A Large-Scale Sentiment Analysis of AI vs. Conventional Music Videos

Stéphane Gauvin*
Department of Marketing
Université Laval
Québec, CANADA G1V 0A6
stephane.gauvin@fsa.ulaval.ca

Abstract

Prior experimental research shows a robust AI-authorship bias, where consumers express lower appreciation for creative works when they are attributed to AI. Our analysis of music videos posted on YouTube Official Artist Channels (OAC) suggests a more nuanced pattern. We analyzed a corpus of more than 100 million comments left on over 4 million videos and found that comments posted on AI in Music (AIM) videos were, on average, more positive than comments left on conventional OAC videos. A notable exception was AI covers, where sentiment was significantly lower. Furthermore, although comments explicitly referencing AI were substantially less positive than those that did not, average sentiment towards AI remained nonetheless positive. Our findings suggest that audience resistance to AI music depends less on the mere presence of the technology, and more on whether AI is perceived as facilitating the creation of original music versus imitating or plagiarizing existing human works.

1 INTRODUCTION

Artificial intelligence (AI) is the latest in a long line of technologies to provoke ambivalence within the music community. Innovations such as the Moog synthesizer, the drum machine, and Auto-Tune were initially met with resistance. Eventually, however, both artists and the public embraced these technologies as tools rather than threats to the art of music (for reviews, see Diaz, 2009; Gibb, 2019; Katz, 2010; Leff, 2004; Tapper, 2024).

The current debate surrounding AI in Music (AIM) is arguably more profound. Because generative technologies have the potential to fundamentally disrupt the creative process, many industry stakeholders express more concern than enthusiasm.

Few contemporary artists have fully integrated AI into their creative workflows, even though there are notable exceptions. Grimes, for example, used XR technology to replace traditional chroma keying with AI-generated real-time visual effects. She has also developed a voice model, distributed via Elf.tech, allowing creators to use her vocal twin in exchange for a 50-50 revenue-sharing agreement (Hoppenjans, 2025).

Other voices express ambivalence or outright opposition. In 2024, the Artist Rights Alliance published an open letter that, while acknowledging AI's "enormous potential to advance human creativity," raised alarms "against the predatory use of AI to steal professional artists' voices and likenesses, violate creators' rights, and destroy the music ecosystem" (Artist Rights Alliance, 2024). This sentiment is echoed by emerging artists like Violet Maye Grohl, who recently expressed her desire for authentic human connection through music and her frustration at being "in competition with a computer," emphasizing the "magical experience" of human-to-human studio collaboration (Kerrang! Radio, 2026), and by students at the Berklee College of Music who have protested against courses on AI technologies, claiming that the value of learning from talented professionals is cheapened by the presence of "plagiarized content" (Pressman, 2026).

Academic research also shows overwhelming evidence of negative AI bias among listeners. However, results are more nuanced when it comes to the relative appreciation of AI works. In experiments where subjects are blind to authorship, findings are mixed: some listeners prefer AI works (Ferreira et al., 2023), others find them inferior (Zhou & Kawabata, 2023), and some find AI virtually indistinguishable from human efforts (Zlatkov et al., 2023).

To date, the bulk of research contrasting AI and human compositions has relied on controlled experiments using small subject pools, a methodology that is perfectly adequate to study the presence of label bias, but poses significant limitations when comparing AI- and human-created works: researchers then face the intractable challenge of sampling works that are representative of the actual state of the art, either AI- or human-generated.

Our study addresses these limitations with data drawn from a large corpus of real-world behavior. We analyze user sentiment across more than 107 million comments left on more than 4.6 million videos published on YouTube Official Artist Channels during the past year. Complementing small-scale controlled experiments with a large-scale analysis of observational data provides a novel perspective on how audiences actually perceive and engage with AI-generated or AI-altered content. We address the self-selection issue and its implications in the discussion section.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature; Section 3 describes our data and sentiment analysis models; Section 4 presents our empirical findings alongside qualitative vignettes that illustrate the diversity of the AI-music landscape; Section 5

*We wish to acknowledge the invaluable contribution of our research assistants, Jennifer Neault and Yiwei Zhou

concludes with a discussion of key insights and directions for future research.

2 RELEVANT LITERATURE

Research at the intersection of AI and music is largely experimental and can be categorized by its primary objectives: (1) investigating *authorship bias* by comparing the evaluation of identical works attributed either to AI or to humans, and (2) examining the *comparative appreciation* of distinct works, some actually created by AI and others by humans. Tables 1 and 2 list illustrative experimental studies in these areas.

2.1 Experimental research on AI-authorship labels

The bulk of recent research on authorship bias shows that identical or similar works receive lower evaluations when attributed to AI rather than humans (Table 1). Only two studies report null or context-dependent effects: Chia et al. (2025) and Tigre Moura & Maw (2021). Chia et al. speculate that their results come from the fact that they tested pop-music on university students, while Tigre Moura & Maw's manipulation involved priming subjects with a description of the creative process. These exceptions suggest that negative sentiment towards AI-generated music may not be systematic.

