

From Pixels to Personas: Tracking the Evolution of Anime Characters

Rongze Liu¹, Jiaxin Pei^{2,3}, Jian Zhu¹

¹University of British Columbia

²Stanford University

³University of Texas at Austin

celt313@student.ubc.ca, pedropei@stanford.edu, jian.zhu@ubc.ca

Abstract

Anime, originated from Japan, is one of the most influential cultural products in modern society and is especially popular among younger generations. The popularity of anime reflects important cultural evolutions in our society. Despite existing research on anime as a cultural phenomenon, we still have a limited understanding of how anime really evolves over the years. In this study, using a large-scale multimodal dataset of anime characters from an anime review site, we applied computational methods that integrate textual, visual, and production features of anime characters with online popularity traces. By combining LLM-extracted personality features with avatar features, we identify recurring personality archetypes and visual tropes with their temporal evolution over the past decades. We found that the target audience of anime has undergone a systematic shift from children to a maturing audience of teenagers and young adults over time. Character design has been undergoing *moe*-ification, with softer or sexualized female traits becoming increasingly prominent since the 2000s. Some personality archetypes are often visually predictable, yet audiences also tend to prefer less conventionalized characters. Finally, we reveal that visual signals play a more dominant role than personality traits in shaping audience preferences, with features such as *moe*-style faces and mechanical designs contributing greatly to popularity. These findings offer insights into the broader dynamics of anime’s cultural and creative practices.

Code — <https://github.com/Arc-Celt/pixels2personas>

Introduction

Anime, together with Manga and video games, constitute the most important cultural products of contemporary Japan (O’Hagan 2007). Artistically, anime and Manga represent a distinct visual style and aesthetics in contemporary Japanese culture and national image (Norris 2009). Anime has a profound impact on the popular cultures in East Asia, and the rest of the world, giving rise to spin-off merchandise like figurines (Condry 2009), anime tourism (Okamoto 2015), Internet memes (Saito 2017), etc., and new digital forms like VTubers (Zhao et al. 2025). In 2025, Netflix reports that fifty percent of global viewers on its own platform now watch anime (Brzeski 2025). Despite its global impact, anime has

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

received surprisingly little attention in digital humanities and computational social sciences, with very few studies focusing on the computational analysis of anime and manga (Douglass, Huber, and Manovich 2011; Manovich 2012), as compared to other media like movies, music, and games.

Unique and appealing characters are one of the most long-lasting attractions in anime. In prior research, characters in Japanese manga and anime are often referred to as “*kyara*” (Japanese transliteration of characters), to distinguish them from conventional characters (Hōsei and Bergstrom 2023; Gō and Nakamura 2011). *Kyara* is a character that is identifiable even beyond the original story context. Such characters can persist beyond anime, manga, and video games, in merchandise like figurines, key chains, and CDs, and in fan-made works like fan fiction and cosplay. However, as hundreds of characters are churned out by the anime industry every year, not all of them will persist beyond the airing, except for a few preferred by the audience.

Research Questions We conceptualize anime characters as cultural products with multimodal design elements encompassing personality traits, distinct visual features, and elaborate voice acting. Anime has been easily recognizable as being ‘anime-esque’ for its distinguishing art style (Suan 2017). Even within the surface uniformity of anime style, there still exists a large space for unique character design. While we know that character design matters (Iijima, Wada, and Yamada 2016), we lack large-scale, data-driven evidence of how personality and visual style capture audience attention. By utilizing the online crowd-sourced preference data from *MyAnimeList*, an anime review site, our study progresses from macro to micro: first examining historical production trends to establish the evolution context, then identifying dominant character design patterns across periods, and finally exploring how these design elements are systematically associated with audience preference. Specifically, we ask the following research questions:

- **RQ1:** What is the historical trend of anime production?
- **RQ2:** What are the most common character archetypes in specific periods in the past decades?
- **RQ3:** What are the common visual tropes in character creation? How do the artistic styles of anime characters change over time?

- **RQ4:** Are there any systematic associations between personality traits and visual features?
- **RQ5:** How do personality and visual styles predict (mostly Western) audience preference, as measured by crowd-sourced favorite counts?

Contributions To answer these research questions, we curated a comprehensive anime database, covering over 20k animes in the past decades and over 150k unique characters. Large-scale computational methods were applied to analyze the temporal trends of anime characters. Our study makes the following contributions.

- **We uncovered core personality archetypes and recurring visual tropes, tracing their historical evolution and revealing their associations.** Our results reveal that character personality archetypes have shifted toward greater psychological complexity, indicating a shift toward a maturing audience of teenagers and young adults. Anime character design has been significantly influenced by *moe*-ification, the design trend targeting *moe* (affectionate responses to fictional characters) (Galbraith 2014), with soft or sexualized female visual traits becoming increasingly prevalent during recent decades. Moreover, personality archetypes are often visually predictable, especially for certain flat archetypes, suggesting that personality and visual features operate symbiotically for coherent character design.
- **We demonstrated how personality and artistic designs jointly shape audience preferences within Japanese anime’s established design conventions.** While both artistic styles and personality traits contribute to audience appeal, visual features exhibit greater predictive power. Characters that feature feminine traits like *moe*-style faces, long stylish hair, and slender proportional bodies, as well as elaborate mechanical designs like armor or *mecha* elements, tend to achieve higher popularity. This illustrates how anime’s cultural production strikes a balance between creative innovations and cultural conventions through mixing and matching of recognizable anime-esque elements (Suan 2017).

Related Works

Anime

Anime is often created not only for its own sake but as part of the interconnected cultural products. Media mix is a commercial strategy adopted by the Japanese companies to develop transmedia franchises that span anime, manga, video games, spin-off toys, and other franchised toys (Steinberg 2012; Kopylova 2023; Napier 2011; Kinoshita 2024; Steinberg 2023). For example, the *Mobile Suit Gundam* franchise extends its TV anime series with novelizations, model kits, and life-sized statues, which help sustain and expand its cultural impact (Kinoshita 2024). Such a strategy also exerts an influence on the visual style of anime character design (Kopylova 2023; Watzky 2023). Many analyses or overviews of Japanese anime tend to focus exclusively on a few influential hits like those from the Ghibli Studio (Lun-

ning 2011). Yet the scope has rarely been expanded to include wider ranges of anime created across decades.

Anime Characters

The artistic style of anime characters has been investigated in the humanities and social sciences, covering the creation pipeline (Condry 2009), racial perception (Lu 2009; Fennell et al. 2013), popularity (Tang et al. 2024), voice dubbing (Howell 2007), parasocial attachment (Ramasubramanian and Kornfield 2012) and cosplay (Winge 2006). Some character design has also been criticized for pandering to the *Otaku* community, a subculture of devoted anime and manga fans (Azuma 2009) that are mostly heterosexual males, perpetuating female stereotypes (Ting 2019; Lunning 2011). Yet, some anime, like *Sailor Moon* series, once regarded as objectifying, was also argued to subvert patriarchal and gender stereotypes (Yatron 2022).

Anime characters have been argued to be *Mukokuseki* (meaning ‘nationless’ or ‘stateless’), as the physical appearance of many characters does not look Japanese but more like abstract human bodies, depending on the creators and genre tropes (Ruh 2014). Many classic anime characters will also become the foundation for the creation of new characters for the *Otaku* community to consume, referred to as ‘database consumption’ (Azuma 2009), or ‘database fantasy’ (Ruh 2014). While the interconnectivity of anime character design and tropes has long been recognized by scholars, it is mostly approached from a qualitative perspective with case studies and impressionistic analyses. Such analyses can be complemented by large-scale computational analysis, which can serve to augment the discovery of shared cultural resources in the design of anime characters.

Computational modelling of anime characters in the Computer Vision Communities tend to focus on generating anime avatars from text or image prompts through training on large-scale datasets (e.g., Jin et al. 2017; Li et al. 2021; Huang et al. 2023; Cao et al. 2023). Neural networks have also been trained to recognize the distinct artistic style of anime characters across different anime works (Li et al. 2022). However, these computational studies only center on predictions through neural networks that are usually uninterpretable. There still lacks a comprehensive study to understand the design and the cultural affordance of anime character design.

Method

Data Collection

The anime data was gathered from the anime review site MyAnimeList¹, one of the most comprehensive websites for anime. Each month, the site is visited by over 12 million unique users worldwide (75% male), encompassing 200 countries across the globe (25% from the US)².

Data were scraped and parsed with customized scripts. In addition to the anime main pages, we further retrieved the

¹<https://myanimelist.net/>

²<https://myanimelist.net/advertising>

pages for each individual character, if available. These individual character pages usually contain textual descriptions of anime characters with varying details, contributed by community members or other public sources like Wikipedia. We also retrieved the high-resolution character avatars if available. We parsed the crawled webpages to extract meta information (years and user ratings), and text descriptions of anime and characters. Voice is not analyzed in this work due to the lack of audio data. All anime shows aired after July 1, 2025 were excluded. We adhered to ethical practices of data collection. The anime database only contains public data without sensitive information. We set an interval of 10 seconds between consecutive requests to avoid burdening the website server. The whole crawling process took around two weeks to complete.

Analyzing Character Personalities

Character Labeling Anime characters on MyAnimeList are usually accompanied by unstructured free text descriptions. LLMs have also been shown to exhibit strong capabilities to understand literary characters (Jang and Jung 2024; Amalvy, Labatut, and Dufour 2025; Papoudakis, Lapata, and Keller 2024; Abrams 2025). So we prompted an LLM to summarize personality traits in a few adjectives, given the crowd-sourced character descriptions on MyAnimeList. Sample prompts and outputs are available in the Appendix.

We utilized Qwen3 (Yang et al. 2025), a series of state-of-the-art open-sourced LLMs, to systematically annotate the character descriptions into structured representations. The Qwen3-32B-FP8 was deployed through vLLM (Kwon et al. 2023) in FP8 precision, and the model performed inference in non-thinking mode.

