

Augmenting Content Analysis in the Era of Streaming Video: Harnessing AI for Comprehensive VoD Research

Stéfany Boisvert

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
boisvert.stefany[AT]uqam.ca¹

Dave Anctil

COLLÈGE JEAN-DE-BRÉBEUF
dave.anctil[AT]brebeuf.qc.ca

Abstract

Streaming services have profoundly transformed the audiovisual industry, reshaping both production and distribution practices as well as viewing habits. Notably, video-on-demand (VoD) services have greatly expanded the number of series produced annually. Yet despite the extensive volume of audiovisual productions available on VoD services, most scholarly work continues to prioritize case studies or limit their scope to a small corpus of texts. This article critically examines artificial intelligence (AI)-assisted content analysis as a methodological avenue that could allow scholars to analyze extensive corpuses of audiovisual productions available on streaming services. Using multimodal generative algorithms and other integrated digital tools, such as the large language model (LLM) Gemini and the platform Google AI Studio, we will show how AI-assisted analysis might enable more thorough understandings of media production within the VoD landscape. Drawing on the results of test analyses conducted with Gemini, this article also critically addresses the epistemological and methodological challenges of AI-augmented content analysis.

Keywords: Video-on-Demand, Content Analysis, Multimodal AI, Large Language Models (LLMs), Gemini.

Introduction²

Streaming services have profoundly transformed the audiovisual industry, reshaping both production and distribution practices as well as viewing habits. Notably, VoDs have greatly expanded the number of series produced annually. This surge in content has resulted in major disruptions across audiovisual industries, exemplified by the widely analyzed phenomena of the Peak TV era in the United States (2016–2022) and the “streaming wars.”³

Yet despite the extensive volume of audiovisual productions available on VoD services, most scholarly work continues to prioritize case studies or limit their scope to a small corpus of texts.⁴ The few studies that have conducted textual/content analyses of a large “corpus of originals”⁵ have predominantly adopted quantitative methods or, when employing qualitative approaches, relied on content categorization strategies that circumvent the need for extensive viewing, such as genre classification.⁶

Such methodological limitations are understandable, given the significant workload associated with conducting content⁷ analyses. The segmentation of series and their extended duration make their study particularly labor-intensive. With streaming services, these challenges are further compounded by the relentless demand for fresh content, as they operate outside the traditional television logic of seasonal programming and require continuous updates to their catalogs. Nevertheless—and precisely because of the overabundance of content in the era of VoD—it is perhaps more critical than ever to explore new methodological approaches capable of addressing the demands of analyzing vast corpuses.

Large-scale content analyses could, among other benefits, uncover a wider diversity of narrative and aesthetic trends within series offered on VoD platforms, highlight cross-cutting themes, or produce more conclusive findings—both quantitative and qualitative—regarding a streaming service’s commitment to diversity and inclusion. Additionally, conducting large-scale content analyses could encourage scholars to include in their corpus “routine” productions, which, while “not [being] dissected by critics, nonetheless make important contributions to the cultural role of screen storytelling.”⁸ As Manovich observes, this approach could thus help deconstruct the legitimist biases often underlying our corpus selection.⁹ In this light, examining a wider selection of productions could yield more definitive insights into narrative and aesthetic trends, or a platform’s overall production strategies, while circumventing biases related to the legitimacy or popularity of selected works.

Building on these observations, this article will critically examine artificial intelligence (AI)-assisted analysis as a methodological avenue that could allow TV/screen scholars to analyze extensive corpuses of audiovisual productions available on streaming services. Using multimodal generative algorithms and other integrated digital tools, such as the large language model (LLM) Gemini and the platform Google AI Studio, we will show how AI-assisted analysis might enable more thorough understandings of media production within the VoD landscape. We propose the concept of *AI-augmented analysis* to redefine the emerging collaborative methodologies between AI and researchers. Unlike the type of “automated analysis” carried out by earlier AI technologies, the generative and interpretative capabilities of LLMs give more nuanced insights and create exploratory perspectives

that go beyond predefined algorithmic outputs. Drawing on the results of test analyses conducted with Gemini, this article also critically addresses the epistemological and methodological challenges of *AI-augmented content analysis*, offering insights into future customized methodologies.

Enhancing Media Studies through Automated Analysis

Over the past two decades, research initiatives have emerged to develop methodologies leveraging computational methods and AI for the analysis of audiovisual productions. Among the most notable efforts are **Cinemetrics** (2005), created by Yuri Tsivian and Gunars Cijvans, and, shortly thereafter, the Shot Logger software developed by Jeremy Butler (2007). These tools enabled the data visualization of various “stylistic patterns.”¹⁰ More recently, the Audiovisual Cinematic Toolkit for Interaction, Organization, and Navigation (**ACTION**), an open-source Python platform, has further expanded the possibilities of data visualization for both audiovisual and user-generated content.

Other researchers in TV, screen, and critical media studies have also adopted automated analysis methods, leveraging algorithms and digital tools to perform more detailed and comprehensive examinations of texts through various AI applications. Automated analysis has been employed for a wide range of audiovisual data extraction and processing tasks, including scene segmentation, object detection, speech and facial recognition, character analysis, dialog transcription, and thematic, aesthetic, or narratological analysis. While the first computational methods were mostly confined to quantification, recent advancements have enabled the integration of both quantitative and qualitative approaches.

These advancements have been particularly valuable for studying diversity in audiovisual productions¹¹ and facilitating the analysis of longer formats, such as television series. Automated analysis has indeed proven useful in compiling data that are difficult to track manually, such as screen time and speaking time. This has allowed for precise quantified findings regarding, for instance, the overrepresentation of men compared to women in audiovisual content.¹²

Furthermore, AI-assisted analysis has demonstrated potential for narratological and discourse analyses across extensive corpuses of series, offering broader insights into narrative structures and trends¹³ or linguistic patterns.¹⁴ For instance, Bost et al. utilize dynamic conversational networks to generate video summaries centered on specific characters.¹⁵ This technique has been developed in a context where binge-watching practices and extended intervals between seasons have made it difficult for audiences to retain key narrative details, and could ultimately affect their engagement with and interpretation of ongoing storylines.¹⁶

However, despite these promising advancements in AI-assisted analysis, it is crucial to recognize that such initiatives remain on the periphery of TV/screen studies. The underutilization of AI in this field can be attributed primarily to the relatively recent development of

AI capabilities for analyzing multimodal content, such as videos. Moreover, the reliance on specialized programming languages like Python presents a significant barrier.

Compatibility issues with operating systems add another layer of complexity to the analytical process. Furthermore, while Python tools can recognize voice, text, and images, thereby enabling the analysis of extensive samples of texts, manually modifying the outputs they generate remains a significant challenge. In other words, while automated and deterministic analyses of audiovisual productions have been achieved, performing an augmented analysis—where researchers can refine AI-generated findings or manually adjust specific data points—proves to be far more intricate and resource-intensive.

