

Analysis of User Preferences between AI-generated and Hand-drawn Artwork: A Case Study of pixiv.net

Yiqing Hu¹, Rahila Mammadova Huseynova¹

¹University of Warsaw, Faculty of Economics Sciences

Abstract

The emergence of AI drawing models, such as Stable-Diffusion (SD) and Midjourney, has led to a significant influx of AI-generated images in the online art community. In this study, we collected daily rankings of the top 50 AI-generated images and hand-drawn artwork from pixiv.net. Our objective was to explore the factors/variables influencing the ranking of these images and examine the preferences of ordinary users towards AI-generated images versus hand-drawn artwork using an ordered choice model. The ordered choice model allows us to analyze the ordinal nature of the rankings and investigate the impact of various factors on the preferences of users. The findings from this study will contribute to a better understanding of the impact of AI-generated artwork on user preferences and provide insights for the development of AI-assisted creative platforms.

Keywords: Ordered Choice Model, Artificial Intelligence, Generative AI

Introduction

Generative AI models have revolutionized various industries, including the labor market, with their transformative capabilities[1]. However, the influence of AI-generated artwork on user preferences within the online art community remains a topic of significant interest. As the prevalence of AI drawing models, such as Stable-Diffusion (SD) and Midjourney, continues to grow, it becomes crucial to investigate the factors influencing the rankings of AI-generated images and understand user preferences in comparison to hand-drawn artwork. Therefore, our study aims to explore these factors and preferences through the application of an ordered choice model.

The primary hypothesis of our study is that AI-generated artwork significantly influences user preferences and rankings within the online art community. We expect that certain factors, such as the content of AI-generated images, their novelty, and the interaction between users and AI algorithms, will play a crucial role in shaping these preferences. Additionally, we propose a secondary hypothesis that user demographics, art genre preferences, and previous exposure to AI-generated artwork will also impact their rankings and preferences.

Our primary hypothesis focuses on comparing the effects of AI-generated and man-made art on rankings within the online art community. To investigate this, we analyze various user behaviors such as liked rates, bookmarked rates, and the content indicated by artwork tags. By comparing these features between AI-generated and man-made artwork, we can gain valuable insights into the unique characteristics and influences of AI-generated creations. This analysis allows us to

assess whether AI-generated art has a distinct impact on user preferences and rankings compared to traditional hand-drawn art.

Additionally, we propose a secondary hypothesis that explores which user behaviors and specific content attributes contribute to higher rankings for AI-generated artwork. By examining factors such as user interactions, engagement patterns, and the content of the artwork itself, we aim to identify the elements that attract users and lead to greater popularity and preference for AI-generated art.

The outcomes of this study hold significant implications for various stakeholders in the art community. Artists and creators can gain insights into the factors that make AI-generated artwork more appealing to users, helping them refine their creative processes and incorporate AI-assisted techniques effectively. Art enthusiasts and platform developers can utilize the findings to enhance user experiences and curate content that aligns with user preferences. Furthermore, policymakers and researchers can leverage this knowledge to shape policies and regulations that promote the integration of AI-generated art while preserving the diversity and quality of traditional art forms.

In this study, we utilize a comprehensive data set obtained from [pixiv.net](https://www.pixiv.net), a prominent platform where artists share their artwork. The dataset covers a period starting from November of the previous year, capturing the continuous influx of AI-generated works. It includes essential information such as tags, Views, Likes, Bookmarks, and Comments for the Top 50 artworks or pictures each day. This rich data set allows us to examine the dynamic relationship between AI-generated artwork and user preferences over time, providing valuable insights into the evolving landscape of the online art community.

To analyze the data, we employ established econometric techniques[2], including logistic regression (MASS::polr) and probit regression (oglmx::ologit.reg and ordinal::clm). We also conduct Goodness-of-fit tests, such as the Hosmer-Lemeshow, Lipsitz, and Pulkstenis-Robinson tests, to assess the initial performance of the models. Variable selection is conducted using the lrttest for Likelihood and anova for the general-to-specific method [3]. The evaluation of the models includes examining the impact of various variables on the rankings of AI-generated images and conducting Marginal Effects analysis specifically for the AI-generated model.

By employing this comprehensive analytical framework, our study aims to provide a nuanced understanding of the intricate dynamics between AI-generated artwork and user preferences. The subsequent sections will present the empirical findings, contributing to the literature on the effects of AI-generated artwork within the online art community.

Literature review

Data

In this study, we developed a web spider to crawl the dataset from [pixiv.net](https://www.pixiv.net), a prominent online platform where artists share their artwork. Pixiv.net is widely recognized and highly regarded within the online art community, making it an ideal source for collecting data related to AI-generated artwork and user preferences.

Pixiv.net provides a rich and diverse collection of artwork, including both AI-generated and hand-drawn creations. Artists from various backgrounds and genres contribute to the platform, resulting in a vast repository of artistic expressions.

The dataset we collected covers a specific period, starting from November of the previous year and spanning a continuous influx of AI-generated works. It comprises essential information related to the artworks, such as tags, views, likes, bookmarks, and comments. Specifically, we focused on the Top 50 artworks or pictures each day, ensuring a comprehensive representation of the most popular and engaging content within the online art community.

By utilizing the data from pixiv.net, we are able to examine the dynamic relationship between AI-generated artwork and user preferences over time. This allows us to gain valuable insights into the evolving landscape of the online art community and understand the factors that influence the rankings and preferences of AI-generated images compared to hand-drawn artwork.

Samples

We collected the data spanning from October 31, 2022, to May 15, 2023 from the top list of AI-generated and man-made image pages. After de-duplicating same image pages which may appear in top with differnet and different day, we gathered the samples:

- Number of all samples: 14576
- Number of samples of AI-generated Artworks: 8092
- Number of samples of Hand-drawn or man-made Artworks: 6484

Variables

Variable	Type	Description
pid	integer	artworkpage id
date	integer	date of being top50
like_rate	numeric	ratio of liked amount of viewed amount
mark_rate	numeric	ratio of bookmarked number of viewed number
is_comic	factor	whether the artwork is comic
is_ai	integer	whether the artwork is generated by AI
is_Genshin	factor	whether the artwork is about Genshin
is_Honkai	factor	whether the artwork is about Honkai
comments	integer	comment amount
views	numeric	viewed amount in thousand
rank	numeric	dependent variable
top_cnt	integer	how many times being top50 for the same artwork
date_diff_day	integer	date difference between created and being top
like_rate2	numeric	power of like_rate
mark_rate2	numeric	power of mark_rate

Covariates

Our research focuses on investigating the impact of various independent variables on the rankings of daily top 50 artworks on pixiv.net. The primary objective of our study is to develop an evaluation model that effectively captures the factors influencing these rankings.

