Behind the Camera and on the Screen: Gender in Movies*

Jérémie Haese[1]

[1]ETH Zurich

November 4, 2025

Preliminary and incomplete draft.
Please do not circulate.

Abstract

This paper studies gender representation in the film industry and how behind-the-camera roles influence casting, marketing, and content outcomes. I use machine learning methods on a dataset of over 500,000 movies produced between 1900 and 2024 to measure gender dynamics in creative labor and media output. Women comprise 34% of lead roles, 15% of directors, 24% of producers, and 16% of screenwriters. Male directors exhibit a 12 percentage points higher probability of working with same-gender leads, demonstrating homophily effects. The #MeToo movement influenced these patterns: female directors increased same-gender casting while male directors decreased it. Promotional materials show gradual convergence in gender representation. While female-directed films initially appear more thematically distinctive, this disappears when controlling for genre, indicating differences reflect gender concentration in certain genres.

Keywords: Movies, Gender, Computer Vision, Natural Language Processing
JEL Codes: L82, J16, N30, C45

*Haese (corresponding author): jeremie.haese@gess.ethz.ch. I thank Giovanna d'Adda, Clément Bellet, Emanuele Cagnolo, Luise Eisfeld, Franziska Kaiser, Hong Luo, Yabra Muvdi, Christian Peukert, Martin Quinn, Rob Seamans, Alexander Staub, and Christian Zehnder for their helpful comments. The paper has benefited from the feedback of participants at the 9th Monash-Paris-Warwick-Zurich Text-as-Data Workshop and the doctoral seminars at HEC Lausanne and NYU Stern School of Business.

1 Introduction

Gender inequality in the labor market has declined in most countries, but significant gender gaps persist in certain professions and sectors (Bertrand, 2020; Goldin, 2023). Fields like STEM and top executive positions remain heavily male-dominated (Bertrand, 2018; Cimpian et al., 2020). One factor frequently cited as perpetuating these gender disparities is the existence of social norms and stereotypes about gender roles and abilities (Akerlof and Kranton, 2000; Bertrand, 2020; Carlana, 2019).

Popular culture and media play a powerful role in shaping these stereotypes through the representations and portrayals of men and women. The images, narratives, and framings in movies, TV shows, books, advertising, and other media can reinforce traditional gender role conceptions (Adukia et al., 2023; Ash et al., 2021; Bellet and Pocchiari, 2024; Kearney and Levine, 2015; Kerkhof and Reich, 2023; Luo and Zhang, 2024; Mulvey, 1975; Riley, 2022).

The movie industry, with 4.8 billion admissions and 1.8 billion subscribers to SVOD platforms in 2024[1], has a substantial impact on societal perceptions. Understanding how genders are represented in movies can shed light on the role of stereotypes in society and how popular culture reinforces them. Moreover, it can inform discussions on the importance of diverse representation, as the presence of women role models has been shown to help reduce gender gaps in traditionally "masculine" fields (Porter and Serra, 2020; Riley, 2022).

This study aims to address two key questions: First, how have genders been represented in movies? Second, to what extent do patterns of gender representation in movies reflect the gender composition of key decision-making roles in the industry? By examining these questions through an analysis of director-lead actor relationships, promotional materials, and content characteristics, this research seeks to contribute to a better understanding of how gender dynamics in creative leadership positions relate to representation patterns in the movie industry, and by extension, the cultural narratives presented to global audiences.

To address these questions, I compile and analyze a comprehensive dataset of over 500,000 movies produced globally between 1900 and 2024. The analysis integrates three main sources of data: (i) movie credits, with a focus on the gender of directors, producers, screenwriters, and lead actors; (ii) movie promotional material, specifically posters; and (iii) movie content, captured through synopses. Connecting these three dimensions allows me to examine how gender representation behind the camera

[1] See https://www.rte.ie/entertainment/2025/0519/1513629-global-cinema-attendance-drops-in-2024/ and https://nscreenmedia.com/svods-grow-subs-ad-revenue-competition/

relates to casting decisions, visual marketing, and narrative content. In particular, movie posters serve as a standardized and widely available form of promotional media, offering insight into prevalent societal norms and values at a given time.

This study employs a multifaceted approach, combining computer vision techniques to analyze movie posters, natural language processing to analyze movie content, and econometric methods to examine the relationships between industry demographics and on-screen representations. This comprehensive methodology allows for a nuanced understanding of gender representation across different aspects of movie production and marketing.

My findings reveal that while female representation in key roles such as directors, producers, and screenwriters has increased over time, the majority of these roles remain male-dominated, influencing the dynamics of gender inclusion in casting and content creation. Notably, male directors display higher levels of gender homophily, showing approximately a 12 percentage points higher probability of working with leads of their own gender compared to female directors. The #MeToo movement, which emerged in 2017 as a widespread response to sexual harassment and gender-based abuse in the workplace, appears to have influenced these patterns, with cross-country evidence suggesting that female directors became more likely to cast female leads following the movement, while male directors showed the opposite trend. Analysis of promotional materials shows a gradual convergence in gender representation on movie posters, though disparities remain. Regarding thematic content, while female-directed films initially appear more thematically distinctive, this relationship becomes statistically insignificant when controlling for genre, indicating that apparent differences in content distinctiveness are driven by genre composition rather than director gender itself.

These findings indicate gradual changes in gender representation and inclusion in the movie industry and highlight structural factors that may influence equity outcomes. The observed patterns of gender homophily suggest that increasing female representation in leadership roles could lead to more diverse casting through network effects and changing industry norms. The #MeToo movement's apparent impact on female directors' casting choices demonstrates how social movements can influence industry practices, potentially accelerating progress toward more equitable representation (Luo and Zhang, 2021, 2024). However, the null finding regarding thematic distinctiveness indicates that female directors are not inherently creating more distinctive narratives, but may be concentrated in genres that typically diverge from mainstream content. The progress in gender representation on promo-

tional materials reflects shifting societal attitudes but emphasizes the need for continued advocacy to ensure equitable representation. By examining the intersection of marketing practices, social movements, and societal norms, this research provides insights for understanding the mechanisms behind persistent gender disparities and potential pathways for change in media representation.

This study contributes to the literature by combining an extensive dataset of over 500,000 movies with advanced machine learning techniques, including computer vision and natural language processing, to analyze gender representation comprehensively. Unlike previous research that often focuses on smaller samples or specific genres, this work provides a longitudinal analysis spanning more than a century of filmmaking. By systematically linking workforce demographics with on-screen and promotional representation, the study reveals how gender homophily operates in creative industries and demonstrates the importance of controlling for structural factors like genre when analyzing content differences. The analysis of the #MeToo movement as a quasi experiment provides novel evidence on how external social shocks can influence industry practices and gender dynamics in cultural production.

2 Background and related literature

2.1 Gender and the movie industry

Research on gender representation in film reveals persistent biases across multiple dimensions. Studies analyzing on-screen representation find systematic underrepresentation of women, gendered differences in visual framing and dialogue, and limited thematic range for female characters (Istead et al., 2022; Kumar et al., 2022; Haris et al., 2023; Mazières et al., 2021). In the U.S., women have remained significantly underrepresented in key production roles including directors, writers, and producers (Smith et al., 2023).

The #MeToo movement[2] catalyzed notable industry shifts, with producers previously associated with Harvey Weinstein becoming more likely to collaborate with female writers post-scandal, while institutional reforms such as the Academy's diversification of Oscar voters led to strategic increases in hiring women and minorities for visible roles (Baldassari, 2023; Luo and Zhang, 2021). Luo and Zhang (2024) show that production teams' responses to #MeToo differ by gender composition, with all-male teams developing more female-protagonist projects while female-inclusive teams prioritized breaking into male-dominated domains. Audience dynamics further complicate these patterns, with female-

[2]The #MeToo movement is a social movement against sexual abuse and harassment. It gained widespread attention in 2017 when the hashtag #MeToo went viral, encouraging individuals worldwide to share their experiences and highlight the prevalence of sexual misconduct. The movement has significantly influenced discussions about gender dynamics in various sectors, including the film industry. See: https://en.wikipedia.org/wiki/MeToo_movement.

led films experiencing greater evaluation dispersion and systematic bias from male reviewers, though platform design changes have helped mitigate some biases (Aguiar, 2024; Stroube and Waguespack, 2024). Cultural factors also matter, as films portraying gender roles consistent with local stereotypes tend to perform better in markets where such stereotypes persist (Michalopoulos and Rauh, 2024).

I expand on this literature by providing comprehensive analysis of gender representation across over 500,000 movies, linking workforce demographics with on-screen and marketing portrayals to understand how structural industry changes influence representation trends.

2.2 The impact of reducing the gender gap in executive roles

The question of what happens when the gender gap is reduced in executive roles has garnered significant attention in recent research, revealing both promising advancements and complex tensions in workplace dynamics. Studies consistently find that female managers reduce gender gaps in promotions and wages, particularly through their involvement in hiring and promotion decisions (Kunze and Miller, 2017; Hirsch, 2013; Flabbi et al., 2019). However, this research also reveals nuanced effects: while female leaders benefit women at the top of organizations, impacts on lower-wage workers can be mixed, and increased female peer competition may paradoxically harm some women's advancement prospects.

The implementation of gender quotas provides additional insights, with Norway's corporate board quota significantly reducing gender earnings gaps within boards while having limited broader spillover effects (Bertrand et al., 2019; Yang et al., 2019). State dependence in hiring practices suggests that organizations appointing female managers are more likely to continue doing so, indicating potential for positive feedback loops in promoting diversity (Bossler et al., 2020).

I build on this discourse by examining the movie industry as a unique cultural context, showing how increased female representation in directorial roles correlates with distinct casting patterns, highlighting the broader societal implications of female participation in creative leadership.

2.3 The role of media and culture in identity and representations construction

Media plays a dual role in perpetuating traditional gender norms and catalyzing social change. Research demonstrates persistent gender disparities and stereotypical portrayals across various media formats, from children's books to news coverage to advertising (Adukia et al., 2023; Ash et al., 2021; Bellet and Pocchiari, 2024; de Courson et al., 2024; Kerkhof and Reich, 2023). However, media exposure can also drive progressive change, as evidenced by cable television's positive impact on women's

status in rural India (Jensen and Oster, 2009). Historical analysis reveals that traditional cultural narratives portraying women in submissive roles correlate with persistent contemporary gender inequality (Michalopoulos and Xue, 2021).

I contribute to this literature and the emerging field employing computer vision and natural language processing for studying identity construction in popular culture (Adukia et al., 2023; Ash et al., 2021; Bellet and Pocchiari, 2024; Kerkhof and Reich, 2023; Michalopoulos and Rauh, 2024; Tan and Wang, 2024; Voth and Yanagizawa-Drott, 2023) by providing comprehensive quantitative analysis of gender representation evolution in cinema, linking workforce demographics with content creation and marketing materials to understand how industry changes influence cultural narratives.

3 Data

I curated a large and novel dataset on movies produced since cinema's birth from diverse sources encompassing visual, textual, and metadata, allowing for a multi-faceted examination of gender dynamics in the movie industry. Descriptive statistics are provided in Table 1.

3.1 Movie metadata

The raw dataset includes detailed metadata for approximately 500,000 movies produced worldwide between 1900 and 2024, sourced from The Movie Database (TMDB). As shown in Panel A of Table 1, the sample is comprehensive and covers movies across multiple genres, with drama (33.1%), comedy (20.2%), and documentary (16.6%) being the most represented categories. US productions constitute 23% of the sample, reflecting the global scope of the dataset.

The metadata provides essential contextual information for each movie, including release date, synopses, genre classifications, production countries, and production companies information. These details are instrumental in examining trends in movie production and thematic content over time and across different regions and genres.

3.2 Movie credits

Comprehensive cast and crew information for each movie is also sourced from TMDB. This data details the individuals involved in the production of each movie, with a particular focus on key creative and leadership roles. As shown in Panel B of Table 1, the dataset includes 591,096 director credits, 462,258 lead role credits, 361,162 producer credits, and 145,738 screenwriter credits.