These limitations underscore the challenges inherent in achieving a more interactive and adaptable approach to analyse content. However, recent advancements in AI, particularly the rise of LLMs, offer new possibilities. Indeed, with the easy access to LLM services, a new method of *augmented analysis* using a multimodal model becomes possible. Other research fields have already begun exploring these possibilities across large volumes of textual, audio, and video content, for example, in health, psychology, and biomedicine.¹⁷ These new methods of content analysis are considerably simpler to initiate, as they operate through textual prompting, allowing researchers to harness the power of deep learning algorithms without requiring extensive programming, data science, and machine learning expertise.

From Automated to Augmented Analysis: LLMs and MLLMs for Video Analysis

The advent of LLMs represents a paradigm shift in AI. Their neural network models language patterns, thus enabling applications ranging from translation to conversational agents. Powered by transformer-based architectures, these models also have the capacity to process vast amounts of textual data. Unlike earlier AI systems, which relied heavily on task-specific rules or narrowly defined statistical models, LLMs learn general patterns and relationships from their huge training dataset.¹⁸ This capacity for abstraction and context-aware prediction, therefore, makes them capable of addressing complex linguistic tasks, such as nuanced text interpretation, question answering, and contextual dialog generation.^{19,20} Transformer-based architectures use an *attention mechanism* to process all elements of a text sequence simultaneously, rather than sequentially. This enables LLMs to capture long-range dependencies and relationships in language more effectively.²¹

Where classic software programs execute predefined tasks with deterministic outputs, transformer models are designed to generate, interpret, and adapt dynamically to the context of the research, enabling a dialectical and iterative interaction between researchers, their research content, and the AI itself. For researchers in media studies, this means that AI can now engage with texts in a manner similar to their own methodologies—posing questions, offering interpretations, and even generating new avenues of exploration. This could therefore allow for a new form of “co-intelligence” where researchers refine AI-generated

insights, identify nuances, and challenge assumptions, much as they would in their own critical reading and viewing practices.²²

Multimodal large language models (MLLMs) extend this capability beyond written text, integrating all types of data, including images, video, and audio. They achieve this by encoding inputs from different modalities—such as visual scenes, spoken dialogs, or environmental sounds—into a shared representational space, allowing for seamless reasoning across diverse forms of information. This multimodal capability is critical for video analysis, where understanding content requires synthesizing temporal, visual, and sonic inputs.²³

Thus, we must understand why LLMs, and especially the new MLLMs, differ fundamentally from previous AI systems used in automated video analysis. Traditional approaches often relied on handcrafted features and domain-specific algorithms.²⁴ Transformer algorithms, by contrast, operate using the self-attention mechanism that allows to weigh the relevance of every part of the input relative to every other part. This enables them to capture complex dependencies and relationships within and across modalities. For video analysis, transformers process spatial information from individual frames, temporal dynamics across sequences, and semantic cues from audio or text—all within a unified framework. This holistic processing of inputs marks a fundamental departure from the piecemeal and task-specific nature of older AI systems.²⁵

To understand this difference in capabilities, imagine you are analyzing a hypothetical scene from *Stranger Things* (Netflix, 2016–). The scene features a conversation in a dimly lit room where Eleven, the protagonist, is arguing with Hopper, her adoptive father, while the background music swells and lights flicker ominously.

In traditional AI systems, (1) a *vision model* might focus on the flickering lights to detect changes in brightness, but it would not connect those to the tension in the conversation. (2) A distinct *audio model* might analyze the rising music tones to detect emotion, but it would not link them to the characters' facial expressions. (3) A *natural language processing* (NLP) algorithm might analyze their dialog but completely miss how visuals and sounds amplify the scene's tension.

Now, consider how an MLLM could analyze this scene. The self-attention mechanism enables the model to simultaneously weigh and connect all the elements: (1) The dialog's tone and content is connected to Eleven's facial expression of frustration, such as a furrowed brow. (2) The flickering lights are linked to the ominous music, amplifying the overall suspense of the scene. (3) The visual framing, such as the camera zooming in on Eleven's face, is connected to her spoken words, emphasizing her emotional state.

In essence, MLLMs can detect and interpret that Eleven's anger is heightened not only by her words but also by her expression and the way the lights and music reflect her emotional turmoil. The self-attention mechanism connects these dots, weighing the importance of each element (dialog, visual cues, and music) relative to the others, and interprets the scene as a cohesive emotional and narrative moment.

For TV scholars, this is akin to what they already do when analyzing *mise-en-scene*, cinematography, and sound design *in context*. The transformer's self-attention mechanism

mimics this integrated perspective, making it a tool that could augment the researcher's analysis of nuanced, multimodal content:

Action and Event Recognition: MLLMs can identify complex actions and events by analyzing the interplay between visual motion patterns and contextual clues from audio or subtitles. Their ability to model temporal dependencies allows them to construct an interpretation of sequences of actions and predict future events.

Video Captioning and Summarization: By combining visual understanding with textual interpretation, MLLMs can generate contextually accurate descriptions of video content. This capability enhances accessibility and makes large video datasets more navigable.

Contextual Search and Retrieval: MLLMs can allow for more complex search functions by correlating textual queries with visual or auditory content. For example, a user could search a video library with a prompt like “show me scenes with dramatic sunsets and melancholic music,” and the model would identify relevant clips.

Behavioral and Sentiment Analysis: By synthesizing facial expressions, tone of voice, and scene context, MLLMs can assess emotions, motivations, and interpersonal dynamics in video content.

Theoretically, MLLMs can redefine what AI-assisted video analysis can achieve. In a technical research context, they have demonstrated their ability to connect visual and textual information, making them effective in video processing.²⁶ For scholars in TV/screen studies, MLLMs could also be used as “coworkers” for analyzing bigger video corpuses and identify character demographics, track narrative structures, and even analyze stylistic patterns. Incidentally, they would reduce the manual labor needed for viewing and annotating content, enabling researchers to focus on patterns and trends that were previously unobservable due to scale constraints.

But technical research also shows that these systems struggle with abstract temporal concepts.²⁷ For now, challenges include biases toward short-term patterns, difficulties in handling abstract concepts like causality or event duration, and limited datasets focused mainly on surface-level tasks like action recognition and video captioning. This, in turn, is explained by existing training datasets that often lack detailed annotations about event order or causality, restricting their usefulness for long-term temporal analysis.²⁸ As we will see in the next section, the tests we conducted have highlighted several potentialities of MLLMs for video analysis but also some of these limitations.

Empirical Insights from a Testing Phase

General Findings

To offer a critical yet practical reflection on the potential and challenges of analyzing TV series with an MLLM, we conducted a testing phase using Gemini 1.5 Pro, an MLLM developed and deployed by Google in 2024. Tests were conducted on a few episodes of a Canadian

teen series that had been cataloged as part of a larger research project on French-language original series produced for Canadian streaming services.²⁹ In other words, the selected series had already been viewed in its entirety, allowing us to validate the accuracy of the MLLM's analysis.

