

COMP 370 Final Project – Movie Release Data Project

Daphne Freda Gavras, Sophie Happi, Ruben Hari,

McGill University School of Computer Science

3480 Rue University, Montréal, QC H3A 2A7

daphne.gavras@mail.mcgill.ca, sophie.happi@mail.mcgill.ca, ruben.hari@mail.mcgill.ca

Abstract

This study examines the media coverage of *Gladiator II* to analyze the reporting dynamics surrounding the film. Specifically, it addresses two key questions: the volume of coverage received by *Gladiator II* relative to five other films released within a similar timeframe and the dominant topics within that coverage. Using a large dataset collected through NewsAPI, the analysis applies a refined typology and TF-IDF computation to explore patterns in media narratives. Contrary to initial expectations, *Gladiator II* accounts for only 14% of the overall media coverage, with early-stage releases like this film focusing predominantly on promotion and narrative intrigue. These strategies aim to build audience anticipation while connecting the film to the legacy of its predecessor. On a broader scale, this study offers valuable insights on the dynamics of film media narratives, revealing how factors such as genre, release timing, and marketing strategies shape news reporting patterns.

Introduction

Why We Chose Our Movie

In selecting our movie, we considered 3 main factors: release date, popularity, and critical response.

Gladiator II is the sequel in the *Gladiator* franchise, with the first movie coming out in May 2000. Our movie was released on November 22, 2024. This was the most influential factor in our movie picks because we are sourcing our data from NewsAPI, which only allows us to select data within a month of the current date (this is discussed more in the Data section of our report). Thus, this date meant that we would be creating our report during the 2 weeks prior to release and 2 weeks post release, which is when many articles related to this movie will be released.

While the release date was the primary factor in our movie selection, we also needed to consider its popularity. We wanted to choose a movie that, a) was popular enough to generate a sufficient number of articles for our analysis, and b) whose level of popularity was uncertain. The first consideration is straightforward—without enough articles, we couldn't conduct a meaningful analysis. The second consideration is more

nuanced. While we needed a popular movie, we didn't want one where we could already predict its success (for example, we would have known that *Barbie* was one of the year's biggest hits). Our goal was to uncover the movie's popularity through our data analysis. By selecting *Gladiator II*, we approached the project without preconceived notions about its success, avoiding any biases that might have arisen if we were already fans of the movie or conversely disliked it.

Finally, while numerous factors influenced our decision, the third primary factor for choosing *Gladiator II* was critical response. Similarly to popularity, we avoided delving too deeply into the critical response beforehand but sought a movie likely to generate a diverse range of opinions. As a sequel, *Gladiator II* naturally invited comparisons to its predecessor, *Gladiator*. Sequels often spark mixed reactions, with some favoring the original and others preferring the follow-up. This made *Gladiator II* an appealing choice, as we expected varied critical responses, even though we were unaware of the specific nature of those reactions at the time.

It is also worth noting that *Gladiator II* was not our initial movie pick. We had initially chosen *The Menendez Brothers* as our movie, but after attempting to pull articles, we found that many were focused on the Menendez brothers' trial rather than referencing the movie itself. This led us to selecting movies that were not based on recent real-life events.

Why We Chose Our Other 5 Movies

After picking our first movie, we picked 5 other movies to compare it to.

1. *Deadpool & Wolverine*
2. *We Live in Time*
3. *Smile 2*
4. *Venom: The Last Dance*
5. *Wicked*

The selection of five movies to compare with *Gladiator II* aimed to provide a comprehensive snapshot of the

cinematic landscape during October and November 2024. Our chosen films span a variety of genres, including superhero action (*Deadpool & Wolverine*), romantic drama (*We Live in Time*), horror (*Smile 2*), sci-fi thriller (*Venom: The Last Dance*), and musical fantasy (*Wicked*). The mix includes both sequels in well-established franchises (*Deadpool & Wolverine*, *Venom*, *Smile 2*) and standalone films (*We Live in Time*, *Wicked*). This diverse selection reflects a deliberate effort to examine the dynamics of media coverage across different genres, audience expectations, and franchise histories.

We decided on these for similar reasons to the first movie — release data, popularity, and critical response. We looked at other movies released around the same time as *Gladiator II* and selected 5 other movies based around that time as well. We did have to adjust these, namely if a movie wasn't popular enough and wasn't pulling any articles. We considered less the critical response for these movies because, even if the response about an individual movie isn't varied, we know the response about this group of movies would be varied.

