How Individual Traits and Language Styles Shape Preferences In Open-ended User-LLM Interaction

ANONYMOUS AUTHOR(S)

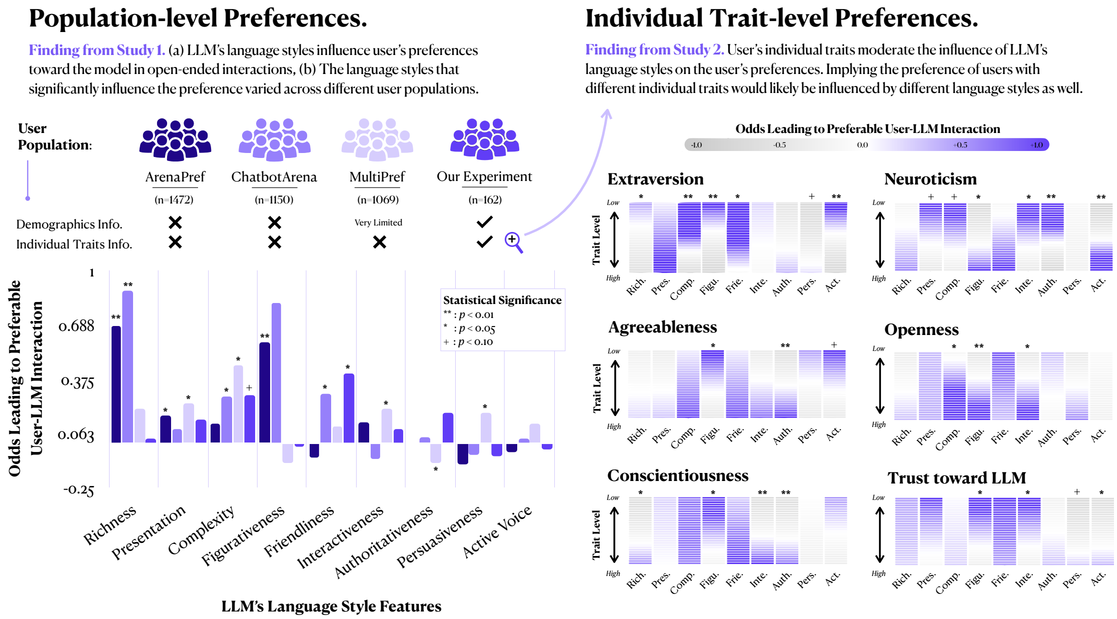


Fig. 1. Our studies explore how the LLM's language styles and user's individual traits influence user's preferences toward the LLM in open-ended interaction. In **Study 1**, we conducted exploratory study on the direct influence of LLM's language styles on user's preferences. In **Study 2**, we conducted experimental study on the moderating effects of user's individual traits on the influence of LLM's language styles on user's preferences.

CCS Concepts: • **Human-centered computing** → **User studies**; **Empirical studies in HCI**.

Additional Key Words and Phrases: Human-AI Interaction, LLM, User's Preference, Personality Traits, Personalization

1 INTRODUCTION

What makes an interaction with the LLM more preferable for the user? It is intuitive to assume that information accuracy in the LLM's responses would be one of the influential variables. While sometimes that is indeed the case [8, 9], recent studies have found that *inaccurate* LLM's responses could still be preferable when they are perceived to be more friendly and not admitting its limitation [12], more verbose and grammatically correct [21], or well-articulated [9]. Each of these variables interestingly falls under the category of *linguistic style* [10], as it represents a linguistic feature with a communicative purpose, implying the LLM's language style might be significantly influencing the user's preferences.

Understanding the influence of LLM's language style on user's preferences is crucial, particularly in *open-ended interaction* mode—where users deliberately converse and exchange new information and opinions with the LLM, as it could lead to double-edged consequences. Having the LLM conversing with the style that maximizes user's preferences might be valuable in improving overall user experiences [19, 20]. But, it also means the users would be more susceptible to accepting information from the LLMs that might be misinformed or hallucinated [1, 17]. How does the relationship between the LLM's language style and user's preference really work? Are all users being influenced similarly by the same language styles? Or perhaps, user's personal factors such as their demographics and individual traits also play role in shaping their preferences? Our long-term objective is to address these research inquiries. As a starting point, we translated these initial inquiries into the following research questions:

- **RQ.1**: How do the LLM's language styles influence the user's preferences toward their interaction with the model?
- **RQ.2**: How do the user's individual traits moderate the influence of the LLM's language styles on the user's preferences?