To conduct the analysis, episodes were uploaded to Google Drive. While some MLLMs allow video analysis via a hyperlink—such as referencing a video on an external website like YouTube³⁰—uploading videos proved useful to better control error margins, for instance by eliminating extraneous elements that could skew the analysis. To prevent the model from becoming “confused” regarding the narrative structure of the episodes, the importance of characters, key dialogs to transcribe, or the screen time of each character, we deemed it preferable to preemptively cut certain segments, such as the opening credits (composed of a succession of images without narrative coherence) and closing credits (which included pre-views of the next episode). Once the episodes were slightly shortened using a video editing software, they were uploaded to our Google cloud and then to the Google AI Studio platform for analysis with Gemini Pro.

For the testing phase, several aspects were explored: An initial step involved asking the model to identify (and thus distinguish) recurring characters by specifying the following characteristics: approximate age, gender, ethnicity/race, and sexual orientation, if it was explicitly mentioned or alluded to. Once characters were correctly identified and listed, the model was asked to specify “relevant identity traits” for each character, as well as “physical characteristics (hair color/length, clothing style, etc.) to help differentiate them more clearly.” Finally, the model was instructed to draft a summary of the story based on analyzed episodes and, in so doing, highlight “key themes” and “major narrative arcs.”³¹

This testing phase yielded promising results, enabling a swift and efficient compilation of specific data. With well-parameterized prompts, such as “Only consider characters who are present for at least X minutes and who speak,” the AI model successfully identified all relevant characters for further analysis. Additionally, facial recognition was executed seamlessly and accurately, facilitating the collection of basic identification data. For example, the gender (male/female) of each character was easily determined, alongside their approximate age. The AI model thus accurately categorized the protagonists as “late teens (school setting)” and provided similar demographic insights for secondary characters, such as parents and teachers. MLLMs also demonstrate accuracy in sentiment analysis, a method commonly employed in marketing and clinical psychology to computationally assess opinions and emotions based on visual and auditory cues.³² During our pretest phase, the model's outputs regarding key character traits and their reactions or motivations were precise, meaning that they corroborated the textual analysis that had been previously conducted manually.

Narrative analysis and targeted summaries could also be refined to concentrate on specific characters. While researchers can independently generate such observations, using AI has been useful to enrich the process by compiling extensive data about individual characters from diverse perspectives. This could therefore also facilitate insightful cross-references to identify narrative trends or recurring themes within larger corpuses. For example, sentiment analysis might enable a more nuanced examination of whether similar stereotypes—be they related to gender, race, or other variables—or story arcs are reproduced across

multiple series.³³ Such an approach could offer researchers a broader and more comprehensive view, allowing them to interrogate patterns of representation and thematic continuity across many audiovisual productions.

Limitations and Methodological Issues in Character Identification

Some identification data are admittedly more challenging, if not impossible, for Gemini 1.5 Pro to determine accurately. For instance, while the model can often identify a character's race based on phenotypes, determining whether a character is racialized (non-white) becomes more complex in certain situations, especially for mixed-race individuals. Additionally, Google has implemented filters, which seems to result in the model defaulting to identify characters as "white" unless facial features or skin color is clearly distinctive. This means that the model does not, of course, consider the fact that race is a social construct whose perception varies depending on the sociocultural context. However, due to the simplicity of MLLMs' written instructions, it is possible to quickly correct cases where a character has been mistakenly categorized as "white," enabling the model to account for more nuanced racial identities.

Similarly, classifications can be easily adjusted for characters whose gender identity cannot be inferred from visual or verbal analysis alone, such as trans, non-binary, or two-spirited characters. The model can also identify other contextual data regarding characters' identities, such as social class or religion, based on elements like settings or costumes, as well as cultural or national affiliations. In our analysis, for instance, the model easily recognized that several characters "speak French with a Québécois accent."

Yet during our testing phase, other technical challenges have been identified, which showed the current limits of AI-augmented video analysis. Particularly, the model still has difficulty accurately synchronizing text with images, which resulted in errors in identifying character names. This means that character identification sometimes needs to be adjusted "manually" through prompting.

While these types of adjustments are relatively easy to make, current limitations of MLLMs nevertheless point to issues when conducting critical or intersectional analyses. As mentioned earlier, the analysis involves vigilant monitoring as it sometimes requires manual data correction, for instance in cases when we need to modify a character's racial, gendered, or sexual identity. Obviously, although *time-consuming*, manual adjustments can potentially make the analysis much more precise and more collaborative; in other words, the analysis can be augmented, rather than merely automated, through a simultaneous consideration of data derived from manual and AI analysis.

However, the need to manually modify identification proves that the training data for MLLMs still often lead them to reproduce fixed, binary, and cisnormative identity markers. For example, following what we explained before about the limitations of current MLLMs on temporal reasoning tasks (see *From Automated to Augmented Analysis: LLMs and MLLMs for Video Analysis*), the model still primarily identifies a character's identity based on speech and facial recognition, rather than on dialog, which is typically what enables a more complex identification of gender, racial identity, or sexual orientation outside of binary frameworks. It is

therefore necessary to circumvent ethical issues related to gender attribution, since “data on gender identities that fall beyond the binary [e.g., trans, non-binary, etc.] is a complex question for big-data research.”³⁴

“I will endeavor to keep the characters straight”: Analytical Constraints of Proprietary Models and Their “Safety” Settings

Additionally, filters and “safety settings” can lead the model to adopt an overly cautious stance when identifying character traits or, conversely, to overinterpret certain behaviors as potentially offensive. Indeed, while MLLMs hold the potential to analyze character behaviors with greater nuance, restrictions imposed by the companies themselves limit the model’s ability to process “prohibited” behaviors, including those flagged as “violent” or “sexual.” Under the guise of “protecting” users, such restrictions can inadvertently suppress content associated with marginalized communities, such as LGBTQ+ discussions on sexuality and gender, which may be deemed “offensive.” Similarly, narratives that critique violence and discrimination might be excluded. These “safety settings” thus not only hinder the model’s capacity to engage with content addressing critical social issues but also restrict its ability to analyze how these issues are represented in films and TV series. To give another example, the present limitations of MLLMs in interpreting narrative nuances mean that an AI might recognize a high frequency of female characters in leadership roles but fail to analyze whether these portrayals reinforce or challenge traditional gender stereotypes. It might also identify recurring motifs across films but miss the intertextual references or the historical context that gives those motifs deeper meaning. In other words, the model often struggles to differentiate between a depiction that endorses a behavior or representation and one that critically interrogates it, resulting in a constrained analysis—even when parameters (e.g., temperature and prompting) are adjusted to permit content categorized as “prohibited” or “unsafe.”

Furthermore, filters designed to limit the analysis of content deemed “sexual” have had an unexpected impact on our pretest analysis. Since most expressions of affection—ranging from hugs, physical closeness, and kisses to more explicit depictions of sexual activity—are categorized as sexual behaviors, the model often relies on these behaviors to infer a character’s sexual orientation. Consequently, this leads to inaccuracies, with some characters being erroneously identified as bisexual or queer simply because their displays of affection are misinterpreted as indicators of sexual attraction.