To address the issue of excessively detailed rankings ranging from 1 to 50, we have categorized the rankings into five levels by equally dividing the range. This categorization allows for a more manageable and meaningful analysis. The formula used to categorize the ranks is as follows:

$$\text{tier} = \left\lfloor \frac{\text{rank} - 1}{10} \right\rfloor + 1$$

By categorizing the ranks, we transform the dependent variable into an ordered, discrete, and continuous variable with five levels. This enables us to better understand the impact of the independent variables on the categorized rankings.

Method/Model

Given the nature of the categorized rank as an ordered, continuous, and discrete variable, we employ an ordered choice model in our analysis. Specifically, we utilize both the ordered logit and ordered probit models to examine the relationships between the independent variables and the categorized rankings effectively.

In our analysis, we utilized the R programming language to fit a logit model to the data. Specifically, we employed the logit function `polr`¹, which is suitable for estimating ordered choice models.

To assess the feasibility and goodness-of-fit of the baseline models, several Goodness-of-Fit tests were conducted. The tests employed were the Hosmer-Lemeshow test, the Lipsitz test, and the Pulkstenis-Robinson test. These tests are commonly used in empirical research to evaluate the fit of logistic regression models and ordered choice models.

The Hosmer-Lemeshow test examines the agreement between the observed and predicted probabilities by dividing the data into several groups based on predicted probabilities and assessing the differences between the observed and expected frequencies within each group. A non-significant p-value indicates a good fit between the model and the observed data.

The Lipsitz test is specifically designed for ordered choice models and assesses the proportional odds assumption. This assumption states that the relationship between the predictors and the response variable is constant across all levels of the response variable. Violation of this assumption may indicate the need for a different modeling approach. The Lipsitz test calculates a chi-square statistic to evaluate the fit of the model to the proportional odds assumption.

The Pulkstenis-Robinson test is another test for assessing the proportional odds assumption in ordered choice models. It examines the parallelism of the cumulative logits across response categories. A non-significant result suggests that the model satisfies the proportional odds assumption.

Based on the results of the Goodness-of-Fit tests, if the baseline models do not demonstrate a satisfactory fit, model optimization techniques will be employed to improve the fit and meet the Goodness-of-Fit criteria.

Furthermore, the Brant test will be used to determine the appropriate model type for our ordered choice model. The Brant test examines the assumption of proportional odds and helps decide whether to use a Logit or Probit model. The test assesses the parallel regression assumption by comparing the coefficients from the Logit and Probit models. If the coefficients are not significantly different, indicating parallel regression, the Logit model will be selected due to its more straightforward interpretation. On the other hand, if the coefficients are significantly different, suggesting non-parallel regression, the Probit model will be chosen.

¹Please refer to the Appendix of the paper for the relevant code pertaining to the `polr` algorithm. for logit model.

The selection of the appropriate model type (Logit or Probit) is crucial as it affects the interpretation of the estimated coefficients and the overall model performance. By conducting the Brant test, we ensure that the chosen model type aligns with the underlying assumptions of the data and provides valid and reliable results for our analysis.

By employing these modeling techniques, we aim to gain valuable insights into the influence of the independent variables on the rankings of artworks on pixiv.net. These methodological considerations adhere to the standards expected in academic research and contribute to the advancement of knowledge in the field.

Results

Goodness-of-fit tests

Hosmer-Lemeshow and Lipsitz tests

To assess the appropriateness of our model, we conducted Hosmer-Lemeshow and Lipsitz tests. These tests examine whether the form of our model adequately fits the data. If either of these tests indicates that the model is inappropriate, it suggests the presence of issues that need to be addressed.

The H0 (null hypothesis) of the Lipsitz test and logitgof test is that the form of our model is appropriate for the data. If either test indicates that the H0 cannot be rejected, it implies that the model may have a problem, and corrective measures are needed. Only when both tests indicate that the model is appropriate can we confidently state that the model is suitable for our analysis. we imply Lipsitz and logitgof tests² in R to check the hypothesis.

Lipsitz goodness of fit test for ordinal response models

X-squared = 38.3353316764042, p-value = 0.0000151747246179834

Hosmer and Lemeshow test (ordinal model)

X-squared = 35.7491570936594, p-value = 0.00192116136889442

The Lipsitz test was performed, and the resulting p-value is 0.0000002942, which is less than the significance level of 5%. Therefore, we cannot reject the H0, indicating that the form of our model has a problem.

Similarly, the logitgof test (Hosmer-Lemeshow test) was conducted with the model's fitted values, and the resulting p-value is 0.00007257, also less than the significance level of 5%. Consequently, we cannot reject the H0, suggesting that the form of our model has a problem.

These goodness-of-fit tests provide valuable insights into the adequacy of our model. Their outcomes indicate the presence of issues that need to be addressed to improve the model's fit to the data. By recognizing and addressing these problems, we can refine our model and enhance its reliability for further analysis. These methodological considerations align with the rigorous standards expected in academic research, contributing to the robustness of our findings.

²Please refer to the Appendix of the paper for the relevant code pertaining to the lipsitz and Lipsitz tests.

Pulkstenis-Robinson tests

In addition to the Hosmer-Lemeshow and Lipsitz tests, we also conducted Pulkstenis-Robinson tests to assess the goodness of fit of our model. These tests are particularly suitable when dealing with models that include dummy variables, such as our model, which incorporates the dummy variables “is_comic,” “is_Genshin,” and “is_Honkai.” The Pulkstenis-Robinson tests provide valuable insights into the model’s fit and the impact of additional predictors.

The Pulkstenis-Robinson chi-squared test evaluates the null hypothesis that there is no significant departure from the expected frequencies based on the ordinal logistic regression model. In simpler terms, this test assesses whether the model fits the observed data adequately, without evidence of lack of fit. The Pulkstenis-Robinson deviance test, on the other hand, compares the more complex model (our model with additional predictors) to a simpler baseline model to determine if the additional predictors significantly improve the model fit.

Pulkstenis-Robinson chi-squared test

X-squared = 62.5396996656291, p-value = 0.0343243225533081

Pulkstenis-Robinson deviance test

X-squared = 66.3154376991681, p-value = 0.0164308098416701

Upon conducting the Pulkstenis-Robinson tests³, we obtained the following results:

Pulkstenis-Robinson chi-squared test: The p-value was calculated as 0.03315, which is less than the significance level of 5%. Therefore, we reject the null hypothesis and conclude that the model does not fit the observed data well.

Pulkstenis-Robinson deviance test: The p-value was found to be 0.01776, also below the significance level of 5%. Consequently, we reject the null hypothesis, indicating that the additional predictors in the more complex model significantly improve the model fit.