Data

Clearly, to understand the news reporting happening about a film, we need to collect news articles to analyze. We decided to use NewsAPI to source our data as this was a free and convenient option. However, there is a major limitation when using NewsAPI: you can only select articles that were released a month from when you are pulling your data. Initially, we had picked movies that were released across the year, but we soon realized that this created a bias against the movies released much earlier. Because we still wanted to use NewsAPI, we decided to reselect movies, that were all released much more recently.

Apart from *Deadpool & Wolverine*, all our movies were released for theaters in either October or November 2024. We kept our choice of *Deadpool Wolverine*, as it was still relatively recent, and its new release on a digital platform (November 12th) has prompted many articles to be written about it, allowing for analysis to be conducted. The release dates of our movies were as follows:

- *Gladiator II*: November 22, 2024
- *Deadpool & Wolverine*: July 26, 2024
- *We Live in Time*: October 11, 2024
- *Smile 2*: October 18, 2024
- *Venom: The Last Dance*: October 25, 2024
- *Wicked*: November 20, 2024

Upon investigating the final dataset which contained 607 datapoints, we observed distinct patterns in the media coverage distribution among the selected films. *Venom: The Last Dance* and *Wicked* received the majority of the articles, with a combined total of 52% of all articles. *Deadpool & Wolverine* followed, with 19% of the articles, boosted by its re-release as mentioned earlier. *Gladiator II*, the focal point of this analysis, accounted for 14% of the total articles. While this is a smaller proportion than anticipated, given the popularity of its predecessor, as well as the high-profile actors involved, it remains sufficiently large for statistical analysis. The film's late release date and overlap with our data collection period explains its lower overall representation. Lastly, *Smile 2* and *We Live in Time* represented 10% and 5% of the data respectively. The distribution of articles across each movie is represented in Figure 1.

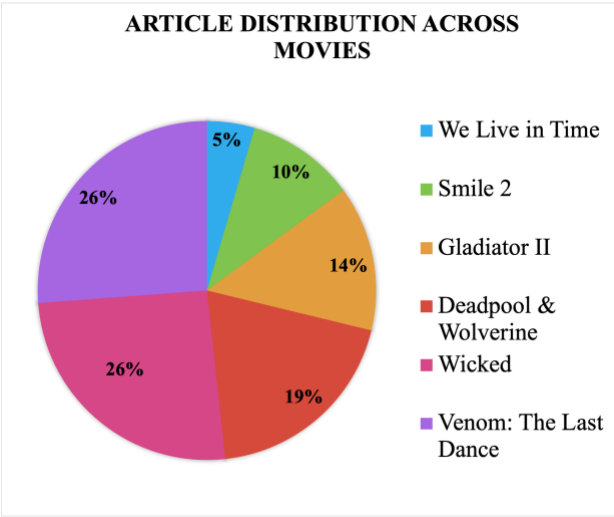


Figure 1: Relative volume of media coverage per movie.

Minimizing Biases

To ensure the objectivity of our data, we made efforts to minimize bias in our selection process. Our considerations are listed below.

Selection Bias in Articles:

One way we minimize bias is through the careful selection of sources. Despite efforts to ensure diversity in the sources, the choice of publications could still introduce bias. Some publications may have a particular editorial stance, such as focusing on music production or providing critical reviews, which could skew the analysis. To mitigate this, we aim to reflect multiple viewpoints by including a variety of

sources, including entertainment magazines and newspapers, rather than relying too heavily on a single type of coverage. A balanced mix of reviews—focused on opinions about the film’s quality—and feature articles or interviews—emphasizing production, stars, or behind-the-scenes aspects—is crucial. Feature articles tend to highlight promotional or industry-related angles, while reviews focus on artistic evaluation and audience reception. By applying filters that distinguish between these types of content, we ensure a more comprehensive and balanced view.

Regional Bias:

All articles were filtered to be in English, facilitating the analysis since it is a universal language and allows for accurate annotation by all group members. However, this language filter introduces a regional bias. Articles in English are more likely to come from publications based in English-speaking countries, such as the United States, Canada, or the United Kingdom, meaning they are more likely to reflect the cultural values and societal norms for films of these regions. While this bias is inherent, it is somewhat mitigated by the fact that the selected films are produced by English-speaking countries, which provides a clearer context for the articles’ focus. Had the articles been sourced from non-English-speaking countries, the films might have been interpreted differently due to varying cultural values and expectations, which could lead to misrepresentations or skewed evaluations.