This issue became evident during our tests. In a teen series like *L’Académie*, displays of affection between female friends occur frequently, which led the model to incorrectly identify the main character, Agathe, as bisexual, despite the series’ story arcs and dialogs clearly establishing her as heterosexual. When engaging further with the model to correct this information, the AI provided a response that highlighted this limitation in analytical capabilities: “My initial interpretation was incorrect, mistaking close female friendships for romantic relationships. [. . .] I am still learning to differentiate between platonic and romantic interactions, and to correctly identify characters and their orientations based on subtle cues.”

While the heteronormativity embedded in most audiovisual productions—and consequently in the training data used to build these models—might suggest that MLLMs would default to

categorizing most characters as heterosexual, our analysis revealed the opposite: Gemini frequently overestimated the presence of queer characters in the series. Complications also arose when attempting to modify certain data. For example, after instructing the model to change the sexual orientation of the main character from “bisexual” to “heterosexual,” the model correctly implemented the adjustment but added “I will endeavor to keep the characters straight”. While this output is open to interpretation, it raises concerns that the model may have interpreted the instruction as a directive to categorize all characters as heterosexual by default, rather than applying the change solely to the specified character. Such a misinterpretation risks introducing a heteronormative bias, potentially skewing the analysis in the opposite direction.

Writing a prompt instructing the model to prioritize narrative and dialog when identifying sexual orientation could potentially mitigate some misidentification issues, although surely not all. This challenge therefore highlights the necessity of approaching AI-assisted analysis as a collaborative process within a genuine framework of interpretive co-creation between MLLMs and researchers. Such a collaborative approach is essential for conducting advanced, nuanced, and potentially intersectional content analyses.

Moreover, given the constraints imposed by “safety settings” on proprietary models—intended to align with sociopolitical sensitivities around representations of sexuality and diversity—the use of an open-source MLLM could offer significant advantages. These include the ability to perform analyses on internal and secure servers and the elimination of restrictive filters that can skew or limit the scope of the analysis. Furthermore, it must be reminded that proprietary AI models and services often come with substantial costs, an issue that open-source multimodal models could help address.³⁵

More generally, our pretests remind us of the importance of crafting meticulously detailed prompts (analytical instructions) and of the persisting challenge of prompt engineering: even minor adjustments in prompts can lead to important variations in results. To prevent this, more nuanced and elaborate instructions beyond the conventional input-output (IO) method are necessary, to ensure effectiveness in analysis.³⁶ Another significant challenge involves standardizing prompts to ensure consistency across a large corpus of audiovisual productions. The decision by some scholars in TV/screen studies to publish their list of prompts and explain their results could, in this sense, prove highly useful for developing common, effective analytical frameworks for studying large samples of series or films.

Methodological and Epistemological Reflection: Toward AI–Human Collaboration

MLLMs present both challenges and opportunities for studying content offered on VoD services. These tools promise unprecedented capabilities in analyzing video content at scale, from identifying patterns across entire “original content” to interpreting recurring motifs. Yet, their true potential and limitations can only be fully understood through active exploration and engagement. In this section, we argue that media scholars must cultivate AI literacy

through hands-on practice with these tools to understand what they can and cannot do, where human interpretation remains indispensable, and where AI-human collaboration can surpass human analysis.

AI literacy involves more than a theoretical understanding of how machine learning models function. Other domains of application have shown that it involves practical, hands-on experience using generative AI tools within one's domain of expertise.³⁷ Media scholars could thus learn a lot from applying MLLM tools to video annotation, as well as thematic and narrative analysis, but in essence, exploration through practice remains essential for understanding both their strengths and limitations. Much like early digital humanities scholars who experimented with computational methods to unlock new textual interpretations, media scholars today need to engage with MLLMs in their workflows to discover what new questions they can ask and what new insights they can generate. That being said, working well with generative AI will involve more than simply mastering technological skills: It will mostly require a reflective and critical approach to digital methodologies, echoing Bernhard Rieder and Theo Röhle's (2023) notion of digital *Bildung*, defined as an informed, self-reflexive, and critical engagement regarding "the actual concepts and methods expressed and made operational through computational methods."³⁸ In this sense, our approach should not be seen as a purely instrumental technique but as a practice encouraging scholars to critically reflect on AI's epistemological potentials and limitations, thereby fostering a deeper methodological awareness and informed analytical creativity within media studies.

Keeping this in mind, we recognize that generative AI models might excel at certain tasks that were previously time-consuming—or even impossible—for human analysts, particularly in pattern recognition and large-scale quantitative analysis. As Esposti and Pescatore have argued, the most important contribution of AI for screen studies is that it "allow[s] for the processing of large amounts of data quickly and efficiently, which would be difficult or impossible to do manually [. . .], leading to new insights and discoveries" (p. 24). For example, with enough access to computation, MLLMs can track the screen or speaking time of various demographic groups across dozens of films or series and generate reports on representation disparities. The manual measurement of screen time or speaking time "can require 10 to 20 times the viewing duration of the materials,"³⁹ which explains why such measurements remain rare in TV/screen studies. MLLMs can also identify visual and auditory motifs, such as recurring symbols, color schemes, or sound effects.

However, as we showed earlier, current AI services remain limited in their ability to interpret cultural and narrative nuances. As highlighted in recent benchmarks like Video-MME (Multi-Modal Evaluation benchmark of MLLMs in Video analysis), the performance of current MLLMs also degrades with video length, pointing to challenges in handling long-form content typical of streaming platforms.⁴⁰ Handling temporal reasoning remains the single most significant barrier to AI video content analysis. Additionally, while current frontier models demonstrate unprecedented capabilities to understand the many modalities of videos, their struggles with long-term dependencies and abstract temporal concepts, such as causality and event sequencing,⁴¹ present challenges to analyze serialized streaming productions or thematic content categorizations. This raises important issues for video analysis, especially since scripted series often prioritize *nonlinear narratives*, where

understanding temporal relationships like flashbacks and parallel storylines is critical. Better temporal reasoning is also essential for understanding *characters' identity evolution* over episodes and seasons. Current MLLMs favor short-term dependencies and lack annotations for long-term temporal relationships or abstract semantic links. These gaps are not only technical but also ethical. If AI models cannot reason effectively about temporal dimensions, they may reinforce biases by favoring surface-level patterns or misrepresenting underrepresented narratives.

Enriched fine-tuning datasets⁴²—with diverse temporal, sociocultural, and demographic annotations—could bridge these gaps. Enhancing these models to better reason over longer narratives could also enable more comprehensive analyses of platform strategies, such as “binge-watchable” series or thematic seriality. Incidentally, these limitations we highlighted point to aspects of the analysis that can benefit more directly from human interpretation, namely, the contextual evaluation of collected data and the assessment of the moral/ideological orientation of story arcs.

By engaging directly with AI tools in their analysis, media scholars can better understand where human interpretation is indispensable and where AI can provide additional or new perspectives. AI-assisted video analysis should therefore evolve toward an “augmented” research framework that emphasizes informed collaboration between human scholars and AI models and services. This human–AI collaboration should free media scholars to focus on interpretive and critical thinking, contextual understanding, cultural awareness, and ethical judgment—capabilities that current AI services lack.