Model Validation To validate our model outputs, we randomly sampled 50 character instances. For each instance, we created a candidate pool of 21 shuffled keywords: the model-generated personality keywords for that character, distractor keywords randomly sampled from other characters’ keyword sets, and a “None of the above” option. Using the Potato annotation tool (Pei et al. 2022), two of the authors annotated all 50 samples independently, with no prior knowledge of the samples. For each character, annotators were asked to read the biography and select all keywords from the candidate pool that accurately described the character’s personality. We then measured the agreement between human selections and model-generated keywords using metrics of precision, recall, F1-score, and Jaccard similarity.

As shown in Table 1, both annotators achieved high agreement with the model-generated keywords, as indicated by the high average agreement scores across all metrics. This consistency suggests that the model outputs are generally reliable and align well with human judgments.

Clustering Character Personalities We extracted the 1024-dimensional embeddings for the personality traits of each character using Qwen3-Embedding-0.6B (Zhang et al. 2025). Each embedding vector was L2-normalized to unit length. Then we applied Uniform Manifold Approximation and Projection (UMAP) (McInnes, Healy, and Melville

Annotator	Precision	Recall	F1	Jaccard
1	0.907	0.671	0.747	0.639
2	0.767	0.656	0.688	0.566

Table 1: Mean agreement metrics between LLM-extracted personality keywords and human annotations for 50 random samples: Precision (fraction of model keywords validated by humans), Recall (fraction of human-selected keywords generated by the model), F1-score, and Jaccard similarity (intersection over union of selected keyword sets). They demonstrate substantial alignment between LLM extractions and human judgments.

2018), a nonlinear dimensionality reduction technique, to reduce the 1024-dimensional vectors to two dimensions.

We constructed a distance-thresholded graph where nodes represent characters and edges connect characters within a specified distance threshold (5th percentile). Edge weights were computed as $w_{ij} = 1/(1 + d_{ij})$, where d_{ij} is the Euclidean distance between characters in the UMAP space. To identify groups of characters with similar personality profiles, we performed community detection on the 2D UMAP embeddings using the Leiden algorithm (Traag, Waltman, and van Eck 2019), yielding a partition with high modularity. The top 10 nodes with the highest degree centrality were selected as representatives of the personality archetype associated with their community.

Analyzing Avatars

Avatar Processing We extracted the latent representations of character avatars with DINOv2 (Oquab et al. 2024), a vision transformer (ViT) encoder model (Dosovitskiy et al. 2020) pretrained on large-scale images with self-supervision. DINOv2 is capable of extracting fine-grained visual features from raw images as represented by embedding vectors for many downstream tasks like image classification and clustering (Oquab et al. 2024). Specifically, we used `dinov2-base3`, an 82M parameter variant for all feature extractions. All original avatar images were rescaled to 224×224 to fit DINOv2’s native resolution, and then all rescaled images were transformed into 768-dimensional DINOv2 embeddings.

Avatar Tagging Genders of characters are not explicitly labeled in the character page. So we utilized the `Camie Tagger4`, a state-of-the-art anime image tagger that has been trained on over 7 million anime images with over 70k tags. We extracted the gender tags from the predicted outputs. Each avatar is labeled as male, female, or other (for non-binary or non-humanoid characters).

Finetuning DINOv2 To investigate the temporal changes in character art style design, we finetuned DINOv2 to classify the years of the characters. All images were labeled according to their anime’s release year into 10 temporal pe-

³<https://huggingface.co/facebook/dinov2-base>

⁴<https://huggingface.co/Camais03/camie-tagger>

Hyperparameter	Value
Training Epochs	200
Min Epochs	10
Learning Rate	3e-5
Batch Size	20
Weight Decay	0.05
Early Stopping Patience	10

Table 2: Fine-tuning hyperparameters for dinov2-base.

Fields	Values
Number of Anime	27,783
Number of Main Characters	26,722
- with text descriptions	18,002
Number of Supporting Characters	111,777
- with text descriptions	54,274
Number of Avatars	126,347
Release Year Range	1917 - 2025

Table 3: General Dataset Statistics

riods, then partitioned with stratified sampling as 40% for training, 10% for validation, and 50% for testing.

We appended a 3-layer MLP classifier to the DINOv2 backbone with ReLU activation and dropout regularization. As DINOv2 was pretrained on a vastly different visual domain than anime avatars(Oquab et al. 2024), we implemented progressive unfreezing: train the head only first, then unfreeze the last 2 blocks at epoch 8, and expand to the last 4 blocks at epoch 12.

RQ1: Historical Trends in Anime

Generally speaking, anime production has expanded significantly historically, while the medium itself has grown more mature, more complex, and has benefited from a wider range of sources. Table 3 summarizes the statistics of our dataset. In general, we have collected over 27,000 animes with over 130k characters, spanning 108 years. We can also observe several general trends in anime production. In Figure 1, the number of anime produced per year has increased steadily since the 1950s, but the increase picked up momentum around the 2000s and peaked after 2015. Together with this increase, the structures of anime have also become richer and more complex in recent decades. While the number of main characters barely changes, the number of supporting characters has undergone a steady increase. This trend may stem from the modernization of the anime industry, which enables larger supporting casts, as well as market shifts toward more complex storytelling.

Figure 2 suggests that the target audience for anime has shifted toward a more adult demographic over the past few decades. The most notable trends include a rapid increase in the number of PG-13-rated anime and a corresponding decline in G-rated (All Ages) and PG-rated (Children) titles. Additionally, anime rated R and R+ experienced a short period of growth during the 2000s but began to decline after

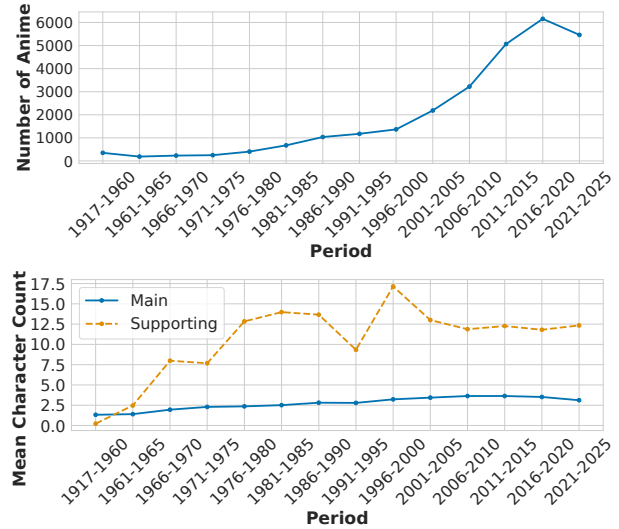


Figure 1: Number of anime per period (top) and average number of characters per anime over time (bottom). Anime production has grown rapidly, with supporting characters steadily increasing, indicating richer story structures and more complex narratives.

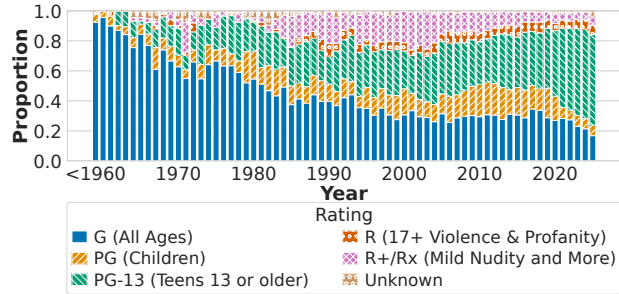


Figure 2: Proportion of PG ratings by year. The decline in G-rated anime alongside the substantial rise in PG-13 content demonstrates a clear shift toward more mature audiences.

around 2010. These changes might reflect a maturing audience, and are likely also fueled by the shifting production model by streaming platforms like Netflix, which allows for adult-like content oriented at a global audience (Noh 2024).

RQ2: Character Personality Archetypes

We investigate recurring patterns in character personality designs by detecting common archetypes, analyzing their role and gender distributions, and tracing temporal trends.

Identifying Character Archetypes

Based on the Leiden community detection results on the UMAP-reduced personality embeddings, **we are able to cluster all characters into 8 major personality archetypes that recur in anime**. Each archetype corresponds to a cluster of characters sharing similar personality profiles, characterized by the top 10 most central characters and their

Archetype	Main / Supporting	Traits	Characters
Energetic Extroverts	44.4% / 55.6%	cheerful, energetic, outgoing	Mini Yaemori, Gokuu Son
Steadfast Leaders	32.2% / 67.8%	calm, responsible, leader-like	Roy Mustang, Saber
Cold-hearted Antagonists	22.5% / 77.5%	cold-hearted, arrogant, cunning	Dio Brando, Frederica Bernkastel
Passionate Strivers	36.3% / 63.7%	passionate, dedicated, determined	Kyoujurou Rengoku, Aaeru
Kind Caregivers	36.1% / 63.9%	kind, compassionate, considerate	Rem, Haruno Yukinoshita
Justice Keepers	41.5% / 58.5%	sense of justice, loyal	Setsuna F. Seiei, Convoy
Reserved Introverts	61.7% / 38.3%	introverted, timid, reserved	Mio Akiyama, Sakura Adachi
Arrogant Tsunderes	50.3% / 49.7%	tsundere, arrogant, sarcastic	Mikoto Misaka, Taiga Aisaka

Table 4: Character Personality Archetypes.

dominant personality keywords. Table 4 summarizes these archetypes and their role distributions.

Anime creators prefer different archetypes for main and supporting characters, with main characters tending to be more complex and multidimensional. Reserved Introverts, Arrogant Tsunderes, Energetic Extroverts, and Justice Keepers are more likely to be main characters than other archetypes. The Energetic Extroverts and Arrogant Tsunderes are common archetypes in *Shonen* (teenagers) manga and anime; yet, characters with intense emotional conflicts and internal activities have also become increasingly prominent. These archetypes allow for character growth throughout the anime and provide stronger self-identification with the audience. Comparatively, supporting characters tend to be Cold-hearted Antagonists, Steadfast Leaders, Kind Caregivers, and Passionate Strivers. These archetypes tend to be flat characters by design.