In the following sections, we will answer our RQs through a series of exploratory and experimental user studies.

2 STUDY 1: Exploratory Study on Style and Preference in User-LLM Interaction

To answer **RQ.1**, we conduct an exploratory study on 3 *preference-alignment* datasets: ArenaPref [2], MultiPref [14], and ChatbotArena [22]. These secondary datasets contain a wide-variety of real-world User-LLM interactions, each representing different user populations. However, they don't contain user-specific information, such as the user's individual traits. Nonetheless, they have been prevalent in shaping the research of the Human-LLM preference alignment [5].

User-LLM Interaction Data Selection. Each instance in the dataset is composed of a user's query, 2 different LLM's responses, and the user's binary preference for the responses. We focused only on open-ended interaction scenarios and found that including only queries with interrogative prefixes (e.g. *what*, *how*, *are*) and without math, code, or computation keywords, effectively filters out non-open-ended scenarios. To minimize confounding variables, we further constrain our instances to only those that are in English, involve single-turn interactions, and have response pairs that are semantically similar by measuring their text embeddings' cosine similarity.

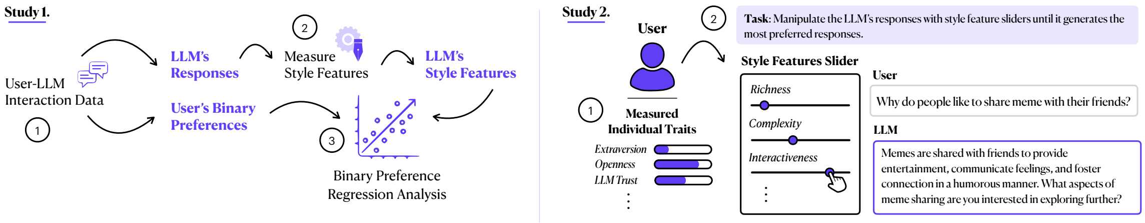
Stylistic Features Measurement. Drawing from works in the linguistic analysis of the LLMs and language style in general [4, 11, 12], we define 9 style features to be measured: information *richness*, information *presentation*, vocabulary *complexity*, usage of *active voices*, *figurativeness*, *friendliness*, *interactiveness*, *authoritativeness*, and *persuasiveness*. We implemented an NLP pipeline to measure the intensity level of each style feature. Depending on the implicitness of the style, we either measure them via rule-based NLP algorithms or neural-based models (details in App. A.1, A.2).

Binary Preference Regression Analysis. To analyze the influence of the measured LLM's style features on user's preferences, we performed a binary preference regression analysis [15] on each user population. Let $x_a, x_b \in \mathbb{R}^9$ be the style feature pairs of response a and b, and y be the preference toward a or b. We defined the independent variables as the difference between the style features, $x = x_a - x_b$, and the dependent variables as $y \in \{0, 1\}$, where $y = 1$ if the user prefers response a, and $y = 0$ otherwise. Our preference regression model is then defined as: $y = logit(\beta_0 + \sum_{i=1}^{9} \beta_i x_i)$.

2.1 Exploratory Study Findings

We examined the parameters of the fitted regression models, particularly the odds associated with the statistically significant ($p < 0.05$) style feature's coefficient, $1 - \exp(\beta_i)$, which describes the change of the user's preference resulting from an increase in a style feature's intensity. We visualize the result in Fig 1, with detailed results in App. A.3.

LLM's Language Style Does Influence User's Preference. We found that there are at least 3 statistically significant style features influencing user's preferences across user populations. In ArenaPref population, an increase in *Richness* (↑88.6%),



Task: Manipulate the LLM's responses with style feature sliders until it generates the most preferred responses.

Why do people like to share meme with their friends?

Memes are shared with friends to provide entertainment, communicate feelings, and foster connection in a humorous manner. What aspects of meme sharing are you interested in exploring further?

Fig. 2. Methodological Overview of **Study 1** (Left) and **Study 2** (Right). We attached the experimental GUI for Study 2 in App. B.1.

Complexity (↑26.9%), and *Friendliness* (↑28.9%) in the LLM's responses elevate user's preference. In ChatbotArena, it was *Richness* (↑68.3%), *Presentation* (↑16.0%), and *Figurativeness* (↑58.1%). Meanwhile, MultiPref seemed to be influenced by a more diverse set of styles: with an increase in *Presentation* (↑23.7%), *Complexity* (↑45.9%), *Interactiveness* (↑20.0%), *Persuasiveness* (↑17.9%), and a decrease in *Authoritativeness* (↓11.1%), likely elevate the user's preference.

