*Corresponding Author's Email: simoneantoninolamela@gmail.com
Proceedings of The International Conference on Social Sciences in the 21st Century
Vol. 2, Issue. 1, 2025, pp. 38-47
DOI: https://doi.org/10.33422/ics21.v2i1.1157

ISSN: 3030-1092 online

From Theaters to Streaming: The Influence of Algorithms on Film Culture and Consumption

Simone Antonino La Mela*, Elvira Celardi

University of Catania, Department of Political and Social Sciences, Italy

Abstract

This study explores the socio-cultural evolution of cinema in the digital age, emphasizing the shift from collective theatrical viewing to personalized, algorithm-driven consumption on streaming platforms and social media. Between January and December 2024, we collected and pre-processed 331,354 English-language user reviews from Letterboxd via a Python-based web-scraping pipeline. Reviews were cleaned (stop-word removal, lemmatization, tokenization), truncated to 512 tokens, and then subjected to sentiment classification using the CardiffNLP "Twitter-XLM-RoBERTa-base-sentiment" model. Subsequently, we developed a prototype recommendation system employing collaborative filtering, deep-learning embeddings, and cosine-similarity metrics to simulate real-world content suggestions. Our sentiment analysis reveals a near balance of positive (108,280), neutral (116,011), and negative (107,063) reactions to films depicting artificial intelligence, underscoring both audience fascination and ethical ambivalence. The recommendation model demonstrates how personalized suggestions can accelerate content discovery while fostering filter bubbles that reinforce existing preferences and limit serendipitous encounters. Despite these algorithmic constraints, online film communities (e.g., Letterboxd) sustain cinema's traditional social function through collective discussion and user-driven curation. These findings contribute to debates on media democratization, algorithmic governance, and cultural stratification, and they offer practical insights for designing more transparent, user-centric recommendation engines that balance personalization with diversity.

Keywords: artificial intelligence; sentiment analysis; recommender systems; digital cinema; audience reception

1. Introduction

Bauman (2012) argues that in consumer societies, shared dependence on consumption—particularly the “universal dependence on shopping”— is the *conditio sine qua non* for individual freedom, most notably the freedom to “be different,” to “have an identity”. This insight offers a compelling lens through which to understand how social media platforms, via highly personalized experiences, enable users to assert their identities by selecting content that aligns with their tastes and preferences (Liu et al., 2022).

Visual culture plays a parallel role in shaping identity. Harper (1993) distinguishes between *visual sociology with images* (the production of images as data) and *visual sociology of images* (the analysis of existing imagery), recognizing both as valuable methodologies for investigating cultural and political dynamics. A century earlier, Altenloh (2018) had already demonstrated that access to film varied according to class, gender, and age, revealing structural inequalities and differentiated opportunities for cultural formation—dynamics that remain relevant today (Sorlin, 2017; Bentivegna & Boccia Artieri, 2019).

With the rise of streaming platforms and social media, *liquid cinema* has gained prominence (Casanelles, 2024; Han, 2022). In this paradigm, film viewing becomes pervasive, asynchronous, and personalized, dissolving the boundaries between physical media (film reels, VHS, DVDs) and digital formats (streaming). However, beneath this apparent democratization of access, hierarchies of taste and symbolic barriers endure, reflecting persistent inequalities in cultural capital (Bourdieu, 2001; Alpini, 2021).

The advent of recommender systems marks a significant shift in the logic of content discovery. These algorithms analyze demographic factors (such as age and gender), consumption habits, and expressed preferences to filter overwhelming content and provide tailored suggestions (Ricci et al., 2010; Ellis, 2002). Typically structured around three main phases—data collection, model training, and recommendation generation—these systems rely on four primary approaches: content-based, collaborative, context-aware, and hybrid (Isinkaye et al., 2015; Jayalakshmi et al., 2022).

- Content-Based Filtering recommends content by analyzing the characteristics of items a user has liked in the past. For example, films are tagged with keywords that define their genre, main characters, and other attributes. The system matches these tags with the user’s profile–constructed from previously liked items, to suggest similar content. This method does not rely on other users' data and adapts quickly to changes in user preferences. However, it can produce overly similar recommendations, limiting diversity, and performs poorly with new users due to a lack of initial data (cold start problem).
- Collaborative Filtering addresses some of these limitations by using the behavior and preferences of other users. It identifies patterns in user interactions: item-based filtering finds items often liked together, while user-based filtering matches users with similar tastes to recommend unseen items. It can provide unexpected yet accurate recommendations, known as serendipitous recommendations. However, it struggles with sparse data, where user-item interactions are limited across large datasets, reducing the accuracy and coverage of suggestions.