Film Selection Criteria:

In selecting the movies for this analysis, we aimed to choose films with broad global appeal, ensuring that they would attract extensive coverage from a wide range of sources. This selection strategy helps ensure a more diverse representation of the films in the media. Additionally, we focused on films featuring well-known casts, as this increases the likelihood that a variety of publications would show interest in covering them. By choosing films with broad international appeal and established actors, we maximize the chances of receiving coverage from different types of publications and critics.

Filters to Minimize Focus on Specific Aspects:

Another consideration was to minimize our focus on a single aspect of the movie that might bias the selection of articles. For example, a musical might receive a disproportionate amount of coverage regarding its music, while an action film would not. To address this, we added an equal number of filters that apply to each genre, but also, we used filters that are relevant to all movies, regardless of the specific film characteristics. Additionally, we avoided the use of subjective terms like “good” or “bad,” which could skew the analysis toward certain types of reviews. By doing this, we aim to capture both positive and negative reviews, ensuring

a balanced view of how the films were received without any bias toward judgmental language.

Methods

Initial Attempts

When we first began this project, we planned to analyze a different selection of films, including one released in November 2022. However, after running a sample test with NewsAPI, we discovered that the tool only allows data collection for articles published within a one-month lookback period. Attempting to collect articles for a film released two years earlier proved ineffective and made it difficult to gather the data required for our analysis. We decided to collect the data on November 27th, which only meant that articles published from October 27th to November 27th could only be selected. Consequently, we pivoted to researching and selecting movies that were primarily released in Autumn 2024. Our movie selections underwent another phase transition upon a second test was run. As outlined in the introduction, section “*Why We Chose Our Other 5 Movies*”, it was important to select movies that were popular, to be able to collect articles for all of them.

Data Collection Methods

Once we finalized our new selection of films, the next step was to collect over 500 articles using NewsAPI while minimizing potential biases in the data collection process. Our approach involved running small tests with our Python script, identifying any issues that arose, and refining the script accordingly.

The first issue we encountered was the collection of articles completely unrelated to the films. NewsAPI often retrieved articles containing individual words from a movie’s title, regardless of context. For instance, when querying for *Smile 2*, the API returned articles about dental hygiene and mental health simply because they included the word “Smile.” To address this, we modified our script to append terms such as “movie,” “film,” or “cinema” to the movie titles in our search queries, significantly reducing the number of irrelevant articles.

The second issue involved retrieving articles that mentioned a movie’s title but were not focused on the film itself. For example, when searching for *Gladiator II*, the API often returned articles about Paul Mescal’s personal life or celebrity gossip, simply because he starred in the movie. To tackle this, we designed structured search queries for each film by pairing its title with specific filters grouped into predefined aspect groups. These aspect groups included categories like reviews (e.g., “review,” “critics’ opinion”) and box office performance (e.g., “weekend earnings,”

“ticket sales”). This strategy substantially reduced the number of irrelevant articles, although it did not eliminate them entirely.

We acknowledge that adding specific filters could introduce bias by over-representing certain topics in the retrieved articles. To mitigate this, we curated diverse and comprehensive keywords for each aspect group, ensuring a balanced representation of media coverage. Additionally, the same standardized queries were applied across all movies in the study, ensuring consistency and comparability in the collected data.

Another challenge stemmed from the structure of our Python script, which initially created a bias in coverage against certain movies. We had set a limit of approximately 600 articles for the dataset, and the script sequentially pulled articles based on the order in which movies were listed. This meant that movies listed earlier in the script often dominated the collection, leaving movies listed later with fewer or even no articles once the limit was reached.

To address this, we rewrote the script to implement a round-robin approach, which fetched articles more evenly across all movies. This method iterated through each aspect group, retrieving a batch of articles for every movie during each cycle. If no articles were returned for a specific movie, the script simply skipped to the next query. We also incorporated a while loop to ensure the script continued iterating until the maximum article limit was reached. This approach balanced the article counts across all films, preventing artificial skewing of coverage while adhering to our data constraints.

Open-Coding & Annotation

Once the articles were collected, open coding was conducted on the first 200 articles. Our initial analysis confirmed that the applied filters were effective, as they selected a diverse range of relevant articles. Furthermore, all selected movies were adequately represented, with no concerns about any movie being mentioned too infrequently. Based solely on the article titles and descriptions, we were able to assign two labels to each article: one corresponding to the movie it referenced, and another pertaining to its content. In this section of the report, we outline the categories and their evolution within our typology, with detailed definitions provided in the "Results" section.