Based on the results of these goodness-of-fit tests, it becomes apparent that the initial model for our task is not appropriate. This finding suggests the need for further refinement and consideration of potential issues such as feedback or endogeneity between views and ranks. To explore this, we removed the parameter of ‘views’ from the model and reevaluated the goodness of fit using the Pulkstenis-Robinson tests.

Pulkstenis-Robinson chi-squared test

X-squared = 49.5589337835266, p-value = 0.261168696410505

Pulkstenis-Robinson deviance test

X-squared = 51.0261024770157, p-value = 0.21689897321962

Upon reanalysis, we found that the p-values for both the Pulkstenis-Robinson chi-squared and deviance tests are now larger than 5%. As a result, we cannot reject the null hypothesis anymore, indicating that the model without ‘views’ is deemed appropriate.

These rigorous evaluations of the model’s fit contribute to the reliability of our findings. By recognizing and addressing the limitations and potential endogeneity issues, we can enhance the robustness and validity of our evaluation model.

³Please refer to the Appendix of the paper for the relevant code pertaining to the pulkrob tests.

Odds Assumption Test

Brant's test

To assess the validity of the proportional odds assumption, we conducted Brant's test. This test specifically examines the parallel regression assumption in the context of the polr model. The null hypothesis of Brant's test is that the parallel regression assumption holds.

Test for	X2	df	probability
Omnibus	67.41	24	0
like_rate	24.02	3	0
mark_rate	23.95	3	0
is_comic1	0.73	3	0.87
is_Genshin1	3.56	3	0.31
is_Honkai1	1.72	3	0.63
comments	15.3	3	0
top_cnt	27.15	3	0
date_diff_day	8.96	3	0.03

H0: Parallel Regression Assumption holds

By applying Brant's test⁴ to our model, we obtained the following results:

- Omnibus test: The test statistic (X2) was calculated as 67.41, with a corresponding p-value of 0. As a result, we reject the null hypothesis, indicating that the parallel regression assumption does not hold for our model. This finding suggests that the ordered logit model is not suitable for our analysis.

Furthermore, it is important to consider the variables associated with the rejection of the test, as indicated by their probabilities. Variables with probabilities higher than 5% are considered connected to the rejection of the test.

Given that the proportional odds assumption does not hold, we need to explore alternative models. In this case, we opted for the ordered probit model by using clm algorithm⁵, which does not rely on the proportional odds assumption.

To validate the appropriateness of the ordered probit model, we conducted the Pulkstenis-Robinson tests for goodness of fit. The results of these tests confirmed that the probit model is suitable for our analysis.

Pulkstenis-Robinson chi-squared test

X-squared = 46.0821543216366, p-value = 0.386130385657324

Pulkstenis-Robinson deviance test

X-squared = 47.4174753379652, p-value = 0.335063478598168

⁴Please refer to the Appendix of the paper for the relevant code pertaining to the Brant's test.

⁵Please refer to the Appendix of the paper for the relevant code pertaining to the clm algorithm.

Therefore, based on the rejection of the parallel regression assumption in Brant’s test, we transitioned from the ordered logit model to the ordered probit model, ensuring a more accurate and reliable evaluation of the factors influencing the categorized rankings.

In the case of man-made artworks, we followed the same procedures as described earlier. However, the original variables did not meet the goodness-of-fit test criteria, with p-values below the 5% threshold. This indicates that the model with the original variables did not adequately fit the observed data for man-made artworks.

Pulkstenis-Robinson chi-squared test

X-squared = 51.4177635339398, p-value = 0.10656549021287

Pulkstenis-Robinson deviance test

X-squared = 52.8360582705313, p-value = 0.0841035069317887

After excluding the ‘comments’ parameter and incorporating additional variables such as like_rate2, mark_rate2, and mark_rate3, our model for the man-made artworks demonstrates improved performance.

The exclusion of the ‘comments’ parameter addresses the issue of endogeneity, which may have been present in the original model. By removing this variable, we mitigate the potential bias caused by the reciprocal relationship between comments and the ranks of the artworks.

Furthermore, the inclusion of higher-order terms, such as the power and cube of mark_rate, allows us to capture potential nonlinear associations between the independent variable (like_rate) and the dependent variable (ranks). This consideration accounts for the possibility of non-linear patterns in the data, enhancing the model’s ability to accurately represent the underlying relationship between like_rate and ranks. The introduction of these new variables enables a more flexible and nuanced analysis, potentially leading to improved goodness-of-fit test results.

By incorporating these adjustments into the model, we aim to refine our understanding of the factors influencing the rankings of man-made artworks. These methodological refinements align with the rigorous standards expected in academic research and contribute to the advancement of knowledge in the field.

R2 statistics

fitting null model for pseudo-r2

llh	llhNull	G2	McFadden	r2ML
-11958.42802746	-13021.45804846	2126.06004201	0.08163679	0.23105516
r2CU				
0.24068770				

fitting null model for pseudo-r2

llh	llhNull	G2	McFadden	r2ML
-11949.44891673	-13021.45804846	2144.01826346	0.08232635	0.23275975
r2CU				
0.24246335				

fitting null model for pseudo-r2

1lh	1lhNull	G2	McFadden	r2ML
-7841.7220799	-10430.6426770	5177.8411941	0.2482034	0.5500220
r2CU				
0.5729761				

In analyzing the R2 values for the logit and probit models, it is important to note that the McFadden R2 alone cannot be interpreted in isolation. Comparing the McFadden R2 values of different models⁶ provides a basis for understanding the relative goodness-of-fit among the models.

For the AI-generated samples, the McFadden R2 of the probit model is 0.0823, while the McFadden R2 of the logit model is slightly lower at 0.0816. This suggests that both models have a similar level of explanatory power in capturing the variations in the data.

On the other hand, for the man-made artworks samples, the McFadden R2 of the probit model is notably higher at 0.248. This indicates that the probit model explains a larger proportion of the variances in the rankings of man-made artworks compared to the logit model.

It is worth noting that the McFadden R2 values, while providing a measure of the models' fit, do not offer insight into the significance or magnitude of individual independent variables. Therefore, additional analysis is necessary to interpret the effects and significance of the variables included in the models.

Overall, the comparison of McFadden R2 values suggests that the probit model performs relatively better in explaining the rankings of both AI-generated samples and man-made artworks samples. However, further examination of the model coefficients and statistical tests is required to gain a comprehensive understanding of the factors influencing the rankings in each case.

Model Selection

Likelihood Ratio Test

We conducted a likelihood ratio test to compare the unrestricted model, which includes the independent variables, with the restricted model, which only includes a constant term (`as.factor(tier)~1`). The purpose of this test was to determine if the inclusion of the independent variables in the model significantly improved its fit.