Moreover, through our experimentation with Gemini 1.5 Pro and other AI models, we learned that developing AI literacy in researchers involves a process of exploration-as-play, where scholars test AI tools on familiar tasks to see what they can achieve. Indeed, beyond basic data compilation, our testing phase with Gemini revealed unexpected applications, particularly in facilitating exploratory research approaches.⁴³ One of the most notable outcomes was the model’s ability to identify elements within the datasets that were not initially apparent.

While some findings were merely anecdotal, others—such as specific dialogs, images, and narrative details—yielded useful insights by highlighting aspects that had been overlooked during manual analysis. For instance, one of the central characters in *L'Académie* (Wendy) was associated with a distinct clothing detail: She was identified as “wearing a sushi-patterned sweater.” Initially dismissing this as a “hallucination,” we revisited the footage and confirmed that the character indeed wore such a sweater in one scene. While this detail might appear trivial, it was actually useful for refining the character’s portrayal. Costumes, often overlooked in character analysis, result from deliberate decisions by costume designers and prop managers and can provide valuable insights into a character’s personality traits. In this case, the sweater contributed to Wendy’s depiction as a vibrant and playful individual who seeks to stand out and exhibits a subtly nonconformist attitude. Similarly, the fact that another character in the series wore a “Females for the Future” T-shirt in the first episode did not go unnoticed. Given the extensive length of series, manual analysis inevitably misses certain potentially significant narrative, visual, and sonic details. Therefore, MLLMs can unearth additional insights into character identities or power dynamics.

Another promising method involved requesting the AI model to compose critical and inter-sectional summaries. As an approach increasingly mobilized in critical media studies, intersectionality involves analyzing identity axes as “axes of oppression or privilege”⁴⁴ that are co-constitutive.⁴⁵ In other words, this implies approaching social relations as “interwoven and simultaneous.”⁴⁶ As Melissa Harvey argues, “[b]uilding a movement for a more just world entails an intersectional approach premised not on a single issue but on a broad vision, large-scale collaboration, and democratic inclusion.”⁴⁷ Yet, even with a detailed analytical framework, manual textual analysis cannot fully address all identity variables simultaneously, especially considering their intersections and their complex implications. It is therefore common for intersectional analyses to overlook certain identity axes, even though their consideration would have been relevant. The constraint of manual analysis limits the ability to comprehensively cross-reference data, thus restricting the breadth of conclusions that can be drawn about both advances and ongoing marginalization within audiovisual content.⁴⁸

During our testing phase, we asked Gemini 1.5 Pro to generate such a critical intersectional summary for the series (*L'Académie*), which highlighted class dynamics that had been previously overlooked. The analysis indeed aligned with the manual intersectional analysis previously conducted, notably by evaluating the representation of sexual and racial diversity, while also highlighting issues that we had not focused on, such as those related to social class and bodies. Indeed, while the question of class privilege had somewhat flown “under our radar,” the model expanded the analysis from this perspective: “Agathe visibly benefits from a certain economic privilege, as suggested by her father’s intervention to secure her a private room at *L'Académie*. At this stage, the series does not appear to question the socio-economic disparities between the characters.”

This observation proved highly significant for our analysis, as socioeconomic privileges are frequently represented in teen series but rarely made explicit or critically examined. Here, the model drew attention to a short moment of dialog that supports the analysis of class privilege—something we had overlooked in the manual analysis due to the brevity of the scene. In a similar vein, while highlighting the importance of sorority and feminist issues in this series, the model noted that it seems to reinforce certain gender norms: “*The emphasis on physical appearance, particularly in the scene where the friends comment on (. . .) unshaven legs, reinforces a certain social pressure tied to beauty standards.*”

This testing phase therefore allowed us to determine that the contribution of AI-assisted analysis is not limited to quickly compiling data, as it can also enrich qualitative analysis by adding interesting details that might have been overlooked in manual analysis. Simply put, it can uncover blind spots and support a more dialectical approach when analyzing series. Employing such synthetic analyses across multiple series and then combining them using the AI model could thus yield substantial insights into the representational trends of a specific VoD service, allowing us to better understand its production strategies.

Moreover, AI-augmented analysis, we found, can also serve to incorporate *divergent thinking* into our research, thereby challenging and reassessing certain data points for their relevance. For instance, analysis can be enhanced using Chain of Thought Self-Consistency (CoT-SC) prompting: This technique involves prompting the model to generate various outputs. In other words, the analysis is conducted by mobilizing a “mixture of experts” (e.g., experts

in gender studies, narratology, semiotics, and esthetics) to uncover diverse interpretations of the same TV series or film. In so doing, whether by exploring avenues of analysis and interpretation that we had not identified or by developing multiple analyses simultaneously, AI-augmented analysis holds potential for **ostranenie** (1917),⁴⁹ i.e., an ability to defamiliarize us from our objects of study, to render “[o]ur basic cultural concepts and ways of organizing and understanding cultural data sets foreign to us so that we can approach them anew.”⁵⁰ This has already been highlighted by some researchers in cultural analytics, and our testing phase has led us to identify this “destabilizing” form of qualitative analysis as one of AI’s most interesting affordances to better understand content distributed by VoDs. As Masson argues,

Embracing [digital tools’] potential [for *ostranenie*] requires that one uses one’s tools not to solve existing scholarly problems, but to raise new questions, trigger new ideas, or as a prompt to try out alternative perspectives on the same objects [. . .].⁵¹

This means that, in addition to detailed prompts to address specific analytical questions, it could be useful to create very general prompts asking the model, for instance, to identify recurring (stylistic or narrative) patterns in the corpus and then explore the potential usefulness of these findings. In this way, instead of merely recognizing the polysemy of media texts as an epistemological foundation, we could actively test the heuristic value of a polysemy “in action.” This approach could also enhance our understanding of how diversity is incorporated, in front of and behind the camera, and the prevailing ideological frameworks within a specific platform’s original content.

Conclusion

This article proposed a reflection and an attempt to implement a new method of analyzing audiovisual productions that would be “augmented” using a multimodal AI model. In contrast to older methods of computer-assisted analysis, MLLMs offer new potentials, among other things, the possibility to conduct a more directly collaborative analysis between the AI and researchers.

We obviously acknowledge the importance of addressing ethical concerns regarding commercial AI tools, including exploitative labor practices, copyright, and environmental impacts. However, these specific ethical issues have already been extensively discussed in other recent publications, and addressing them here would have exceeded our paper’s intended scope. Therefore, we deliberately focused on exploring diversity and inclusion—ethical and societal issues that remain underexamined.⁵²

While highlighting the challenges and limitations of AI-assisted analysis—and thus the necessity of establishing a co-analysis method rather than a study conducted entirely by AI, our main objective was to demonstrate how such a method could prove useful in research on VoD services. In other words, while the method described here could apply to any type of audiovisual corpus, we think that its primary utility would be in enabling the analysis of larger samples of content offered on streaming services. This could lead to broader yet more detailed insights into the production strategies of these media industries or to a better understanding of diversity and inclusion initiatives in the streaming era.