Archetype Patterns by Gender

Figure 3 illustrates how personality archetypes are distributed across gender. Arrogant Tsunderes, Reserved Introverts, Energetic Extroverts, and Kind Caregivers are predominantly female, while archetypes such as Cold-hearted Antagonists and Justice Keepers have a higher proportion of male characters. **Anime storytelling is shaped by traditional gender norms, where males are consistently depicted as tenacious, dominant, and competent, while females are more often portrayed as dependent, emotional, and supportive** (Bresnahan, Inoue, and Kagawa 2006).

Temporal Changes in Character Personality Design

The distribution of personality archetype design has remained generally stable over time, with minor rise and fall of certain archetypes. In Figure 4, among the main characters, Energetic Extroverts and Passionate Strivers have consistently been dominant, with slight fluctuations, reflecting the enduring demand for determined and outgoing protagonists who actively confront challenges and inspire others with their enthusiasm. Noticeably, however, the share of Reserved Introverts has gradually increased, especially in recent decades. This indicates a growing appreciation for psychologically complex protagonists, who may not possess the traits of classic heroes, but are flawed and

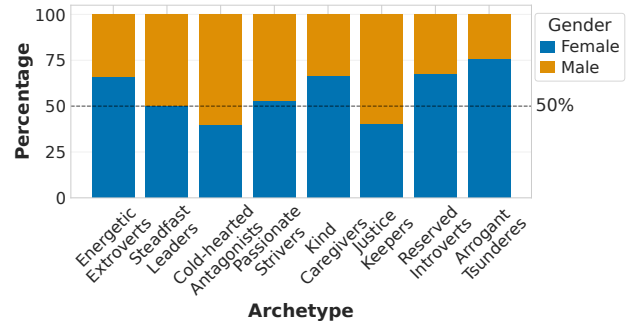


Figure 3: Gender share within each archetype. Female characters are modestly overrepresented in Arrogant Tsunderes, Reserved Introverts, Energetic Extroverts, and Kind Caregivers, whereas Cold-hearted Antagonists and Justice Keepers are more frequently associated with males.

constantly growing, making them more relatable to audiences (Tang et al. 2024). Arrogant Tsunderes, though a small group among main characters, are on the rise. Such characters are both resistant and appealing, with this contrast becoming increasingly prominent in character designs.

The distribution of archetypes among supporting characters also reveals several notable trends. Cold-hearted Antagonists have consistently occupied the largest share over time, but their prevalence has generally decreased over the decades. This suggests a pattern similar to that of protagonists: while anime still relies on antagonistic forces to drive conflict and plot, supporting characters are now depicted with greater diversity and complexity, surpassing traditional villain figures. Support-oriented roles such as Steadfast Leaders and Kind Caregivers have maintained a stable presence over time. They often serve as the steady backbone of a cast, playing a crucial role in aiding protagonists and fostering their growth. Meanwhile, the increase in both Energetic Extroverts and Reserved Introverts indicates a need for enriched interactions and contrasting perspectives, which help drive the protagonists' growth.

RQ3: Character Avatar Archetypes

Visual conventions in character avatars are analyzed through principal component analysis (PCA) on avatar embeddings and interpretation of the top components, alongside their

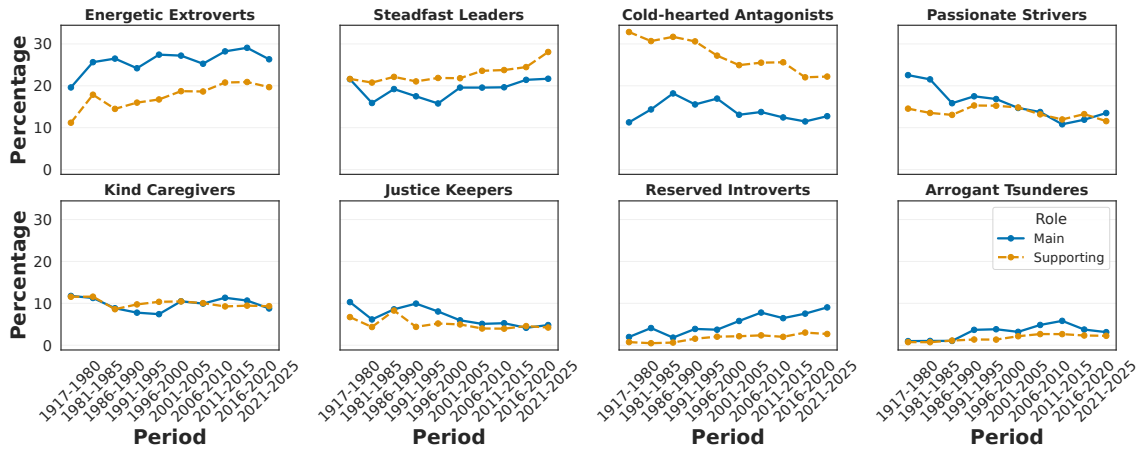


Figure 4: Temporal trends of personality archetypes among anime characters by role. Main characters have consistently favored active and outgoing types, though psychologically complex archetypes like Reserved Introverts and Arrogant Tsunderes have risen modestly. Supporting characters have become more diverse, with fewer antiheroes and more individualized personalities.

temporal shifts.

Dimensions of Character Design

Anime characters are often designed with recurring visual tropes. We decomposed visual character design into interpretable orthogonal dimensions by applying PCA on the DINOv2 embeddings of character avatars. As reflected in the top and bottom character examples in Figure 5, each principal component captures a distinct aspect of character design, which can be interpreted by examining the stylistic features of characters at both extremes. PC1 contrasts large, muscular males with fierce expressions and gentle, smiling females, representing a spectrum from aggressive, masculine features to approachable, feminine ones. PC2 separates realistic human males with angular facial features from flat, non-humanoid characteristics, spanning from realistic human forms to abstract or mechanical designs. PC3 distinguishes tall, slender figures with angular faces from round, cartoonish designs, capturing maturity versus child-like proportions. PC4 contrasts ordinary office workers with goddess-like fantasy figures, representing a spectrum from everyday to fantastical designs. PC5 differentiates bright, vibrant portrayals of young females in light clothing and simple settings from dark, shadowed depictions of serious male characters, reflecting a transition from luminous visuals to more intense and dark visual styles. PC6 contrasts detailed portraits of mature males with mustaches and formal attire with highly abstract avatars, including silhouettes, blurred figures, and animals, capturing a spectrum from recognizable characters to non-character visual elements. PC7 distinguishes armored, masked characters with sharp, angular designs from rounded, vintage-style female faces, representing a spectrum from mechanical, masked figures to classic, old-fashioned face designs.

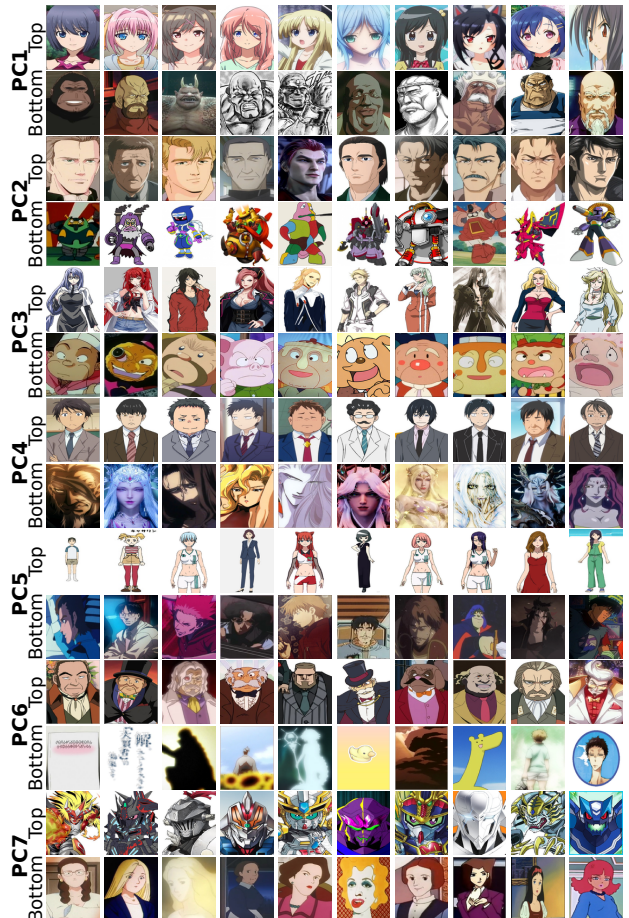


Figure 5: Examples of the top 10 (most positive) and bottom 10 (most negative) characters for each of the top 7 principal components (PCs) derived from avatar embeddings.

Model	F1 Score	Accuracy
Baseline (random)	0.09	9.7%
Baseline (majority)	0.03	17.9%
DINOv2 (frozen backbone)	0.05	11.8%
DINOv2	0.39	39.3%

Table 5: Period classification performance using original character avatars across different models.

Temporal Changes in Art Styles

The art styles of anime characters have shifted systematically over time, with strong trends of *moe*-ification. Character avatars were split using stratified sampling (40% training, 10% validation, 50% testing) and used to finetune DINOv2 for predicting character periods. The F1 score on the test set is 0.39, substantially outperforming the majority vote baseline (F1 = 0.03). The confusion matrix in Figure 15 in the appendix demonstrates that temporal periods of anime character avatars are distinguishable by visual features. While each period has its own distinctive features, there is also a clear sense of continuity in the development of anime character art over time, as reflected by the fact that most errors occur between adjacent periods.

The temporal shifts in dimensions of anime avatar design are illustrated in Figure 6. While most PCs are stable over time, PC1, PC3, and PC5 exhibit a marked increase over the years. As outlined earlier, PC1 reflects a shift from exaggerated masculinity toward softer, more approachable female traits. PC3 indicates a trend in body shapes toward slenderness and more exaggerated proportions, moving away from cartoonish forms. In addition, PC5 illustrates a trend toward brighter, more luminous visual designs. High values of these PCs are often associated with *moe* female character design, suggesting an upward trend to incorporate more *moe* elements in anime characters.