2.2 Experimental research on perceived quality

A second body of research evaluates the comparative quality of actual AI-generated compositions against human-made counterparts (Table 2). This literature yields mixed results. Crucially, aesthetic appreciation appears highly dependent on the underlying generative architecture. For instance, Ferreira et al. (2023) found that while pre-transformer models were rated markedly lower than human composers, advanced transformer-era models actually outscored human-created music. Among the other comparative studies listed in Table 2, findings remain split between lower appreciation towards AI, and inconclusive differences. Ultimately, this demonstrates that the intrinsic quality of AIM can be perceived as inferior or superior to human work.

2.3 Survey research

Table 3 lists published academic survey research on this topic. The results are either inconclusive or indicative of a negative AI bias. In addition to these academic studies, we found two large-scale industry surveys. A survey conducted by the Music Audience Exchange (MAX, 2026) on 14,000 respondents reports that 40% felt disappointed and 37% felt deceived when informed that a song was AI-generated. Conversely, 30% expressed appreciation for AI innovation in music. The other, a Deezer/IPSOS survey of 9,000 respondents from 8 countries, found that "97% couldn't tell the difference between fully AI-generated music and human-made music in a blind test with two AI songs and one real song." (Deezer, 2025)

2.4 Summary

Taken as a whole, the extant literature reveals a pervasive negative bias against AI. This bias is consistent with the *expressionist* concept of art, which defines art primarily as the sharing of human emotions. From this perspective, the observed bias is arguably less about AI itself than it is about fundamental beliefs regarding the meaning of art. A similar negative bias against animal art is well documented (Łagodzka, 2024). When the focus is on the quality of the works themselves, the findings are mixed. As Tigre Moura and Maw (2021) observe:

[...there is] one overall finding of great relevance: as long as listeners enjoy a song, the awareness that the composition has been created by AI without human participation has a minor effect on their perception of the music piece. (p. 143)

3 DATA AND MODELS

We have assembled a comprehensive database containing metadata for most music videos published on YouTube's Official Artist Channels (OAC)[1]. Because YouTube does not provide a definitive public directory of OACs, the list of channels has been built through search. We estimate that our data account for approximately 95% of the entire population of OAC videos according to the random search technique developed by Zhou et al. (2011).

Just as there is no formal registry for OACs, there is no standardized method for identifying AI-generated or AI-altered music videos. We operationalized different types of AIM, based on video descriptions:

- **Altered**: Videos containing the *Altered or Synthetic content* disclosure mandated by YouTube. Originally designed to combat deepfakes, this label distinguishes realistic AI representations from clearly artificial CGI.
- **#aimusic**: Videos containing the *#aimusic* hashtag. Unlike mandated disclosures, publishers use this at their discretion to improve search and visibility within the recommendation algorithm.
- **Derivative**: Videos featuring tags like *#aicover, #aivoice,* or *#aicoversong*. These denote AI-driven "reimaginings" or synthetic voice models.
- **Audio tags**: Videos containing specific generator tags (*#suno, #udio, #stableaudio*, etc.), indicating the audio was generated entirely or in part by AI tools.
- **Video tags**: Videos containing visual generator tags (*#lumaai, #runwaygen3, #soraai*, etc.), denoting that visuals were generated via AI rather than traditional camera work.

A video featuring tags from multiple categories (e.g., an AI-generated cover song featuring AI-generated visuals) is coded as belonging to all applicable categories. A pairwise correlation analysis of these binary indicators showed minimal overlap across the corpus, with all pairwise Phi correlations ($\phi < 0.25$), ensuring that multicollinearity does not compromise the estimation of independent effects in our statistical models. To ensure reproducibility, the exhaustive list of tags used for each category is provided in Appendix A.

Table 4 presents the distribution of observations across these categories. Metrics are reported for *All Time*, from YouTube's inception in 2005, and for the *Past Year*, covering the 12-month period from April 1, 2025 up to March 31, 2026. The metrics include:

- **Number of videos** is the total count of publicly listed videos at the time of this writing.

[1] See https://bit.ly/AIM-2026-01

Table 1: Experimental research on identical works with different labels.