For each individual, I collected their names, gender, roles in the movie, and in the case of actors, their cast order and character names. The gender classification process employed a two-tiered approach

Table 1: Descriptive Statistics

Panel A: Movie Characteristics					
Variable	N	Mean	St. Dev.	Min	Max
Release Year	575,584	1994.181	29.458	1900	2024
US	575,584	0.230	0.421	0	1
Action	575,584	0.061	0.240	0	1
Adventure	575,584	0.034	0.181	0	1
Animation	575,584	0.059	0.235	0	1
Comedy	575,584	0.202	0.402	0	1
Crime	575,584	0.053	0.224	0	1
Documentary	575,584	0.166	0.372	0	1
Drama	575,584	0.331	0.471	0	1
Family	575,584	0.037	0.190	0	1
Fantasy	575,584	0.030	0.171	0	1
History	575,584	0.024	0.153	0	1
Horror	575,584	0.071	0.256	0	1
Music	575,584	0.041	0.198	0	1
Mystery	575,584	0.027	0.163	0	1
Romance	575,584	0.077	0.266	0	1
Science Fiction	575,584	0.028	0.166	0	1
TV Movie	575,584	0.043	0.203	0	1
Thriller	575,584	0.069	0.253	0	1
War	575,584	0.016	0.126	0	1
Western	575,584	0.015	0.120	0	1
Panel B: Gender Representation by Role					
Female Share in:					
Directors	591,096	0.153	0.360	0	1
Lead Roles	462,258	0.343	0.475	0	1
Producers	361,162	0.239	0.427	0	1
Screenwriters	145,738	0.158	0.365	0	1
Panel C: Movie Poster Characteristics					
Year	369,934	2001.806	19.570	1950	2023
Width	369,934	950.747	475.834	500	2000
Height	369,934	1407.788	709.607	750	3000

Notes: This table presents descriptive statistics for the movie sample. Panel A reports statistics for movie characteristics, where all genre variables are indicator variables. Panel B reports the share of female personnel in key production roles. The sample covers movies from 1900 to 2024. The number of observations for directors (N = 591,096) exceeds the number of unique movies because some movies have multiple directors. Similar variations in sample sizes across roles reflect different patterns of multiple credit assignments and data availability. Panel C reports descriptive statistics for movie poster characteristics.

beginning with direct TMDB gender labels when available, which covered 61% of cases. For the remaining individuals, I utilized the genderize.io API to classify their gender based on their first name, enabling me to cover 99% of the individuals present in my sample. To evaluate the reliability of this inference method, I conducted a validation exercise using a stratified random sample comprising 500 female and 500 male first names drawn from the TMDB dataset. The genderize.io predictions were

compared against the ground truths (TMDB data). The resulting performance metrics demonstrate a high degree of accuracy: overall classification accuracy was 85% (the proportion of correct predictions out of all predictions), with a precision of 0.88 (the proportion of names predicted as female that were actually female) and recall of 0.81 (the proportion of actual female names correctly identified) for female names, and a precision of 0.82 and recall of 0.89 for male names. These results suggest that the API performs well across both genders, although slightly more conservatively for female names. A detailed confusion matrix is provided in Figure A.1, offering further insight into the classification performance and error distribution.

Panel B reveals persistent gender disparities across all roles, with women comprising only 15% of directors, 34% of lead roles, 24% of producers, and 16% of screenwriters.

3.3 Visual marketing data

Visual marketing data consists of movie promotional posters collected from TMDB. For each movie in the sample, I collected the official English-language poster. These posters provide a snapshot of how movies are marketed and positioned to audiences, offering valuable insights into the evolving visual representation of gender in movie promotion.

The dataset includes a total of 369,934 posters, with an average width of 950.75 pixels and an average height of 1407.79 pixels. The temporal span of the poster dataset covers movies released from 1950 to 2023.

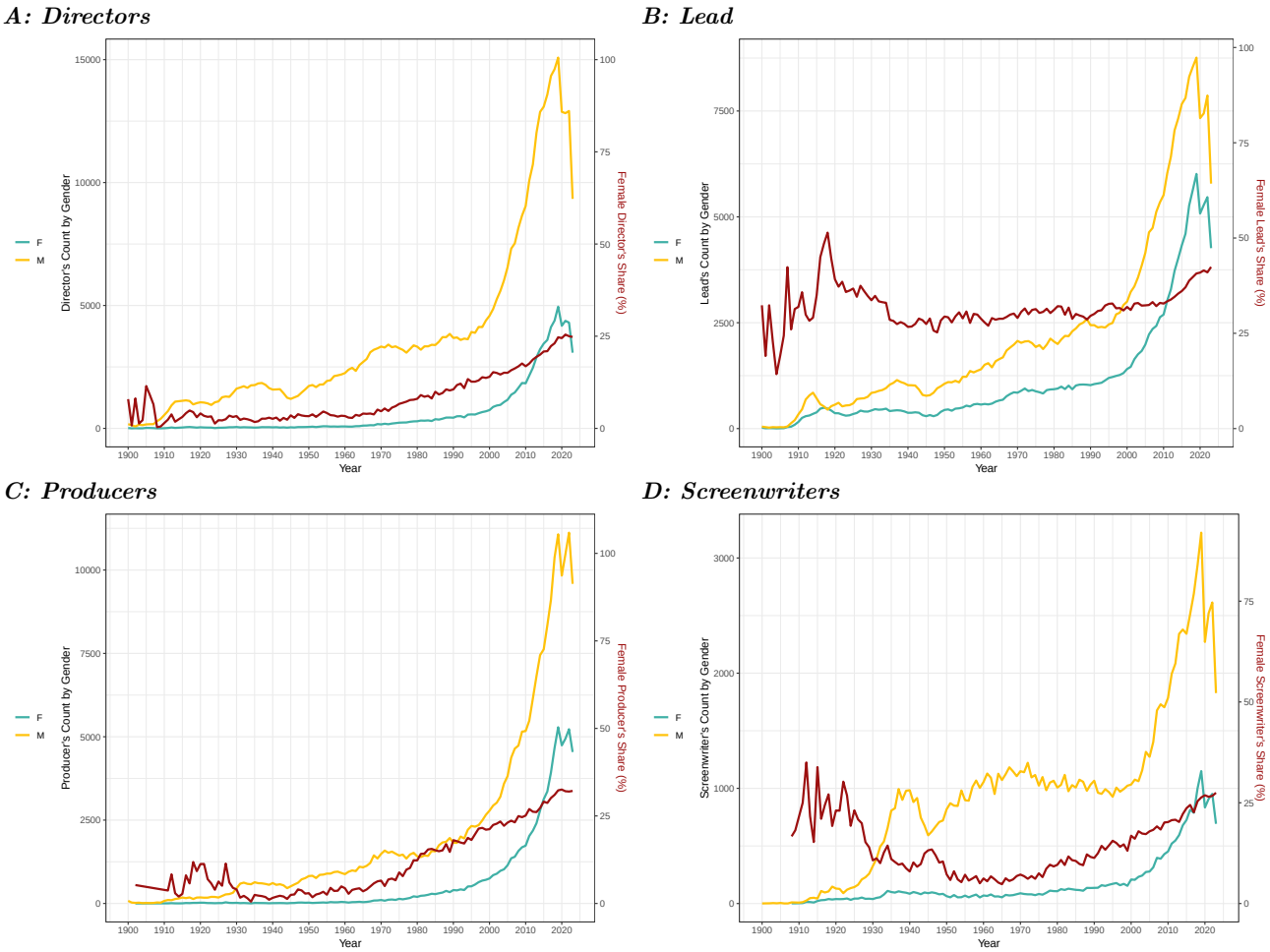
By analyzing the composition and visual elements of these posters, the study investigates gender representation in movie marketing, including trends in the inclusion and prominence of male and female characters over time. This analysis enables the examination of how marketing practices reflect and potentially reinforce gender disparities in the industry.

4 Gender trends in the movie workforce

4.1 Descriptive evidence

The composition of the movie workforce's key roles –encompassing directors, producers, screenwriters, and leading actors– has historically exhibited substantial gender imbalances. Figure 1 presents the share of male and female professionals in key roles over time. Panel A focuses on directors, Panel B on lead actors, Panel C on producers, and Panel D on screenwriters. Although the share of female professionals has increased noticeably in each category, male professionals have consistently constituted the majority for most of the observed period. In 2024, women represented 25% of directors, 42% of

Figure 1: Gender representation in the movie workforce

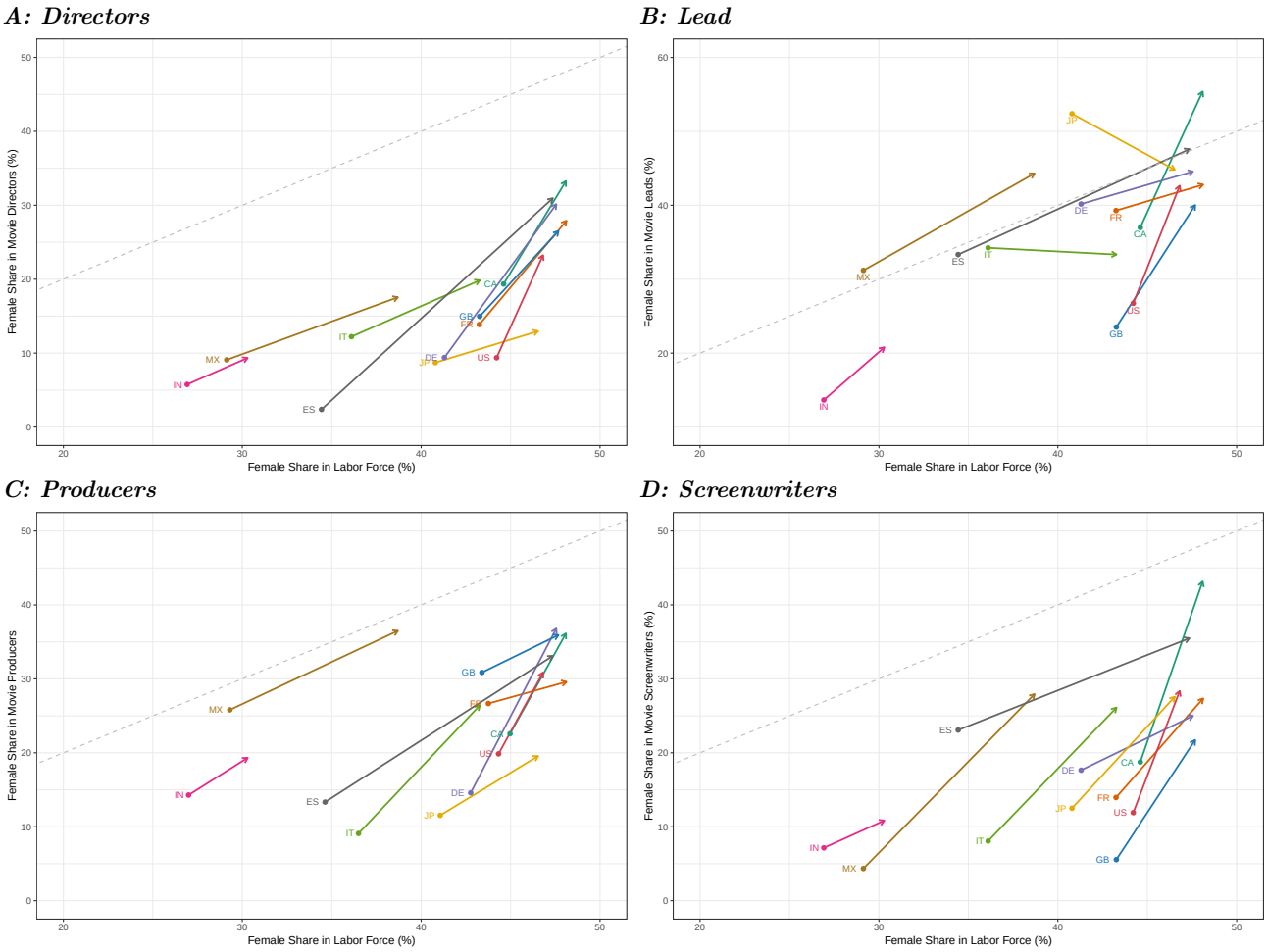


Note: Each panel shows the yearly count of male (yellow line) and female (teal line) representatives in a given role (directors, lead actors, producers, and screenwriters). The red line shows the share of female personnel, calculated as the ratio of female-associated personnel to the total number of individuals working in that role for all movies produced in a given year.

lead roles, 33% of producers, and 29% of screenwriters.