- Context-Based Filtering introduces an additional layer by considering the context in which content is consumed. This method assumes that users with similar preferences in one context will likely agree in others. It identifies shared characteristics across situations to inform recommendations. A more advanced form, context-aware recommender systems (CARS), adjusts suggestions based on variables like time of day, user mood, or company, for instance, avoiding long movies after a tiring day or suggesting romantic films when users are with a partner.
- Hybrid Systems combine two or more previous methods to mitigate their perspective weaknesses and enhance recommendation quality. By integrating multiple data sources and analytical techniques, hybrid models deliver more accurate and personalized suggestions. These systems are particularly effective in overcoming issues like cold start or data sparsity.

Despite their potential, each method has problems: cold start, accuracy, diversity, scalability, sparsity, and synonymy. Modern recommendation platforms often integrate social media accounts to enrich user profiles, reducing cold start problems and enhancing personalization. Hybrid and context-aware systems are increasingly adopted to deliver tailored, effective, and diverse recommendations.

However, relatively few studies have combined user sentiment analysis with recommender system simulation to experimentally assess the real-world impact of algorithms on viewer behavior. This study addresses that gap through a large-scale sentiment analysis of over 330,000 user reviews on Letterboxd and the development of a prototype recommendation engine. In doing so, it investigates how personalized recommendations influence content discovery and shape dynamics within online film communities.

2. Methods

This study adopts a quantitative methodology to investigate how audiences emotionally respond to cinematic representations of artificial intelligence (AI), with a particular focus on science fiction films. Through large-scale data collection and sentiment analysis, the research aims to identify recurring patterns in viewer perception and emotional engagement with the theme of AI in contemporary cinema.

2.1. Data Collection and Preprocessing

User-generated content was collected from *Letterboxd*, a widely used social media platform dedicated to film reviews and community interaction. A Python-based web scraper was developed and executed using Google Colab, leveraging the `requests`”, `BeautifulSoup`”, and `pandas` libraries to extract structured data for approximately 180 films. These titles were selected from the Wikipedia page *“List of Artificial Intelligence Films”*, ensuring thematic coherence while minimizing subjective selection bias. The final corpus included both internationally recognized and lesser-known productions, such as *Teri Baaton Mein Aisa Uljha Jiya* (India, 2024), *Kalki 2898 AD* (India, 2024), and *Invasion* (Russia, 2020).

For each film, reviews were extracted in order of popularity, up to a maximum of 256 pages per title (corresponding to 3,072 reviews). When fewer pages were available, all reviews were retrieved to preserve dataset completeness. Metadata fields such as reviewer username, review text, publication date, number of likes, and film title were systematically recorded. The initial raw dataset contained over 500,000 reviews.

Textual preprocessing was conducted to ensure linguistic consistency and model compatibility. This involved:

- Removal of stop words, emojis, and special characters;
- Normalization of case and punctuation;
- Lemmatization of tokens to reduce words to their base forms;
- Filtering for English-language content resulted in a final dataset of 331,354 reviews.

2.2 Sentiment Analysis

To evaluate the emotional tone of each review, sentiment classification was performed using the model "*cardiffnlp/twitter-xlm-roberta-base-sentiment*", available via the Hugging Face *Transformers* library. This multilingual model is based on the XLM-RoBERTa architecture, an extension of RoBERTa, which itself improves upon BERT by incorporating a robust pretraining method and a bidirectional attention mechanism. Trained on 198 million social media posts, the model is optimized for sentiment detection in short, informal texts.

Given the length and complexity of film reviews compared to tweets, input sequences were truncated at 512 tokens to align with the model's maximum input length. Each review was assigned one of three sentiment categories: positive, neutral, or negative.

While the model demonstrated strong performance in identifying positive and negative sentiments, a relatively high frequency of neutral classifications was observed. This phenomenon can be attributed to two factors: (1) the model's original training on short Twitter texts, and (2) the more nuanced or ambiguous tone often present in long-form film reviews. Nevertheless, performance evaluations indicated acceptable levels of precision and recall for the intended analysis.

To optimize processing, the dataset was randomized and divided into four equal subsets. Each segment was analyzed independently, and the results were stored in separate Excel files. These were subsequently merged into a single dataset for comprehensive evaluation. Additional metadata, such as genre tags and cast members, was parsed into structured fields to support further correlation analysis.