Source	Stimulus	Brief Description	N	Bias
Raj et al. (2026)	Text	16 experiments on impact of ChatGPT vs. human labels.	3,590	−
Agudo et al. (2022)	Music	Sensitivity/emotion ratings across human/AI label groups.	250	−
Ansani et al. (2025)	Music	Crossover design; identical classical piano labeled human or AI.	120	−
Chia et al. (2025)	Music	Blind listening; identical pop songs randomly labeled human/AI.	64	n.s.
Hong et al. (2022)	Music	Impact of anthropomorphism on AI musician acceptance.	222	−
Hong et al. (2022b)	Music	Expectancy Violation Theory applied to AI classical/EDM ratings.	299	−
Shank et al. (2022)	Music	Identical excerpts; participants told composer was human or AI.	451	−
Stammer et al. (2025)	Music	Test 15 music genres using AI/Soundtrack/Human labels.	178	−
Tigre Moura (2021)	Music	Music tracks described as AI-composed vs. human-composed.	86	n.s.
Bellaiche et al. (2023)	Visual	Measured beauty/worth of identical paintings labeled human vs. AI.	150	−
Chiarella et al. (2022)	Visual	Authorship labels and psychophysiological data for abstract art.	110	−
Di Dio et al. (2023)	Visual	Authorship knowledge effect on beauty/liking (blind vs. primed).	95	−
Messingschlager (2023)	Visual	Artworks labeled as fictional human artist vs. creative AI.	381	−
Neef et al. (2024)	Visual	AI-generated images labeled as human-made vs. AI-generated.	952	−
White & Leon (2026)	Various	Multi-study (5) measurement of awe/empathy via art labels.	1,598	−

Table 2: Experimental research on different works.

Source	Stimulus	Brief Description	N	Bias
Adkins et al. (2023)	Music	Blind study of loops; rating originality/liking (method hidden).	31	−
Déguernel et al. (2022)	Music	Traditional music practitioners guessed authorship (blind).	26	−
Ferreira et al. (2023)	Music	Musicality test for five excerpts; AI vs. human pass rate.	117	+
Iskandar et al. (2025)	Music	Ratings after informing half of participants of origin.	122	n.s.
Pasquier et al. (2016)	Music	Naïve rating of string quartets on emotional/natural dimensions.	60	n.s.
Ramos & Vent. (2025)	Music	Quality/creativity test; AI origin revealed post-initial evaluation.	50	n.s.
Zlatkov et al. (2023)	Music	Preference ranking of 5 human vs. 5 computer-composed excerpts.	163	n.s.
Chamberlain (2018)	Visual	Categorization of 60 images to see if features drove aesthetics.	65	n.s.
Demmer et al. (2023)	Visual	2x2 grid design testing intentionality regardless of origin.	48	−
Zhou & Kawa. (2023)	Visual	Eye-tracking during blind evaluations; authorship revealed after.	34	−

Table 3: Published academic survey research on AI-generated music.

Source	Stimulus	Brief Description	N	Bias
Candusso (2024)	Music	Public sentiment/openness to AI applications in music.	113	n.s.
Correia (2024)	Music	Survey on AI music growth, consumption habits, and comfort.	183	−
Li (2024)	Music	Social perception/acceptance of AI music within a specific fan base.	441	n.s.
Tigre Moura (2021)	Music	General stance on AI music concept from public and professionals.	446	−

- **Views / video** is the average number of views received by a video.
- **Comments / video** is calculated similarly.
- **Sentiment** ranges from -1 to 1, calculated as the average (mean) difference between the probability that a comment is positive and the probability that it is negative. A sentiment of 1 would indicate universally positive comments, while -1 would indicate universally negative comments.
- To isolate the specific impact of AI awareness on audience reception, we further subdivided sentiment analysis. The **AI** subset calculates sentiment exclusively from comments where the specific word *AI* is present, whereas the **No AI** subset calculates sentiment when a comment makes no explicit reference to *AI*. We note that this strict lexical filtering serves as a conservative lower bound; it undercounts the true volume of AI-aware comments by omitting tool-specific references (e.g., "Suno", "generative model") as well as localized terminology across the 82 languages in our corpus.

Sentiment was inferred via mmBERT models (Marone et al., 2025), trained on a stratified corpus of 90,000 comments randomly selected across six languages (Arabic, Chinese, English, French, Spanish, and Russian). The training corpus was built with consensus labels generated by Gemini 3.1 Pro and Qwen 3.5, followed by human validation.

To mitigate the risk of positivity bias due to the nature of music video comments, where the positive-to-negative ratio is approximately 20:1, we employed a two-stage training process. First, we trained an initial model on 100,000 randomly selected machine-labeled comments. We applied these preliminary models to classify 1 million comments and used the resulting inferences to build a balanced sample of predicted positive, negative, and neutral comments. This balanced sample formed the 90,000-comment training set for our final models. We hired teams of three fluent graduate research assistants for each of the six languages used in training, tasked with labeling 3,000 comments per target language (18,000 total). Retaining only the instances where the coders agreed, this process resulted in a high-agreement validation dataset of 11,549 comments with consensus human labels, which we used to evaluate the accuracy of the final mmBERT models.

We used two separate binary models, one for positive and the other for negative sentiment, to fine-tune the inference probability thresholds. Empirical testing resulted in an optimal probability threshold of 0.5 for positive inferences and 0.99 for negative inferences; comments falling under these thresholds were classified as neutral or undetermined.