The patterns of gender representation vary substantially across movie genres, revealing important structural factors that influence industry dynamics. Figure A.2 presents the evolution of female representation in key roles disaggregated by genre from 1900 to 2020. Several notable patterns emerge from this analysis. First, certain genres exhibit consistently higher female representation across roles. Documentary filmmaking has relatively high female participation, with female directors accounting for approximately 25–30% in some periods. Family and romance genres also show higher female representation, particularly among lead actors, where the female share frequently reaches or exceeds 50%. By contrast, genres such as action, war, and westerns show consistently low female representation

Figure 2: Change in female participations in film roles vs. labor force participation (1990–2024)



Note: Data from the World Bank and TMDB. Each panel shows the change in female representation in a key film industry role between 1990 (1991 for producers) and 2024, plotted against female labor force participation rates (% of total population ages 15-64) in the ten largest market in terms of movie production. Arrows indicate the direction and magnitude of change from 1990 (tail) to 2024 (head). The dashed 45-degree line represents perfect correlation between workforce participation and film role participation.

across roles, with female directors typically below 10% in recent years. Crime and thriller genres fall between these extremes, with female representation increasing over time but remaining below that observed in other genres. These genre-specific patterns have important implications for interpreting aggregate trends in gender representation. The concentration of female professionals in specific genres suggests that apparent differences in creative output may reflect genre selection rather than inherent differences in creative approach.

Figure 2 compares the evolution of female participation in major film industry roles with changes in female labor force participation in the ten largest national film markets between 1990 and 2024. The 45-degree parity line in each panel offers a reference for proportional representation: countries below

the line reveal sectors in which gender inclusion in film has failed to keep pace with broader economic trends. While all roles show upward trends in female participation, the magnitude of change varies substantially by role and country.

In particular, the gaps between workforce and on-screen representation have narrowed most visibly for lead actors, whereas roles behind the camera (especially directing and screenwriting) continue to lag behind national labor force trends. For instance, even in countries like the U.S., Great Britain, and France where female labor force participation exceeds 45%, the share of women among directors and screenwriters remains under 30% in 2024. This persistent underrepresentation highlights the specific structural barriers women face in gaining creative authority within the industry, even as broader labor market gender gaps have narrowed.

Overall, this figure illustrates that progress toward gender equity in the movie workforce has been uneven and role-dependent, and that labor market inclusion alone is insufficient to ensure equal participation in high-visibility or decision-making positions in media production.

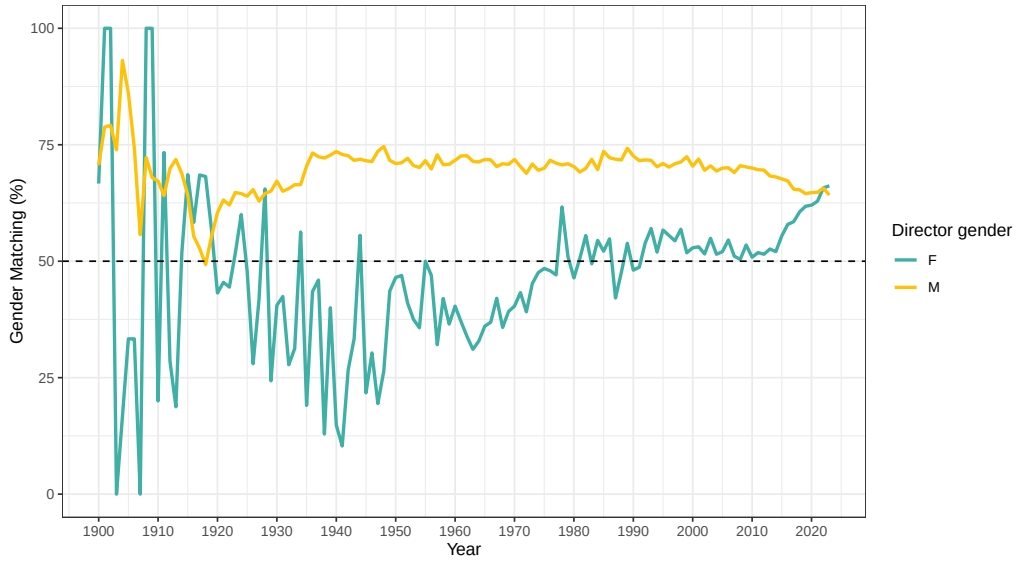
4.2 Do directors choose a lead that matches their gender?

4.2.1 General trends

An important dimension of these workforce dynamics is the degree of gender homophily, defined here as the tendency for directors to pick a lead who shares the same gender (hereafter called *gender homophily*). Figure 3 plots the evolution of gender homophily toward lead actors that match directors' gender. Early in the 20th century, these measures fluctuate widely, in part reflecting the smaller number of total productions. As the industry matures and structures, male-directed productions consistently exhibit relatively high and stable levels of gender homophily, remaining above 0.7 for most of the period. Female-directed productions, by contrast, vary more substantially, but in recent decades show an upward trend, converging toward the range observed for male-directed movies and even reaching a higher level following the #MeToo movement.

To examine these patterns more rigorously, a fixed-effects regression approach is employed. In this specification, the dependent variable Homophily_{it} is a dummy variable taking a value of 1 if the director's gender matches lead gender for movie i released in year t. The key explanatory variable is MaleDirector_{it}, a dummy variable that equals 1 if the director of movie i is male and 0 otherwise. In addition, the model includes year fixed effects γ_t to control for time-specific shocks or trends, and a vector of genre fixed effects δ_g to account for systematic differences across various movie types.

Figure 3: Average gender matching percentage between director's gender and lead gender over time



Note: This figure presents two lines – one for female directors (teal) and one for male directors (yellow). The horizontal axis represents release years. The vertical axis shows the gender matching percentage, indicating how frequently the director's gender aligns with the lead actor's gender, ranging from 0% (no gender match in the given year, e.g.: female directors working with only male leads) to 100% (full gender match in the given year, e.g.: female directors working with only with female leads). The dashed horizontal line at 50% serves as a reference point, i.e. a situation where directors work half the time with a lead matching their gender.

Table 2: Homophily by director's gender

	Homophily		
	(1)	(2)	(3)
Constant	0.56***		
	(0.01)		
MaleDirector	0.12***	0.12***	0.11***
	(0.01)	(0.02)	(0.02)
Year FE	No	Yes	Yes
Genre FE	No	No	Yes
Observations	434,360	434,360	434,360
R^2	0.01	0.01	0.04

Note: The omitted category is female directors. Standard errors are clustered at the year level in parentheses. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Formally, the regression equation can be expressed as:

$$\text{Homophily}_{it} = \beta\, \text{MaleDirector}_{it} + \gamma_t + \delta_g + \varepsilon_{it}.$$

Regression results, presented in Table 2, further highlight the observed pattern. For all specifications, controlling for time trends and genre-specific factors or not, the coefficient on MaleDirector_{it} is positive

and statistically significant. This indicates that movies led by male directors are associated with higher levels of gender homophily in the choice of lead actors, even after accounting for temporal changes and differences across movie genres. In other words, the likelihood of having a lead that matches the director's gender is higher when the director is male. Male directors show between 11 and 12 percentage points higher probability of working with leads of their own gender compared to female directors.

4.2.2 The effects of #MeToo

Cross-country variation in director homophily: The global rise of the #MeToo movement in late 2017 sparked widespread discussion of gender bias and representation in film.

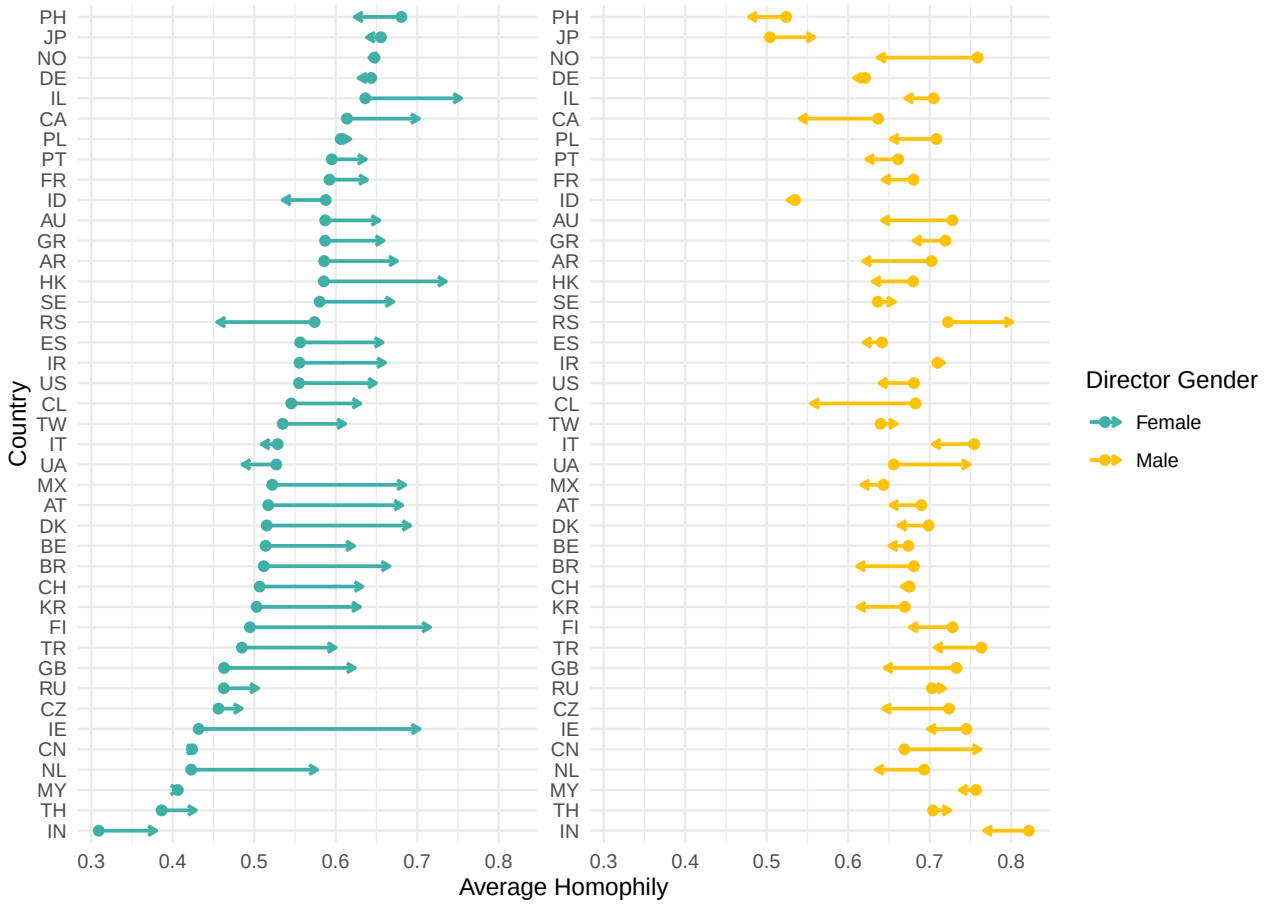
Figure 4 illustrates the change in average gender homophily among film directors before and after the MeToo movement, disaggregated by director gender and country. Among female directors, most countries show an increase in average homophily, suggesting that female directors became more likely to cast female leads following the movement. This trend is particularly pronounced in countries like Ireland, Finland, and Denmark. In contrast, male directors exhibit more stable patterns, with smaller shifts in homophily across the same period. While some countries such as France and Spain show slight decreases, others like Germany and Switzerland exhibit little to no change. The U.S. (the largest production country in the sample) follow a similar pattern to most countries, an increase in homophily for female directors and a decrease though smaller for male directors. The overall divergence between the two panels suggests that female directors responded more dynamically to the post-MeToo context in terms of casting behavior, potentially reflecting increased awareness or agency in shaping on-screen gender representation.