Likelihood ratio test

```
Model 1: tier ~ like_rate + mark_rate + is_comic + is_Genshin + is_Honkai +
  comments + top_cnt + date_diff_day
```

```
Model 2: as.factor(tier) ~ 1
```

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	12	-11949			
2	4	-13022	-8	2144	< 0.00000000000000022 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

After running the `lrtest` test⁷, test likelihood ratio test yielded a p-value close to zero, indicating strong evidence against the null hypothesis ($H_0: \beta_1 = \beta_2 = 0$), which represents the condition

⁶Please refer to the Appendix of the paper for the relevant code pertaining to the pR2 test.

⁷Please refer to the Appendix of the paper for the relevant code pertaining to the `lrtest`.

of restriction in the restricted model. Therefore, we reject the null hypothesis and conclude that the independent variables in the unrestricted model are jointly significant.

Based on the results of the likelihood ratio test, we can confidently choose the unrestricted model over the restricted model. This confirms that the inclusion of the independent variables improves the model’s ability to explain and predict the rankings of the artworks on pixiv.net.

Variable Power Transformations

To investigate the potential nonlinear relationships between the independent variables and the dependent variable, we performed a test on the inclusion of power terms in the model. We compared the unrestricted model, which includes the original independent variables (like_rate, mark_rate, is_comic, is_Genshin, is_Honkai, comments, top_cnt, date_diff_day), with an extended model that incorporates additional power terms (like_rate² and mark_rate²).

Likelihood ratio tests of cumulative link models:

```

                                formula:
probit_ai          tier ~ like_rate + mark_rate + is_comic + is_Genshin + is_Honkai + comments + t
probit_ai_power    tier ~ like_rate + mark_rate + is_comic + is_Genshin + is_Honkai + comments + t
                                link:  threshold:
probit_ai          probit flexible
probit_ai_power    probit flexible

                                no.par   AIC logLik LR.stat df Pr(>Chisq)
probit_ai          12 23923 -11949
probit_ai_power    14 23909 -11940  18.048  2  0.0001205 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The analysis using anova to measure likelihood ratio test⁸ indicated a p-value of 0.00012, which is less than the predetermined significance level of 5%. Consequently, we reject the null hypothesis that the power terms have coefficients of zero, suggesting that the two models are significantly different. Hence, including the power terms in the model (probit_ai_power) yields a better fit than the model without the power terms (probit_ai).

These findings highlight the importance of considering nonlinear associations between the independent variables and the dependent variable. By including the power terms, we capture potential nonlinear patterns and enhance the model’s ability to accurately represent the complex relationship between the independent variables (such as like_rate and mark_rate) and the dependent variable (tier). This refined model provides a more comprehensive understanding of the factors influencing the rankings of the artworks on pixiv.net.

Variables Selection

To select the most relevant variables for our AI model, we employed the general-to-specific method. We began by examining the coefficients of the variables in the extended model (probit_ai_power).

⁸Please refer to the Appendix of the paper for the relevant code pertaining to the anova which is a example of variables selection.

The p-values associated with ‘like_rate2’ and ‘is_Honkai1’ were found to be less than 5%, indicating that these variables were statistically insignificant for the model.

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
1 2	-1.3606197	0.1131953	-12.0201	< 0.00000000000000022 ***
2 3	-0.6567886	0.1126169	-5.8321	0.000000005686 ***
3 4	-0.0728850	0.1124906	-0.6479	0.5170546
4 5	0.5668039	0.1126528	5.0314	0.000000497345 ***
like_rate	-12.2456027	3.7323739	-3.2809	0.0010391 **
mark_rate	10.3938217	2.1123227	4.9206	0.000000880009 ***
is_comic1	0.8637065	0.2396678	3.6038	0.0003155 ***
is_Genshin1	-0.1283309	0.0265285	-4.8375	0.000001339375 ***
is_Honkai1	-0.0568993	0.0486479	-1.1696	0.2421902
comments	-0.0327167	0.0030684	-10.6623	< 0.00000000000000022 ***
top_cnt	-0.5618988	0.0197977	-28.3820	< 0.00000000000000022 ***
date_diff_day	-0.2652163	0.0266711	-9.9440	< 0.00000000000000022 ***
like_rate2	5.4639612	12.4129091	0.4402	0.6598158
mark_rate2	-10.0575579	4.0767402	-2.4671	0.0136434 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Likelihood ratio tests of cumulative link models:

```

      formula:
probit_ai_honkai tier ~ like_rate + mark_rate + is_comic + is_Genshin + comments + top_cnt + date_diff_day
probit_ai_power  tier ~ like_rate + mark_rate + is_comic + is_Genshin + is_Honkai + comments + top_cnt + date_diff_day
      link: threshold:
probit_ai_honkai probit flexible
probit_ai_power  probit flexible

```

	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
probit_ai_honkai	13	23908	-11941			
probit_ai_power	14	23909	-11940	1.3683	1	0.2421

To refine the model further, we employed the process of elimination using the anova test. First, we compared the extended model (probit_ai_power) with a simplified model that excluded the ‘is_Honkai’ variable (probit_ai_honkai). The resulting p-value of 0.24 was greater than the predetermined significance level of 0.05, suggesting that the two models were not significantly different. Consequently, we selected the simpler model without the ‘is_Honkai’ variable.

Likelihood ratio tests of cumulative link models:

```

      formula:
probit_ai_honkai_liked2 tier ~ like_rate + mark_rate + is_comic + is_Genshin + comments + top_cnt + date_diff_day
probit_ai_honkai        tier ~ like_rate + mark_rate + is_comic + is_Genshin + comments + top_cnt + date_diff_day
      link: threshold:
probit_ai_honkai_liked2 probit flexible
probit_ai_honkai        probit flexible

```

	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
probit_ai_honkai_liked2	12	23906	-11941			
probit_ai_honkai	13	23908	-11941	0.2185	1	0.6402

Next, we compared the selected model (probit_ai_honkai) with another variant that excluded the ‘like_rate2’ variable (probit_ai_honkai_liked2). Again, the resulting p-value of 0.24 was greater than 0.05, indicating that the two models were not significantly different. Hence, we chose the model without the ‘like_rate2’ variable for its simplicity.

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
1 2	-1.3509496	0.1061138	-12.7311	< 0.00000000000000022 ***
2 3	-0.6472379	0.1055069	-6.1346	0.0000000008942 ***
3 4	-0.0634168	0.1053794	-0.6018	0.5473275
4 5	0.5762096	0.1055595	5.4586	0.0000000494062 ***
like_rate	-10.8387234	0.8874689	-12.2131	< 0.00000000000000022 ***
mark_rate	9.6129323	1.1346571	8.4721	< 0.00000000000000022 ***
is_comic1	0.8625165	0.2396125	3.5996	0.0003206 ***
is_Genshin1	-0.1201951	0.0256356	-4.6886	0.0000027958857 ***
comments	-0.0326257	0.0030563	-10.6750	< 0.00000000000000022 ***
top_cnt	-0.5623914	0.0197902	-28.4177	< 0.00000000000000022 ***
date_diff_day	-0.2648563	0.0266642	-9.9330	< 0.00000000000000022 ***
mark_rate2	-8.3242784	2.0056964	-4.1503	0.0000335421048 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Pulkstenis-Robinson chi-squared test

X-squared = 29.0277179938486, p-value = 0.219089440858309

Pulkstenis-Robinson deviance test

X-squared = 29.3916394849468, p-value = 0.205739139344962

By performing the coeftest on the final model (probit_ai_honkai_liked2), we confirmed that all remaining parameters were statistically significant. Furthermore, we conducted the goodness-of-fit tests using the Pulkstenis-Robinson method. The results indicated that the model exhibited an appropriate fit to the observed data, as the p-values for both the chi-squared and deviance tests were greater than 5%.

In summary, through the general-to-specific method, we arrived at a refined model (probit_ai_honkai_liked2) that includes the significant variables of ‘like_rate’, ‘mark_rate’, ‘is_comic’, ‘is_Genshin’, ‘comments’, ‘top_cnt’, and ‘date_diff_day’. This selection process ensures that our model includes only the most relevant variables, enhancing its interpretability and reliability for analyzing the rankings of man-made artworks on pixiv.net.

For the sample of man-made artworks, we applied the same general-to-specific method to refine our model. We removed the ‘mark_rate’, ‘like_rate2’, and ‘is_Genshin’ variables based on their statistical insignificance and potential lack of relevance to the ranking of man-made artworks on pixiv.net.

The resulting optimized model (probit_man_opt) included the variables ‘like_rate’, ‘mark_rate2’, ‘mark_rate3’, ‘is_comic’, ‘is_Honkai’, ‘top_cnt’, ‘date_diff_day’, and ‘views’. We performed the Pulkstenis-Robinson chi-squared and deviance tests to assess the goodness-of-fit of the model. The p-values associated with both tests were greater than 5%, indicating that the model exhibited an appropriate fit to the observed data for man-made artworks.

Additionally, we conducted the coeftest on the optimized model to assess the significance of the remaining variables. The results confirmed that all variables included in the model were statistically significant, further validating their relevance to the ranking of man-made artworks on the platform.

Pulkstenis-Robinson chi-squared test

X-squared = 37.9873162251723, p-value = 0.0463934885297833

Pulkstenis-Robinson deviance test

X-squared = 35.0811699793141, p-value = 0.0867447217886955

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
1 2	-4.58414802	0.09111831	-50.3098	< 0.00000000000000022 ***
2 3	-3.55961479	0.08409368	-42.3292	< 0.00000000000000022 ***
3 4	-2.73963392	0.08099761	-33.8236	< 0.00000000000000022 ***
4 5	-1.87019618	0.08022134	-23.3130	< 0.00000000000000022 ***
like_rate	-34.46126887	1.53665218	-22.4262	< 0.00000000000000022 ***
mark_rate2	103.36942846	7.48610465	13.8082	< 0.00000000000000022 ***
mark_rate3	-153.75420543	17.19604929	-8.9413	< 0.00000000000000022 ***
is_comic1	0.10163062	0.04147155	2.4506	0.014288 *
is_Honkai1	-0.40599044	0.14081225	-2.8832	0.003949 **
top_cnt	-0.82444377	0.01655719	-49.7937	< 0.00000000000000022 ***
date_diff_day	-0.18450164	0.02659360	-6.9378	0.0000000000004369 ***
views	-0.00755315	0.00023125	-32.6618	< 0.00000000000000022 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

In summary, through the general-to-specific method, we arrived at an optimized model (probit_man_opt) for the analysis of man-made artworks. This model incorporates the significant variables of ‘like_rate’, ‘mark_rate2’, ‘mark_rate3’, ‘is_comic’, ‘is_Honkai’, ‘top_cnt’, ‘date_diff_day’, and ‘views’. By selecting these variables, we ensure that our model focuses on the most influential factors and provides a more accurate representation of the ranking process for man-made artworks on pixiv.net.

Findings

Estimation

To Compare the results of the two models using an academic perspective, we observe both similarities and differences in the qualitative analysis of the variables’ effects on the tier outcome in the

ordered choice model of probit using ologit.reg algorithm⁹.

In analyzing the results, it is important to focus on the estimated coefficients rather than the threshold parameters. The estimated coefficients provide valuable insights into the effects of each variable on the tier rankings in both the AI-generated and man-made sample models.

Ordered Logit Regression

Log-Likelihood: -11950.87

No. Iterations: 4

Mcfadden's R2: 0.08221698

AIC: 23925.75

	Estimate	Std. error	t value	Pr(> t)
like_rate	-18.1555930	1.5093519	-12.0287	< 0.00000000000000022 ***
mark_rate	15.9244659	1.9369456	8.2214	< 0.00000000000000022 ***
is_comic1	1.3458658	0.4003875	3.3614	0.0007755 ***
is_Genshin1	-0.2026204	0.0431979	-4.6905	0.000002725 ***
comments	-0.0574955	0.0054985	-10.4566	< 0.00000000000000022 ***
top_cnt	-0.9279942	0.0339249	-27.3544	< 0.00000000000000022 ***
date_diff_day	-0.4484277	0.0456119	-9.8314	< 0.00000000000000022 ***
mark_rate2	-13.7399807	3.4202912	-4.0172	0.000058895 ***

----- Threshold Parameters -----

	Estimate	Std. error	t value	Pr(> t)
Threshold (1->2)	-2.32661	0.18294	-12.7181	< 0.00000000000000022 ***
Threshold (2->3)	-1.10682	0.18109	-6.1118	0.0000000009849 ***
Threshold (3->4)	-0.14967	0.18068	-0.8283	0.4075
Threshold (4->5)	0.91167	0.18117	5.0321	0.0000004850831 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ordered Logit Regression

Log-Likelihood: -7830.092

No. Iterations: 5

Mcfadden's R2: 0.2493184

AIC: 15684.18

	Estimate	Std. error	t value	Pr(> t)
like_rate	-60.88099012	2.70923032	-22.4717	< 0.00000000000000022 ***
mark_rate2	180.05258395	13.00417722	13.8457	< 0.00000000000000022 ***
mark_rate3	-265.14080776	29.76503623	-8.9078	< 0.00000000000000022 ***
is_comic1	0.20006618	0.07225621	2.7688	0.005626 **
is_Honkai1	-0.67472830	0.24402266	-2.7650	0.005692 **
top_cnt	-1.41484643	0.03061475	-46.2145	< 0.00000000000000022 ***
date_diff_day	-0.31372497	0.04616324	-6.7960	0.00000000001076 ***
views	-0.01468009	0.00049416	-29.7069	< 0.00000000000000022 ***

----- Threshold Parameters -----

	Estimate	Std. error	t value	Pr(> t)
Threshold (1->2)	-8.09823	0.16930	-47.833	< 0.00000000000000022 ***
Threshold (2->3)	-6.32671	0.15447	-40.957	< 0.00000000000000022 ***

⁹Please refer to the Appendix of the paper for the relevant code pertaining to the anova which is a example of variables selection.