Finally, to ascertain the zero-shot cross-lingual transfer accuracy, we trained a pair of mmBERT models in five languages and evaluated the accuracy (F1-score) on the Chinese corpus that had not been used in training. This resulted in an F1-score of 0.96, a confirmation of the impressive cross-lingual transfer capability of multilingual encoder-based architectures. Upon retraining the models with the full corpus, in-language F1-scores ranged from a low of 0.97 (Arabic) to a high of 0.99 (French and Spanish). It should be noted that these F1-scores represent the models' performance on high-consensus validation data. Excluding neutral or ambiguous cases where human experts lacked consensus, these scores reflect the model's accuracy in identifying clear sentiment signals.

The final sentiment values reported in Table 4 were generated across all comments in our dataset, spanning 82 different languages.

4 RESULTS

Our discussion of the results focuses on the data covering the past year, when AIM took off: roughly one-fifth of all OAC videos in our database have been published in the past year, whereas nearly 90% of the videos in the *Altered* category and approximately 65% in the *#aimusic* category were published in the past year. AIM is still a marginal phenomenon, accounting for barely 0.22% of OAC video production from our past-year sample.

AIM videos do not appear to be hugely popular, but the differences in average views per video are misleading. The mean number of views received by a typical OAC video is much larger than that of an *AIM* video, but this is due to a few runaway OAC hits with views in excess of 1 billion, something that has yet to happen to an AIM video. The distribution of OAC views per video shows an extreme right-skewness (skewness = 234), making it unreasonable to use mean values as a tool to compare distributions. A similar situation affects the average number of comments, as the OAC mean is inflated by massive historical outliers such as BTS's "Dynamite" with over 10 million comments.

Table 5 presents several distribution indicators for the types of AIM identified heretofore. In order to facilitate interpretation, we include a global AIM category that incorporates all types of AIM. In addition to the means, Table 5 shows the median, a more appropriate indicator of central tendency for asymmetric distributions, Cliff's δ, a measure of distributions overlap, and P_{sup}, the probability that a random observation drawn from the OAC distribution is larger than one drawn from the category of interest (Cliff 1993). Results of the Mann-Whitney test are omitted since they systematically returned values close to zero due to the large sample sizes.

We can see that the ratio of the OAC vs AIM mean views is close to 3, whereas it drops to 1.4 if we consider the medians. Moreover, Cliff's δ means that both distributions show considerable overlap (a value of 1 would mean that all OAC values are larger than AIM and -1 would indicate the reverse, with a value of 0 meaning perfect overlap). The P_{sup} of 0.547 is perhaps more intuitive. It means that in terms of viewership, a randomly selected OAC video has only a marginal advantage over an AIM video, roughly equivalent to a coin flip.

We also find two AIM categories with viewership odds that are better than OAC's: *Altered* and *Derivative*, where medians are 6 and 2 times larger than OAC's, respectively.

AIM being a recent phenomenon, we ran the same analyses on videos published during the month of March 2026. We can see that the distribution of AIM views has become practically identical to that of OAC, and that all AIM categories are edging closer to OAC, including *Altered*, which dominates OAC, with perhaps the exception of *#aimusic* where the median has dropped considerably.

In summary, AIM appears to be moving closer to OAC as far as viewership is concerned. As of March 2026 there is essentially no difference between the two distributions, and the *Altered* and *Derivative* categories outperform typical OAC releases.

Third, we can see that the average sentiment of comments that include the word 'AI' is systematically lower than the average sentiment of comments that do not include this word. Yet, the sentiment of comments explicitly referring to AI remains positive (recall that this value is the difference between the probability that a comment is positive vs negative). Even in the lowest-performing category (a sentiment of 0.253 for OAC videos published last year), positive sentiment significantly outweighs negative sentiment. If we exclude neutral and ambiguous comments to look strictly at the balance of polarized opinions, a sentiment of 0.253 indicates that there are roughly 1.7 positive comments for every negative comment explicitly discussing AI.

Fourth, we can see that the sentiment across all AI categories, with the notable exception of *Derivative*, stands above that of OAC videos.

We used a Generalized Linear Model to assess the significance of these differences, where OAC is specified as the intercept to facilitate interpretation.

Table 6 shows the results obtained by a main-effect model and a first-order interaction model. It should be noted that the low McFadden R2 is not surprising: the purpose of the GLM is not to explain the total variance of user sentiment (which is inferred from the text content using the fine-tuned mmBERT classifiers), but rather to provide a robust comparison of means across different comment categories. While there is considerable heterogeneity in individual comment sentiment, the massive sample size allows us to show systematic and statistically significant differences in sentiment across video categories. Results confirm that comments referring to AI itself have a significantly lower sentiment (-0.6495, $p < 0.001$) compared to other comments. More importantly, sentiment is significantly higher across all AIM video categories with the exception of *Derivative*, where sentiment is lower. All parameters are highly significant due to the large sample size.

The first-order interaction model (AI in comment x AIM category) has no impact on the model fit. However, it highlights the fact that across all video categories, comments referring to AI are more positive than similar comments left on OAC videos: every interaction term is positive, including *Altered,* which only approaches significance due to a very small number of observations. We discuss this finding in the next section.