RQ4: Associations Between Personality Traits and Avatars

While commonly assumed, the extent to which visual designs of characters encode personality traits lacks empirical evidence. Despite the variation in the styles of individual artists, **our analysis revealed systematic associations between the artistic rendering of avatars and personality traits.** We performed PCA on the DINOv2 embeddings of avatars and selected the top 100 PCs as image features. Using stratified sampling, we partitioned the data into 70% training, 10% validation, and 20% test sets. Then Random Forest and an XGBoost models were trained to predict personality archetypes from the avatars. Both models exceeded baselines (random: F1 0.11/Acc 12.2%; majority: F1 0.05/Acc 22.8%), with random forest achieving 0.32 macro-F1 and 40.7–41.0% accuracy across 20–100 PCs, and XGBoost performing similarly and peaking at 100 PCs (F1 0.34, Acc 41.2%). Table 6 suggests that personality archetypes are highly predictable from the visual features of avatars with only 20 PCs.

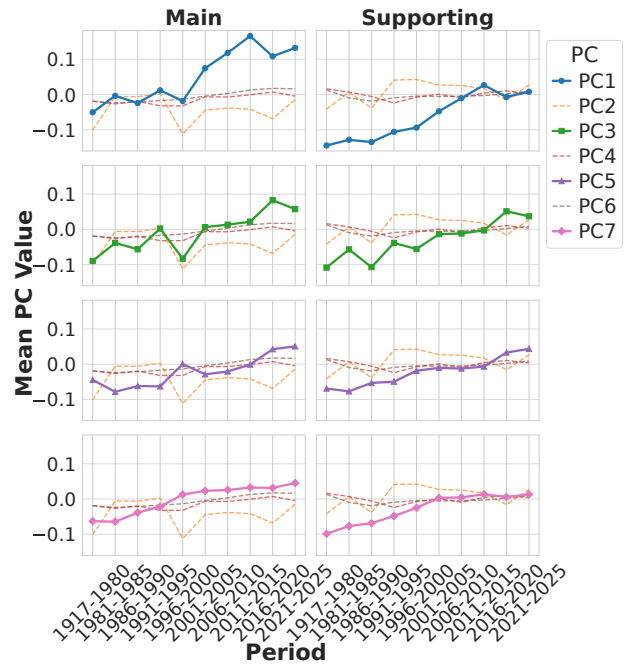


Figure 6: Mean values of the top 7 principal components of avatar embeddings across different periods. PC1, PC3, PC5, and PC7 show an increasing trend, reflecting greater use of *moe* elements in art style designs.

The confusion matrix in Figure 16 in the Appendix further indicates that the personality archetypes are systematically associated with the art design. Justice Keepers, Energetic Extroverts, and Cold-hearted Antagonists are among the most predictable archetypes from the visual features, suggesting that there are highly conventional and frequently reused visual tropes for designing these character types.

RQ5: Modeling User Preferences

Prior research has suggested that anime characters’ personality traits have an impact on the popularity of female characters but not male characters (Tang et al. 2024). Here we attempt to model user preferences using the crowd-sourced ratings on MyAnimeList. We examined the distribution of favorite counts across different personality archetypes on MyAnimeList in Figure 7. Despite their small share among all archetypes, Reserved Introverts and Arrogant Tsunderes have the highest mean and median favorite counts, with wider upper tails. This may be because both archetypes are predominantly female, and traits such as being easily shy, tsundere, and contrasting personalities tend to be more popular among male audiences. While character design relies on many recurring tropes, some combinations of these tropes might make characters more popular. Motivated by this hypothesis, we design experiments to test whether character popularity is predictable from personality and visual design.

We predicted anime character popularity (measured by the number of user favorites) using PCs derived from per-

Model	F1 Score	Accuracy
Baseline (random)	0.11	12.2%
Baseline (majority)	0.05	22.8%
Random Forest (PC=20)	0.32	40.9%
Random Forest (PC=50)	0.32	41.0%
Random Forest (PC=100)	0.32	40.7%
XGBoost (PC=20)	0.30	36.3%
XGBoost (PC=50)	0.34	39.9%
XGBoost (PC=100)	0.34	41.2%

Table 6: Personality archetype classification performance using PCs of DINOv2 avatar embeddings. XGBoost substantially outperforms baselines, suggesting that visual features could indicate personality archetypes.

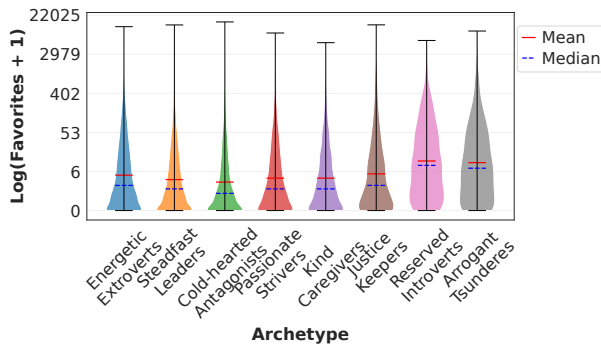


Figure 7: Distribution of log-scaled favorite counts by personality archetype.

sonality and avatar features. For each character, we used the top 20, 50, or 100 PCs as predictors and tested different feature sets: personality PCs only, avatar PCs only, both sets combined, and each of these with extra metadata (year, gender, role, anime favorites). Missing values in PCs were imputed with zero, and entries with missing metadata were dropped. Since favorites are non-negative and heavily right-skewed, we applied a $\log(1 + x)$ transform on the target favorites to further stabilize variance. For prediction, Poisson regression and Histogram-based Gradient Boosted Trees (HGBT) with Poisson loss, using default model hyperparameters. The data were randomly split into train and test sets (90%/10%). The predictions were evaluated with mean absolute error (MAE) and mean Poisson deviance (MPD).

To investigate feature importance and provide interpretable insights into the best performing model (HGBT using Top 50 PCs of Avatar and Personality as well as metadata), we employed SHapley Additive exPlanations (SHAP) (Lundberg and Lee 2017) to quantify the feature contributions, as shown in Figure 8. Metadata features of anime favorites, roles, and release year demonstrate high importance. **This aligns with expected patterns that characters from more popular anime series and main characters from more recent years naturally gain more audience attention and favorability.** In addition to metadata, the character design is also associated with audience preference.

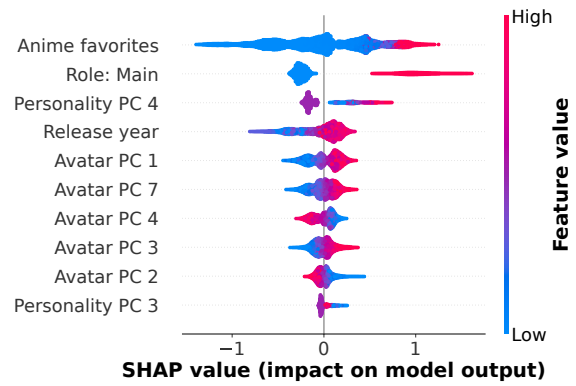


Figure 8: Beeswarm plot showing SHAP effect distributions for the top 10 features influencing character popularity prediction using HGBT (top 50 PCs from avatars and personalities plus metadata), ranked by importance.

Specifically, **visual design plays a much more crucial role than personality features**, likely because anime characters typically follow a fixed set of personality archetypes with limited variation. As illustrated in the figure, PC1, PC7, PC4, PC3, and PC2 show significant contrasts in SHAP values across different feature values. Based on our previous interpretation of these principal components of avatar visual features, characters are generally more favored when they exhibit softer, rounder, and more feminine facial structures (PC1), probably catering to the male preference for female characters with neotenous, *moe*-style features (Matsui 2022). Besides, those who wear masks, armors, or helmets (PC7) also demonstrate higher popularity, as this delicate and elaborate design of characters makes them more visually intriguing and cool to audiences. Artistic designs featuring proportionally slender bodies, oval faces (PC3), long hair, and goddess-like aesthetics (PC4) are also highly favored, potentially reflecting curvaceous female features intended to appeal to male viewers (Reysen et al. 2017; Liu et al. 2020).

Discussions

Here we would like to discuss in depth the implications of these empirical findings.

Anime for a Maturing Audience Our analysis indicates a systematic shift in anime’s target audience, moving from one primarily for children to a more maturing audience of teenagers and young adults (the *Shonen* population) in recent decades (**RQ1**). Such a shift has also been manifested in multiple aspects. The narrative has become more complex, with more supporting characters and more psychologically complex main characters. For example, the decline of manipulative anti-heroes signifies a deviation from the simple plots driven by the struggle against evil. The rise of complex and flawed protagonists like emotional introverts also suggests a move towards more relatable characters with intense emotional conflicts. Similarly, in terms of visual design, the characters are also designed with more idealized and exaggerated body ratios (Avatar PC3) and more sexualized de-

Features	PCs	Model	MAE	MPD
Baseline (mean)	–	–	1.190	2.166
Both PCs + Metadata	20	Poisson	0.897	1.416
Both PCs + Metadata	20	HGBT	0.669	0.946
Both PCs + Metadata	50	Poisson	0.886	1.385
Both PCs + Metadata	50	HGBT	0.668	0.945
Avatar PCs only	20	Poisson	1.022	1.725
Avatar PCs only	20	HGBT	1.015	1.693
Avatar PCs only	50	Poisson	1.001	1.666
Avatar PCs only	50	HGBT	1.008	1.673
Personality PCs only	20	Poisson	1.119	1.997
Personality PCs only	20	HGBT	0.972	1.590
Personality PCs only	50	Poisson	1.110	1.979
Personality PCs only	50	HGBT	0.970	1.585
Metadata only	–	Poisson	0.994	1.651
Metadata only	–	HGBT	0.779	1.204

Table 7: Results of predicting log-transformed favorite counts using avatar PCs, personality PCs, and metadata (anime favorites, gender, role, year). Lower MAE and MPD suggest better performance.

signs (Avatar PC3 and PC5). Yet anime with more explicit content for older adults is in decline, after a brief peak in the 2000s. Contemporary anime is mostly defined by the *Shonen* anime (Drummond-Mathews 2010).