```

Threshold (3->4) -4.87244    0.14657 -33.244 < 0.00000000000000022 ***
Threshold (4->5) -3.38503    0.14296 -23.678 < 0.00000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

In the AI-generated sample model, an increase in the variable `mark_rate2` (square of bookmarked rate) leads to a increase in the probability of achieving tier-1 and an decrease in the probability of achieving tier-5. This suggests that higher like rates are associated with higher tiers, indicating a positive relationship between `like_rate` and the rank.

On the other hand, in the man-made sample model, an increase in the variable `mark_rate2` is associated with a decrease in the probability of achieving tier-1 and an increase in the probability of achieving tier-5. Therefore, higher bookmarked rates in the man-made sample model are indicative of a lower likelihood of being ranked in tier-1 and a higher likelihood of being ranked in tier-5.

Furthermore, the presence of additional variables in each model contributes to their differences. In the AI-generated sample model, variables such as `is_Genshin1` (related to the game “Genshin”) and `comments` play significant roles in determining the tier outcome. Conversely, the man-made sample model includes variables such as `mark_rate3` (mark rate cubed), `is_Honkai1` (related to the series games “Honkai”), and `views`, which are not considered in the AI-generated sample model. These variations in variable inclusion reflect the specific characteristics and influences within each sample.

The difference in the relationship between `mark_rate2` and the tier variable in the AI-generated and man-made models could be attributed to several factors.

Firstly, it’s important to consider the underlying characteristics and composition of the AI-generated and man-made samples. These samples may have distinct patterns and characteristics, leading to variations in how `mark_rate2` influences the tier variable. The AI-generated sample may have a different distribution or range of `mark_rate2` values compared to the man-made sample, which could result in diverse effects on the tier variable.

Secondly, the AI-generated and man-made samples may differ in terms of the content or context of the artworks. The relationship between `mark_rate2` and the tier variable could be influenced by various factors such as the subject matter, style, or themes of the artworks. It’s possible that `mark_rate2` has a stronger impact on the tier variable in one sample due to specific characteristics or preferences associated with AI-generated or man-made artworks.

Additionally, the modeling approach and other variables included in the models could contribute to the differences in the effect of `mark_rate2` on the tier variable. The inclusion of different variables or the use of alternative modeling techniques in the AI-generated and man-made models may interact with `mark_rate2` differently, leading to contrasting results.

In summary, while both models share similar relationship between variables like `like_rate` and the probability of tier-1, the influence on tier-5 differs. The additional variables in each model highlight the unique factors affecting the tier outcome in the respective sample. By analyzing these differences and similarities, researchers can gain insights into the nuanced dynamics and factors driving the rank variation between AI-generated and man-made samples in the context of the ordered choice model. The discrepancies between these two models can be attributed to various factors, such as differences in data sources, model training techniques, or the inclusion/exclusion of certain variables. These variations highlight the complex nature of modeling ordered choice outcomes and the impact of different factors on the ranking of content in AI-generated and man-made scenarios. Further

analysis and research are needed to explore these differences and their underlying causes in more detail.

Marginal Effects

To analyze the Marginal Effects of the AI-generated artworks model, we examine the output of the code provided. The marginal effects represent the change in the probability of each outcome category based on a unit change in the corresponding independent variable, while holding other variables constant. We choose `margins.oglm`¹⁰ to imply marginal effects.

Marginal Effects on Pr(Outcome==1)

	Marg. Eff	Std. error	t value	Pr(> t)
is_comic1	-0.12007167	0.01965635	-6.1085	0.0000000010054454188 ***
is_Genshin1	0.02897413	0.00626190	4.6271	0.0000037090669989739 ***
like_rate	2.56780597	0.21740855	11.8110	< 0.000000000000000022 ***
mark_rate	-2.25225022	0.27635217	-8.1499	0.00000000000000003641 ***
comments	0.00813178	0.00079029	10.2896	< 0.000000000000000022 ***
top_cnt	0.13124931	0.00512074	25.6309	< 0.000000000000000022 ***
date_diff_day	0.06342263	0.00651788	9.7306	< 0.000000000000000022 ***
mark_rate2	1.94329122	0.48475411	4.0088	0.0000610233102153661 ***

Marginal Effects on Pr(Outcome==2)

	Marg. Eff	Std. error	t value	Pr(> t)
is_comic1	-0.13736230	0.03293880	-4.1702	0.000030429511333428 ***
is_Genshin1	0.02015730	0.00428592	4.7031	0.000002561839346160 ***
like_rate	1.82537624	0.16139650	11.3099	< 0.000000000000000022 ***
mark_rate	-1.60105712	0.20057330	-7.9824	0.0000000000000001435 ***
comments	0.00578064	0.00057929	9.9789	< 0.000000000000000022 ***
top_cnt	0.09330120	0.00453415	20.5774	< 0.000000000000000022 ***
date_diff_day	0.04508524	0.00478939	9.4136	< 0.000000000000000022 ***
mark_rate2	1.38142743	0.34634688	3.9886	0.000066473883081489 ***

Marginal Effects on Pr(Outcome==3)

	Marg. Eff	Std. error	t value	Pr(> t)
is_comic1	-0.06648587	0.03519228	-1.8892	0.0588627 .
is_Genshin1	-0.00295741	0.00086095	-3.4351	0.0005924 ***
like_rate	-0.23330746	0.04873351	-4.7874	0.0000016894 ***
mark_rate	0.20463648	0.04644807	4.4057	0.0000105440 ***
comments	-0.00073884	0.00016001	-4.6174	0.0000038861 ***
top_cnt	-0.01192514	0.00233452	-5.1082	0.0000003253 ***
date_diff_day	-0.00576250	0.00124357	-4.6338	0.0000035899 ***
mark_rate2	-0.17656487	0.05546227	-3.1835	0.0014550 **

Marginal Effects on Pr(Outcome==4)

	Marg. Eff	Std. error	t value	Pr(> t)
is_comic1	0.06158231	0.01069189	5.7597	0.000000008425301512 ***