Table 4: Descriptive statistics by AIM category.

Variable		OAC	Altered	#aimusic	Derivative	Audio tags	Video tags
Number of videos							
All time		21,357,107	68	7,829	2,150	7,202	2,823
Past year (2025-04-01 to 2026-03-31)		4,624,628	61	5,083	1,270	4,495	1,252
Views / video							
All time		405,648	41,187	22,306	55,997	19,472	75,248
Past year		70,835	44,088	24,964	61,883	23,543	10,712
Comments / video							
All time		85.9	25.3	10.6	18.5	10.5	46.1
Past year		31.2	26.4	9.3	14.0	10.5	4.2
Sentiment							
All time	All	0.890	0.977	0.918	0.835	0.932	0.925
	No AI	0.890	0.986	0.921	0.837	0.936	0.926
	AI	0.391	0.500	0.650	0.583	0.609	0.716
Past year	All	0.901	0.975	0.915	0.831	0.923	0.944
	No AI	0.902	0.985	0.923	0.831	0.927	0.949
	AI	0.253	0.500	0.547	0.600	0.615	0.732

Table 5: Comparative view statistics by AIM category for the past year and past month.

	Past year (2025-04-01 to 2026-03-31)					Past month (March 2026)				
AIM category	N	Mean	Median	Cliff's δ	P_{sup}	N	Mean	Median	Cliff's δ	P_{sup}
OAC	4,624,628	70,835	1,365	–	–	668,902	38,748	1,191	–	–
AIM	10,256	24,129	959	0.094	0.547	1,244	14,728	1,169	0.014	0.507
Altered	61	44,088	8,383	-0.485	0.257	16	47,990	2,349	-0.333	0.333
#aimusic	5,083	24,964	901	0.109	0.554	558	13,697	626	0.140	0.570
Derivative	1,270	61,883	2,974	-0.220	0.390	155	27,604	4,660	-0.308	0.346
Audio tags	4,495	23,543	876	0.117	0.559	555	11,556	1,166	0.053	0.527
Video tags	1,252	10,712	445	0.264	0.632	143	14,185	1,863	-0.093	0.454

Note: P_{sup} (Probability of Superiority) is the probability that the views of a random OAC video are higher than those of a random video from the AIM category.

Table 6: Generalized Linear Model (GLM) for comment-level sentiment with first-order interactions.

		Main Effect Model			First-order Interaction Model		
Variable	Comment N	Coeff.	S.E.	z	Coeff.	S.E.	z
OAC	107,005,029	0.9016	0.0000	21531.48***	0.9016	0.0000	21531.48***
Main Effects							
AI in comments	48,187	-0.6495	0.0020	-329.20***	-0.6545	0.0020	-329.20***
#aimusic	36,645	0.0194	0.0024	8.01***	0.0161	0.0024	7.00***
Altered	1,318	0.0887	0.0119	7.44***	0.0852	0.0121	7.00***
Derivative	12,795	-0.0748	0.0038	-19.45***	-0.0750	0.0039	-19.00***
Audio tags	36,071	0.0237	0.0024	9.72***	0.0207	0.0025	8.00***
Video tags	16,006	0.0336	0.0034	9.80***	0.0304	0.0034	9.00***
Interaction Effects							
#aimusic × AI in comment	605	–	–	–	0.1414	0.0201	7.00***
Altered × AI in comment	28	–	–	–	0.1582	0.0828	2.00*
Derivative × AI in comment	30	–	–	–	0.3411	0.0798	4.00***
Audio tags × AI in comment	462	–	–	–	0.2426	0.0222	11.00***
Video tags × AI in comment	98	–	–	–	0.3742	0.0457	8.00***
Fit statistics							
McFadden R2		8.74E-04			8.78E-04		
Likelihood ratio Chi2		1.09E+05			1.09E+05		
p-value		0.00E+00			0.00E+00		

Note: Comment N reports the number of comments in each category. Because AIM categories are not mutually exclusive, category-specific counts may overlap. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Taken together, these results: (1) provide strong support for the idea that negative comments referring to AIM are prevalent; (2) with the important caveat that there are more positive than negative comments, with an average sentiment of 0.253 across all OAC video; and (3) the first-order interaction model shows that the average sentiment of comments about AI, when left on an AIM video, is significantly higher than that of comments referring to AI left on other OAC videos, a pattern consistent with the idea that music fans are more likely to watch AIM videos when their attitude towards AI is positive.

More importantly, viewer sentiment is consistently higher across nearly all AIM categories compared to conventional OAC works. The sole exception is *Derivative* (e.g., AI covers), where sentiment drops to 0.827, compared to 0.901 for OAC. This discrepancy is even more pronounced when considering that conventional, non-AI covers typically enjoy a positivity premium. OAC videos tagged with #cover receive an average sentiment of 0.94, likely because artists generally select popular works to reinterpret. This sentiment gap of more than 0.10 is consistent with the possibility that audiences respond more favorably to AI uses framed as original creation or restoration, while penalizing derivative uses that raise concerns about imitation, appropriation, or authorship.