#MeToo impact on director homophily: I employ an event study framework to examine how the #MeToo movement affected gender matching between directors and lead actors. The primary specification estimates

$$\text{Homophily}_{it} = \sum_{k} \beta_k \mathbf{1}[t - t_{\text{MeToo}} = k] + \alpha_i + \delta_g + \gamma_c + \varepsilon_{it}, \tag{1}$$

where Homophily_{it} indicates whether director i's film in year t features a same-gender lead actor, α_i represents director fixed effects, δ_g genre fixed effects, γ_c country fixed effects, and year 2017 serves as the reference period. I estimate this model separately for male and female directors rather

Figure 4: Change in average director homophily before and after MeToo by gender and country

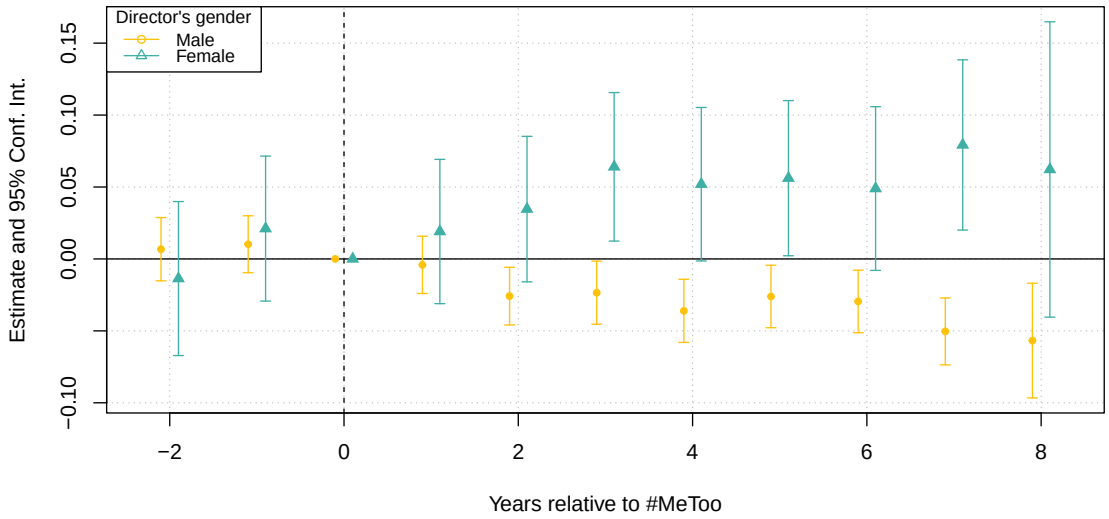


Note: This figure depicts the change in average gender homophily among film directors before and after the start of the #MeToo movement (the reference point is October 2017) for the period 2010–2024. Homophily is defined as the share of films in which the gender of the lead actor matches the gender of the director. For each country and director gender, average homophily was computed across years and visualized as a directional arrow from the pre- to post-MeToo period. Points indicate pre-MeToo homophily levels, while arrows show the direction and magnitude of change. The data are faceted by director gender, and countries are ordered by the pre-MeToo homophily levels for female directors to facilitate interpretation. The sample is restricted to countries with at least 200 movies in each country-gender-period cell, resulting in a sample of major film-producing countries with sufficient directorial activity.

than pooling as both groups are affected by #MeToo but potentially in an heterogeneous manner. The director fixed effects absorb all time-invariant director characteristics including talent, networks, and baseline gender preferences, ensuring identification comes from within-director changes in casting behavior around #MeToo. This approach requires directors to have multiple films in the sample, with post-MeToo effects identified from 14,579 directors active both before and after 2017 (2,686 female and 11,893 male directors).

Figure 5 provides support for the parallel trends assumption. In the pre-period (2015-2016), coefficients for both male and female directors fluctuate around zero and are not statistically significant, with no

Figure 5: Changes in director's homophily after #MeToo by gender and number of movies directed



Note: This figure shows event study coefficients examining changes in gender homophily around the #MeToo movement (October 2017). The figure shows the OLS estimates of the β_k coefficients obtained from $\text{Homophily}_{it} = \sum_k \beta_k \mathbf{1}[t - t_{\text{MeToo}} = k] + \alpha_i + \delta_g + \gamma_c + \varepsilon_{it})$. The vertical line at $t = 0$ marks the #MeToo timeline. Coefficients represent the estimated impact on the probability of same-gender director-lead matching relative to the pre-period baseline. 95% confidence intervals are displayed, with standard errors clustered at the director level.

evidence of differential trending between genders. Significant effects emerge only from period $k = 2$ (2019) onward with male directors showing persistent negative effects starting two years post-#MeToo, while female directors exhibit positive and significant effects in years three, five, and seven. This delayed response aligns with film production timelines, as movies typically require 2 to 4 years from greenlight to theatrical release. The absence of any significant effects in period $k = 1$ (2018) further validates the identification strategy, as films released in 2018 would have been cast and produced before #MeToo, making immediate effects implausible. This pattern suggests the estimates capture genuine behavioral responses to #MeToo rather than continuation of pre-existing trends or spurious timing coincidences.

To complement the event study's year-by-year effects, I also estimate the average treatment effect using the following specification:

$$\text{Homophily}_{it} = \beta \cdot \text{Post}_t + \alpha_i + \delta_g + \gamma_c + \epsilon_{it}, \tag{2}$$

where Post_t equals one for years after 2017, providing an average treatment effect of MeToo on within-

Table 3: Effect of MeToo on homophily by director's gender

	Homophily					
	Female Directors			Male Directors		
	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.05***	0.04***	0.04***	-0.03***	-0.03***	-0.03***
	(0.01)	(0.01)	(0.01)	(0.005)	(0.005)	(0.005)
Director FE	Yes	Yes	Yes	Yes	Yes	Yes
Genre FE	No	Yes	Yes	No	Yes	Yes
Country FE	No	No	Yes	No	No	Yes
Directors	2,689	2,689	2,689	11,896	11,896	11,896
Observations	8,530	8,530	8,530	44,423	44,423	44,423
R^2	0.43	0.43	0.44	0.40	0.42	0.42

Note: This table presents separate regressions for female directors (columns 1-3) and male directors (columns 4-6), examining the effect of the #MeToo movement on gender homophily in director-lead casting decisions. The dependent variable is an indicator equal to 1 if the director's gender matches the lead actor's gender. Post is an indicator equal to 1 for movies released after October 2017 (the start of the #MeToo movement). Columns (2) and (5) include director and genre fixed. Columns (3) and (6) add country fixed effects. Standard errors are clustered at the director level. The sample period covers 2015-2024. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

director casting choices. The split sample approach allows me to test whether MeToo generated heterogenous responses among female directors and male directors.

Table 3 presents the estimates of this specification. The results reveal that female directors exhibit a 4-5 percentage point increase in homophily after the #MeToo movement across all specifications. In contrast, male directors show a 3 percentage point reduction in homophily following #MeToo. These magnitudes are economically meaningful: from a pre-#MeToo baseline of 59.7%, the 4 percentage point increase for female directors represents a 6.7% relative rise in their propensity to cast female leads. Similarly, male directors' 3 percentage point decrease from their baseline rate of 66.4% constitutes a 4.5% relative decline in casting male leads. The 7 percentage point divergence between male and female directors' responses, with women increasing same-gender casting while men decrease it, suggests #MeToo fundamentally reshaped gender matching patterns in the movie industry.

5 Gender inclusion in promotional material

The evidence thus far documents distinct patterns of gender homophily in casting decisions, with male directors showing significantly stronger same-gender matching tendencies. These production-level patterns, however, represent only one dimension of how gender manifests in the film industry. Marketing materials, particularly movie posters, serve as the primary visual communication between films and potential audiences, potentially reflecting or diverging from the gender composition observed in casting. Examining whether promotional materials display similar gender-based patterns would

Figure 6: Examples of movie posters with detected faces classified by gender

(a) Poster 1 **(b)** Poster 2 **(c)** Poster 3 **(d)** Poster 4

Note: Each poster shows the detected faces with their bounding boxes, the label classifying the face as female (f) or male (m), and the probability associated to the prediction.

reveal whether the documented casting tendencies extend to how films are presented to audiences.

To analyze how the inclusion of women has evolved over time on movie posters, I employ a machine learning approach using computer vision models for face detection and gender classification.

The analysis pipeline first runs the RetinaFace model (Deng et al., 2020) to detect all faces present in a given movie poster image. The detected face images are then sequentially passed to a custom gender classifier, which predicts whether each face belongs to a female or male. This process allows me to quantify the prominence and inclusion of women and men on movie posters by statistics such as the proportion of poster faces that are female/male and the relative positioning/size of female vs. male faces within posters over time. Examples of the pipeline output are provided on Figure 6. Additional details on the methodology and the performance of the face detection and gender classification models are provided in Appendix A.2.1. By applying this face detection and gender classification pipeline to the full dataset (369,934 movie posters from 1950-2023), I can systematically measure how the representations of gender have evolved across decades and genres in this influential form of popular media marketing. After applying the pipeline, I identified 251,469 movie posters where I detected at least one face.

Figure A.3 shows the trends in the number and the relative size of detected faces by gender from 1950 to 2024. Conditioning on the presence of at least one face on the movie poster, male faces consistently outnumber female faces on posters throughout the observed period. Nevertheless, the

gap has narrowed since the 2010s. Relative face size shows the opposite trend. Female faces have been more prominent in terms of size throughout the observed period.

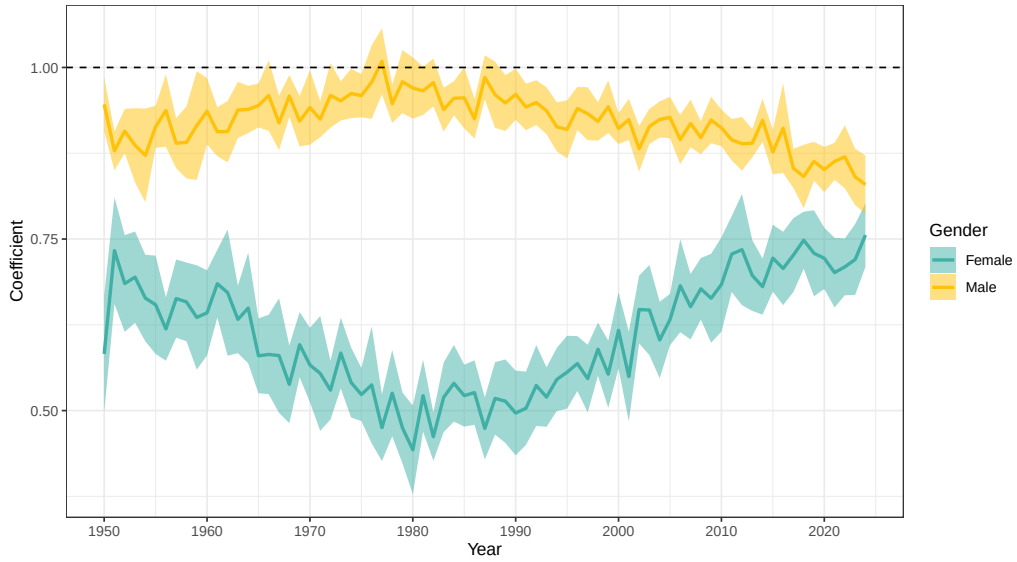
Figures A.4 and A.5 explore the spatial positioning of faces on posters across decades, conditional on face presence. On Figure A.4, detected faces consistently appear centered horizontally and slightly above the vertical midpoint for both genders, consistent with conventional movie poster designs. Differences between male and female spatial distributions are subtle and stable over time, suggesting that framing conventions remain relatively consistent across decades. Figure A.5 visualizes the spatial differences in detected male and female face locations across decades. Overall, differences remain subtle, with most regions showing minimal divergence in detection frequency (indicated by white areas). Yellow regions denote a greater relative frequency of male face detections, typically concentrated near the upper-central portion of the frame, while teal regions indicate more frequent female detections, often located slightly lower. These spatial asymmetries are modest but observable in the 1950s, 1960s, and from the 1990s onward. By contrast, the 1970s and 1980s exhibit the least distinguishable gender differences in spatial positioning, suggesting a temporary flattening of framing distinctions during that period.

