV.” The Geena Davis Institute on Gender in Media.

How to cite this study: Baruah, Sabyasachee, Digbalay Bose, Meredith Conroy, Shachi Dave, Sriram Giridharan, Sagar Gubbi, Rajat Hebbar, Michele Meyer, Shrikanth S. Narayanan, Romeo Pérez, Susanna Ricco, Komal Singh, Krishna Somandepalli, & Hao Zhao. 2023. “Reflecting India: An intersectional and longitudinal analysis of popular scripted television from 2018 to 2022.” The Geena Davis Institute on Gender in Media.

THANK YOU!

The authors would like to thank Bhumika Joshi, Candice Schumann, Marco Andreetto, Utsav Prabhu, Avinash Pandey (President, IAA India Chapter), Nina Elavia Jaipuria (Chairperson, Women Empowerment Committee - IAA India Chapter), and Megha Tata (Co-Chairperson, Women Empowerment Committee & Immediate Past President, IAA India Chapter). The authors would also like to thank Sofie Christensen, Summer van Houten, Melanie Lorísdóttir, Alexis Romero-Walker, and Lena Schofield for contributing to the asset collection. We also would like to thank Getty Images for the images in this report.

REPORT AUTHORS

Sabyasachee Baruah, University of Southern California

Sabyasachee Baruah is a Computer Science Ph.D. student in the Signal Analysis and Interpretation Laboratory (SAIL), at University of Southern California. After receiving his master’s degree in Computer Science and Engineering from Indian Institute of Technology, Kharagpur, India, he joined SAIL in 2018 to work in natural language analysis of media content. His research interests include using text and multimodal cues to advance media story understanding and create character representations.

Digbalay Bose, University of Southern California

Digbalay Bose is a 5th year Ph.D. student in the Signal Analysis and Interpretation Laboratory (SAIL) at University of Southern California (USC). Prior to USC, he received his master’s degree in Electrical Engineering from the Indian Institute of Technology, Bombay in 2016. He worked as a Research Software Engineer at IBM Research, India from 2016 to 2018. Currently, his research interests include multimodal machine learning for media-centered applications and affective computing.

Meredith Conroy, Geena Davis Institute on Gender in Media

Meredith Conroy is the Vice President of Research and Insights at the Geena Davis Institute, and holds a Ph.D. in Political Science from University of California, Santa Barbara.

Shachi Dave, Google Research

Shachi Dave is a Software Engineer in the Natural Language Understanding group at Google Research India. She received her master’s degree in Computer Science from University of Southern California. Currently, she works on fairness in generative models.

Sriram Giridharan, Google

Sriram Giridharan received his master's degree in Electrical and Computer Engineering from the Viterbi School of Engineering at University of Southern California. Currently, he works as a Software Engineer in the Google Drive team.

Sagar Gubbi, Google Research

Sagar Gubbi is a postdoctoral researcher in the Natural Language Understanding group at Google Research India. He has a Ph.D. in Electrical Communication Engineering from the Indian Institute of Science, Bangalore. Currently, he works on safety in generative models and grounding language in UI.

Rajat Hebbar, University of Southern California

Rajat Hebbar is a Ph.D. student in the Signal Analysis and Interpretation Laboratory (SAIL) at University of Southern California. Previously, he received his master's degree in Electrical and Computer Engineering at USC in 2018. He works on audio representation learning for multimedia applications in his research.

Michele Meyer, Geena Davis Institute on Gender in Media

Michele Meyer, Senior Director of Research & Methodologies at the Geena Davis Institute, is a mixed-methods social scientist who specializes in the representation of marginalized groups in new, social, and entertainment media. She holds a Ph.D. in Media and Communication from UNC Chapel Hill.

Shrikanth (Shri) Narayanan, University of Southern California

Shrikanth (Shri) Narayanan is University Professor and Niki & C. L. Max Nikias Chair in Engineering at the University of Southern California (USC), and holds appointments as Professor of Electrical and Computer Engineering, Computer Science, Linguistics, Psychology, Neuroscience, Otolaryngology-Head and Neck Surgery, and Pediatrics. Shri leads the Signal Analysis and Interpretation Laboratory (SAIL) at USC Viterbi School of Engineering, which focuses on developing engineering approaches to understand the human condition and in creating machine intelligence technologies that can support and enhance human experiences.

Romeo Pérez, Geena Davis Institute on Gender in Media

Romeo Pérez is the Senior Project Manager of Research and Insights at the Geena Davis Institute, and focuses on representations and perceptions of queer people in entertainment media.

Susanna Ricco, Google Research

Susanna Ricco is a Senior Staff Software Engineer in Google Research. She leads a team focused on building fairer and more inclusive computer vision systems and advancing responsible development and use of next-generation technologies. She is particularly interested in leveraging computational techniques to understand how individuals and communities are portrayed in multimedia content. She holds a Ph.D. in Computer Science from Duke University and is a proud alum of Harvey Mudd College with a bachelor's degree in Mathematics - Computer Science.

Komal Singh, Google Research

Komal Singh is a Senior Program Manager in Google Research, author of children's STEM books, board member, and a TEDx speaker. At Google, her work in Responsible AI and ML-Fairness is dedicated to making AI systems and products work equitably for people and communities. She is interested in innovative ways to apply technology to measure representation and improve inclusivity in mainstream media. Prior to Google, Komal worked in technology consulting at Accenture. Her STEM book series dedicated to diversifying characters for underrepresented children, have been critically acclaimed. Komal holds a master's degree in Computer Science.

Krishna Somandepalli, Google Research

Krishna Somandepalli received his Ph.D. in Electrical and Computer Engineering from the Viterbi School of Engineering at University of Southern California (USC). Prior to USC, he received his master's degree in Electrical and Computer Engineering from the University of California at Santa Barbara. Currently, he works at Google Research at the intersection of affective computing and human-centric understanding in multimedia content.

Hao Zhao, Google

Hao Zhou is a software engineer in the Tensorflow model analysis team at Google. Currently, he is focusing on building model analysis tools and bulk inference solutions for large language models. Prior to Google, he received his Ph.D. in mechanical engineering from California Institute of Technology and his bachelor's degree in mechanical engineering from Peking University.

ABOUT THE GEENA DAVIS INSTITUTE

Since 2004, the Geena Davis Institute has worked to mitigate unconscious bias while creating equality, fostering inclusion and reducing negative stereotyping in entertainment and media. As a global research-based organization, the Institute provides research, direct guidance and thought leadership aimed at increasing representation of marginalized groups within six identities: gender, race/ethnicity, LGBTQIA+, disability, age, and body type. Because of its unique history and position, the Institute can help achieve true onscreen equity in a way that few organizations can.
Learn more at seejane.org.