3. Results

This section presents the findings of the sentiment analysis performed on six key films featuring representations of artificial intelligence (AI), as well as an overview of general trends emerging from a dataset of over 330,000 reviews. The films were selected based on thematic relevance, cultural significance, and diversity of narrative approach.

3.1 Sentiment Analysis by Film

The sentiment analysis of selected science fiction films reveals distinct emotional patterns in audience responses to different representations of artificial intelligence. Each film offers a unique narrative and aesthetic interpretation of AI, reflected in its sentiment distributions.

❖ *Metropolis (1927)*

This silent film classic introduces the humanoid robot Maria and presents a dystopian vision of an industrial future. The sentiment analysis shows most positive reactions, highlighting the film’s cultural and historical impact. A smaller proportion of negative responses is likely linked to the film's age and stylistic distance from modern cinematic conventions.

Tab. 1: Sentiment distribution - Metropolis

	Count by label
Negative	379
Neutral	796
Positive	837

Source: Author's elaboration based on data from Letterboxd

❖ *2001: A Space Odyssey (1968)*

Kubrick’s masterpiece features the sentient AI HAL 9000 and explores the relationship between humanity and technology. The sentiment analysis indicates a dominant positive sentiment, reflecting the film’s visionary character, technical innovation, and philosophical depth.

Tab. 2: Sentiment distribution - 2001: A Space Odyssey

	Count by label
Negative	494
Neutral	986
Positive	1087

Source: Author's elaboration based on data from Letterboxd

❖ *The Matrix (1999)*

In *The Matrix*, humanity is unknowingly trapped in a simulated reality controlled by AI. The analysis shows a balanced distribution of positive and neutral sentiments, suggesting strong appreciation for its themes, execution, and the complexity of its narrative structure.

Tab. 3: Sentiment distribution - The Matrix

	Count by label
Negative	531
Neutral	1081
Positive	1076

Source: Author's elaboration based on data from Letterboxd

❖ *Wall-E (2008)*

This animated film presents emotionally expressive AI characters in a future shaped by ecological collapse. The sentiment data shows a majority of positive reviews, likely influenced by the film’s emotional appeal, optimistic tone, and critique of consumer society.

Tab. 4: Sentiment distribution - Wall-E

	Count by label
Negative	521
Neutral	860
Positive	1184

Source: Author's elaboration based on data from Letterboxd

- *Her (2013)*

Her imagines a near future in which people develop emotional relationships with AI-powered operating systems. While positive and neutral sentiments dominate, some reviews express negative reactions driven by concerns over isolation and emotional dependency on machines.

Tab. 5: Sentiment distribution - Her

	Count by label
Negative	669
Neutral	937
Positive	893

Source: Author's elaboration based on data from Letterboxd

- *Ex Machina (2015)*

Ex Machina raises ethical and philosophical questions through the character of Ava, a highly realistic and autonomous AI. Most reviews show positive sentiment, recognizing the film's narrative depth and tension. The small portion of negative sentiment is likely due to its unsettling tone and confined setting.

Tab. 6: Sentiment distribution - Ex Machina

	Count by label
Negative	379
Neutral	796
Positive	837

Source: Author's elaboration based on data from Letterboxd

3.2 General Sentiment Trends

The overall sentiment analysis of the full dataset comprising 331,354 reviews shows a predominance of neutral and positive labels, with negative sentiment forming the smallest proportion. This differs from the six selected films, which generated significantly more positive sentiment. The discrepancy can be attributed to the broader dataset's inclusion of lesser-known or lower-quality productions that did not resonate as strongly with audiences.

Two key factors explain the high rate of neutral sentiment:

1. Many user reviews express ambivalence, mixing praise and critique, resulting in emotionally neutral classifications.

2. The sentiment model may misclassify complex expressions such as sarcasm, irony, or film quotes as neutral due to limitations in contextual understanding.

The low percentage of negative sentiment in the six highlighted films further supports their cultural value and critical success. In contrast, the general dataset includes films with varying levels of production quality, leading to a more balanced distribution across sentiment categories.

The findings suggest that audiences are generally receptive to films dealing with artificial intelligence. Emotional narratives, anthropomorphized AI, and ethical dilemmas appear to engage viewers deeply, particularly within romantic, dramatic, or dystopian contexts.

3.3 Recommender System

The second part of this study involved the development of a content-based recommender system, specifically designed to work with the 180 films previously analyzed through sentiment analysis. The objective was to simulate a realistic system capable of suggesting similar movies based on user input and extracted film metadata.