But of greater consequence is the fact that sentiment towards creative, non-derivative AIM videos is systematically superior compared to that of conventional OAC videos.

To illustrate the diversity of AIM reception, we highlight three prominent cases that leverage AI for nostalgia and emotional resonance rather than mere imitation. First, the Beatles' *Now and Then* is perhaps one of the best known cases of AIM. The video used AI-assisted audio restoration to incorporate John Lennon's archival vocals. A vocal minority criticized the digital intervention as "uncanny" or "disrespectful," but the majority of AI-aware comments expressed profound gratitude for the opportunity to experience the band together one last time. A similar embrace of nostalgia appears in channels producing wholly original, AI-generated tracks in retro or niche styles. For example, *My Sweet Classroom Crush*, the most watched *Altered* video, uses algorithms to create 1950s-style videos, while *Acem Kizi*, the most watched *Audio tags*, generates "Anatolian Psych Folk" using Suno, playing under a static image. Both videos have amassed millions of views alongside overwhelmingly positive sentiment. In a qualitative analysis of comments on *Acem Kizi*, listeners frequently expressed disbelief at the emotional depth of synthetic vocals, often arguing that these culturally resonant, AI-generated folk tracks possess more aesthetic value than contemporary human-produced pop. These vignettes suggest that audiences are highly receptive to AI when it builds on a body of work of nostalgic value, in sharp contrast with the negative attitude towards derivative AI covers.

5 DISCUSSION

Our results are based on the analysis of more than 4.6 million OAC videos and 107 million comments, including more than 150,000 comments on AIM videos.

We find negative bias towards AI, as expected, with two important caveats. First, although the average sentiment of comments explicitly referring to AI is lower than the general average, the average remains positive.

Second, with the notable exception of *Derivative* videos, every AIM category exhibits significantly higher sentiment than non-AIM OAC. This suggests that when AI is used *as a creative tool* it is well received, but when AI is *imitating* humans, it is not.

In addition, interaction effects between explicit references to AI and AIM video categories are positive across all AIM categories. There are at least two potential explanations for this: (1) self-selection bias or (2) an increased appreciation for AIM when seen in action.

Regarding self-selection, potential viewers holding negative attitudes toward AIM are likely to ignore these videos, whereas viewers with a positive prior attitude are likely to actively search and engage with them. Selection can also be passive, as YouTube's recommendation engine amplifies attitudes by surfacing videos consistent with them.

Regarding increased appreciation, our qualitative observations that many comments express surprise and gratitude suggest that some viewers may revise their attitudes, in particular when there is a positive latent attitude that is activated by nostalgia or emotional resonance.

These findings specifically describe the reception of explicitly labeled, AI-related musical content by audiences engaging with Official Artist Channels on YouTube. OAC videos are a highly curated, non-random population, such that a higher average AIM sentiment does not imply that AIM is intrinsically superior: it may indicate that the current pool of published AIM productions is more performative, better aligned with listeners' preferences. An intriguing possibility is that curation is facilitated by generative AI tools. Human artists must invest significant effort over a long period of time to craft a song, often requiring collaboration with other creators. In contrast, AI tools allow creators to rapidly iterate and evaluate numerous variations, potentially changing the role of the artist from creator to curator of work fragments suggested by AI tools.

It should also be noted that our conclusions are based on data from the easily discoverable fraction of the actual AIM body of work. We used a very simple method to identify AIM and are certainly missing a large number of videos that may well be qualitatively different from those identified on the basis of their description. We are actively working on the development of better classifiers for that very purpose. Furthermore, many highly successful AIM videos are produced by non-OAC channels, either because these creators have not reached critical mass or because their work aligns more closely with visual "AI Art" than traditional music releases.

Finally, while the proliferation of bot-generated text on social media is a valid platform-wide concern, we find no evidence to suggest that it introduces bias in our results. Longitudinal analysis of our corpus reveals no anomalous spikes in comment volume or sentiment following the widespread availability of LLMs. This is a likely consequence of YouTube's aggressive policy regarding comment spam (more than 1.5 billion comments were taken down in Q1 2026 alone), and of the absence of significant incentives to deploy bots for textual praise compared to view-count manipulation. In addition, there is no theoretical reason to expect bot activity to be systematically higher or more positive on AIM videos than on traditional OAC releases.

AIM currently represents a very small niche within the broader YouTube music ecosystem, a niche that is growing rapidly. In light of our findings, creators who integrate AI into their workflows to produce original content have reasons for optimism.

REFERENCES

Adkins, S., Sarmento, P., & Barthet, M. (2023, April). LooperGP: A loopable sequence model for live coding performance using GuitarPro tablature. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)* (pp. 3-19). Cham: Springer Nature Switzerland.