The relationship between the gender of credited actors and their inclusion on promotional movie posters provides a lens into how marketing materials reflect societal attitudes toward gender representation. I estimate the following model:

$$\text{PosterFaces}_{it} = \gamma_t \cdot \text{Credits}_{it} \times \text{Year}_t + \delta_g + \epsilon_{it},$$

to examine whether the number of male and female faces included on promotional posters (PosterFaces_{it}, specified as $\text{MalePosterFaces}_{it}$ and $\text{FemalePosterFaces}_{it}$) can be explained by the number of male and female actors credited in movies (Credits_{it}, represented as MaleCredits_{it} and $\text{FemaleCredits}_{it}$). Fixed effects for genres (δ_g) control for genre-specific marketing norms, while the interaction term captures the evolution of this relationship over time. To illustrate the notion I aim to measure, Figure 6 - Poster 1 depicts two characters, one female and one male for the movie "A Streetcar Named Desire". The first two actors in casting order are Vivien Leigh (female) and Marlon Brando (male), so the two estimated γ would be equal to 1, showing a proportional representation of the cast on the movie poster. This approach highlights whether and how the correlation between credited roles and poster representation has changed, reflecting potential shifts in societal attitudes or industry practices regarding gender

Figure 7: Relationship between movie credits and faces on posters over time



Note: The figure plots the coefficients from regressions of the number of male and female actors credited in movies on the number of male and female faces on movie posters, respectively, across the years 1950–2023. The coefficients reflect the strength of the relationship between credits and poster representation, with values close to one indicating proportional representation. Separate regression models were run for male and female credits interacted with year dummies and genre fixed effects were included to control for genre-specific marketing practices. Shaded areas around the lines represent 95% confidence intervals.

representation.

Figure 7 shows the results of the regression. Throughout much of the observed period, the coefficients for male credits remain relatively stable below but close to one, indicating a consistent relationship between credited male actors and their representation on posters. This suggests that male actors have historically been reliably featured in marketing materials in proportion to their roles in movies, reflecting their perceived importance in driving audience interest.

In contrast, the coefficients for female credits are generally lower than those for males, particularly in the 1980s and the 1990s. This disparity points to a historical underrepresentation of female actors on posters relative to their credited roles. Such underrepresentation may reflect gendered marketing practices that prioritized male visibility. Over time, however, the gap between the coefficients for male and female credits narrows, particularly after 2000. This convergence suggests a shift in industry practices, with increasing recognition of the value of female actors in promotional materials.

6 Gender of directors and the positioning of movie content

The analysis of promotional materials reveals systematic patterns in how gender is visually represented to audiences. Yet these marketing decisions may reflect deeper differences in the content being

produced. A natural question emerges: do films created by directors of different genders exhibit distinct thematic characteristics? To investigate this, I turn to analyzing the narrative content itself, examining whether director gender correlates with measurable differences in story positioning.

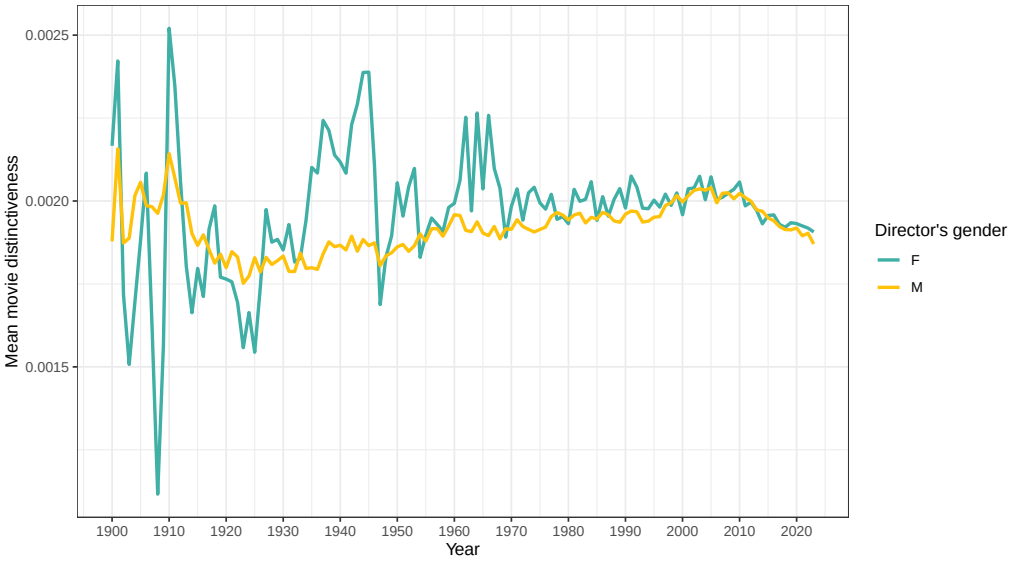
A key avenue of inquiry within the broader discourse on gender and media production is whether and how a director's gender correlates with the thematic content of the movies they create. To investigate this relationship, I rely on natural language processing techniques to quantify a movie's narrative "distance" from previously produced content, treating this measure as an indicator of the movie's thematic distinctiveness. This framework allows me to pose the following research question: does the director's gender systematically align with differences in the thematic positioning of a movie relative to the established cinematic corpus? The analysis thus centers on examining potential patterns in content distinctiveness as a function of director gender, offering insight into underlying dynamics of cultural production.

A central goal of this analysis is to quantify the extent to which each movie's content diverges from previously produced movies. To capture this notion of distinctiveness, I apply natural language processing and embedding models. I use the Longformer model (Beltagy et al., 2020) to convert movie synopses into numerical representations that capture their full narrative content without truncation. For each new movie, I calculate a baseline by averaging the numerical representations of all previously released films, creating a reference point that represents the established content norm at that time. I then quantify movie distinctiveness as the distance between the new film's representation and this historical baseline, where greater distances indicate more substantial narrative departures from conventional content patterns. This approach provides a systematic method for measuring creative and thematic innovation in cinema by comparing each film against the accumulated body of prior cinematic work. Additional information regarding the methodology are provided in Appendix A.2.2.

Before turning to regression analysis, a descriptive exploration illustrates broad patterns in content distinctiveness over time. Aggregating movies by director gender and calculating the average distinctiveness score by year reveals notable trends. Figure 8 plots the average distinctiveness of movies directed by men and women from the early 1900s to the present.

Early in the history of cinema, the relatively small number of female-directed movies exhibit, on average, higher distinctiveness scores. While the early data are noisy due to fewer observations, the pattern suggests that movies by female directors were often less similar to the prevailing content

Figure 8: Average content distinctiveness of movies by director gender over time



Note: This figure plots the mean distinctiveness scores of movies, as defined by their narrative distance from previously released movies, separately for male- and female-directed movies. Each data point represents the average distinctiveness for all movies released in that year.

baseline, potentially reflecting distinct thematic interests or less conventional narratives. Over the mid-twentieth century and into the contemporary period, the gap in distinctiveness between male- and female-directed movies tends to shrink. By the late twentieth and early twenty-first centuries, the difference in average distinctiveness by director gender becomes less pronounced, though female-directed movies still appear marginally more distant from the existing content norm on average.

A series of regression models are estimated to systematically assess the relationship between a movie's distinctiveness and the director's gender. The baseline specification takes the form:

$$\text{Distinctiveness}_{it} = \beta \cdot \text{MaleDirector}_{it} + \gamma_t + \delta_g + \varepsilon_{it},$$

where $\text{Distinctiveness}_{it}$ is the distinctiveness measure for movie i released in year t, MaleDirector_{it} is an indicator variable that equals one if the director is male, γ_t and γ_g are respectively year and genre fixed effects. Standard errors are clustered by year to account for potential correlations within a given release year.

Table 4 presents the relationship between director gender and movie distinctiveness. The baseline specification shows a negative correlation between male directors and distinctiveness. However, this relationship weakens when accounting for year fixed effects and becomes statistically insignificant

Table 4: Movie distinctiveness by director's gender

	Distinctiveness		
	(1)	(2)	(3)
Constant	0.001967***		
	(0.000003)		
MaleDirector	-0.000027***	-0.000020***	0.000003
	(0.000003)	(0.000004)	(0.000004)
Year FE	No	Yes	Yes
Genre FE	No	No	Yes
Observations	377,223	377,223	377,223
R^2	0.0002	0.007	0.04

Note: The omitted category is female directors. Standard errors are clustered at the year level in parentheses. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

when controlling for genre. The addition of genre fixed effects suggests that gender repartition by genre, rather than director gender, may be the primary driver of thematic distinctiveness. This finding suggests that institutional and genre constraints may be more influential in determining content choices than director characteristics.

7 Beyond movies: gender dynamics in television series

The patterns documented in movies, from workforce composition through casting decisions to content positioning, raise questions about their generalizability across media formats. Television series production operates under different structural constraints: longer production timelines, collaborative writing rooms, and sustained character development across multiple episodes. These differences may create distinct patterns of gender representation. Examining television provides a comparative lens to understand which patterns are specific to film versus reflecting broader dynamics in visual media production.

7.1 Data and methodology

To extend the analysis beyond movies, I compiled a comprehensive dataset of television series using the same data source and methodological approach employed for the movie analysis. Television series data were sourced from The Movie Database (TMDB), which provides extensive metadata for both movies and television content, ensuring consistency in data collection and processing procedures across the two media formats.

The television dataset includes detailed information on series production spanning from 1950 to 2024, encompassing both traditional broadcast television and modern streaming platform content. For each

series, I collected comprehensive metadata including release dates, genre classifications, production countries, and detailed cast and crew information.

Following the same methodology as the movie analysis, I focus on two key roles that parallel the movie study: showrunners (the television equivalent of directors) and lead actors. Showrunners serve as the primary creative executives responsible for the overall vision, writing, and production oversight of television series, making them the natural counterpart to movie directors in terms of creative authority and decision-making power.

The gender classification process for television personnel followed the identical two-tiered approach used for movies. I began with direct TMDB gender labels when available, supplemented by the genderize.io API for gender classification based on first names for individuals lacking explicit gender information. This consistent methodology ensures comparability between the movie and television analyses and maintains the same coverage rates achieved in the movie dataset.

7.2 Evidence

The television industry provides an important counterpoint to the movie industry for understanding gender dynamics in entertainment media. Television production operates under different structural constraints, with series requiring sustained creative leadership over multiple seasons and episodes, potentially creating different incentives for gender representation compared to standalone movie projects.

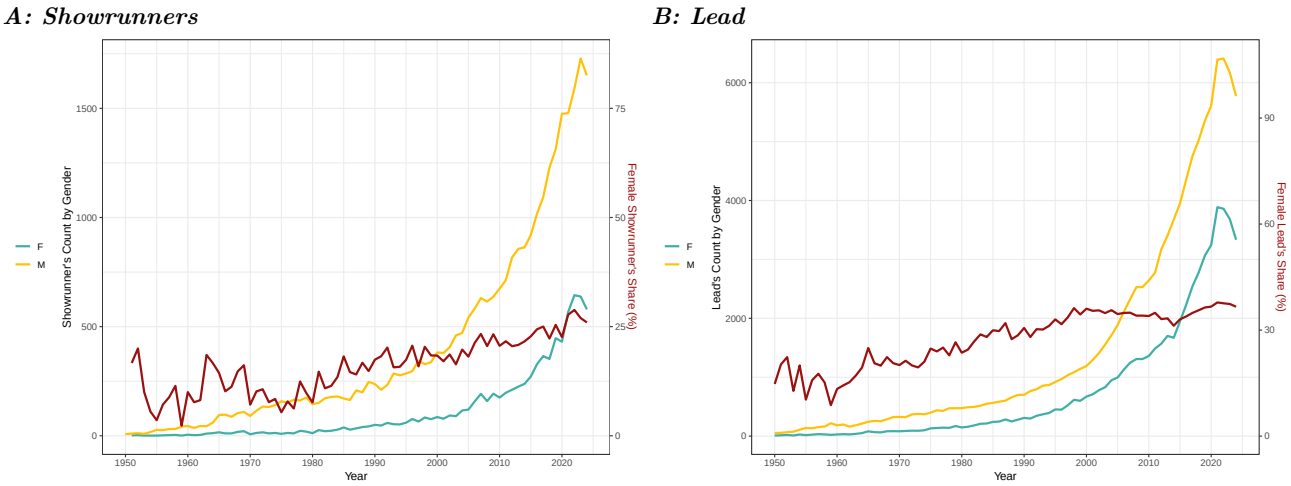
Figure 9 depicts patterns in television gender representation over time. Panel A shows that female representation among showrunners experienced a steady growth from the mid-1970s. By the end of the observation period, female showrunners account for approximately 25% of the total.

The evolution of female lead actors in television series, shown in Panel B, follows a slightly different trajectory. Female representation among television leads started to grow in the 1960s, reaching 30% in the mid-1990s. From 2000, the female share of lead in TV series has plateaued and fluctuated between 30 and 40%.

The cross-media comparison in Figure 10 reveals striking differences between television and movie industries. Panel A demonstrates that television has historically provided slightly more opportunities for women in creative leadership roles compared to movies.