Moe-ification of Anime Characters A prominent trend emerging from the data is the tendency of *moe*-ification (RQ2 & 3). In terms of personality traits, two archetypes, Reserved Introverts and Arrogant Tsunderes, have witnessed a noticeable rise in recent decades. These two archetypes, dominated by female characters, rank among the most popular archetypes. The appeal of introverts draws viewers through quiet charm and vulnerability, while tsundere characters display contrasting traits of arrogance and hidden warmth, an attractive feature known as `gap-moe`. The visual features that have undergone the most significant changes are all features related to *moe* or sexualized female traits (Avatar PC1 and PC3). The *moe*-ified visual features are also strongly predictive of audience preference on MyAnimeList, where the audience is mostly outside Japan and consists of 75% males. This suggests a more universal appeal of *moe* design among the male audience that transcends national boundaries.

Commercially, this has been known to be a deliberate choice of the designers driven by economic gains (Galbraith 2014). In the media-mix strategy (Steinberg 2012; Yatron 2022), anime is only a part of the commercialization scheme, and popular characters created in the anime will be further turned into various merchandise. Through commodifying affection, the anime studios can push for maximum profits from the primary consumer base, the Otaku community (McVeigh 1996), who would form a strong attachment to *moe* characters and are willing to consume character-related spin-offs. Yet, new from prior studies, our analysis

quantitatively reveals that the *moe*-ification trend becomes increasingly prominent after 2000, especially among main characters, despite the international expansion of audience (Noh 2024). This reflects an ongoing structural shift in the anime industry, a shift towards less diversity in content design and more pandering to the male Otaku community. It may create a self-reinforcing feedback loop, as the prevailing design conditions the preferences of the expanding market worldwide to align with such tastes, thereby shaping audience preferences across diverse cultural contexts, as well as global animation production practices.

Despite the commercial success, it is also hard to omit the gender biases and female objectification in such *moe* character design. This is coupled by the fact that female characters are more likely to be associated with stereotypical personality traits like naivety, shyness, supportiveness, and emotional volatility. Such highly conventional and objectifying portrayals of femininity, with some infantilized as immature, innocent, and docile (Avatar PC1) and others sexualized as highly sensual and curvaceous (Avatar PC3 and PC5), have long been criticized (Ting 2019; Lunning 2011; Reysen et al. 2017). Sustained exposure to stereotypical gender representations may contribute to adverse effects on self-image and mental health, as well as the normalization of sexism, especially among young female audiences (Ward and Grower 2020; Santoniccolo et al. 2023; Cho et al. 2025). However, the *moe* design cannot always be reduced to male-pandering, as some argue that the *moe* design can also be empowering, subverting societal norms, and asserting queer identities (Akgün 2020; Yatron 2022; Saito 2014).

Creativity in Cultural Products Japanese anime distinguishes itself from comics, cartoons, and other cultural products through the recognizable ‘anime-esque’ traits from a database of conventionalized models (Suan 2017; Ruh 2014; Azuma 2009). Therefore, anime creativity is often confined to local variations of the global uniformity and repetition of anime-esque elements (Suan 2017). In our analysis, the relative proportion of different personality archetypes are generally stable over time, reflecting an embrace of such uniformity. We also map out these local variations of anime character design and how they shift over time with large-scale computational methods. Only 20 PCs are sufficient to explain 57.88% of variance of character avatars.

Personality archetypes are often predictable from visual features, especially for certain flat archetypes (RQ4). This aligns well with empirical evidence from existing research that facial features are a significant factor in personality perception among the anime audience (Öze and Esengöl 2025), and that viewers develop learned expectations about how specific visual features correspond to personality (Chen and Chen 2015; Nakashima, Fujimoto, and Yamanishi 2023). Yet our findings indicate that the audience seems to prefer personality archetypes that are less predictable from their visual designs, or less conventionalized characters.

Our modeling of audience preferences demonstrates that audience preferences are driven by the integration of artistic styles and personality traits. Yet the visual features play a more dominant role than personality traits, even after

controlling for the meta-information like role and anime popularity (RQ5). Feminine traits such as *moe*-style faces, long stylish hair, and slender proportional bodies are associated with higher popularity. Characters with elaborate mechanical designs, like armor and giant humanoid robots (or *Mecha*), also contribute to popularity. Despite the large variation in character design, audience preference is consistently linked to certain aesthetic choices from the anime database (Ruh 2014; Azuma 2009). We also found these two design elements, *moe* and *Mecha*, have exhibited a growing trend in the visual design. Our findings also align with the observation that anime creativity is more on mixing and matching of conventional anime-esque elements (Suan 2021), a modular production style that ensures these artistic designs retain their appeal regardless of genre boundaries, the expanding creator base, or changing market trends.

Conclusions

In this study, we analyze the character design in Japanese anime with large-scale multimodal data from an anime review site. By combining textual, visual, and production features of anime characters with online popularity traces, we are able to reveal the historical trend of anime as a genre. Our results suggest that anime designers do utilize recurring tropes in constructing character personalities and designing their visual representations. The design has been undergoing *moe*-ification since the 2000s, suggesting increased gendered objectification of females in the modern anime production industry. This phenomenon might be worth further examination to understand the causes and its influence on the audience. Overall, our research demonstrates that large-scale analysis of online cultural artifacts can provide insights into cultural analytics.

Limitations

This study is still limited in several ways. First, the data were gathered from MyAnimeList, which could potentially be biased towards English cultures and values. Secondly, due to the crowd-sourced nature of the site, the representations of anime in MyAnimeList are biased towards popular and contemporary works. Thirdly, due to the limitation in data, we were not able to analyze the vocal design, which is also critical to the appeal of anime characters (Starr 2015). In addition, the incompleteness of genre and theme metadata precluded analyses across different genres. Finally, some findings may be affected by mediating factors that are unobserved in the curated dataset (e.g., cultural context, content rating mechanisms, and temporal factors), and should be interpreted with caution. Future research could potentially mitigate these biases with richer multimodal data from more diverse cultural communities.

Acknowledgments

We thank the reviewers and the area chair for their helpful comments. This research was enabled in part through the computational resources provided by Advanced Research Computing at the University of British Columbia and the Digital Research Alliance of Canada. The research activities

were also supported by the NSERC Discovery Grant and the CFI JELF Grant awarded to JZ.

References

- Abrams, G. 2025. Literary characters and GPT-4: from William Shakespeare to Elena Ferrante. *Digital Scholarship in the Humanities*, 40(1): 1–14.
- Akgün, B. 2020. Mythology *moe*-ified: classical witches, warriors, and monsters in Japanese manga. *Journal of Graphic Novels and Comics*.
- Amalvy, A.; Labatut, V.; and Dufour, R. 2025. The Role of Natural Language Processing Tasks in Automatic Literary Character Network Construction. In *Proceedings of the 31st International Conference on Computational Linguistics*, 8462–8473.
- Azuma, H. 2009. *Otaku: Japan's database animals*. U of Minnesota Press.
- Bresnahan, M. J.; Inoue, Y.; and Kagawa, N. 2006. Players and Whiners? Perceptions of Sex Stereotyping in Animé in Japan and the US. *Asian Journal of Communication*, 16(2): 207–217.
- Brzeski, P. 2025. Netflix Says 50 Percent of Global Users Now Watch Anime, Reveals Expanded Slate. <https://www.hollywoodreporter.com/tv/tv-news/netflix-anime-slate-2025-1236307714/>. The Hollywood Reporter. Accessed: 2026-03-19.
- Cao, Y.; Meng, X.; Mok, P.; Liu, X.; Lee, T.-Y.; and Li, P. 2023. Animediffusion: Anime face line drawing colorization via diffusion models. *arXiv preprint arXiv:2303.11137*.
- Chen, M.-H.; and Chen, I.-P. 2015. The Relationship Between Personalities and Faces of Manga Characters. *The Comics Grid: Journal of Comics Scholarship*, 4.
- Cho, L. Y.; Dougherty, M. R.; Roh, H. J.; and Harriger, J. A. 2025. Sexual Objectification, Sexualization, and Appearance-Related Themes in Anime: A Content Analysis. *Sex Roles*, 91(11): 77.
- Condry, I. 2009. Anime creativity: Characters and premises in the quest for cool Japan. *Theory, Culture & Society*, 26(2-3): 139–163.
- Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*.
- Douglass, J.; Huber, W.; and Manovich, L. 2011. Understanding scanlation: How to read one million fan-translated manga pages. *Image & Narrative*, 12(1): 190–227.
- Drummond-Mathews, A. 2010. What Boys Will Be: A Study of Shonen Manga. *Manga: An Anthology of Global and Cultural Perspectives*, 62.
- Fennell, D.; Liberato, A. S.; Hayden, B.; and Fujino, Y. 2013. Consuming anime. *Television & New Media*, 14(5): 440–456.
- FORCE11. 2020. The FAIR Data principles. <https://force11.org/info/the-fair-data-principles/>.