Population-level Preference Influenced by Different LLM's Language Style. While the LLM's language style does have influence on user's preference, it varies across different user populations. This variation implies the presence of confounding variables outside of ones we've controlled in §2a, and possible moderating variables such as the user's demographics and traits that are unknown in the datasets, which we are about to investigate in the next study.

3 STUDY 2: Experimental Study on Individual Traits, Style, and Preference in User-LLM Interaction

To answer **RQ.2**, we conducted an experimental study involving human users. We deliberately collect and study the role of user's individual traits, a variable that tends to be missing in the current User-LLM preference datasets and study.

Experimental Design and Procedure. We designed a within-subject experiment where each user went through 2 stages. In the first stage, we asked them to fill-in questionnaires that measure their individual traits. In the second stage, we asked them to interact with an LLM via the provided interface, they will be given the ability to manipulate the LLM's language styles, and asked to manipulate the LLM's responses to their preferences. The core methods in this experiment follows [7, 18], which we designed to *sample* the LLM's language style that maximizes the user's preferences.

User Participants. Our user participants pool are based in the UK, use English as their primary language, within the age of 20-30, balanced by their sex, and are daily users of LLM services (e.g. OpenAI, Anthropic). We recruited 10 users from the Prolific platform, where each user contributed 60 samples of preferences. After filtering, we have a total of 162 valid preference samples to analyze. Detailed user's statistics is reported in App. B.2. It is important to interpret the rest of this findings cautiously, as we have not covered wider demographic diversity and larger sample sizes.

User's Individual Traits Measurement. We measure 2 sets of user's individual traits: their personality traits and trust toward the LLMs. To measure their personality traits, we administer the 10-item measure of the Big-5 personality dimensions [6] to the user, which measure the user's level of *Extraversion*, *Conscientiousness*, *Neuroticism*, *Openness*, and *Agreeableness*. To measure their trust toward the LLMs, we administer the ChatGPT Trust Scale [3] to the user.

Stimuli Design, Style-varying LLM's Responses. We first defined 3 queries for this study and prompt a LLM (OpenAI's GPT-4o-Mini) with a factual context to provide a baseline response. To craft responses in a variety of styles, we implemented a zero-shot style transfer pipeline following [16], which we designed to modify the baseline response to convey a given style features. As we have 3 queries and 9 style features with 3 intensity levels, we synthesized a total of $3 * 3^9 = 59,049$ LLM's responses as the possible stimuli for the users. Details of the prompts is attached in App B.3.

Sampling Preference-eliciting Style with People. Collecting user's preferences in a similar fashion as the datasets used in Study 1 (§2) would be prohibitively costly for us to do. We instead adopted *gibbs sampling with people* [7] method to effectively sample the LLM's responses with style features that maximize each user's preferences. Let $g(v_1, ..., v_9)$ be the LLM's response parametrized by the style features, the user will be asked to iteratively manipulate the intensity of $v_i \in \{1, 2, 3\}$ while others are fixed, then choose which intensity generates LLM's responses that they prefer the most. In the end, we would have chains of style features that converged toward a style combination the user prefer the most.

Moderated Binary Preference Regression Analysis. To analyze the moderation effect of user's individual traits, we expand the regression analysis in Study 1 to include the traits as moderator variables. For each measured individual trait, z_k, we performed moderated preference regression model defined as: $y = logit(\beta_0 + \sum_{i=1}^{9} \beta_i x_i + \sum_{j=1}^{9} \beta_j x_i z_k)$.

3.1 Experimental Study Findings

For each trait z_k, We examined the shift of odds associated with the statistically significant ($p < 0.05$) trait-moderated style feature's coefficient, $1 - exp(\beta_i + \beta_j z_k)$. We visualize the result in Fig. 1 (Right).

