Data Analysis

The study employs a combination of descriptive statistical analysis, non-parametric exploratory methods, and panel regression techniques to analyze the collected data. All analyses were conducted using Python-based statistical scripts.

Descriptive Pre–Post Analysis of AI compatibility

To examine whether genre characteristics associated with generative AI compatibility changed over time, a descriptive pre–post analysis was conducted using Spotify data from 2018 to 2024. Average values of the SI and CI were compared before and after 2020, which serves as a reference point for the broader availability of AI music tools.

Mean differences between the pre-2020 and post-2020 periods were evaluated using two-sample t-tests. The test statistic is given by:

$$t = \frac{\bar{X}_{post} - \bar{X}_{pre}}{\sqrt{\frac{s^2_{post}}{n_{post}} + \frac{s^2_{pre}}{n_{pre}}}}$$ **(image)**

where $\bar{X}$ (**image**) denotes the sample mean, s^2(**image**) the sample variance, and n the number of observations in each period.

Associated p-values were used to assess statistical significance. Results are interpreted using conventional thresholds: p-values below 0.05 indicate strong statistical significance, values between 0.05 and 0.10 indicate moderate significance, and values above 0.10 are considered statistically insignificant. This analysis is descriptive and does not imply causality.

LOWESS Analysis of AI compatibility and Popularity

To explore the relationship between AI compatibility and song popularity at the track level, the study applies a locally weighted scatterplot smoothing (LOWESS) method. Each song is assigned a composite AI compatibility score, defined as:

$$AICompatibility_g = \frac{SI_g + CI_g}{2}$$ **(image)**

where SI_g(**image**) and CI_g (**image**) are the substitution and complementarity indices for genre g. Song popularity is measured using Spotify's standardized popularity score (0–100).

LOWESS estimates a smoothed conditional mean of popularity by fitting local regressions weighted by proximity in the exposure space. This method does not produce coefficients, t-values, or p-values and is used purely as an exploratory, descriptive tool to visualize potential nonlinear patterns between AI compatibility and popularity.

Descriptive Diversity Analysis Using Entropy Scores

To assess changes in stylistic diversity over time, the study computes Shannon entropy scores based on Spotify genre distributions. For each year t, genre shares are calculated as:

$$p_{g,t} = \frac{\text{tracks in genre } g \text{ in year } t}{\text{total tracks in year } t}$$ **(image)**

The entropy score is then computed as:

$$H_t = -\sum_g p_{g,t} \log(p_{g,t})$$ **(image)**

Higher entropy values indicate a more even distribution of output across genres, while lower values indicate greater concentration. Entropy scores are compared across years to describe how overall genre diversity evolved during the study period. This analysis is descriptive and does not involve hypothesis testing.

Descriptive Diversity Analysis Using HHI and Gini Index

To examine how commercial recognition is distributed among songwriters, this study conducts a descriptive diversity analysis using the Billboard Hot 100 dataset. Chart rankings reflect overall success but do not directly capture how evenly recognition is shared across contributors. To address this, the analysis examines the concentration of lyricist credits in the Top 100 over time, using the Herfindahl–Hirschman Index (HHI) and the Gini Index.

For each period, lyricist shares are calculated as the proportion of total lyricist credits attributed to each individual. The HHI is computed as:

$$HHI = \sum_{i=1}^{N} s_i^2$$ **(image)**

where s_i is the share of lyricist credits held by lyricist i. Higher HHI values indicate greater concentration among a smaller number of lyricists, while lower values indicate a more dispersed distribution.

To measure inequality across the full distribution, the Gini Index is calculated as:

$$Gini = 1 - \sum_{i=1}^{N} (s_i + s_{i-1})(p_i - p_{i-1})$$ **(image)**

where lyricists are ordered by increasing share, and p_i represents the cumulative proportion of lyricists. The Gini Index ranges from 0 to 1, with higher values indicating greater inequality and lower values indicating a more equal distribution.