Finally, as we attempted to demonstrate, one of the most interesting contributions of AI-augmented analysis might be its potential for *ostranenie*, its ability to partially defamiliarize us from our corpus. Some of the model's responses will undoubtedly prove irrelevant, but others could be unexpectedly useful in approaching certain trends (aesthetic, narrative, and thematic) or motifs in productions distributed on VoD services. In other words, media scholars trying to augment their analysis of video productions might incidentally find their own "sushi sweater."

¹ Stéfany Boisvert is a professor at the École des médias (School of Media) of the University of Quebec at Montreal (UQAM). Codirector of the Laboratory on consumer and media culture in Quebec (LaboPop), she is also a regular member of the Interuniversity Center on Literature and Culture in Quebec (CRILCQ), the Quebec Network in Feminist Studies (RéQEF) and the Research Group on the Emergence and Formation of Media Identities (GRAFIM). Her research focuses on the development of SVOD services in Canada, new forms of serialization, as well as diversity in media productions. She published in journals such as *SERIES*, *Critical Studies in Television*, *Convergence*, *Feminist Media Studies*, and the *Canadian Journal of Film Studies*, and co-edited the book *Une télévision allumée: les arts dans le noir et blanc du tube cathodique* with Viva Paci (2018).

² This article was written with the support of a Fonds de recherche du Québec—Société et culture (FRQSC) Research Support for New Academics, awarded to the first author (Stéfany Boisvert).

³ Amanda D. Lotz, *Media Disrupted: Surviving Pirates, Cannibals, and Streaming Wars* (MIT Press, 2021); Robert Alan Brookey, Jason Phillips, and Tim Pollard, *Triaging the Streaming Wars* (Taylor & Francis Groups, 2023).

⁴ A similar observation has been made by Lotz and Lobato: Amanda D. Lotz and Ramon Lobato, eds., *Streaming Video: Storytelling across Borders* (New York University Press, 2023), 6.

⁵ Lotz and Lobato, *Streaming Video*, 6.

⁶ cf. Adelaida Afilipoaie, Catalina Iordache, and Tim Raats, "The 'Netflix Original' and What It Means for the Production of European Television Content," *Critical Studies in Television* 16, no. 3 (2021): 304–25; Tatiana Hidalgo-Marí, Jesús Segarra-Saavedra, and Patricia Palomares-Sánchez, "In-Depth Study of Netflix's Original Content of Fictional Series: Forms, Styles and Trends in the New Streaming Scene," *Communication & Society* 34, no. 3 (2021): 1–13.

⁷ "Content analysis" is often reserved for quantitative analyses, while "textual analysis" is typically linked to qualitative approaches. However, we use the term "content analysis" throughout this article for all analytical approaches that can be employed to study the content of audiovisual productions.

⁸ Lotz and Lobato, *Streaming Video*, 7–8.

⁹ Lev Manovich, "Cultural Analytics, Social Computing and Digital Humanities," in *The Datafied Society*, ed. Mirko Tobias Schäfer and Karin van Es (Amsterdam University Press, 2017), 60.

¹⁰ Christian Gosvig Olesen, "Towards a 'Humanistic Cinematics'?" in *The Datafied Society*, ed. Mirko Tobias Schäfer and Karin van Es (Amsterdam University Press, 2017), 39.

- ¹¹ Giogio Avezzi and Marta Rocchi, eds., *Audiovisual Data: Data-Driven Perspectives for Media Studies* (Bologna: Media Mutations Publishing, 2023), 11.
- ¹² David Doukhan, Géraldine Poels, Zohra Rezgui, and Jean Carrière, "Describing Gender Equality in French Audiovisual Streams with a Deep Learning Approach," *VIEW* 7, no. 14 (2018): 103–122; David Doukhan, Zohra Rezgui, Géraldine Poels, and Jean Carrière, "Estimer automatiquement les différences de représentation existant entre les femmes et les hommes dans les médias," *DAHLIA* (June 2019); Laetitia Biscarrat, David Doukhan, and Ange Richard, "Quantifier les inégalités de genre dans les médias: approches computationnelles," *3e Congrès International de l'Institut du Genre* (July 2023).
- ¹³ Marta Rocchi and Guglielmo Pescatore, "Modeling Narrative Features in TV Series: Coding and Clustering Analysis," *Humanities and Social Sciences Communications* 9, no. 1 (2022): 1–11.
- ¹⁴ Monika Bednarek, "On the Usefulness of the Sydney Corpus of Television Dialogue as a Reference Point for Corpus Stylistic Analyses of TV Series," in *Telecinematic Stylistics*, ed. Christian Hoffmann and Monika Kirner-Ludwig (Bloomsbury, 2020), 39–61.
- ¹⁵ Xavier Bost, Serigne Gueye, Vincent Labatut, Martha Larson, Georges Linarès, Damien Malinas, and Raphaël Roth, "Remembering Winter Was Coming: Character-Oriented Video Summaries of TV Series," *Multimedia Tools and Applications* 78, no. 24 (2019): 35373–99.
- ¹⁶ Bost et al., "Remembering Winter Was Coming," 35374.
- ¹⁷ e.g., Danielle Hitch, "Artificial Intelligence Augmented Qualitative Analysis: The Way of the Future?" *Qualitative Health Research* 34, no. 7 (June 2024): 595–606.
- ¹⁸ Dimitri Coelho Mollo and Raphaël Millière, "The Vector Grounding Problem," *arXiv preprint* (2023).
- ¹⁹ Jiaqi Wang et al., "A Comprehensive Review of Multimodal Large Language Models: Performance and Challenges across Different Tasks," *arXiv preprint* (2024).
- ²⁰ Due to the fast-paced evolution of multimodal AI models and tools, essential technical research often appears only on *arXiv*, as immediate publication and citation of new papers have become the standard practice in these fields.
- ²¹ Ashish Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems*, *arXiv* (2017).
- ²² Ethan Mollick, *Co-Intelligence: Living and Working with AI* (Penguin Random House, 2024).
- ²³ Wang et al., "A Comprehensive Review of Multimodal Large Language Models"; Yin Shukang et al., "A Survey on Multimodal Large Language Models," *National Science Review* 11, no. 12 (December 2024).
- ²⁴ For instance, older systems for action recognition or object detection in videos used techniques like optical flow analysis, convolutional neural networks (CNNs) for spatial data, and recurrent neural networks (RNNs) for sequential data. While effective for specific tasks, these methods are limited by their reliance on predefined rules and often struggled to generalize across tasks or adapt to new domains.
- ²⁵ Yin Shukang et al., "A Survey on Multimodal Large Language Models."