Initially, for the movie label, we focused on the six selected movies. However, as we delved deeper into the data, we found that while our filters mostly pulled relevant articles, they were not flawless. Occasionally, some articles mentioned our keywords but did not directly relate to the

selected movies. To address this, we introduced two additional labels: "Unrelated to Selected Movies" and "Unrelated to Movies." This allowed for more accurate classification and ensured a clearer distinction between articles that were truly relevant and those that were not.

In the initial pass of the data, we identified 16 content categories. After further discussion, we decided to consolidate several of these, narrowing them down to 11. Namely, categories like comparison, popularity, ranking, and review were merged under the broader label of "Review". Similarly, we combined categories related to plot and genre into a single "Plot" category. We then had categories specific to actors, directors, and production (both producers as people and production cost), and another one called release. However, these were too vague and all-consuming as a category, that we reformed into "Crew", "Logistics", and "Interviews".

Several content labels proved sufficient from the first pass, namely, "Characters", "Critics", "Box Office". "Music", "Promotion", and "Merchandise". The rationale behind the formation of these labels, along with the distinctions between them, can be found in the "Results" section.

Once these categories were finalized, each group member annotated the dataset. The results from all three members were then statistically analyzed, and the labels were assigned accordingly. In some cases, this process prompted us to refine and redefine our typology to ensure greater consistency and accuracy in the labeling. This iterative approach helped improve the precision of our categorization and better aligned the data with the evolving typology.

Analysis

Once the annotation phase was completed using our typology, we proceeded to calculate the term frequency-inverse document frequency (TF-IDF) for our topics and dataset. While the findings of this analysis will be discussed in the "Results" section, here we outline our computational approach. The analysis was divided into two phases.

Phase 1 focused on preparing the dataset to ensure that formatting inconsistencies did not impact the TF-IDF calculations. First, we eliminated case sensitivity by applying the `.lower()` method to the text. We also removed any non-alphabetic characters that were not part of a word. Additionally, we deleted truncated words (e.g., "di...", "text...") and patterns such as `[+xxxx chars]` that frequently appeared at the end of rows to indicate the character count of the article.

Phase 2 involved grouping articles by their typology categories, computing the TF-IDF scores using the

TfidfVectorizer() function, and extracting the top 10 results for each category. These results were then exported to a separate CSV file for further analysis.

Results

Final Typology Definitions:

- **Review:** Articles in which the focus is a review of the movie; in other words, the author's opinion of the quality of the movie. Positive example: "*Gladiator II* is a messy successor that delivers better spectacle than sensibility." This is included as its title gives a summary of the author's thoughts on the film. Negative example: "The best new movies to stream this November." While this suggests a positive opinion of the movie, it does not say anything more than that. Its focus is more on promoting the listed movies, so it is excluded. Edge Case Example: "The Monkey Is the Secret Hero of *Gladiator II*". While this seems like a catchy hook for a review, it isn't clear based solely on the title. Once you investigate the description, its focus is on examining the plot of the movie, rather than giving an opinion on the movie's quality, so this is excluded.
- **Critics:** Similar to review, however the article must specifically talk about critics. Positive example: "Critics say *Gladiator II* is 'gobsmacking' and 'loopy'." Evidently the article is about critic's thoughts on the movie, so it is included. Negative example: "Paul Mescal's '*Gladiator II*' character appears in the first movie, here's a refresher on who Lucius is". This article focuses on the character Lucius and not on a critic's opinion, so it is excluded. Edge Case Example: "*Gladiator II* will entertain you, but *Gladiator* is the better movie. Here's why". While this is a review piece, it does not specify that it is from critics, so it would be excluded from this category, and instead be included review.
- **Interviews:** Articles that focus on a real individual involved in the film, and their words. If an article mentions an individual, but doesn't use their words in the article, it is excluded. Positive example: "Ariana Grande says she's planning a career shift". This article mentions an actor in *Wicked*, and something they have said, so it's included. Negative example: "'*Wicked*' Sets New Record With \$114 Million USD in Opening Weekend Box Office". This mentions *Wicked*, but not any individuals involved. Edge Case Example: "Review: Paul Mescal aims for the big leagues in *Gladiator II* — so is it any good?". This mentions an actor in *Gladiator*, but it doesn't reference that actor's words, so it is excluded.
- **Crew:** Articles that focus on a real individual involved in the film, but that don't use their words. Essentially this includes articles that talk about actors, directors, and producers, but is not an interview. Positive example: "Paul Mescal chickened out of his initial meeting with Denzel Washington on *Gladiator II*". This talks about the actors but does not include their own words. Negative example: "How Much *Gladiator II* Needs To Make To Become A Box Office Hit". No mention of any individual. Edge Case Example: "Tom Hanks And Robin Wright's Here Opens Flat At Weekend Box Office". Mentions people's name, but only as clickbait. The focus is on Box office performance, so it is excluded.
- **Characters:** Articles that focus on a movie character. Notably, it should not focus on the real-life actor, but rather the character which they play. Positive example: "*Wicked* Character Guide?". This article's focus is on the who the fictional characters are, and how they relate to each other. Negative example: "*Venom: The Last Dance* set to win weekend box office, but there's a battle for second spot". No mention of a character. Edge Case Example: Ryan Reynolds Explains Ghost Rider Absence in *Deadpool & Wolverine*. While characters are briefly mentioned in the title, it is from an individual related to the movie's words. Upon further inspection, the character was only mentioned once, and the rest of the article is a regular interview, hence it is excluded from the character category, and included in the interview category.
- **Plot:** Articles which focus is the story of the movie. Can be what happens, prequels, deleted scenes, or alternative endings. Positive example: "How *Gladiator II* Connects to the Original *Gladiator*". Explaining how a sequel connects is exploring the plot. Negative example: "The best movies new to streaming this November." Focus is on promoting the movie, and not about the story. Edge Case Example: "*Gladiator II* will entertain you, but *Gladiator* is the better movie. Here's why". Mentions the prequel, but is comparing their quality, and not the story, thus it is excluded from plot, and included in review.
- **Box office:** Articles that focus on the movie's box office performance, or overall generated revenue. Positive example: "'*Wicked*' soars at the box office with an estimated \$164 million worldwide". Clearly talks about box office performance. Negative example: "'*We Live in Time*' Star Andrew Garfield in Gucci at the 2024 LACMA Art & Film Gala". No mention of box office performance. Edge Case Example: "*Gladiator II* team-up heads for \$200M opening weekend". Does not specifically mention "box office" but mentions sales, so it is included as box office.

- **Music:** Articles that focus on the soundtrack, music production, composers, or singing. Positive example: “Wow, There’s Finally Footage of Singing in *Wicked*”. Focuses on the singing in *Wicked*. Negative example: “Here’s how ‘*Wicked*’ pays homage to the original Broadway musical in a star-studded cameo.” Doesn’t mention music, but rather an actor, so it would be excluded. Edge Case Example: “*Wicked* unabashedly caters to people who like musicals and no one else (for the better).” Mentions being a musical but, this is merely a hook to comment on the quality of the film, so it is excluded, and instead included in review.
- **Logistics:** Articles that focus on release dates, streaming platforms, or production efforts. Positive example: “*Deadpool & Wolverine* Just Got A Disney Plus Streaming Date And It’s Not Far Away”. Talks about a streaming date. Negative example: “Here Are the Incredible A-List Comic Book Cameos That Didn’t Make It Into *Deadpool & Wolverine*”. Focus is on cameos, so would be excluded, and instead included under crew. Edge Case Example: “What to Watch This Weekend: *Deadpool & Wolverine*, *Thelma*, *Cobra Kai*, and more”. Talks about what to watch, but not when to watch it, thus would be excluded, and rather included under promotion.
- **Promotion/marketing:** Any articles that focus on promoting or marketing the movie to viewers and encouraging you, the reader, to watch it. Includes media hype, “what to watch” articles, and teasers/trailers, without talking about the plot or what the general audience thinks about the movie. Positive example: “The best new movies to stream this November”. Talks about what movies you should be watching, hence promoting them. Negative example: “A legendary musician wrote a bonkers *Gladiator II* movie that was never made”. Talks about alternative story, not promoting it. Would be under plot. Edge Case Example: “*Wicked* Is a Rousing, Gorgeous, Slightly Drawn-Out Musical Adventure”. Talks so fondly that you might think it is an ad or promotion, but is more a review of the quality, where the article just has a high opinion of the movie. Thus, it is excluded and would be under review.
- **Merchandise:** Any articles that focus on a movies merchandise. Positive example: “New *Gladiator II* Popcorn Bucket Has a Cool Feature We Haven’t Seen Any Previous Movie Do”. Talks about the popcorn bucket specifically for *Gladiator II*, so would be included. Negative example: The First Reactions to The *Wicked* Movie Have Critics United”. Does not mention merch, so would be excluded. Edge Case Example: “I tried Starbucks’ green and pink ‘*Wicked*’-themed drinks. I can’t stop thinking about the one without coffee.” Seems to be

a coffee review, but this themed item is technically merch, so it would be included.