The system was constructed using Python within the Google Colab environment. The initial phase consisted of scraping data from the main *Letterboxd* pages of each film, using the "requests" and "BeautifulSoup" libraries. This process retrieved key metadata such as the film's title, director, release year, tagline, cast members, and genre tags. These attributes were then organized into structured datasets using the "pandas" library and exported into Excel files for easy processing and storage.

Following data collection, the next stage focused on modeling. A content-based recommendation engine was designed, relying on deep learning and cosine similarity to identify and rank similar titles. The model was built using "Keras" and "Sklearn", incorporating an architecture based on three core layers: the input layer ingests encoded film and user features; the embedding layer reduces dimensionality and refines feature relationships; the output layer generates predicted similarity scores. The data was first preprocessed using "LabelEncoder", which converted categorical variables (such as director, cast, and genre) into numerical representations. Each embedding vector was set to a dimension of four, and additional dense layers were used to combine information and produce final predictions. The final model architecture included eight input features, twelve hidden layers, and one output node configured for regression, which provided a similarity score between titles.

Once the model was trained on the complete film dataset, a similarity matrix was generated using the "cosine_similarity()" function. This allowed the system to identify the most closely related films in the collection based on content features. The recommendation function, named "consiglia_film()", accepts multiple input criteria, including title, director, cast member, or genre. Based on the input provided, the system calculates similarity rankings and returns the five most relevant films, displaying their metadata (title, tagline, year, etc.). If the system does not find any matching entries, it falls back on generating five random suggestions from the database. This logic is implemented through a simple conditional structure that selects a random film when no index match is found.

To test the functionality, several explanatory queries were conducted. For example, using "Tom Cruise" as the input actor, the system recommended five films not starring him, but similar in content and thematic proximity. These were *Star Trek: Generations* (1994), *The Artifice Girl* (2022), *Screamers* (1995), *The Iron Giant* (1999), and *RoboCop* (1987). This confirmed that the system does not simply retrieve films containing the same actor but rather those resembling the overall features of the actor's known roles.

A second query included two input parameters: actor "Tom Cruise" and genre "Drama." The system returned: *Tron* (1982), *Colossus: The Forbin Project* (1970), *Prometheus* (2012), *Transcendence* (2014), and *Superman III* (1983). These results demonstrate the system's ability to combine multiple features and return coherent, content-aligned suggestions.

Each query generates slightly different outcomes depending on the input structure and selected metadata, balancing consistency and novelty. This dynamic behavior mirrors how commercial streaming platforms integrate recommender systems, aiming to fulfill explicit user preferences while enabling serendipitous discovery. In environments with vast content libraries, such mechanisms are essential to ensure user satisfaction and exploration, adapting flexibly to evolving tastes and needs.

4. Discussion

The results of this study emphasize how the current cinema and audiovisual media landscape is significantly influenced by the prevailing themes in public discourse, especially artificial intelligence and algorithmic personalization. These topics, central to cultural imagination and technological innovation, drive film narratives and influence how content is consumed, discussed, and recommended in digital spaces.

While some reviews conveyed apprehension and ethical concerns, many expressed curiosity, admiration, and empathy toward the representation of artificial intelligence. This duality reflects the complexity of the AI theme itself, which spans dystopian control (*The Matrix*), philosophical inquiry (*Ex Machina*), emotional dependency (*Her*), and even robotic tenderness (*Wall-E*).

The sentiment distributions also illustrate how audience perception of AI has evolved, mirroring changes in its narrative depiction—from the prophetic vision of *Metropolis* (1927) to the computational logic of *2001: A Space Odyssey* (1968) and the post-human questions posed by more recent titles. This shift suggests that public understanding of AI is not static but mediated by popular culture in response to technological developments and societal anxieties.

However, the study has some significant limitations. One key constraint was the exclusive focus on English-language reviews. Although this choice was necessary for model consistency and accuracy, it inadvertently excluded non-English-speaking audiences, reducing cultural representativeness and reinforcing an Anglocentric perspective. A future multilingual approach could better capture global sentiment and uncover cultural variations in the perception of AI.

One limitation of automated sentiment analysis is its effectiveness. Although the Cardiff NLP model in this study performs well with short texts, it has difficulty analyzing longer reviews. As a result, it often misclassifies sarcasm, irony, and mixed emotions as "neutral." This tendency leads to an excessive number of neutral labels, which can obscure the nuanced opinions expressed by users. Additionally, sentiment scores do not differentiate between appreciation of the film's theme, technical quality, or emotional resonance, raising questions about interpretative depth.