- Galbraith, P. W. 2014. *Moe Manifesto: An Insider's Look at the Worlds of Manga, Anime, and Gaming*. Tuttle Publishing.
- Gō, I.; and Nakamura, M. 2011. Tezuka Is Dead: Manga in Transformation and Its Dysfunctional Discourse. *Mechademia*, 6(1): 69–82.
- Hōsei, I.; and Bergstrom, B. 2023. The Prince of Tennis in 2-and 2.5 D: Character Growth and Growth into Character. *Mechademia*, 15(2): 32–46.
- Howell, P. 2007. Character voice in anime subtitles. *Perspectives*, 14(4): 292–305.
- Huang, Z.; Xie, H.; Fukusato, T.; and Miyata, K. 2023. Anifacedrawing: Anime portrait exploration during your sketching. In *ACM SIGGRAPH 2023 Conference Proceedings*, 1–11.
- Iijima, Y.; Wada, Y.; and Yamada, M. 2016. Proportion of Facial Features for Constructing a “Moe” Character. In *Advances in Affective and Pleasurable Design: Proceedings of the AHFE 2016 International Conference on Affective and Pleasurable Design*, 155–163. Springer.
- Jang, W.; and Jung, S. 2024. Evaluating llm performance in character analysis: A study of artificial beings in recent korean science fiction. In *Proceedings of the 4th international conference on natural language processing for digital humanities*, 339–351.
- Jin, Y.; Zhang, J.; Li, M.; Tian, Y.; Zhu, H.; and Fang, Z. 2017. Towards the automatic anime characters creation with generative adversarial networks. *arXiv preprint arXiv:1708.05509*.
- Kinoshita, S. 2024. Gundam and the Japanese Media Mix: Novelizations, Model Kits and Statues. *Journal of Anime and Manga Studies*.
- Kopylova, O. 2023. Drawing Works Together: Anime Media Mix and Graphic Style as a Shared Cultural Resource. *Mechademia*, 16(1): 190–211.
- Kwon, W.; Li, Z.; Zhuang, S.; Sheng, Y.; Zheng, L.; Yu, C. H.; Gonzalez, J.; Zhang, H.; and Stoica, I. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th Symposium on Operating Systems Principles*, 611–626.
- Li, B.; Zhu, Y.; Wang, Y.; Lin, C.-W.; Ghanem, B.; and Shen, L. 2021. Anigan: Style-guided generative adversarial networks for unsupervised anime face generation. *IEEE Transactions on Multimedia*, 24: 4077–4091.
- Li, H.; Guo, S.; Lyu, K.; Yang, X.; Chen, T.; Zhu, J.; and Zeng, H. 2022. A challenging benchmark of anime style recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4721–4730.
- Liu, K.; Chang, K.-M.; Liu, Y.-J.; and Chen, J.-H. 2020. Animated Character Style Investigation with Decision Tree Classification. *Symmetry*, 12(8).
- Lu, A. S. 2009. What race do they represent and does mine have anything to do with it? Perceived racial categories of anime characters. *Animation*, 4(2): 169–190.
- Lundberg, S.; and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. *arXiv:1705.07874*.
- Lunning, F. 2011. Under the Ruffles: Shōjo and the Morphology of Power. *Mechademia*, 6(1): 3–19.
- Manovich, L. 2012. How to compare one million images? In *Understanding digital humanities*, 249–278. Springer.
- Matsui, T. 2022. Moe-Phobia: Effect of Users’ Gender on Perceived Sexuality and Likability Toward Manga-Like Virtual Agents. *Frontiers in Psychology*, 13: 752748.
- McInnes, L.; Healy, J.; and Melville, J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
- McVeigh, B. 1996. Commodifying affection, authority and gender in the everyday objects of Japan. *Journal of Material Culture*, 1(3): 291–312.
- Nakashima, F.; Fujimoto, N.; and Yamanishi, R. 2023. Do Appearances Represent Personality of Created Characters. *Procedia Computer Science*, 225: 1776–1785. 27th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES 2023).
- Napier, S. 2011. Manga and anime: entertainment, big business, and art in Japan. In *Routledge handbook of Japanese culture and society*, 226–237. Routledge.
- Noh, S. 2024. Global Media Streams: Netflix and the Changing Ecosystem of Anime Production. *Television & New Media*, 25(3): 234–250.
- Norris, C. 2009. 13 Manga, anime and visual art culture. *The Cambridge companion to modern Japanese culture*.
- Okamoto, T. 2015. Otaku tourism and the anime pilgrimage phenomenon in Japan. *Japan Forum*, 27(1): 12–36.
- Oóhagan, D. M. 2007. Manga, anime and video games: Globalizing Japanese cultural production. *Perspectives*, 14(4): 242–247.
- Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. 2024. DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research Journal*, 1–31.
- Papoudakis, A.; Lapata, M.; and Keller, F. 2024. BookWorm: A dataset for character description and analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, 4471–4500.
- Pei, J.; Ananthasubramaniam, A.; Wang, X.; Zhou, N.; Dedeloudis, A.; Sargent, J.; and Jurgens, D. 2022. POTATO: The Portable Text Annotation Tool. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*.
- Ramasubramanian, S.; and Kornfield, S. 2012. Japanese anime heroines as role models for US youth: Wishful identification, parasocial interaction, and intercultural entertainment effects. *Journal of International and Intercultural Communication*, 5(3): 189–207.
- Reysen, S.; Katzarska-Miller, I.; Plante, C.; Roberts, S.; and Gerbasi, K. 2017. Examination of Anime Content and Associations between Anime Consumption, Genre Preferences, and Ambivalent Sexism. *The Phoenix Papers*, 3: 285–303.
- Ruh, B. 2014. Conceptualizing anime and the database fansyscape. *Mechademia: Second Arc*, 9: 164–175.

Saito, A. P. 2017. Moe and internet memes: The resistance and accommodation of Japanese popular culture in China. *Cultural studies review*, 23(1): 136–150.

Saito, K. 2014. Magic, Shōjo, and metamorphosis: Magical girl anime and the challenges of changing gender identities in Japanese society. *The Journal of Asian Studies*, 73(1): 143–164.

Santoniccolo, F.; Trombetta, T.; Paradiso, M. N.; and Rollè, L. 2023. Gender and Media Representations: A Review of the Literature on Gender Stereotypes, Objectification and Sexualization. *International Journal of Environmental Research and Public Health*, 20(10).

Starr, R. L. 2015. Sweet voice: The role of voice quality in a Japanese feminine style. *Language in Society*, 44(1): 1–34.

Steinberg, M. 2012. *Anime’s media mix: Franchising toys and characters in Japan*. U of Minnesota Press.

Steinberg, M. 2023. Introducing the Media Mix. *Mechademia*, 16(1): 1–11.

Suan, S. 2017. Anime’s performativity: Diversity through conventionality in a global media-form. *Animation*, 12(1): 62–79.

Suan, S. 2021. *Anime’s identity: performativity and form beyond Japan*. U of Minnesota Press.

Tang, R.; Yang, P.; Miyauchi, R.; and Inoue, Y. 2024. The correlation between the MBTI-based personality analysis of anime characters and their popularity. *Behavioral Sciences*, 14(7): 522.

Ting, G. E.-Y. 2019. Gender, manga, and anime. In *The Routledge Companion to Gender and Japanese Culture*, 311–319. Routledge.

Traag, V. A.; Waltman, L.; and van Eck, N. J. 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1).

Ward, L. M.; and Grower, P. 2020. Media and the development of gender role stereotypes. *Annual Review of Developmental Psychology*, 2(1): 177–199.

Watzky, M. 2023. The dynamism of anime images: The case of the ‘Kanada-style’ movement. *Journal of Anime and Manga Studies*, vol. 4.

Winge, T. 2006. Costuming the imagination: Origins of anime and manga cosplay. *Mechademia*, 1(1): 65–76.

Yang, A.; Li, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Gao, C.; Huang, C.; Lv, C.; et al. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Yatron, C. 2022. 30 years later, re-examining the “Pretty Soldier”: A gender study analysis of Sailor Moon. *Journal of Anime and Manga Studies*, vol. 3.

Zhang, Y.; Li, M.; Long, D.; Zhang, X.; Lin, H.; Yang, B.; Xie, P.; Yang, A.; Liu, D.; Lin, J.; et al. 2025. Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models. *arXiv preprint arXiv:2506.05176*.

Zhao, R.; Diep, B.; Pei, J.; Yoon, D.; Jurgens, D.; and Zhu, J. 2025. Who Reaps All the Superchats? A Large-Scale Analysis of Income Inequality in Virtual YouTuber Livestreaming.

In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI ’25.

Öze, N.; and Esengöl, G. 2025. Perception of Anime Cartoon Characters Depending on Their Visual Traits and Facial Features. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 27(1): 82–108.

Paper Checklist

1. For most authors...
 - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **Yes. Our study analyzes anime characters using only public data, acknowledges cultural biases, and involves no privacy or ethical violations.**
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper’s contributions and scope? **Yes.**
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes. See the Method section.**
 - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes. We discuss the potential biases from MyAnimeList’s English-oriented community and how representations of anime are skewed toward popular and contemporary works in the Limitation section.**
 - (e) Did you describe the limitations of your work? **Yes. See the Limitations section.**
 - (f) Did you discuss any potential negative societal impacts of your work? **No. Our research analyzes fictional characters using public data and does not pose societal risks.**
 - (g) Did you discuss any potential misuse of your work? **No. Our findings are limited on fictional characters and do not enable harmful applications or misuses.**
 - (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **NA.**
 - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes.**
2. Additionally, if your study involves hypotheses testing...
 - (a) Did you clearly state the assumptions underlying all theoretical results? **NA.**
 - (b) Have you provided justifications for all theoretical results? **NA.**
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **NA.**
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **NA.**

- (e) Did you address potential biases or limitations in your theoretical framework? *NA*.
 - (f) Have you related your theoretical results to the existing literature in social science? *NA*.
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? *NA*.
3. Additionally, if you are including theoretical proofs...
- (a) Did you state the full set of assumptions of all theoretical results? *NA*.
 - (b) Did you include complete proofs of all theoretical results? *NA*.
4. Additionally, if you ran machine learning experiments...
- (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? *Yes*.
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? *Yes. The data splits are listed in the main text, and the fine-tuning hyperparameters for DINOv2 are summarized in Table 2. More details can be found in the code provided in the GitHub repository.*
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? *No*.
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? *Yes, specified in the Resources section in the appendix.*
 - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? *Yes. As specified in the Method section, our evaluations are tailored to the research questions and directly support our claims.*
 - (f) Do you discuss what is “the cost“ of misclassification and fault (in)tolerance? *NA*.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
- (a) If your work uses existing assets, did you cite the creators? *Yes. We have cited MyAnimeList for data, as well as pretrained models and relevant prior works used for feature extraction and analysis.*
 - (b) Did you mention the license of the assets? *Yes, details about the licenses for the models (Camie Tagger, dinov2-base, Qwen3-32B-FP8, and Qwen3-Embedding-0.6B) are provided in the Licenses section in the appendix.*
 - (c) Did you include any new assets in the supplemental material or as a URL? *No. We do not release any new assets.*
 - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? *NA*.
- (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? *No. All data come from publicly available sources on MyAnimeList and do not contain personal information or offensive content.*
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? *NA*.
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? *NA*.
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
- (a) Did you include the full text of instructions given to participants and screenshots? *Yes. The full human annotation instructions and an interface screenshot are provided in the Appendix.*
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? *NA*.
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? *No. Since the annotation task was performed by two of the authors, no participant compensation was required.*
 - (d) Did you discuss how data is stored, shared, and de-identified? *Yes. Details are provided in the Appendix.*

Licenses

Pretrained models used in this study include Camie Tagger, which is distributed under the GNU General Public License v3.0, as well as DINOv2-base, Qwen3-32B-FP8, and Qwen3-Embedding-0.6B, all of which are released under the Apache License 2.0.