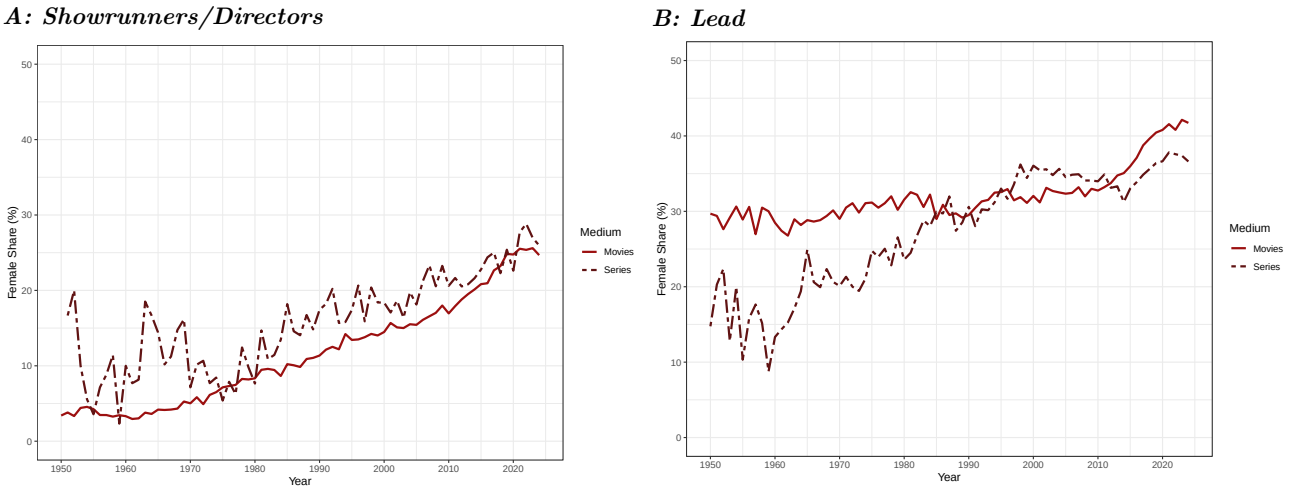
For lead actors, Panel B shows a more complex relationship between the two industries. Until the mid-1980s, movies demonstrate consistently higher female lead representation. While the share of female lead roles have been stable for both media from 2000 to 2010 (with a larger share for TV series

Figure 9: Gender representation in the TV series workforce



Note: Each panel shows the yearly count of male (yellow line) and female (teal line) representatives in a given role (showrunners and lead actors). The red line shows the share of female personnel, calculated as the ratio of female-associated personnel to the total number of individuals working in that role for all series' seasons produced in a given year.

Figure 10: Gender representation in television series vs. movies



Note: This figure compares the evolution of female representation between television series and movies from 1950 to 2024. Panel A shows the share of female showrunners in television series (dashed line) versus female directors in movies (solid line). Panel B presents the share of female lead actors in television series (dashed line) compared to movies (solid line). The female share is calculated as the percentage of women in each role category for all productions released in a given year.

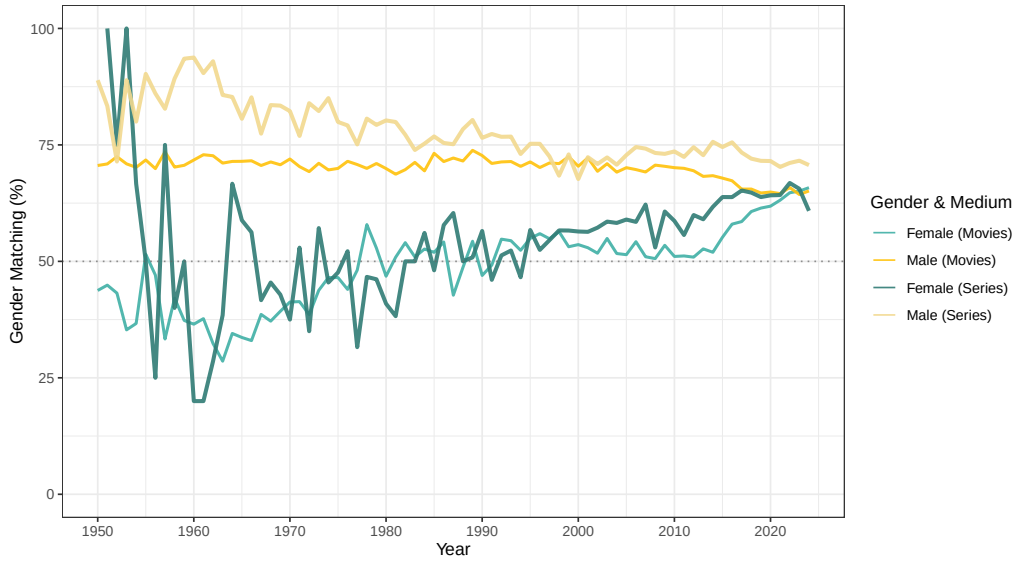
than movies), an inflection point occurred in the beginning of the 2010s with movies experiencing a higher growth in female share of leads, creating divergent trends.

7.3 Gender homophily in television

Figure 11 presents the evolution of gender homophily in television series compared to movies from 1950 to 2024, revealing distinct patterns across the two media.

First, as in film, male showrunners have consistently exhibited higher levels of gender homophily than

Figure 11: Average gender matching between director/showrunner gender and lead actor gender (movies vs. series)



Note: This figure shows the evolution of gender matching between showrunners/directors and lead actors in television series versus movies from 1950 to 2024. The y-axis represents the percentage of productions where the gender of the showrunner (for series) or director (for movies) matches the gender of the lead actor. Female showrunners/directors are shown in teal/green, while male showrunners/directors are shown in yellow. The dashed horizontal line at 50% serves as a reference point, i.e. a situation where directors/showrunners work half the time with a lead matching their gender.

female showrunners. However, in recent years, homophily levels for both male and female showrunners have converged to a similar range, exceeding 60%, indicating that showrunners of both genders tend to work with lead actors matching their gender more than half the time. Notably, this convergence began earlier in television than in film: around the 1990s for female showrunners, compared to the 2010s for female directors in cinema.

Second, male showrunners have shown consistently higher homophily than their male counterparts in film throughout the observed period, and this pattern persists into the 2020s. This suggests that gendered casting preferences among men may be even more pronounced in television series than in movies.

Table 5 reproduces the specifications presented in Section 4.2 using television series data. The baseline specification (column 1) shows that male showrunners are 13 percentage points more likely to cast leads matching their gender compared to female showrunners. This effect remains remarkably stable at 9-13 percentage points across all specifications, even after controlling for year fixed effects (column 2) and genre characteristics (column 3). The robustness of this coefficient across different model specifications suggests that the gender homophily gap is not driven by temporal trends or genre-specific casting

Table 5: Homophily in TV series by director's gender

	Homophily		
	(1)	(2)	(3)
Constant	0.61***		
	(0.009)		
MaleShowrunner	0.13***	0.13***	0.09***
	(0.01)	(0.01)	(0.01)
Year FE	No	Yes	Yes
Genres FE	No	No	Yes
Observations	61,446	61,446	61,446
R^2	0.01	0.02	0.12

Note: The omitted category is female showrunners. This table presents OLS estimates where the dependent variable (Homophily_{it} is a binary indicator equal to 1 if the showrunner and lead actor share the same gender, and 0 otherwise. The sample includes all television series with identified showrunner and lead actor genders from 1950 to 2024. Column (1) presents the baseline specification with only the male showrunner indicator. Column (2) adds year fixed effects to control for temporal trends in casting practices. Column (3) further includes genre fixed effects. Standard errors are clustered at the serie level in parentheses. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

conventions, but rather reflects systematic differences in how male and female showrunners approach lead casting decisions.

8 Discussion and conclusion

This study provides comprehensive empirical evidence on the evolution of gender representation in the movie industry, examining both industry workforce composition and content creation patterns across over 500,000 films spanning more than a century. Using advanced machine learning techniques including computer vision and natural language processing, the findings reveal several key insights about gender dynamics in cultural production and the persistence of representation gaps.

First, the analysis of workforce composition shows that while female representation in key roles has increased over time, their share remains lower than for the male counterpart. The data reveal that male professionals continue to account for the largest share of directorial, production, and screenwriting positions, with women comprising 25% of directors, 33% of producers, and 29% of screenwriters in 2024. Besides, male directors exhibit stronger tendencies toward same-gender casting, showing approximately 11-12 percentage points higher probability of working with leads of their own gender compared to female directors, even after controlling for temporal and genre-specific factors.

Second, the #MeToo movement appears to have catalyzed meaningful changes in industry practices. Cross-country evidence suggests that female directors became more likely to cast female leads following the movement, with a 4 percentage point increase in homophily after October 2017. Male directors

showed the opposite pattern, with a significant reduction in same-gender casting after #MeToo. This differential response demonstrates how external social shocks can influence industry practices and highlights the potential for advocacy movements to accelerate trends toward increased female participation (Luo and Zhang, 2021, 2024).

Third, the examination of promotional materials through computer vision analysis reveals evolving but persistent patterns in gender representation. The relationship between credited roles and poster representation has historically favored male actors, with female actors being systematically underrepresented in marketing materials relative to their credited roles. While this gap has narrowed considerably in recent decades, particularly post-2000, differences persist. This finding demonstrates how marketing practices both reflect and potentially reinforce gender disparities, aligning with broader research on media's role in perpetuating gender stereotypes (Adukia et al., 2023; Ash et al., 2021; Bellet and Pocchiari, 2024; de Courson et al., 2024).

Fourth, the analysis of content distinctiveness initially suggests that films directed by women exhibit higher levels of thematic distinctiveness compared to the historical content baseline. However, this relationship becomes statistically insignificant once genre fixed effects are included, indicating that apparent differences in thematic positioning are driven more by the differences of female and male participation in certain genres. This suggests that female directors are not inherently more likely to create distinctive narratives but may be concentrated in genres that typically diverge from mainstream content. Consequently, while increasing gender diversity in directorial roles can be a target for public funding agencies for broader industry representation, its direct impact on narrative distinctiveness appears to be mediated by genre composition rather than an intrinsic difference in creative approach.

The extension to television series reveals several structural differences between media formats. Television series have historically provided slightly more opportunities for women in creative leadership roles but female share in showrunners and directors roles converge by the 2020s. However, the trajectory for lead actors shows more complex dynamics, with movies experiencing higher growth in female representation in the 2010s, creating divergent trends between the two media.

While this study primarily focuses on the supply side of media production – who participate in the film industry and how they are presented to audiences – it is important to acknowledge the role of audience demand and its interaction with production dynamics. Evidence from the sports industry suggests that improved visibility of women athletes is often followed by notable increases in viewership

and engagement (Nielsen, 2023; Women's Sport Trust, 2023). For example, the 2024 NCAA women's basketball final drew 18.9 million viewers, exceeding the men's final.[3] Expanded coverage of the WNBA and other leagues has coincided with substantial rises in audience metrics.[4] While these associations do not establish causal effects, they point to the possibility that demand for gender-diverse content can remain latent or underrealized when supply is limited. Gender parity, then, is not only a normative concern but also potentially tied to unrealized cultural and commercial opportunities.

These findings have several important implications. The observed patterns of homophily indicate that increasing female representation in leadership positions creates positive feedback loops, facilitating more opportunities for female professionals throughout the industry. For policymakers, the results support arguments for initiatives promoting gender diversity in media production, as such diversity appears to contribute to increase female participation both behind the camera and on the screen.

The binary framework adopted in this study inevitably simplifies the complex and multifaceted nature of gender. By focusing on male and female classifications, this research overlooks non-binary and gender-nonconforming individuals. Similarly, the absence of an intersectional lens limits the study's ability to capture the compounded effects of intersecting identities on representation and inclusion in the movie industry. While these limitations do not detract from the core findings, they highlight the need for future research.

The patterns documented in this research reveal how demographic characteristics of decision-makers have been related to production and marketing outcomes in creative industries. These findings extend beyond film itself: as movies influence cultural narratives and social perceptions globally, the gender patterns documented here help explain how certain perspectives and representations become amplified or diminished in popular culture. The persistence of distinct gender-based patterns across a century of filmmaking, combined with the measurable shifts following events like #MeToo, demonstrates both the durability of these patterns and their potential for change.