Individual Traits Moderate the Influence of Language Style Differently. We found that depending on user's individual traits, certain LLM's language styles influence user's preferences differently. For users with higher level of ↑*Agreeableness*, lower level of ↓*Figurativeness* and higher level of ↑*Authoritativeness* increase their preferences. For users with ↑*Extraversion*, it was ↓*Figurativeness*, ↓*Richness*, ↓*Complexity*, ↓*Friendliness*, and ↓*Active Voice*. Users with ↑*Neuroticism* influenced more by ↑*Figurativeness*, ↑*Active Voice*, ↓*Authoritativeness*, and ↓*Interactiveness*. For users with ↑*Openness*, it was ↑*Figurativeness*, ↑*Complexity*, and ↑*Authoritativeness*. Meanwhile, users with ↑*Trust toward LLMs* are influenced by ↓*Figurativeness*, ↓*Interactiveness*, and ↑*Active Voice*.

Polarizing Effects of Individual Traits. Though we have observed various statistically significant moderation effects of each trait independently, user's individuality is represented by a combination of these traits as whole. In the case of *Extraversion* and *Neuroticism* for example, we can see that these traits moderate user's preference in a polarizing way (Fig. 3). How does these dynamic apply for users with both high or low level of those traits? Which language styles will influence them more? Future studies could explore these questions further with more observational samples and applying techniques such as joint moderation effects or other explainability methods [13].

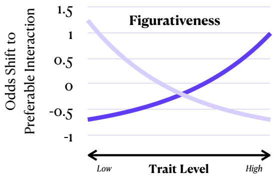


Fig. 3. **Example of Polarizing Moderator Effects**

4 Conclusion, Limitation, and Future Direction

In this paper, we presented our preliminary study on how user's very own individual traits and LLM's language style influence user's preferences in open-ended User-LLM interaction. As a preliminary study, it is important to interpret our findings with caution, given that our samples still need wider demographics diversity and larger sample sizes. Our future direction is to first address these limitations. We are then interested to further investigate the joint effects and possible causal relationship between and *beyond* our variables; Why do users with certain traits are more or less influenced by certain language style? Will the elicited preferences actually translate into better user experiences? Will it increase their susceptibility to LLM's misinformation, hallucination, and other risks?

References

[1] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

[2] Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. 2024. Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. arXiv:2403.04132 [cs.AI]

[3] Avishek Choudhury and Hamid Shamszare. 2023. Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis. *Journal of Medical Internet Research* 25 (2023), e47184.

[4] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Hinrich Schuetze, Pascale Fung, and Massimo Poesio (Eds.). Association for Computational Linguistics, Sofia, Bulgaria, 250–259. https://aclanthology.org/P13-1025/

[5] Shachar Don-Yehiya, Ben Burtenshaw, Ramon Fernandez Astudillo, Cailean Osborne, Mimansa Jaiswal, Tzu-Sheng Kuo, Wenting Zhao, Idan Shenfeld, Andi Peng, Mikhail Yurochkin, Atoosa Kasirzadeh, Yangsibo Huang, Tatsunori Hashimoto, Yacine Jernite, Daniel Vila-Suero, Omri Abend, Jennifer Ding, Sara Hooker, Hannah Rose Kirk, and Leshem Choshen. 2024. The Future of Open Human Feedback. arXiv:2408.16961 [cs.HC] https://arxiv.org/abs/2408.16961

[6] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

[7] Peter Harrison, Raja Marjieh, Federico Adolfi, Pol van Rijn, Manuel Anglada-Tort, Ofer Tchernichovski, Pauline Larrouy-Maestri, and Nori Jacoby. 2020. Gibbs sampling with people. *Advances in neural information processing systems* 33 (2020), 10659–10671.

[8] Yongnam Jung, Cheng Chen, Eunchae Jang, and S Shyam Sundar. 2024. Do We Trust ChatGPT as much as Google Search and Wikipedia?. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.

[9] Samia Kabir, David N. Udo-Imeh, Bonan Kou, and Tianyi Zhang. 2024. Is Stack Overflow Obsolete? An Empirical Study of the Characteristics of ChatGPT Answers to Stack Overflow Questions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 935, 17 pages. doi:10.1145/3613904.3642596

[10] Dongyeop Kang and Eduard Hovy. 2021. Style is NOT a single variable: Case Studies for Cross-Stylistic Language Understanding. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 2376–2387. doi:10.18653/v1/2021.acl-long.185

[11] Junlong Li, Shichao Sun, Weizhe Yuan, Run-Ze Fan, hai zhao, and Pengfei Liu. 2024. Generative Judge for Evaluating Alignment. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=gtkFw6sZGS

[12] Junlong Li, Fan Zhou, Shichao Sun, Yikai Zhang, Hai Zhao, and Pengfei Liu. 2024. Dissecting Human and LLM Preferences. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 1790–1811. doi:10.18653/v1/2024.acl-long.99

[13] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[14] Lester James V. Miranda, Yizhong Wang, Yanai Elazar, Sachin Kumar, Valentina Pyatkin, Faeze Brahman, Noah A. Smith, Hannaneh Hajishirzi, and Pradeep Dasigi. 2024. Hybrid Preferences: Learning to Route Instances for Human vs. AI Feedback. *arXiv* abs/2410.19133 (Oct. 2024).