- **Unrelated to selected movies:** The primary focus of the article is still about movies but doesn’t mention any of our selected movies or our movies are secondary. Positive example: “Moana is Disney’s Biggest Movie on Streaming Ever”. Talks about a movie, but that movie is Moana, which is not a movie we selected. Negative example: “After *Gladiator II*, Ridley Scott Is Taking Paul Mescal to the Apocalypse.” Clearly talks about *Gladiator II*. Edge Case Example: “The best new movies to stream this November”. At first glance it does not mention any of our movies, but in the description is mentions *Gladiator II*, so it would be excluded.
- **Unrelated to movies in general:** Any articles which don’t focus on movies at all. Positive example: “10 Weekend Reads”. Focuses on books, not movies, so would be included. Negative example: “Anora, A Real Pain, Expand, Crunchyroll’s, Overlord & Small Things Like These Specialty Box Office”. Talks about movies, so would be excluded from this category. Edge Case Example: “How the Novelty Popcorn Bucket Came to Rule the Movies”. While it seems to focus on movies, it focuses on the popcorn bucket, so it would be included this category.

Annotation Results

Using the established typology, the annotated data was organized for each movie to allow for comparisons. The topic distribution for our focus movie, *Gladiator II*, is presented in Figure 2.

Figure 3 provides a summary of these findings, offering a direct comparison between *Gladiator II* and other films.

TF-IDF Results

As outlined earlier, we conducted a TF-IDF analysis on our samples to identify the most representative words for each category. These findings alongside our annotation results will be referenced in the following section to aid in interpreting our data.

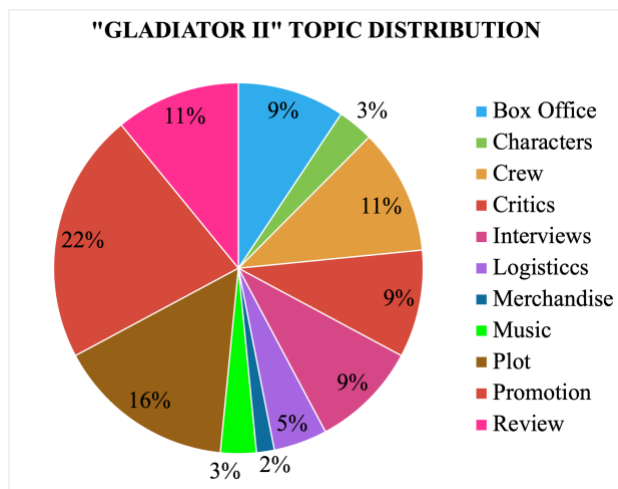


Figure 2: Topic distribution for *Gladiator II*. Note that the majority of samples belong to the Promotion category.

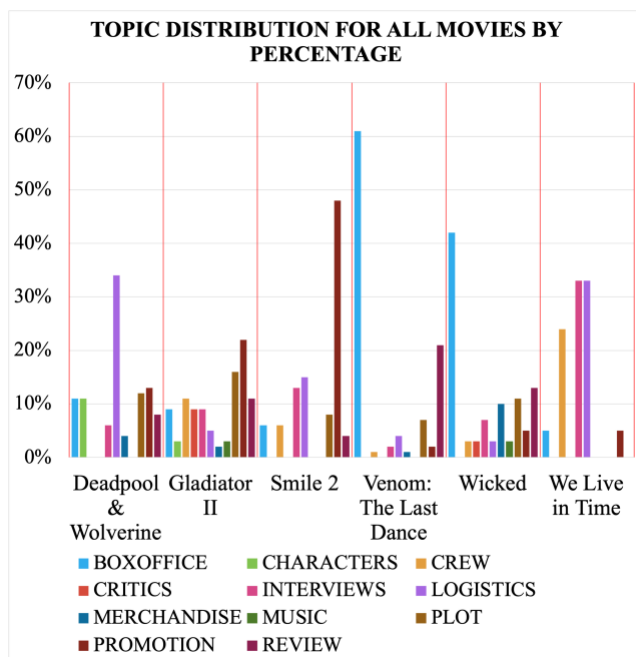


Figure 3: Relative Topic Distribution per movie. Red lines distinguish the distributions between movies.