[3] Associated Press (2024, April 8). *Women's NCAA title game outdraws men's championship.* Retrieved from https://apnews.com/article/b592435cc286c75a7ac9278c97326ad8

[4] Business Insider (2024, April). *How women's college basketball ratings soared.* Retrieved from https://www.businessinsider.com/ncaa-college-womens-basketball-tv-ratings-popularity-wnba-caitlin-clark-2024-4; New York Post (2024, April 16). *Caitlin Clark-led 2024 WNBA Draft shatters TV ratings record.* Retrieved from https://nypost.com/2024/04/16/caitlin-clark-led-2024-wnba-draft-shatters-tv-ratings-record

References

Adukia, A., Eble, A., Harrison, E., Runesha, H. B., and Szasz, T. (2023). "What we teach about race and gender: Representation in images and text of children's books." *The Quarterly Journal of Economics*, *138*, 2225–2285.

Aguiar, L. (2024). "Bad apples on rotten tomatoes: Critics, crowds, and gender bias in product ratings." *Marketing Science*.

Akerlof, G. A., and Kranton, R. E. (2000). "Economics and identity." *The Quarterly Journal of Economics*, *115*, 715–753.

Ash, E., Durante, R., Grebenshchikova, M., and Schwarz, C. (2021). "Visual representation and stereotypes in news media." *Working Paper*.

Baldassari, D. (2023). "Oscar voters so white? hiring effects of an evaluator's diversity intervention." *Working Paper*.

Bellet, C., and Pocchiari, M. (2024). "Quantifying gender representation in tv advertising." *Working Paper*.

Beltagy, I., Peters, M. E., and Cohan, A. (2020). "Longformer: The long-document transformer."

Bertrand, M. (2018). "Coase lecture – the glass ceiling." *Economica*, *85*, 205–231.

Bertrand, M. (2020). "Gender in the twenty-first century." *AEA Papers and Proceedings*, *110*, 1–24.

Bertrand, M., Black, S. E., Jensen, S., and Lleras-Muney, A. (2019). "Breaking the glass ceiling? the effect of board quotas on female labour market outcomes in norway." *The Review of Economic Studies*, *86*, 191–239.

Bossler, M., Mosthaf, A., and Schank, T. (2020). "Are female managers more likely to hire more female managers? evidence from germany." *ILR Review*, *73*, 676–704.

Carlana, M. (2019). "Implicit stereotypes: Evidence from teachers' gender bias." *The Quarterly Journal of Economics*, *134*, 1163–1224.

Cimpian, J. R., Kim, T. H., and McDermott, Z. T. (2020). "Understanding persistent gender gaps in stem." *Science*, *368*, 1317–1319.

de Courson, B., Richard, A., and Bastin, G. (2024). "Gender gaps and stereotypes in the long run. a computational approach to how le monde got (slightly) demasculinized (1944-2024)." *Working Paper*.

Deng, J., Guo, J., Ververas, E., Kotsia, I., and Zafeiriou, S. (2020). "Retinaface: Single-shot multi-level face localisation in the wild." *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5202–5211.

Flabbi, L., Macis, M., Moro, A., and Schivardi, F. (2019). "Do female executives make a difference? the impact of female leadership on gender gaps and firm performance." *The Economic Journal*, *129*, 2390–2423.

Goldin, C. (2023). "Why women won." *Working Paper*.

Haris, M. J., Upreti, A., Kurtaran, M., Ginter, F., Lafond, S., and Azimi, S. (2023). "Identifying gender bias in blockbuster movies through the lens of machine learning." *Humanities and Social Sciences Communications 2023 10:1*, *10*, 1–8.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Hirsch, B. (2013). "The impact of female managers on the gender pay gap: Evidence from linked employer–employee data for germany." *Economics Letters*, *119*, 348–350.

Istead, L., Pocol, A., and Siu, S. (2022). "Evaluating gender bias in film dialogue." *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13286 LNCS*, 403–410.

Jensen, R., and Oster, E. (2009). "The power of tv: Cable television and women's status in india." *The Quarterly Journal of Economics*, *124*, 1057–1094.

Kearney, M. S., and Levine, P. B. (2015). "Media influences on social outcomes: The impact of mtv's 16 and pregnant on teen childbearing." *American Economic Review*, *105*, 3597–3632.

Kerkhof, A., and Reich, V. (2023). "Gender stereotypes in user-generated content." *CESifo Working Paper No. 10578*.

Kumar, A. M., Goh, J. Y., Tan, T. H., and Siew, C. S. (2022). "Gender stereotypes in hollywood movies and their evolution over time: Insights from network analysis." *Big Data and Cognitive Computing*, *6*, 50.

Kunze, A., and Miller, A. R. (2017). "Women helping women? evidence from private sector data on workplace hierarchies." *The Review of Economics and Statistics*, *99*, 769–775.

Luo, H., and Zhang, L. (2021). "Scandal, social movement, and change: Evidence from metoo in hollywood." *Management Science*, *68*, 1278–1296.

Luo, H., and Zhang, L. (2024). "Gender inequality and product choice: Evidence from metoo." *Working Paper*.

Mazières, A., Menezes, T., and Roth, C. (2021). "Computational appraisal of gender representativeness in popular movies." *Humanities and Social Sciences Communications 2021 8:1*, *8*, 1–9.

Michalopoulos, S., and Rauh, C. (2024). "Movies." *Working Paper*.

Michalopoulos, S., and Xue, M. M. (2021). "Folklore." *The Quarterly Journal of Economics*, *136*, 1993–2046.

Mulvey, L. (1975). "Visual pleasure and narrative cinema." *Screen*, *16*, 6–18.

Nielsen (2023). "Women's sports viewership on the rise." Available at: https://www.nielsen.com/insights/2023/womens-sports-viewership-on-the-rise. Accessed August 4, 2025.

Porter, C., and Serra, D. (2020). "Gender differences in the choice of major: The importance of female role models." *American Economic Journal: Applied Economics*, *12*, 226–54.

Riley, E. (2022). "Role models in movies: The impact of queen of katwe on students' educational attainment." *The Review of Economics and Statistics*, 1–48.

Smith, L. S., Pieper, K., and Wheeler, S. (2023). "Inequality in 1,600 popular films: Examining portrayals of gender, race/ethnicity, lgbtq+ disability from 2007 to 2022." *USC Annenberg Inclusion Initiative*.

Stroube, B. K., and Waguespack, D. M. (2024). "Status and consensus: Heterogeneity in audience evaluations of female- versus male-lead films." *Strategic Management Journal*, *45*, 994–1024.

Tan, H. R., and Wang, T. (2024). "Mccarthyism, media, and political repression: Evidence from hollywood." *Working Paper*.

Voth, H.-J., and Yanagizawa-Drott, D. (2023). "Image(s)." *Working Paper*.

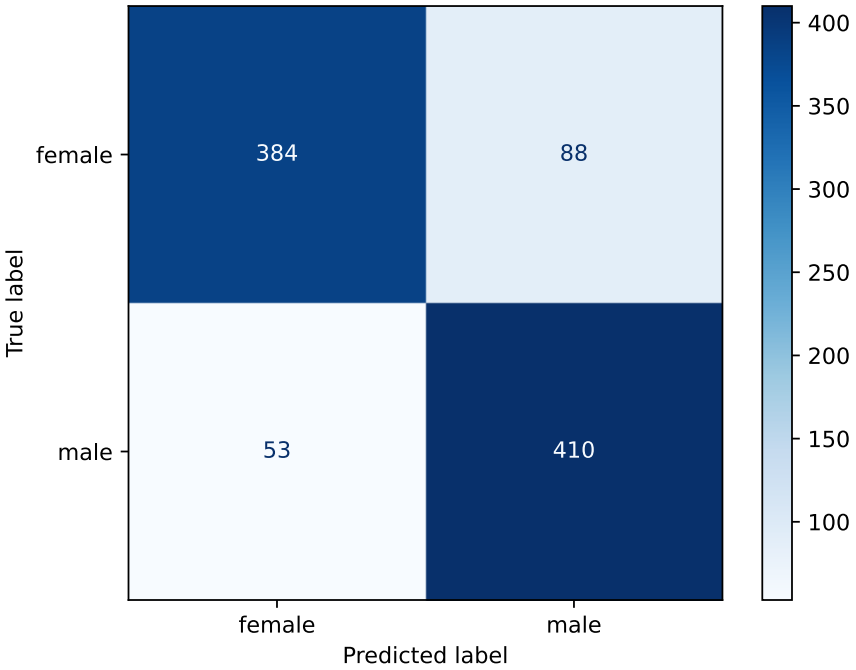
Women's Sport Trust (2023). "Visibility and fandom of women's sport in 2023." Available at: https://www.womenssporttrust.com/latest-research-from-womens-sport-trust-delves-into-visibility-and-fandom-of-womens-sport-in-2023. Accessed August 4, 2025.

Yang, P., Riepe, J., Moser, K., Pull, K., and Terjesen, S. (2019). "Women directors, firm performance, and firm risk: A causal perspective." *The Leadership Quarterly*, *30*, 101297.

A Appendix

A.1 Additional tables and figures

Figure A.1: Gender classification metrics based on first names using genderize.io

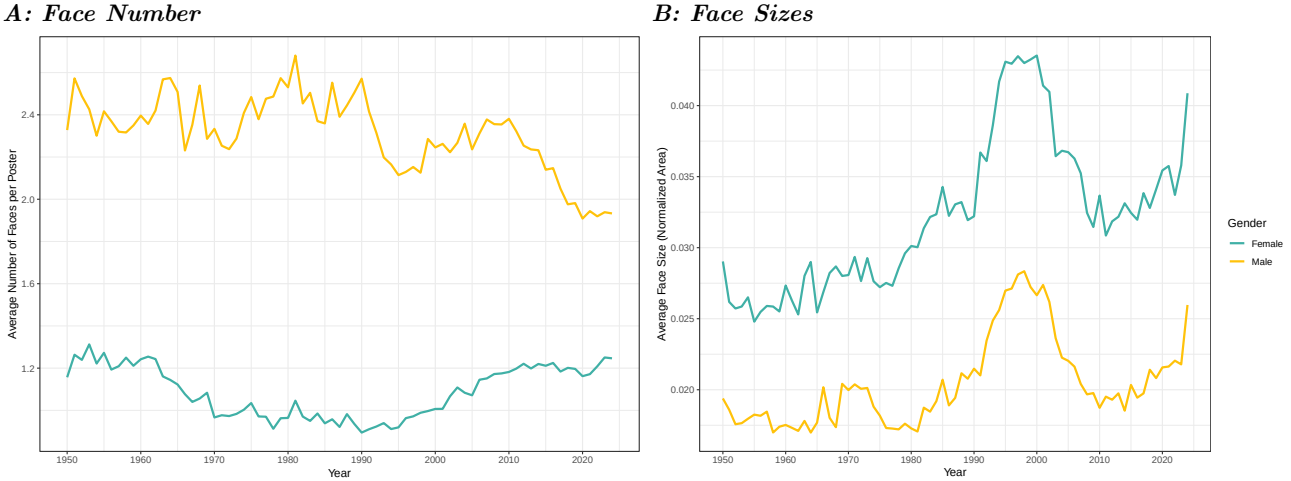


Note: This confusion matrix presents the results of a validation test of the Genderize.io API for inferring gender from first names. A balanced random sample of 1,000 first names (500 female and 500 male), drawn from TMDB data, was used to assess classification accuracy. The matrix compares predicted labels against verified ground truth, showing strong overall performance with an accuracy of 85%, precision of 0.88 for female names and 0.82 for male names, and recall of 0.81 for female names and 0.89 for male names. These results support the reliability of name-based gender inference in the context of the study.