¹⁰Please refer to the Appendix of the paper for the relevant code pertaining to the `margins.oglmx` algorithm for marginal effects.


```

is_Genshin1    -0.01919832  0.00412688  -4.6520  0.000003286933884080 ***
like_rate      -1.71645208  0.15033050 -11.4179 < 0.00000000000000022 ***
mark_rate       1.50551857  0.18779600   8.0168  0.000000000000001086 ***
comments       -0.00543569  0.00054123 -10.0433 < 0.00000000000000022 ***
top_cnt        -0.08773371  0.00391746 -22.3955 < 0.00000000000000022 ***
date_diff_day  -0.04239491  0.00446685  -9.4910 < 0.00000000000000022 ***
mark_rate2     -1.29899466  0.32539166  -3.9921  0.000065491908615994 ***

```

Marginal Effects on Pr(Outcome==5)

	Marg. Eff	Std. error	t value	Pr(> t)
is_comic1	0.26233753	0.09753254	2.6897	0.007151 **
is_Genshin1	-0.02697570	0.00570035	-4.7323	0.0000022200048292901 ***
like_rate	-2.44342267	0.20504356	-11.9166	< 0.00000000000000022 ***
mark_rate	2.14315230	0.26177419	8.1870	0.0000000000000002678 ***
comments	-0.00773788	0.00074251	-10.4212	< 0.00000000000000022 ***
top_cnt	-0.12489166	0.00488016	-25.5917	< 0.00000000000000022 ***
date_diff_day	-0.06035047	0.00618613	-9.7558	< 0.00000000000000022 ***
mark_rate2	-1.84915912	0.46071338	-4.0137	0.0000597777070404874 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

In the case of Outcome==1, which corresponds to tier 1 in the ranking, artworks related to Genshin have a higher probability compared to artworks related to other topics, with a difference of 2.8119 percentage points. This implies that Genshin-themed artworks are more likely to be ranked in the top tier.

Furthermore, an increase a unit in the 'like_rate' variable would lead to an increase in the probability of an artwork being ranked in tier 1 by 108.1806 percentage points. This suggests that higher like rate contribute to a higher likelihood of achieving a top-tier ranking.

Based on the code output, we can analyze the effects of different outcomes on rank and the influence of various variables.

For Outcome==1, which represents the first tier ranking, the following other variables have an impact on the rank: - is_comic1 (whether it is related to comics): Increasing the presence of comic-related artworks significantly reduces the probability of achieving the first tier rank, decreasing it by 11.753 percentage points. - is_Genshin1 (whether it is related to Genshin): Artworks related to Genshin have a higher probability of achieving the first tier rank compared to other themes, increasing it by 2.8119 percentage points. - like_rate: An increase in the like rate of the artwork leads to a higher probability of attaining the first tier rank, with a marginal effect of 1.0818. - mark_rate: Higher mark rates are associated with a lower probability of achieving the first tier rank, resulting in a marginal effect of -1.2748. - comments: The number of comments received has a positive impact on the probability of achieving the first tier rank, with a marginal effect of 0.0080. - top_cnt: Increasing the number of times the artwork appears in the top selection enhances the likelihood of attaining the first tier rank, with a marginal effect of 0.1312. - date_diff_day: Longer time intervals since the artwork's creation date are associated with a higher probability of achieving the first tier rank, resulting in a marginal effect of 0.0632. - like_rate2: The square of the like rate also positively influences the probability of attaining the first tier rank, with a marginal effect of 4.9383.

Similarly, we can analyze the effects of these variables on other outcomes (Outcome==2, Outcome==3, Outcome==4, Outcome==5) by referring to the respective marginal effects provided in the code output.

By analyzing the Marginal Effects, we gain insights into the impact of specific variables on the probabilities of different ranking outcomes. These findings provide valuable information for understanding the factors influencing the rankings of AI-generated artworks on the platform.

Bibliography

- [1] A. Zarifhonarvar, “Economics of ChatGPT: A labor market view on the occupational impact of artificial intelligence,” *SSRN*, 2023, Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4350925
- [2] E. D. Lennart Ante, “The ChatGPT effect on AI-themed cryptocurrencies,” *SSRN*, 2023, Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4350557
- [3] D. F. H. J. Campos N. R. Ericsson, “General-to-specific modeling: An overview and selected bibliography, board of governors of the federal reserve system,” *International Finance Discussion Papers*, no. 838, Aug. 2005.

Appendix

The codes provided in the Appendix are simplified examples for illustration purposes. In practice, additional steps such as data preprocessing, handling missing values, and further diagnostics would be conducted.

By including such explanatory text and providing a clear explanation of the code implementation, readers will be able to follow the code and understand its relevance to the research study.

polr

In this section, we provide the code implementation for fitting a logit model using the MASS::polr function in the R programming language. The logit model is used as an ordered choice model to analyze the factors influencing the rankings of AI-generated artwork.

```
# Convert the 'rank' variable to a factor
tier<-as.factor(ai$rank)

# Fit the logit model using the polr function
logit_ai <- polr(tier~like_rate+mark_rate
                +is_comic+is_Genshin+is_Honkai
                +comments+top_cnt+date_diff_day
                +views
                ,data=ai)
```

In the above code, we convert the ‘rank’ variable to a factor using the as.factor function, ensuring that it is treated as an ordered variable. Next, we fit the logit model using the polr function, specifying the dependent variable tier and the independent variables like_rate, mark_rate, is_comic, is_Genshin, is_Honkai, comments, top_cnt, date_diff_day, and views.

clm

In this section, we present the code implementation and the results for the probit model `probit_ai`. This model examines the influence of various variables on the rankings of AI-generated images.

```
# Convert the 'rank' variable to a factor
ai$tier<-as.factor(ai$rank)

# Fit the probit model for AI-generated images
probit_ai <- clm(tier~like_rate+mark_rate
                +is_comic+is_Genshin+is_Honkai
                +comments+top_cnt+date_diff_day
                , data=ai, link=c('probit'))
```

In the above code, we first convert the ‘rank’ variable in the ai data set to a factor variable named ‘tier’, representing the rankings of AI-generated images. Then, we fit the probit model `probit_ai` to analyze the relationship between the ‘tier’ variable and predictor variables such as `like_rate`, `mark_rate`, `is_comic`, `is_Genshin`, `is_Honkai`, `comments`, `top_cnt`, and `date_diff_day`.

ologit.reg

In this section, we provide the code implementation and the summary results for the probit models `probit_ai` and `probit_man`. These models examine the influence of various variables on the rankings of AI-generated images (`probit_ai`) and hand-drawn artwork (`probit_man`).

```
# Fit the probit model for AI-generated images
probit_ai <- ologit.reg(tier~like_rate+mark_rate
                      +is_comic+is_Genshin
                      +comments+top_cnt+date_diff_day
                      +mark_rate2
                      ,data=ai)

# Fit the probit model for hand-drawn artwork
probit_man <- ologit.reg(tier~like_rate
                      +mark_rate2
                      +mark_rate3
                      +is_comic+is_Honkai
                      +top_cnt+date_diff_day+views
                      ,data=man)

# Summarize the results of the probit models
summary(probit_ai)
summary(probit_man)
```

In the above code, we fit the probit model `probit_ai` to analyze the rankings of AI-generated images, considering variables such as `like_rate`, `mark_rate`, `is_comic`, `is_Genshin`, `comments`, `top_cnt`, `date_diff_day`, and `mark_rate2`. Similarly, we fit the probit model `probit_man` to examine the rankings of hand-drawn artwork, considering variables such as `like_rate`, `mark_rate2`, `mark_rate3`, `is_comic`, `is_Honkai`, `top_cnt`, `date_diff_day`, and `views`.