Discussion

The analysis of news coverage for *Gladiator II* and the selected movies (*Smile 2*, *Venom: The Last Dance*, *Deadpool & Wolverine*, *Wicked*, and *We Live in Time*) underscores significant variations in how media outlets prioritize different topics based on the genre, context, and stage in the film's release cycle. These differences are further influenced by the films' franchise status, the

audience they target, and the marketing strategies employed to generate buzz and engagement.

For *Gladiator II*, the media coverage is heavily skewed towards promotion (22%) and plot/theme (16%), reflecting its recent release and the dual focus on generating excitement around the film's narrative and its connection to the legacy of the original *Gladiator*. This emphasis is characteristic of films in their early release phase, where marketing campaigns and narrative intrigue are essential to building momentum and attracting audiences.

The significant coverage of plot/theme highlights the media's interest in exploring the storyline, themes, and character development of the film. Discussions have centered around how *Gladiator II* balances new narrative elements with the epic tone and legacy of its predecessor, aiming to resonate with both returning fans and new viewers. The prevalence of words "Gladiator," "ridley," and "mescal" from the critics category underscores the media's dual focus: celebrating the film's ties to the original and emphasizing critical reception to build anticipation. Interestingly, box office coverage makes up only 9% of the total, a relatively low proportion compared to promotion. This disparity can be attributed to the short interval between the film's release and the collection of data via NewsAPI. At this early stage, promotional content naturally dominates, as box office performance data may not yet be substantial or central to the narrative.

In contrast, *Venom: The Last Dance* reveals an overwhelming focus on box office (61%), which is expected given the film's status as part of an established and commercially successful franchise. The dominance of words like "million," "weekend," and "box office" highlights the industry's heavy reliance on financial metrics to measure success. As a well-established franchise, *Venom* has built a loyal audience. Because of this, box office performance quickly becomes a defining element of the film's public narrative. The media coverage reflects this priority, with financial success taking center stage and discussions of plot/theme (7%) and promotion (2%) appearing in secondary roles. The significant focus on Box office mirrors the industry's broader trend of emphasizing financial outcomes for blockbuster films, which are seen as key indicators of a film's longevity and potential for sequels or spin-offs.

The case of *Deadpool & Wolverine* reveals a more varied coverage distribution, with a notable focus on logistics (34%) and promotion (13%), and significantly less on box office (5%) and plot/Theme (9%). The increased focus on Promotion highlights the importance of

generating hype for major franchises like the Marvel Cinematic Universe through interviews, sneak peeks, and extensive media coverage. As with other Marvel films, promotional content plays a vital role in maintaining interest and engagement ahead of release, as evidenced by the media's emphasis on keywords like "marvel" and "wandavision.". The heavy focus on Logistics can be attributed to the film's unique release timeline. While the movie originally premiered in theaters in late July, it became available on Disney+ on November 12th, aligning its streaming release with other movies in this analysis. This overlapping timeline provides an additional lens through which to evaluate media coverage.

Since its theatrical debut predates most of the other films examined, the earlier promotion and initial box office blitz have already subsided. As a result, the current media narrative is dominated by logistical updates, such as streaming availability, release dates, and the complexities of integrating *Deadpool & Wolverine* within the Marvel Cinematic Universe. This perspective underscores how extended timelines can shape coverage. For *Deadpool & Wolverine*, the focus has evolved from theater-centric metrics and character discussions to highlighting accessibility via digital platforms. The prominence of words like "streaming" and "marvel" reflects this shift. By contrast, films like *Gladiator II* and *Venom*, still in their early release phase, remain centered on promotional campaigns and box office performance. This comparison highlights how the timing of streaming releases, in conjunction with theatrical premieres, influences the narrative arc of media coverage, offering a dynamic view of how films sustain audience interest across different stages of their lifecycle.

Smile 2 stands out for its heavy promotion coverage (48%), a characteristic trend for horror film sequels that often rely on marketing-driven strategies. Horror films, particularly those with a dedicated fanbase stemming from a successful predecessor, are highly dependent on targeted promotional campaigns to maximize reach and ensure continued interest. The prominence of words like "smile," "horror," and "Halloween" further emphasizes the seasonal marketing push, while box office (6%) coverage remains minimal, as financial performance metrics typically follow in the weeks after release. The film is strategically positioned to capture the Halloween crowd, a time when horror movies sees heightened interest. This seasonal tie-in is a key part of the film's promotional strategy, with media coverage focused on aligning the movie with the broader Halloween marketing ecosystem. While interviews (13%) and logistics (15%) are also covered, they remain secondary to the film's promotional efforts. The emphasis on promotion highlights

how established fanbases and seasonal timing are leveraged to maximize the impact of media coverage.