Figure A.2: Evolution of the female share in key roles in the movie industry by genre

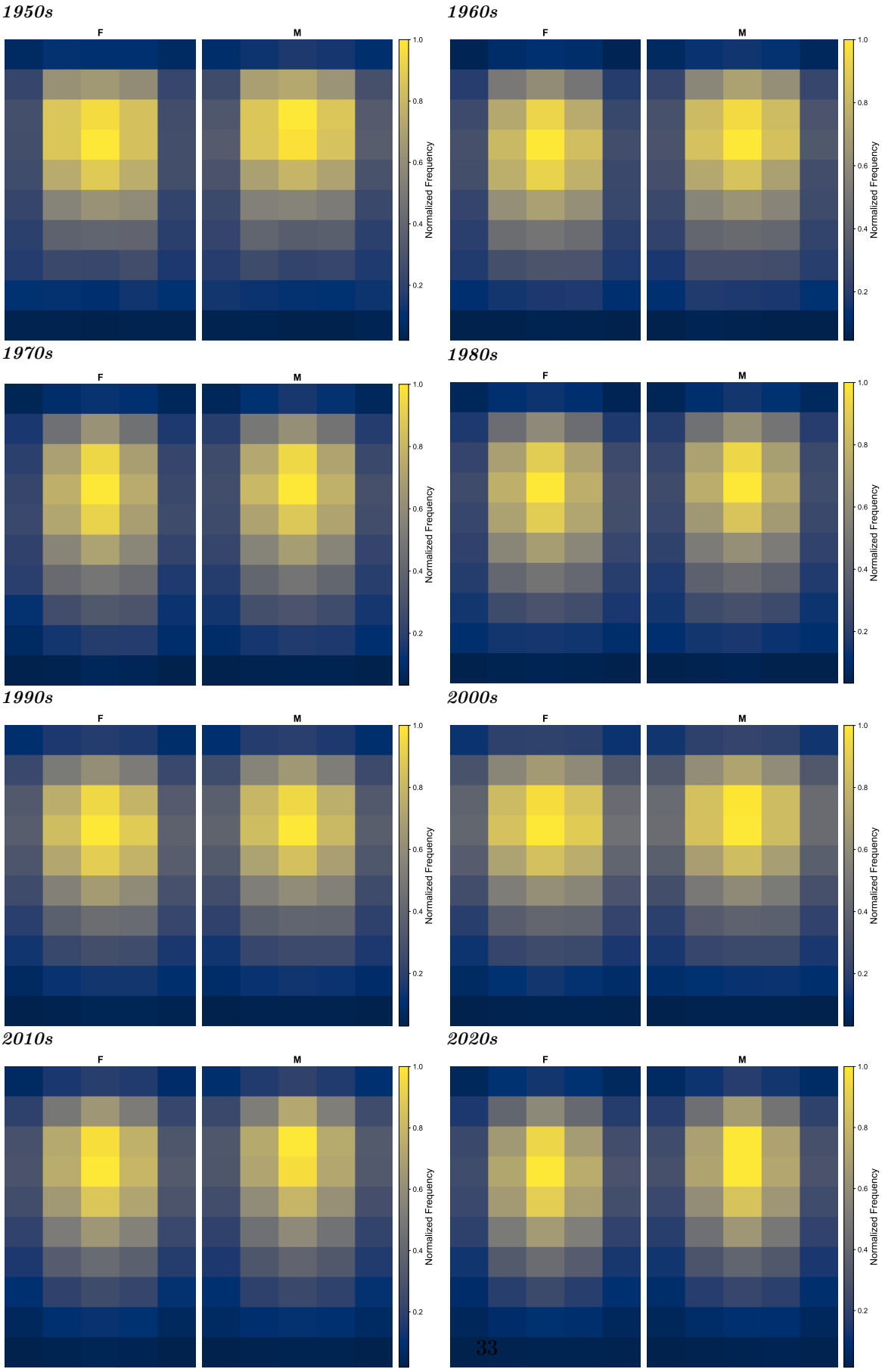
Note: Data collected from TMDB. This figure shows the evolution of female representation in key movie industry roles from 1900 to 2024, broken down by genre. Each panel represents a different genre, with four lines tracking the percentage of women among directors (navy), lead actors (skyblue), producers (gold), and screenwriters (oramge).

Figure A.3: Trends in the number and size of detected faces by gender on movie posters



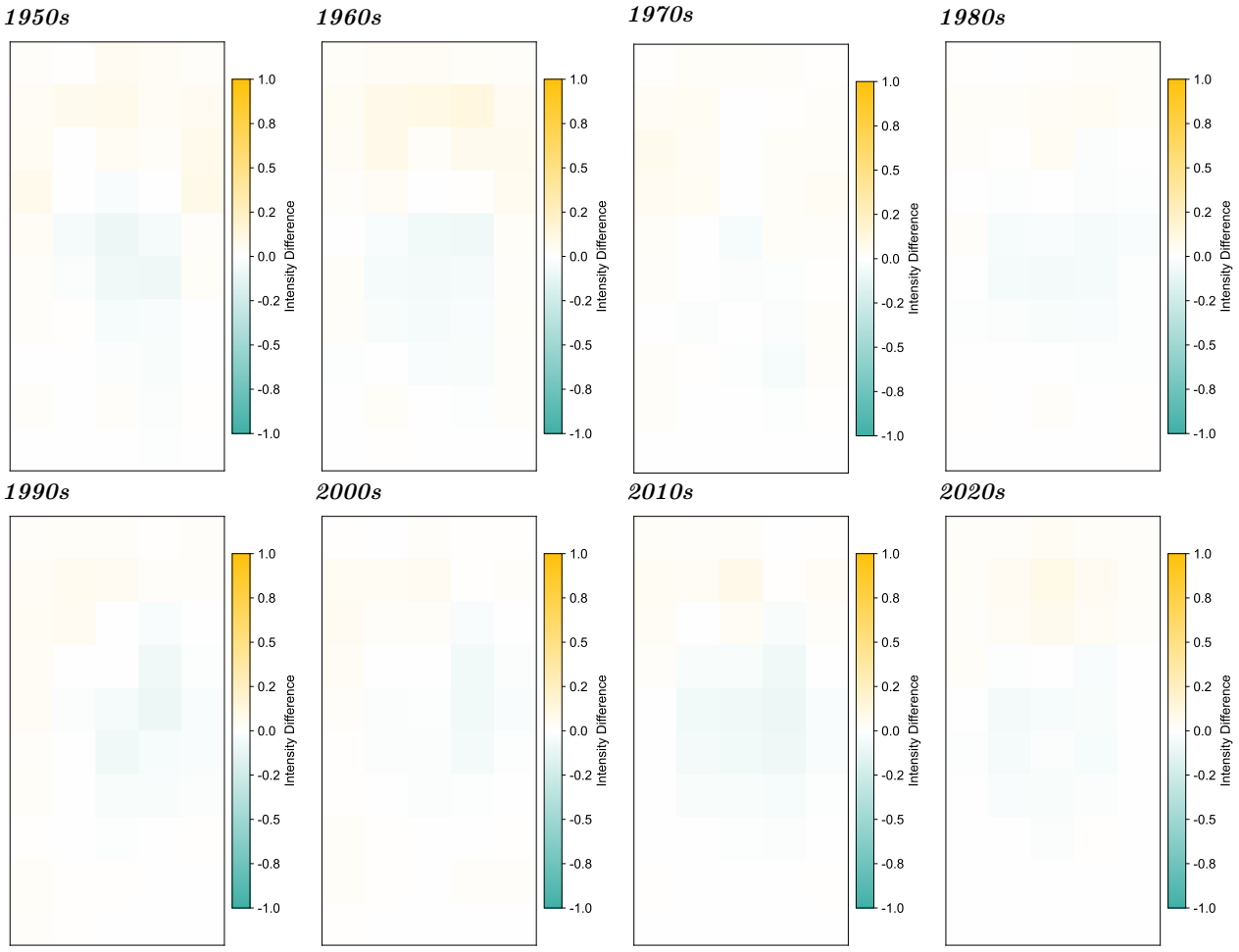
Note: Panel A shows the average number of detected male and female faces per movie poster over time, while Panel B presents the average relative size of those faces, normalized by poster area. Both measures are conditional on the presence of at least one detected face on the poster.

Figure A.4: Heatmap for detected male and female faces by decade



Note: Each panel shows the spatial distribution of detected faces by predicted gender across film frames, aggregated and normalized within each decade. The color intensity reflects the relative frequency of bounding boxes over a standardized 3:2 frame canvas, with brighter regions indicating more frequent placements.

Figure A.5: Spatial differences in detected male and female face locations by decade



Note: Heatmaps represent the difference in spatial distributions of detected male and female faces by decade, as identified by the face detection and gender classification models. Yellow regions indicate higher relative frequency of male face detections, while teal regions indicate higher frequency of female detections. The values are normalized per decade, and white areas reflect little or no difference in detection frequency.

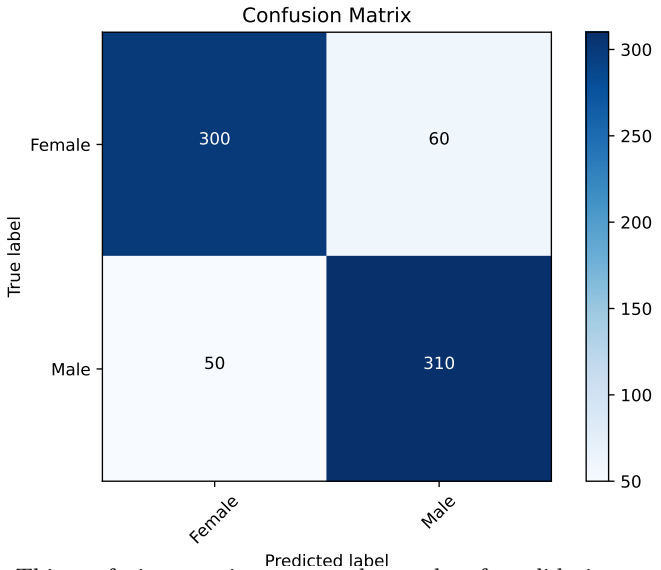
A.2 Additional notes on methods

A.2.1 Detecting faces and classifying gender on movie posters

I utilize the pre-trained RetinaFace model to detect faces in movie poster images. To evaluate RetinaFace's performance on the stylized faces common in posters, I manually labeled a random sample of 157 posters containing 505 faces. On this face-level test set, RetinaFace achieves a precision of 0.92 and a recall of 0.84. These posters are fully disjoint from those used in subsequent gender classification training and evaluation.

Due to the artistic and varied depiction of faces on movie posters, existing pre-trained gender classification models performed poorly on this data. To address this, I trained a custom gender classifier using transfer learning. I adopted the ResNet101 convolutional neural network architecture (He et al., 2016) and fine-tuned it on a curated dataset of face images extracted from 1,480 posters (10 randomly selected per year), each manually labeled as female or male. The classifier was evaluated on a held-out, balanced test set of 720 face images (360 female, 360 male), achieving an overall accuracy of 85%. Precision was 0.86 for female faces and 0.84 for male faces, with corresponding recall values of 0.83 and 0.86, respectively (see Figure A.6 below).

Figure A.6: Gender classification based on face images metrics



Note: This confusion matrix presents the results of a validation test of the custom ResNet101 classifier to predict gender from face images extracted from movie posters. A balanced random sample of 720 faces (360 female and 360 male), randomly drawn from posters, was used to assess classification accuracy. The matrix compares predicted labels against verified ground truth, showing an accuracy of 85%, precision of 0.86 for female faces and 0.84 for male faces, and recall of 0.83 for female faces and 0.86 for male faces.

A.2.2 Measuring movie distinctiveness

To measure movie distinctiveness, I convert synopsis data to a high-dimensional text embedding. Instead of relying on traditional pre-trained models like BERT, which have a limited sequence length, I employ a Longformer model (Beltagy et al., 2020). The Longformer architecture is specifically designed for long text sequences, allowing the full synopsis to be encoded without truncation. This ensures that the entire narrative content of the movie is represented in the embedding space.

For a given focal movie i released at time t, I use all previously released movies (i.e., those with release dates before t) to construct a historical content centroid. Let $\mathcal{I}_t$ be the set of indices corresponding to these earlier movies, and let $\mathbf{e}_j \in \mathbb{R}^d$ denote the embedding of movie j's synopsis. The centroid is computed as the average of these embeddings:

$$\bar{\mathbf{e}}_t = \frac{1}{|\mathcal{I}_t|} \sum_{j \in \mathcal{I}_t} \mathbf{e}_j.$$

This centroid represents the accumulated content "norm" or "baseline" against which new releases can be compared. I then measure the distinctiveness of movie i as the cosine distance between its embedding $\mathbf{e}_i$ and the centroid $\bar{\mathbf{e}}_t$:

$$\text{Distinctiveness}_{it} = 1 - \frac{\mathbf{e}_i \cdot \bar{\mathbf{e}}_t}{\|\mathbf{e}_i\| \|\bar{\mathbf{e}}_t\|}.$$

Higher values of this distance indicate that the focal movie's narrative differs more substantially from the previously established content baseline, thus capturing a dimension of creative or thematic distinctiveness.