Building a prototype recommender system, in addition to sentiment analysis, reveals how algorithmic structures influence the user experience. The system showcased how deep learning and content similarity can effectively personalize recommendations. However, it also brought to light concerns about filter bubbles and reduced exposure to unfamiliar content. This situation highlights broader discussions in digital culture regarding the impact of recommendation systems on cultural consumption and the reinforcement of existing preferences.

Nevertheless, the potential of such systems should not be dismissed. When designed responsibly, recommender algorithms can support discovery, highlight underrepresented films, and enhance user satisfaction. The key challenge is to ensure that personalization does not come at the expense of diversity, critical engagement, or the user's agency.
Ultimately, this study shows that AI-themed cinema—and its reception—offers a valuable lens for examining the intersection of technology, culture, and emotion. As artificial intelligence becomes more embedded in everyday life, cinema's role as a mirror and a mediator of collective imagination becomes increasingly significant. Future research should explore hybrid methods, combining large-scale data analytics with qualitative inquiry, to capture the full spectrum of how audiences respond to the technological narratives that increasingly shape our shared realities.

5. Conclusion

This study examined how artificial intelligence is perceived and emotionally interpreted in cinematic narratives, combining large-scale sentiment analysis with a dedicated content-based recommender system design. The results indicate that AI is not simply received with fear or dystopian concern but frequently inspires curiosity, empathy, and philosophical engagement. Audience reactions to the selected films demonstrate a deep connection with technological themes and their narrative and emotional dimensions.

The recommender system further illustrated the growing influence of algorithmic structures on audiovisual consumption. While such systems enhance personalization and user satisfaction, they also raise ethical questions concerning user autonomy, content diversity, and algorithmic neutrality.

Overall, this research confirms the growing relevance of audiovisual products as lenses through which contemporary society processes and negotiates its relationship with artificial intelligence. Integrating sociological insight with technological methodology proved to be a productive approach for identifying patterns in user sentiment and recommendation dynamics.

Future investigations should explore multilingual and multimodal datasets, incorporate qualitative research, and deepen the examination of AI not only as a narrative theme but also as a digital infrastructure. As artificial intelligence continues to reshape the boundaries of human experience, cinema remains a crucial space for reflection, debate, and cultural literacy.

References

Alpini, S. (2021). *Sociologia del cinema*. Edizioni ETS.

Altenloh, E. (2018). *Verso una sociologia del cinema. Industria e pubblico* (M. Santoro & B. Grüning, A c. di). Mimesis.

Bauman, Z. (2012). *Modernità liquida*. Laterza.

Bentivegna, S., & Boccia Artieri, G. (2019). *Le teorie delle comunicazioni di massa e la sfida digitale*. Laterza.

Bourdieu, P. (2001). *La distinzione. Critica sociale del gusto*. Il Mulino.

Casanelles, S. (2024). *From hyperreality to digital cinema: A theoretical overview. In The hyperorchestra* (pp. 29–44). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-75193-6_2

Ellis, N. C. (2002). *Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition*. Studies in Second Language Acquisition, 24(2), 143–188. https://doi.org/10.1017/S0272263102002024

Han, B.-C. (2022). *Le non cose. Come abbiamo smesso di vivere il reale*. Einaudi.

Harper, D. (1993). *Orizzonti sociologici. Saggio di sociologia visuale*. Sociologia della comunicazione, 10(19).

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). *Recommendation systems: Principles, methods and evaluation*. Egyptian Informatics Journal, 16(3), 261–273. https://doi.org/10.1016/j.eij.2015.06.005

Jayalakshmi, S., Ganesh, N., Čep, R., & Senthil Murugan, J. (2022). *Movie recommender systems: Concepts, methods, challenges, and future directions*. Sensors, 22(13), 4904. https://doi.org/10.3390/s22134904

Liu, M., Kamper-DeMarco, K. E., Zhang, J., Xiao, J., Dong, D., & Xue, P. (2022). *Time spent on social media and risk of depression in adolescents: A dose–response meta-analysis*. International Journal of Environmental Research and Public Health, 19(9), 5164. https://doi.org/10.3390/ijerph19095164

Ricci, F., Rokach, L., & Shapira, B. (2010). *Recommender systems handbook*. https://doi.org/10.1007/978-0-387-85820-3_1

Sorlin, P. (2017). *Introduzione a una sociologia del cinema*. Edizioni ETS.