Resources

All model training and experiments were conducted on an institutional high-performance computing cluster using the 20 GB H100 GPU allocation, providing sufficient capacity for both LLM annotation and Vision Transformer fine-tuning. Other computations, such as community detection and principal component analysis, were executed using up to 16 CPUs and 32GB of RAM on the cluster.

Data Storage and Privacy

All data were stored on the cluster and were only accessible to the authors. The data contains no personally identifiable information, and no de-identification was required.

Disclosure on the Use of Generative AI

We confirm that all text in the manuscript was written by the authors, with no original intellectual contributions generated by AI. We solely utilized AI tools for assisting with research code, grammar and spelling checks, as well as translating obscure vocabulary and rephrasing.

LLM Prompts

LLM Prompt for character personality trait extraction

You are an expert Japanese anime character analyst. Based on the character's biography, provide a structured JSON description.

Describe personality traits using a small set of Japanese anime-style keywords in two languages:

- Japanese: short descriptors commonly used in anime character analysis.
- English: natural English equivalents in the same order.

Gender: M / F / Unknown.

Character JSON: the character JSON filename, including the .json extension.

Instructions:

- Output only the JSON object. Do not include any explanation, commentary, or extra text.
- If the biography contains clear details to infer personality traits, extract them normally.
- If the biography does not provide enough information to infer personality traits, then:
 - Set “personality_keywords” to {“Japanese”: [], “English”: []}.
- For gender inference:
 - Set “gender” to M or F only if clearly stated or implied (e.g., pronouns).
 - Otherwise, set “gender” to “Unknown”.
- Do not guess or assume any information not supported by the biography.

Output Structure:

```
{
  "character_name": "",
  "personality_keywords": {
    "Japanese": [],
    "English": []
  },
  "gender": "",
  "character_json": ""
}
```

Now analyze this character:

Character name: {character_name}

Biography: {biography}

Character JSON: {character_json}

Annotation Interface Screenshot

Keyword Annotation - Test Sample

Lirio(リリオ)

Lirio is a child taken care of by Ricardo. She is aware of Ricardo's bounty hunting profession and is not seemed to be bothered by it. Each time Ricardo goes out to hunt a bounty, she waits patiently for him to return. She is very fond of Nadie and Ellis as well. At the end of the series, Ricardo teaches her how to use a boomerang. She very rarely speaks and instead, communicates with other people by smiling, facial expressions and nodding or shaking her head.

Select all keywords that accurately describe the character's personality based on their biography.

- Slightly Tsundere
- Denial
- Order-Obsessed
- Hierarchical
- Wealthy-raised
- Loves cute things
- Split personality
- Fighting fate
- Reserved
- Mysterious smile
- Adores her sister
- Patient
- Slightly crazy
- Otome
- Quiet
- Eroge Fan
- Dedicated to research
- Kind
- Anti-materialistic
- Tyrant
- None of the above

Move backward

Move forward

Figure 9: Screenshot of the Potato annotation interface used in the model validation task. The example shown is for illustration only and was not part of the annotated set.

Annotation Guidelines

Annotation Guideline for Model Validation

The goal of this task is to assess whether personality keywords extracted by the Qwen3-32B-FP8 model are accurate based on character biographies. For each character, you will be shown:

- A character biography.
- A shuffled list of candidate personality keywords.

Instructions:

- Read the biography carefully.
- Select all keywords that are clearly supported by the biography.
- If none apply, select “None of the above”.

Inter-Annotator Agreement Metrics

Precision	Recall	F1	Jaccard
0.649	0.770	0.705	0.551

Table 8: Inter-annotator agreement metrics between the two independent human annotators for 50 random samples. To compute directional metrics, Annotator 1 is treated as the reference and Annotator 2 as the prediction: Precision (fraction of Annotator 2’s selected keywords chosen by Annotator 1), Recall (fraction of Annotator 1’s selected keywords chosen by Annotator 2), F1-score, and Jaccard similarity (intersection over union of selected keyword sets). Given the great complexity of the multi-label annotation task, these metrics indicate a high degree of inter-annotator agreement.

Yearly Proportion of Anime Sources

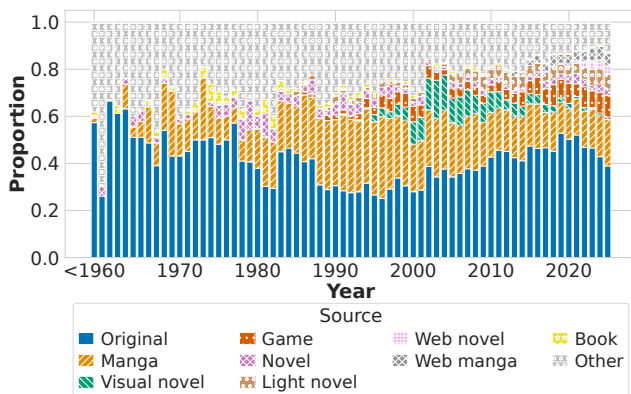


Figure 10: Proportion of source types by year. The sources of anime creation have been diversified in recent years. In addition to original creation and manga adaptation, new anime series are also adapted from games, light novels, web novels, and visual novels.

Character Archetype Community Detection

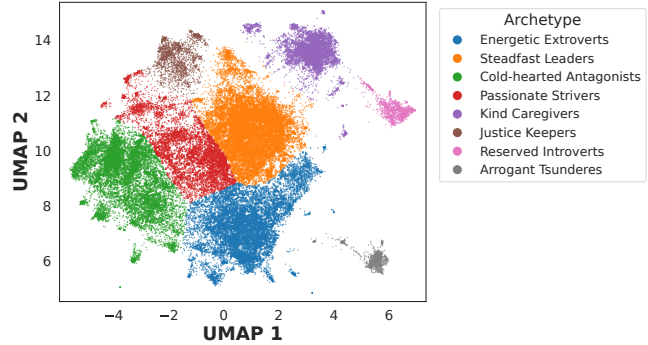


Figure 11: Character archetype communities obtained by applying Leiden algorithm to a distance-thresholded (5th percentile) graph over the 2D UMAP coordinates; marker size reflects node degree centrality. The modularity score achieves 0.72, suggesting a strong community structure with highly separable clusters.

Character Archetypes by Anime Rating

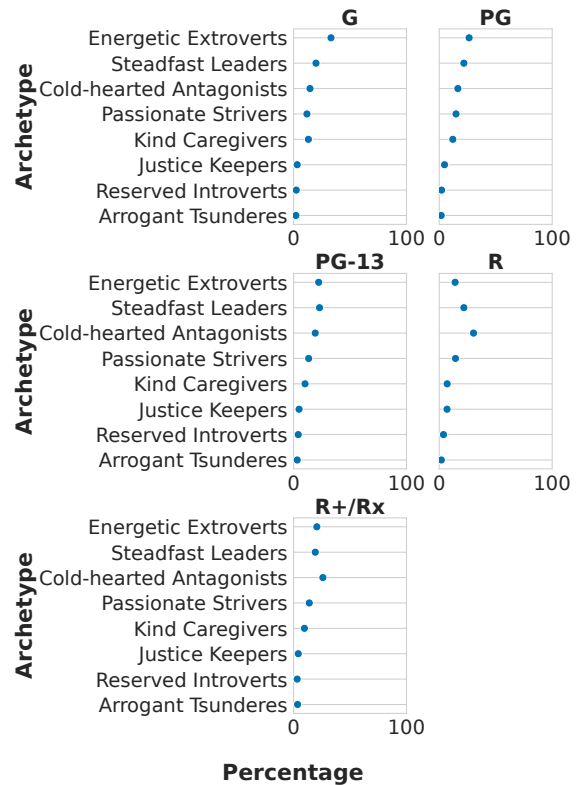


Figure 12: Archetype share across rating categories. More mature anime features darker antagonistic archetypes.