- ²⁶ Wang et al., “A Comprehensive Review of Multimodal Large Language Models.”
- ²⁷ Xi Ding and Lei Wang, “Do Language Models Understand Time?” *arXiv preprint* (2024).
- ²⁸ Ding and Wang, “Do Language Models Understand Time?”; Wang et al., “A Comprehensive Review of Multimodal Large Language Models”; Wang et al., “VisionLLM: Large Language Model is also an Open-Ended Decoder for Vision-Centric Tasks,” *Advances in Neural Information Processing Systems* 36, no. 2688 (2023).
- ²⁹ For a detailed presentation of this research project and the various textual analyses previously conducted, see Stéfany Boisvert, “Streaming Diversité: Exploring the Plurality of Representations within French-Language Scripted Series on Canadian SVOD Services,” *Convergence* 30, no. 4 (2024): 1510–1528.
- ³⁰ The episodes of *L'Académie* were available on YouTube at the time of our pretests and have since been deleted (not only from our computer but also on YouTube). Analyzing TV series may, depending on the laws in effect in each country, require permission from production companies to use digital copies for research purposes.
- ³¹ The analysis was conducted during an exploratory phase. We therefore want to ensure that readers do not interpret these explanations as a “step-by-step guide” or a fixed set of actions to follow but rather as methodological suggestions.
- ³² For example, David L. Morgan, “Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT,” *International Journal of Qualitative Methods*, no. 22 (2023).
- ³³ Wenxuan Zhang et al., “Sentiment Analysis in the Era of Large Language Models: A Reality Check,” *Findings of the Association for Computational Linguistics: NAACL 2024* (Mexico Association for Computational Linguistics): 3881–3906.
- ³⁴ Rosa Barotsi, Mariagrazia Fanchi and Matteo Tarantino, “Constructing an Open, Participatory Database on Gender (In)Equality in the Italian Film Industry,” in Giorgio Avezzù and Marta Rocchi, eds., *Audiovisual Data: Data-Driven Perspectives for Media Studies* (Bologna: Media Mutations Publishing, 2023), 99.
- ³⁵ While we have not yet tested an advanced open-source model, options such as CogVLM2 show considerable promise for future research endeavors.
- ³⁶ Jason Wei et al., “Chain of Thought Prompting Elicits Reasoning in Large Language Models,” *arXiv* (2022); Shunyu Yao et al., “Tree of Thoughts: Deliberate Problem Solving with Large Language Models,” *arXiv* (2023).
- ³⁷ Mollick, *Co-Intelligence*.
- ³⁸ Bernhard Rieder and Theo Röhle, “Digital Methods. From Challenges to Bildung,” in *The Datafied Society*, ed. Mirko Tobias Schäfer and Karin van Es (Amsterdam University Press, 2017), 117.
- ³⁹ Doukhan et al., “Estimer automatiquement,” n. pag.
- ⁴⁰ Chaoyou Fu et al., “Videomme: The First-Ever Comprehensive Evaluation Benchmark of Multi-Modal LLMs in Video Analysis,” *arXiv preprint* (2024).
- ⁴¹ Ding and Wang, “Do Language Models Understand Time?”
- ⁴² Fine-tuning datasets can provide targeted knowledge that aligns and refine an LLM’s capabilities with specific analytical goals, making it better suited for tasks in a specific domain (e.g., TV studies).

- ⁴³ This exploratory process, we think, will be crucial for identifying the future boundaries between human and machine expertise. Beyond practical applications, scholars must not only be aware of the data biases that AI models inherit from training datasets, but they must also discover how AI can be used responsibly, and how to leverage AI against harmful biases or misinformation.
- ⁴⁴ Alison Harvey, *Feminist Media Studies* (Polity, 2020), 20.
- ⁴⁵ Patricia Hill Collins, *Intersectionality as Critical Social Theory* (Duke University Press, 2019); Sirma Bilge, “La pertinence de Hall pour l’étude de l’intersectionnalité,” *Nouvelles pratiques sociales* 26, no. 2 (2014); Erica B. Edwards and Jennifer Esposito, eds., *Intersectional Analysis as a Method to Analyze Popular Culture: Clarity in the Matrix* (New York: Routledge, 2019).
- ⁴⁶ Marlène Coulomb-Gully “Féminin/masculin: question(s) pour les SIC”. *Questions de communication* 1, no. 17 (2010): 169–94.
- ⁴⁷ Harvey, *Feminist Media Studies*, 19.
- ⁴⁸ Stéfany Boisvert, “Streaming Diversité: Exploring Representations within French-Language Scripted Series on Canadian SVOD Services”, 1524.
- ⁴⁹ Viktor Shklovsky, “Art, as Device,” *Poetics Today* 36, no. 3 (2015): 151–74.
- ⁵⁰ Manovich, “Cultural Analytics,” 67.
- ⁵¹ Eef Masson, “Humanistic Data Research: An Encounter between Epistemic Traditions,” in *The Datafied Society*, ed. Mirko Tobias Schäfer and Karin van Es (Amsterdam University Press, 2017), 34.
- ⁵² Environmental impacts of AI should arguably be considered more broadly across the streaming and digital industries as a whole, and not solely by focusing on the new “kids on the block” (i.e., AI tools). Indeed, streaming itself is highly energy-intensive—an environmental cost often overlooked in public discourse due to the widespread societal acceptance and popularity of streaming technology. Cf. Jean-Samuel Beuscart, Samuel Coavoux, and Jean-Baptiste Garroq, “Listening to Music Videos on YouTube,” *Journal of Consumer Culture* 23, no. 3 (2023): 654–71; Laura U. Marks and Radek Przedpelski, “The Carbon Footprint of Streaming Media: Problems, Calculation, Solutions,” in Pietari Kääpä and Hunter Vaughan, *Film and Television Production in the Age of Climate Crisis* (Palgrave Macmillan, 2022).

Bibliography

- Afilipoaie, Adelaida, Catalina Iordache, and Tim Raats. “The ‘Netflix Original’ and What It Means for the Production of European Television Content.” *Critical Studies in Television* 16, no. 3 (2021): 304–25.
- Avezzù, Giogio and Marta Rocchi, eds. *Audiovisual Data: Data-Driven Perspectives for Media Studies*. Media Mutations Publishing, 2023.
- Barotsi, Rosa, Mariagrazia Fanchi, and Matteo Tarantino. “Constructing an Open, Participatory Database on Gender (In)Equality in the Italian Film Industry.” In *Audiovisual Data: Data-Driven Perspectives for Media Studies*, edited by Giogio Avezzù and Marta Rocchi. Media Mutations Publishing, 2023.