We then obtain the summary results for each model using the summary function, which provides information on the estimated coefficients, standard errors, p-values, and other relevant statistics.

lipsitz and logitgof tests

In this section, we provide the code implementation for conducting goodness-of-fit tests on the fitted logit model using the `lipsitz.test` and `logitgof` functions in R. These tests are used to assess the adequacy of the model and evaluate its fit to the observed data.

```
# Perform Lipsitz test for goodness-of-fit
Lipsitz <- lipsitz.test(logit_ai)
cat(paste(Lipsitz$method, '\n'))
cat(paste0('X-squared = ', Lipsitz$statistic,
          ', p-value = ', Lipsitz$p.value,
          '\n\n'))

# Perform Lemeshow test for goodness-of-fit
Lemeshow <- logitgof(tier, fitted(logit_ai), g=5, ord = TRUE)
cat(paste(Lemeshow$method, '\n'))
cat(paste0('X-squared = ', Lemeshow$statistic,
          ', p-value = ', Lemeshow$p.value,
          '\n\n'))
```

In the above code, we must load the necessary libraries, including the ordinal library, which provides the functions for conducting the goodness-of-fit tests. We then proceed to perform the Lipsitz test using the `lipsitz.test` function, which evaluates the overall fit of the logit model. Additionally, we conduct the Lemeshow test using the `logitgof` function, which assesses the calibration of the model by comparing the observed and expected probabilities across groups.

pulkrob tests

In this section, we provide the code implementation for conducting the Pulkstenis-Robinson tests on the fitted logit model using the `pulkrob.chisq` and `pulkrob.deviance` functions in R. These tests are used to assess the significance of specific variables in the model and evaluate their contribution to the overall fit.

```
# Perform Pulkstenis-Robinson chi-square test
chi <- pulkrob.chisq(logit_ai, c("is_comic", "is_Genshin", "is_Honkai"))
cat(paste(chi$method, '\n'))
cat(paste0('X-squared = ', chi$statistic,
          ', p-value = ', chi$p.value,
          '\n\n'))

# Calculate Pulkstenis-Robinson deviance
dev <- pulkrob.deviance(logit_ai, c("is_comic", "is_Genshin", "is_Honkai"))
cat(paste(dev$method, '\n'))
cat(paste0('X-squared = ', dev$statistic,
          ', p-value = ', dev$p.value,
```

```
'\n\n'))
```

In the above code, we first load the necessary libraries, including the ordinal library, which provides the functions for conducting the Pulkstenis-Robinson tests. We then proceed to perform the Pulkstenis-Robinson chi-square test using the `pulkrob.chisq` function, which assesses the significance of the specified variables in the logit model. Additionally, we calculate the Pulkstenis-Robinson deviance using the `pulkrob.deviance` function, which measures the contribution of the variables to the overall deviance of the model.

Brant's test

In this section, we present the code implementation for conducting the Brant test on the fitted logit model using the `brant` function in R. The Brant test is used to assess the assumption of proportional odds in the ordered choice model and helps determine whether the logit or probit model is more appropriate.

```
brant(logit_ai)
```

In the above code, we first load the necessary libraries, including the ordinal library, which provides the `brant` function for conducting the Brant test. We then proceed to perform the Brant test on the fitted logit model `logit_ai`.

pR2 test

In this section, we present the code implementation for calculating the McFadden R2 statistic using the `pR2` function from the `pscl` library in R. The McFadden R2 is a measure of the goodness-of-fit for logistic and probit models in ordered choice analysis.

```
pR2(logit_ai)
pR2(probit_ai)
pR2(probit_man)
```

In the above code, we should load the necessary `pscl` library firstly, which provides the `pR2` function for calculating the McFadden R2 statistic. We then proceed to calculate the McFadden R2 for the logit model `logit_ai`, as well as the probit models `probit_ai` and `probit_man`.

lrtest.

In this section, we present the code implementation for conducting the Likelihood Ratio Test (LRT) to compare the full probit model `probit_ai` with a restricted model `probit_ai_restricted`. The LRT is used to assess the significance of additional variables in the full model compared to a reduced or restricted model.

```
# Fit the restricted probit model
probit_ai_restricted <- clm(as.factor(tier)~1, data=ai)

# Perform the Likelihood Ratio Test
lrtest(probit_ai, probit_ai_restricted)
```

In the above code, we first fit the restricted probit model `probit_ai_restricted`, which includes only the intercept term. We then proceed to perform the Likelihood Ratio Test using the `lrtest` function, comparing the full probit model `probit_ai` with the restricted model.

Example of Variables Selection

In this section, we present the code implementation for applying variable power transformations in the probit model `probit_ai` to examine the impact of including squared variables (`like_rate2` and `mark_rate2`) on the model's fit. The `anova` function is then used to compare the full model `probit_ai` with the model including the squared variables `probit_ai_power`.

```
probit_ai_power <- clm(tier ~ like_rate+mark_rate
                      +is_comic+is_Genshin+is_Honkai
                      +comments+top_cnt+date_diff_day
                      +like_rate2+mark_rate2, data = ai, link=c('probit'))
anova(probit_ai, probit_ai_power)
```

In the above code, we extend the probit model `probit_ai` by including squared variables (`like_rate2` and `mark_rate2`) to capture potential non-linear relationships between these variables and the response variable. We fit the extended model `probit_ai_power` and then use the `anova` function to compare the full model `probit_ai` with the model including the squared variables.

margins.oglmx

In this section, we present the code implementation and the results for the marginal effects analysis of the probit model `probit_ai`. This analysis allows us to examine the differential impact of various variables on the rankings of AI-generated images.

```
probit_ai_power <- clm(tier ~ like_rate+mark_rate
                      +is_comic+is_Genshin+is_Honkai
                      +comments+top_cnt+date_diff_day
                      +like_rate2+mark_rate2, data = ai, link=c('probit'))
anova(probit_ai, probit_ai_power)
```

In the above code, we use the `margins.oglmx()` function to calculate the marginal effects for the `probit_ai` model. This function computes the average marginal effects of each predictor variable on the predicted probabilities of the different rank levels.

The results of the marginal effects analysis provide insights into how changes in each predictor variable affect the probabilities of different rankings for AI-generated images.