Agudo, U., Arrese, M., Liberal, K. G., & Matute, H. (2022). Assessing emotion and sensitivity of AI artwork. Frontiers in Psychology, 13

Ansani, A., Koehler, F., Giombini, L., Hämäläinen, M., Meng, C., Marini, M., & Saarikallio, S. (2025). AI performer bias: Listeners like music less when they think it was performed by an AI. Empirical Studies of the Arts, 43(2), 1137-1161.

Artist Rights Alliance. (2024, April 2). 200+ artists call on AI developers, technology companies, platforms and digital music services to cease the use of artificial intelligence to infringe upon and devalue the rights of human artists. [Press release]. *https://thetrichordist.com/wp-content/uploads/2024/04/araopenletterai-release.pdf*

Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., ... & Seli, P. (2023). Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. Cognitive research: principles and implications, 8(1), 42.

Candusso, S. (2024). *Exploring the impact of generative AI on the music composition market: a study on public perception, behavior, and industry implications* (Doctoral dissertation, Politecnico di Torino).

Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. Psychology of Aesthetics, Creativity, and the Arts, 12(2), 177.

Chia, S., Hartanto, A., & Tong, E. M. (2025). Do Listeners Devalue AI-Generated Pop Music? Exploring Negative Biases in Listeners' Responses to AI-Labelled vs Human-Labelled Pop Music. *Computers in Human Behavior: Artificial Humans*, 100217.

Chiarella, S. G., Torromino, G., Gagliardi, D. M., Rossi, D., Babiloni, F., & Cartocci, G. (2022). Investigating the negative bias towards artificial intelligence: Effects of prior assignment of AI-authorship on the aesthetic appreciation of abstract paintings. Computers in Human Behavior, 137, 107406.

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological bulletin*, *114*(3), 494.

Correia, F. F. (2024). The role of artificial intelligence in music production: A survey on public acceptance. *Perception and Bias (Master's thesis, Universidade NOVA de Lisboa (Portugal))*.

Deezer. (2025). *Deezer/Ipsos survey: 97% of people can't tell the difference between fully AI-generated and human made music – clear desire for transparency and fairness for artists*. Deezer Newsroom. https://newsroom-deezer.com/2025/11/deezer-ipsos-survey-ai-music/

Déguernel, K., Sturm, B. L., & Maruri-Aguilar, H. (2022). Investigating the relationship between liking and belief in AI authorship in the context of Irish traditional music. In *CREAI 2022 Workshop on Artificial Intelligence and Creativity*.

Demmer, T. R., Kühnapfel, C., Fingerhut, J., & Pelowski, M. (2023). Does an emotional connection to art really require a human artist? Emotion and intentionality responses to AI-versus human-created art and impact on aesthetic experience. *Computers in Human Behavior*, *148*, 107875.

Di Dio, C., Ardizzi, M., Schieppati, S. V., Massaro, D., Gilli, G., Gallese, V., & Marchetti, A. (2023). Art made by artificial intelligence: The effect of authorship on aesthetic judgments. *Psychology of Aesthetics, Creativity, and the Arts*.

Diaz, J. (2009). The Fate of Auto-Tune. *Music and Technology*.

Ferreira, P., Limongi, R., & Fávero, L. P. (2023). Generating music with data: Application of deep learning models for symbolic music composition. *Applied Sciences*, *13*(7)

Gibbs, L. E. (2019). Synthesizers, virtual orchestras, and Ableton Live: Digitally rendered music on Broadway and musicians' union resistance. *Journal of the Society for American Music*, *13*(3), 273-304.

Hong, J. W., Fischer, K., Ha, Y., & Zeng, Y. (2022). Human, I wrote a song for you: An experiment testing the influence of machines' attributes on the AI-composed music evaluation. *Computers in Human Behavior*, *131*, 107239.

Hong, J. W., Peng, Q., & Williams, D. (2022b). Are you ready for artificial Mozart and Skrillex? An experiment testing expectancy violation theory and AI music. *New Media & Society*, *23*(7), 1920-1935.

Hoppenjans, K. (2025, May). The singer in the loop: searching for embodiment in Grimes's AI voice clone. In *Ethnomusicology Forum* (Vol. 34, No. 2, pp. 258-263). Routledge.

Iskandar, K. L., Spil, T., & Bukhsh, F. (2025). AI and Music: How do listeners and artists perceive it? An Empirical Study toward the Attitude of Humans to AI Music. In *58th Hawaii International Conference on System Sciences, HICSS 2025* (pp. 4086-4095). IEEE.

Katz, M. (2010). *Capturing sound: How technology has changed music*. Univ of California Press.

Kerrang! Radio. (2026, March 12). *Violet Grohl | What's wrong with AI music* [Video]. YouTube. *https://www.youtube.com/watch?v=z_ppjTEjuPo*

Łagodzka, D. (2024). Can Animals Be Artists? Monkey Painting. *ACADEMIA. The magazine of the Polish Academy of Sciences*, 66-70.