Wicked, as a major musical adaptation, sees significant box office coverage (42%), reflecting the commercial stakes of a high-profile, fan-driven film. The prominence of box office discussions underscores the importance of financial performance for such large-scale productions, where the success of the movie is often measured by its opening numbers and sustained theatrical run. Words like "Mattel," "doll," and "toy" signal the integration of merchandise sales into the promotional strategy. The movie's association with the *Wicked* brand is further extended through merchandise, which comprises 10% of the article coverage. This highlights an alternative promotional approach, leveraging the brand's existing popularity to create additional revenue streams and deepen audience engagement. Additionally, there is some coverage of music (10%), pointing to the film's roots in the stage musical, with much of the media conversation tied to the music and performance legacy of the original production. Plot/theme (11%) and review (13%) discussions are also relatively subdued, highlighting the media's prioritization of commercial success for a movie of this scale.

Finally, *We Live in Time* offers an entirely different media narrative, reflecting its independent and niche status. The film's coverage is dominated by interviews (33%) and plot/theme (24%), emphasizing the film's emotional resonance and the connections between the cast and their work. The focus on plot/theme highlights the film's commitment to storytelling and thematic exploration, which appeals to audiences drawn to more intimate, character-driven narratives. The low emphasis on box office (5%) and promotion (5%) reflects the film's independent nature and the absence of a large-scale marketing campaign typical of blockbuster releases. The substantial focus on Interviews underscores the importance of personal perspectives in shaping the film's media narrative, resonating with a core audience that values authenticity and artistic intent. *We Live in Time* exemplifies how media coverage for independent films diverges significantly from the mainstream blockbusters discussed in this analysis.

The analysis of news coverage across *Gladiator II*, *Smile 2*, *Venom: The Last Dance*, *Deadpool & Wolverine*, *Wicked*, and *We Live in Time* reveals the multifaceted ways in which media narratives are shaped by a film's genre, release timeline, franchise status, and target audience. Each film's unique characteristics drive the media's emphasis on specific topics, ranging from promotional campaigns and box office metrics to plot-driven discussions and interviews with the cast.

Conclusion

The aim of this project was to analyze the media coverage surrounding *Gladiator II* and compare it to five other films released around the same period. Our study sought to understand both the volume of coverage and the dominant topics discussed, shedding light on the dynamics of film media narratives during their release cycles.

Following our analysis, we have established that *Gladiator II* experienced lower media visibility, especially when considered in the greater pool of the six movies chosen for our analysis. Media coverage for *Gladiator II* was lacking compared to other films, making up only 14% of the total volume of movie articles. In contrast, *Venom: The Last Dance* and *Wicked* were the focus of 26% of the articles each, highlighting their stronger presence in the media. Given our robust bias minimization strategies and thorough methodology, we are confident that our data accurately reflects the actual distribution of media coverage each movie garnered between October 27th and November 27th, 2024.

With regards to the reception of *Gladiator II*, the film's promotional efforts have been successful in building excitement, with a significant portion of the media coverage focusing on its connections to the original *Gladiator* film and its potential to meet the high expectations set by its predecessor. The category promotion dominates the topic distribution, reflecting the strong focus on generating buzz prior to the movie's release. A notable portion of the coverage also centers on the plot/theme of *Gladiator II*, as media outlets explore how the sequel expands upon the original story and its legacy. While there have been fewer articles specifically addressing box office returns compared to other films in the analysis, this is largely due to the film's early release stage. As the film's box office performance continues to unfold, it is likely that this topic will become more prominent in future coverage. Further, our analysis has also enabled us to identify notable trends in the quality of coverage for each of the other selected movies.

The broader findings of our analysis shine light on how genre, franchise status, timing, and release cycles directly influence the scope and nature of a film's media coverage. Future research could expand by incorporating platforms like Letterboxd, where user-generated content and reviews significantly influence a film's reputation, especially among movie buffs. Additionally, analyzing social media platforms like TikTok could offer insights into how Gen Z fans drive organic media hype. Through viral "edits" and fan-driven content, these communities have proven instrumental in creating buzz around films, influencing broader media

narratives. Analyzing these platforms would provide a unique perspective on how grassroots digital engagement is transforming the way films gain attention.