These indices together describe the degree of concentration and inequality in lyricist participation on the Billboard Hot 100 over time.

Panel Regression and Difference-in-Differences Analysis

To estimate the relationship between AI compatibility and popularity outcomes, the study employs panel regressions using ordinary least squares (OLS). Genre–month panels are constructed, and regressions include genre fixed effects and month fixed effects to control for time-invariant genre characteristics and standard industry-wide shocks.

The baseline specification is:

$$Y_{g,t} = \alpha + \beta(AIExp_g \times Post_t) + \gamma_g + \delta_t + \varepsilon_{g,t}$$ (**image**)

where $Y_{g,t}$ (**image**) represents the outcome of interest (average popularity or hit probability), $AIExp_g$ (**image**) is genre-level AI compatibility, $Post_t$ (**image**) is an indicator for the post-2020 period, γ_g (**image**) are genre fixed effects, and δ_t (**image**) are time fixed effects.

Standard errors are clustered at the genre level to account for serial correlation.

The statistical significance of the estimated coefficients is evaluated using t-statistics and p-values, interpreted using the thresholds described above.

AI compatibility is marginally associated with a higher probability of producing a hit song

To examine how compatibility with generative AI relates to commercial success after 2020, I analyze both average song popularity and the likelihood of producing hit songs. Comparing genres with different levels of AI compatibility helps clarify whether AI changes overall performance or mainly affects which songs succeed the most.

I first examine changes in average popularity using the following relationship:

$$MeanPop_{g,t} = \alpha + \beta(AIExp_g \times Post_t) + \gamma_g + \delta_t + \varepsilon_{g,t}$$ (**image**)

In this equation, $MeanPop_{g,t}$ (**image**) represents the average Spotify popularity score of songs released in genre g during month t. $AIExp_g$ (**image**) measures how compatible a genre is with generative AI, and $Post_t$ (**image**) indicates the period after 2020. The interaction term captures whether genres with higher AI compatibility experienced different changes in popularity after AI tools became more common.

Table 2: *OLS Regression Results for Average Popularity*

OLS Regression Results						
Dep. Variable:	mean_pop		R-squared:		0.883	
Model:	OLS Adj.		R-squared:		0.881	
Method:	Least Squares		F-statistic		47.49	
No. Observations	7908		Prob (F-statistic):		2.94e-53	
DF Residuals:	7728		Log-Likelihood:		-23333.	
DF Model:	179		AIC:		4.703e+04	
Covariance Type:	cluster		BIC:		4.828e+04	
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	25.6319	0.567	45.215	0.000	24.521	26.743
Post: AI__exp	0.1608	0.440	0.366	0.715	-0.701	1.023
Omnibus:	645.843		Durbin-Watson:		1.135	
Prob(Omnibus):	0.000		Jarque-Bera (JB):		3132.784	
Skew:	0.244		Prob(JB):		0.00	
Kurtosis:	6.045		Cond. No.		372.	

Notes: Standard errors clustered at the genre level (N = 100 genre clusters). Genre and month-year fixed effects included but not reported.

The estimated interaction coefficient (Post: AI_exposure) is small and not statistically significant (coefficient = 0.1608, $p = 0.715$). This means that genres with higher AI compatibility did not see meaningful increases in average popularity after 2020. While AI-sensitive genres may have produced more music, the typical song in those genres did not become more popular. This suggests that generative AI has not raised average performance across the market.

I then turn to the probability of producing hit songs, which focuses on outcomes at the top of the popularity distribution. Hit share is defined as the fraction of songs in a genre–month whose Spotify popularity score is equal to or greater than 60, chosen to identify rare standout outcomes while remaining grounded in the data's own distribution. The following relationship is estimated:

$$HitShare_{g,t} = \alpha + \beta(AIExp_g \times Post_t) + \gamma_g + \delta_t + \varepsilon_{g,t}$$ **(image)**