- Beuscart, Jean-Samuel, Samuel Coavoux, and Jean-Baptiste Garrocq. "Listening to Music Videos on YouTube: Digital Consumption Practices and the Environmental Impact of Streaming." *Journal of Consumer Culture* 23, no. 3 (2023): 654–71.
- Bilge, Sirma. "La pertinence de Hall pour l'étude de l'intersectionnalité." *Nouvelles pratiques sociales* 26, no. 2 (2014): 62–81.
- Biscarrat, Laetitia, David Doukhan, and Ange Richard. "Quantifier les inégalités de genre dans les médias: approches computationnelles." » 3e Congrès International de l'Institut du Genre: No(s) Futur(s) Genre: bouleversements, utopies, impatiences (Jul 2023): (hal-04906745)
- Boisvert, Stéfany. "Streaming Diversité: Exploring Representations within French-Language Scripted Series on Canadian SVOD Services." *Convergence* 30, no. 4 (2024): 1510–28.
- Brookey, Robert Alan, Jason Phillips, and Tim Pollard. *Triaging the Streaming Wars*. Taylor & Francis Groups, 2023.
- Coelho Mollo, Dimitri, and Raphaël Millière. "The Vector Grounding Problem." *arXiv preprint* (2023). <https://arxiv.org/abs/2304.01481>
- Coulomb-Gully, Marlène. "Féminin/masculin: question(s) pour les SIC." *Questions de communication* no. 17 (2010): DOI : 10.4000/questionsdecommunication.383
- Ding, Xi, and Lei Wang. "Do Language Models Understand Time?". *arXiv preprint: arXiv:2412.13845*.2024
- Doukhan, David, Géraldine Poels, Zojra Rezgui, and Jean Carrive. "Describing Gender Equality in French Audiovisual Streams with a Deep Learning Approach." *VIEW* 7, no. 14 (2018): 103–22.
- David Doukhan, Zohra Rezgui, Géraldine Poels and Jean Carrive. "Estimer automatiquement les différences de représentation existant entre les femmes et les hommes dans les médias." journée DAHLIA (June 2019). <https://hal.science/hal-02168148/document>
- Edwards, Erica B., and Jennifer Esposito, eds. *Intersectional Analysis as a Method to Analyze Popular Culture: Clarity in the Matrix*. Routledge, 2019.
- Esposti, Mirko Degli, and Guglielmo Pescatore. "Exploring TV Seriality and Television Studies through Data-Driven Approaches." In *Audiovisual Data: Data-Driven Perspectives for Media Studies*, edited by Giggio Avezzù and Marta Rocchi. Media Mutations Publishing, 2023.
- Chaoyou Fu, Yuhan Dai, Yongdong Luo, Lei Li, Shuhuai Ren, Renrui Zhang, Zihan Wang, Chenyu Zhou, Yunhang Shen, Mengdan Zhang, et al. "Videomme: The First-Ever Comprehensive Evaluation Benchmark of Multi-Modal LLMs in Video Analysis." *arXiv preprint* (2024): arXiv:2405.21075.
- Harvey, Alison. *Feminist Media Studies*. Polity, 2020.
- Hidalgo-Marí, Tatiana, Jesús Segarra-Saavedra and Patricia Palomares-Sánchez. "In-Depth Study of Netflix's Original Content of Fictional Series: Forms, Styles and Trends in the

- New Streaming Scene.” *Communication & Society* 34, no. 3 (2021): 1–13. <https://doi.org/10.15581/003.34.3.1-13>
- Hitch, D. “Artificial Intelligence Augmented Qualitative Analysis: The Way of the Future?” *Qualitative Health Research* 34, no. 7 (June 2024): 595–606. <https://doi.org/10.1177/10497323231217392>.
- Hill Collins, Patricia. *Intersectionality as Critical Social Theory*. Duke University Press, 2019.
- Lotz, Amanda D. *Media Disrupted: Surviving Pirates, Cannibals, and Streaming Wars*. MIT Press, 2021.
- Lotz, Amanda D., and Ramon Lobato, eds. *Streaming Video: Storytelling Across Borders*. New York University Press, 2023.
- Manovich, Lev. “Cultural Analytics, Social Computing and Digital Humanities.” In *The Datafied Society*, edited by Mirko Tobias Schäfer and Karin van Es. Amsterdam University Press, 2017.
- Marks, Laura U., and Radek Przedpelski, “The Carbon Footprint of Streaming Media: Problems, Calculation, Solutions.” In *Film and Television Production in the Age of Climate Crisis: Towards a Greener Screen*, edited by Pietari Kääpä and Hunter Vaughan. Palgrave Macmillan, 2022.
- Masson, Eef. “Humanistic Data Research: An Encounter between Epistemic Traditions.” In *The Datafied Society*, edited by Mirko Tobias Schäfer and Karin van Es. Amsterdam University Press, 2017.
- Mollick, Ethan. *Co-Intelligence: Living and Working with AI*. Penguin Random House, 2024.
- Morgan, David L. “Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT.” *International Journal of Qualitative Methods* no. 22 (2023). <https://doi.org/10.1177/16094069231211248>
- Olesen, Christian Gosvig. “Towards a ‘Humanistic Cinematics?’” In *The Datafied Society*, edited by Mirko Tobias Schäfer and Karin van Es. Amsterdam University Press, 2017.
- Rieder, Bernhard, and Theo Röhle, “Digital Methods: From Challenges to Bildung.” In *The Datafied Society*, edited by Mirko Tobias Schäfer and Karin van Es. Amsterdam University Press, 2017.
- Rocchi, Marta, and Guglielmo Pescatore. “Modeling Narrative Features in TV Series: Coding and Clustering Analysis.” *Humanities and Social Sciences Communications* 9, no. 1 (2022): 1–11.
- Shklovsky, Viktor. “Art, as Device.” *Poetics Today* 36, no. 3 (2015): 151–74.
- Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. “A Survey on Multimodal Large Language Models.” *National Science Review* 11, no. 12 (December 2024). <https://doi.org/10.1093/nsr/nwae403>

- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. "Attention Is All You Need." *Advances in Neural Information Processing Systems*, *arXiv* (2017). <https://arxiv.org/abs/1706.03762>
- Wang, Jiaqi, Jiaqi Wang, Hanqi Jiang, Yiheng Liu, Chong Ma, Xu Zhang, Yi Pan, Mengyuan Liu, Peiran Gu, Sichen Xia, Wenjun Li, Yutong Zhang, Zihao Wu, Zhengliang Liu, Tianyang Zhong, Bao Ge, Tuo Zhang, Ning Qiang, Xintao Hu, Xi Jiang, Xin Zhang, Wei Zhang, Dinggang Shen, Tianming Liu, and Shu Zhang. "A Comprehensive Review of Multi-modal Large Language Models: Performance and Challenges Across Different Tasks." *arXiv* (2024): arXiv:2408.01319.
- Jason, Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou, "Chain of Thought Prompting Elicits Reasoning in Large Language Models." *arXiv* (2022). <https://arxiv.org/abs/2201.11903>
- Wang, Wenhai, Zhe Chen, Xiaokang Chen, Jiannan Wu, Xizhou Zhu, Gang Zeng, Ping Luo, Tong Lu, Jie Zhou, Yu Qiao, and Jifeng Dai. "VisionLLM: Large Language Model is also an Open-Ended Decoder for Vision-Centric Tasks." *Advances in Neural Information Processing Systems* 36, no. 2688 (2023): 61501–13.
- Yao, Shunyu, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. "Tree of Thoughts: Deliberate Problem Solving with Large Language Models." *arXiv* (2023). <https://arxiv.org/abs/2305.10601>
- Zhang, Wenxuan, Yue Deng, Bing Liu, Sinno Pan, and Lidong Bing. "Sentiment Analysis in the Era of Large Language Models: A Reality Check." In *Findings of the Association for Computational Linguistics: NAACL 2024*, 3881–3906. Mexico Association for Computational Linguistics.