Character Archetypes by Anime Source

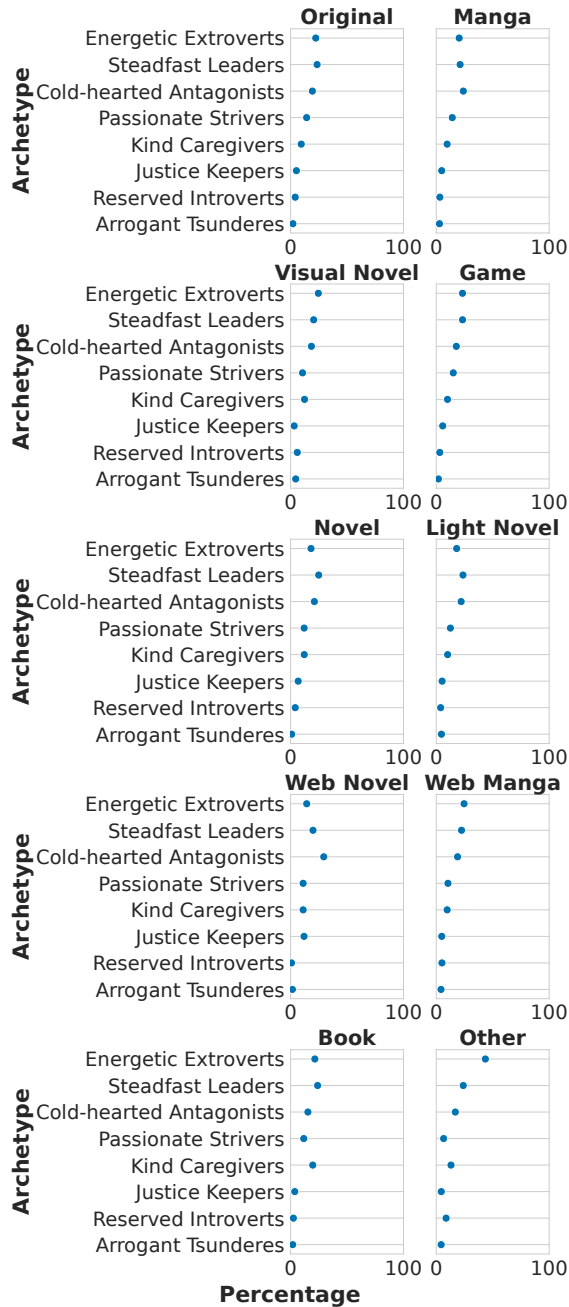


Figure 13: Archetype share across anime sources. Most sources maintain a diverse mix of character types. Web novel adaptations feature more Cold-hearted Antagonists and fewer Energetic Extroverts, reflecting darker narratives, while visual novel and web manga adaptations have more Energetic Extroverts, highlighting more optimistic themes.

Top/Bottom Characters in Top 10 Avatar PCs

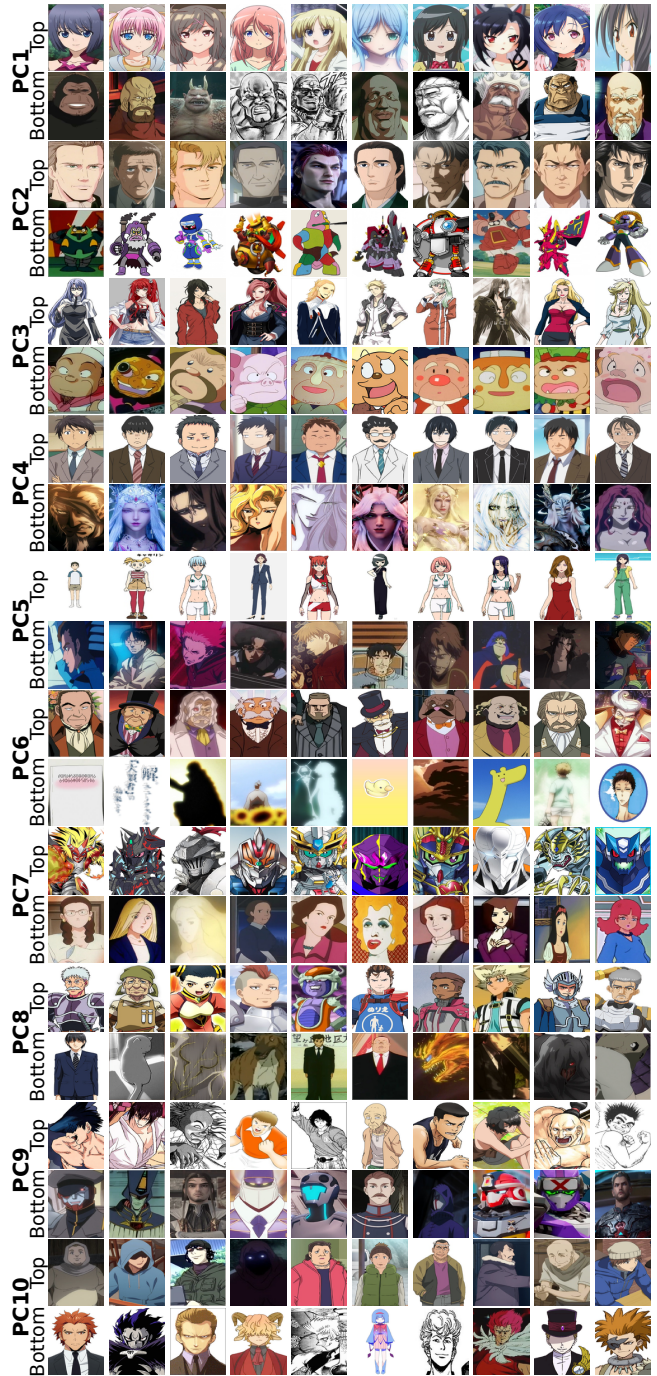


Figure 14: Top 10 and bottom 10 character examples for the first 10 principal components of avatar embeddings.

Confusion Matrix for Temporal Period Prediction

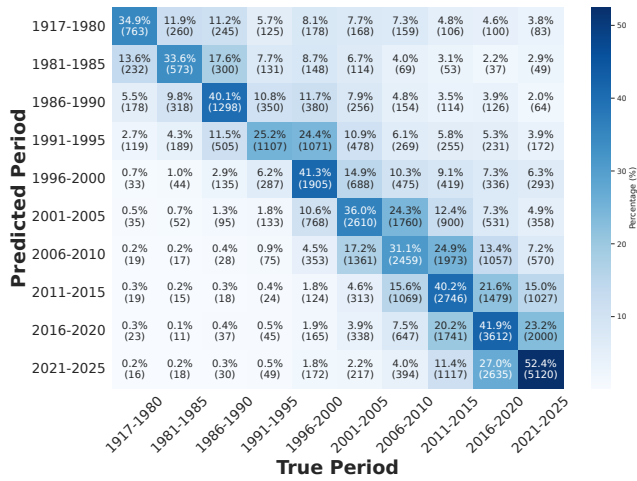


Figure 15: Confusion matrix of period prediction for the original avatars, showing that visual features can distinguish temporal periods, with most errors between adjacent periods.

Confusion Matrix for Personality Archetype Prediction

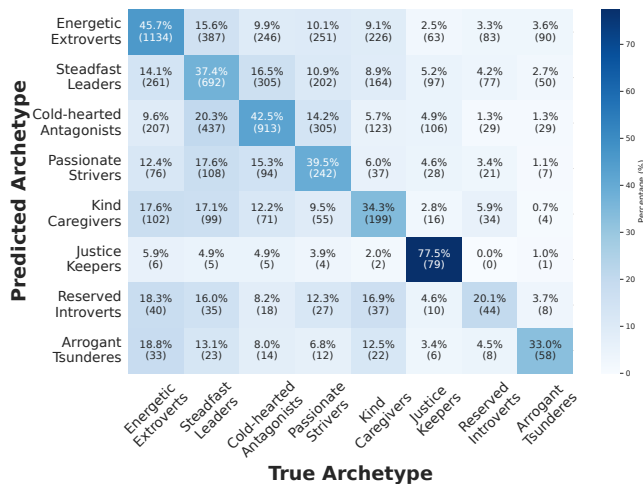


Figure 16: Confusion matrix for personality archetype prediction from the first 100 principal components of DINOv2 avatar embeddings. Justice Keepers, Energetic Extroverts, and Cold-hearted Antagonists are most predictable.

Full Results for User Preference Modeling

Features	PCs	Model	MAE	MPD
Baseline (mean)	–	–	1.190	2.166
Both PCs	20	Poisson	0.984	1.640
Both PCs	20	HGBT	0.881	1.354
Both PCs	50	Poisson	0.963	1.583
Both PCs	50	HGBT	0.880	1.345
Both PCs	100	Poisson	0.941	1.522
Both PCs	100	HGBT	0.879	1.340
Both PCs + Meta	20	Poisson	0.897	1.416
Both PCs + Meta	20	HGBT	0.669	0.946
Both PCs + Meta	50	Poisson	0.886	1.385
Both PCs + Meta	50	HGBT	0.668	0.945
Both PCs + Meta	100	Poisson	0.868	1.343
Both PCs + Meta	100	HGBT	0.669	0.946
Avatar	20	Poisson	1.022	1.725
Avatar	20	HGBT	1.015	1.693
Avatar	50	Poisson	1.001	1.666
Avatar	50	HGBT	1.008	1.673
Avatar	100	Poisson	0.978	1.601
Avatar	100	HGBT	1.007	1.663
Avatar + Meta	20	Poisson	0.921	1.461
Avatar + Meta	20	HGBT	0.724	1.071
Avatar + Meta	50	Poisson	0.911	1.432
Avatar + Meta	50	HGBT	0.723	1.069
Avatar + Meta	100	Poisson	0.893	1.387
Avatar + Meta	100	HGBT	0.724	1.069
Personality	20	Poisson	1.119	1.997
Personality	20	HGBT	0.972	1.590
Personality	50	Poisson	1.110	1.979
Personality	50	HGBT	0.970	1.585
Personality	100	Poisson	1.102	1.964
Personality	100	HGBT	0.969	1.583
Personality + Meta	20	Poisson	0.958	1.575
Personality + Meta	20	HGBT	0.699	1.022
Personality + Meta	50	Poisson	0.953	1.564
Personality + Meta	50	HGBT	0.699	1.022
Personality + Meta	100	Poisson	0.950	1.558
Personality + Meta	100	HGBT	0.699	1.022
Meta only	–	Poisson	0.994	1.651
Meta only	–	HGBT	0.779	1.204

Table 9: Full results of predicting log-transformed favorite counts using avatar PCs, personality PCs, and metadata (anime favorites, gender, role, year).