Leff, J. (2004). Rage against the Machine: How the NLRB Used Section 8 (e) of the National Labor Relations Act to Kill the Virtual Orchestra.

Li, Q., & Ma, Y. (2024, December). A Study on Social Perception and Acceptance of AI Music-Taking JJ Lin's Fans as an Example. In *2024 7th International Conference on Humanities Education and Social Sciences (ICHESS 2024)* (pp. 768-779). Atlantis Press.

Marone, M., Weller, O., Fleshman, W., Yang, E., Lawrie, D., & Van Durme, B. (2025). mmbert: A modern multilingual encoder with annealed language learning. *arXiv preprint arXiv:2509.06888*.

MAX (2026). 35% of Gen Zers Would Stop Liking a Song After Learning It Was Created by AI. Blog post. https://www.max.live/blog/35-of-gen-zers-would-stop-liking-a-song-after-learning-it-was-created-by-ai

Messingschlager, T. V., & Appel, M. (2025). Mind ascribed to AI and the appreciation of AI-generated art. *New Media & Society*, *27*(3), 1673-1692.

Neef, N. E., Zabel, S., Papoli, M., & Otto, S. (2025). Drawing the full picture on diverging findings: adjusting the view on the perception of art created by artificial intelligence. *AI & SOCIETY*, *40*(4), 2859-2879.

Pasquier, P., Burnett, A., Thomas, N. G., Maxwell, J. B., Eigenfeldt, A., & Loughin, T. (2016, June). Investigating listener bias against musical metacreativity. In *Proceedings of the seventh international conference on computational creativity* (pp. 42-51).

Pew Research (2025). Concern and excitement about AI. https://www.pewresearch.org/topic/internet-technology/artificial-intelligence/

Pressman, A. (2026, April 9). Students at Berklee College of Music pay up to $85,000 to attend. Some say the school's AI classes are a waste of money. The Boston Globe.

Raj, M., Berg, J. M., & Seamans, R. (2026). The artificial intelligence disclosure penalty: Humans persistently devalue AI-generated creative writing. *Journal of Experimental Psychology: General*.

Ramos, D. L. T., & Ventayen, R. J. M. (2025, July). Monitoring the Perceived Quality of Artificial Intelligence Composed Music Through Blind Listener Evaluation. In *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)* (pp. 1489-1493). IEEE.

Shank, D. B., Stefanik, C., Stuhlsatz, C., Kacirek, K., & Belfi, A. M. (2023). AI composer bias: Listeners like music less when they think it was composed by an AI. *Journal of Experimental Psychology: Applied*, *29*(3), 676.

Stammer, J. (2024). You can try weirder and weirder things: once mocked synthesisers enjoy new golden era. *The Guardian* 2024-10-12.

Tapper, D., Strauss, H., & Knees, P. (2025). Perception of AI-Generated Music--The Role of Composer Identity, Personality Traits, Music Preferences, and Perceived Humanness. *arXiv preprint arXiv:2512.02785*.

Tigre Moura, F., & Maw, C. (2021). Artificial intelligence became Beethoven: how do listeners and music professionals perceive artificially composed music?. *Journal of Consumer Marketing*, *38*(2), 137-146.

White, M. W., & de Leon, R. P. (2026). Less "awe"-some art: How AI diminishes the empathic power of the arts. *Journal of Experimental Social Psychology*, *122*

Zhou, J., Li, Y., Adhikari, V. K., & Zhang, Z.-L. (2011). Counting YouTube videos via random prefix sampling. Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference - IMC '11. doi:10.1145/2068816.2068851

Zhou, Y., & Kawabata, H. (2023). Eyes can tell: Assessment of implicit attitudes toward AI art. *i-Perception*, *14*(5)

Zlatkov, D., Ens, J., & Pasquier, P. (2023, April). Searching for human bias against AI-composed music. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)* (pp. 308-323). Cham: Springer Nature Switzerland.

APPENDIX

A AIM CATEGORIZATION TAGS

To construct the dataset, the following specific tags and disclosures were queried to operationalize the AI categories:

- **Altered:** Altered or synthetic content (YouTube mandated disclosure)
- **#aimusic:** #aimusic
- **Derivative:** #aicover, #aivoicecover
- **Audio tags:** #aimusicapp, #aimusicgenerator, #aimusicproduction, #aimusicstudio, #aiva, #boomy, #kitsai, #mubert, #sonauto, #soundraw, #soundverse, #stableaudio, #suno, #sunoai, #udio, #udioai, #udiomusic
- **Video tags:** #aianimation, #aiartvideo, #deforum, #diffusionvideo, #ebsynth, #generativevideo, #heydgen, #img2video, #kaiberai, #kling3, #klingai, #kreaai, #latentspace, #ltxstudio, #lumaai, #lumadreammachine, #morphstudio, #neuralframes, #pikaai, #pikalabs, #runwaygen3, #runwayml, #sora2, #soraai, #stablediffusionvideo, #svd, #synthesia, #text2video, #vid2vid, #viggleai