[15] Alexander Peysakhovich, Virot Chiraphadhanakul, and Michael Bailey. 2015. Pairwise choice as a simple and robust method for inferring ranking data. In *WWW 2015 Conference Proceedings*.

[16] Emily Reif, Daphne Ippolito, Ann Yuan, Andy Coenen, Chris Callison-Burch, and Jason Wei. 2022. A Recipe for Arbitrary Text Style Transfer with Large Language Models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 837–848. doi:10.18653/v1/2022.acl-short.94

[17] Pranab Sahoo, Prabhash Meharia, Akash Ghosh, Sriparna Saha, Vinija Jain, and Aman Chadha. 2024. A Comprehensive Survey of Hallucination in Large Language, Image, Video and Audio Foundation Models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (Eds.). Association for Computational Linguistics, Miami, Florida, USA, 11709–11724. doi:10.18653/v1/2024.findings-emnlp.685

[18] Adam Sanborn and Thomas Griffiths. 2007. Markov Chain Monte Carlo with People. In *Advances in Neural Information Processing Systems*, J. Platt, D. Koller, Y. Singer, and S. Roweis (Eds.), Vol. 20. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2007/file/89d4402dc03d3b7318bbac10203034ab-Paper.pdf

[19] Michael Shumanov and Lester Johnson. 2021. Making conversations with chatbots more personalized. *Computers in Human Behavior* 117 (2021), 106627.

[20] Sarah Theres Völkel and Lale Kaya. 2021. Examining user preference for agreeableness in chatbots. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–6.

[21] Minghao Wu and Alham Fikri Aji. 2025. Style Over Substance: Evaluation Biases for Large Language Models. In *Proceedings of the 31st International Conference on Computational Linguistics*, Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert (Eds.). Association for Computational Linguistics, Abu Dhabi, UAE, 297–312. https://aclanthology.org/2025.coling-main.21/

[22] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric. P Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. arXiv:2306.05685 [cs.CL]

A Details on Study 1

A.1 Measurement of Language Stylistic Features

Below we listed the measurement methods for each stylistic feature we have defined. This collection of methods make up our stylistic feature measurement pipeline mentioned in §2.

Style	**Measurement Method**	**Main Tools**
Richness	Measure the frequencies of the nouns, adjectives, conjunctions, coordinating conjunctions, and subordinating conjunctions in the LLM's response.	spaCy's NLP pipeline
Presentation	Measure the presence of various markdown styling, such as bolding, italicizing, and list enumeration formatting in the LLM's response.	RegEx matching
Complexity	Measure the Dale-Chall readability score of the LLM's response.	Textstat library
Figurativeness	Measure the intensity level of figurativeness through zero-shot classification prompting.	OpenAI's GPT-4o-Mini as zero-shot classifier
Friendliness	Measure the intensity level of friendliness through zero-shot classification prompting.	OpenAI's GPT-4o-Mini as zero-shot classifier
Interactiveness	Measure the intensity level of interactiveness through zero-shot classification prompting.	OpenAI's GPT-4o-Mini as zero-shot classifier
Authoritativeness	Measure the intensity level of authorativeness through neural-based classifier model.	BERT-based model trained on Szeged Uncertainty Corpus
Persuasiveness	Measure the discrete intensity level of persuasiveness through zero-shot classification prompting.	OpenAI's GPT-4o-Mini as zero-shot classifier
Active Voice	Measure the frequencies of the linguistic pattern match of active and passive voices in the LLM's response.	spaCy's NLP pipeline and linguistic pattern matching

Table 1. Measurement methods and tools for each language style features.

A.2 Summary Statistics of Measured Styles Across Populations

Below we reported the summary statistics (mean and standard deviation) of the measured stylistic features in the LLM's responses across populations.

	Mean (SD) of Stylistic Features								
	Rich.	**Pres.**	**Comp.**	**Figu.**	**Frie.**	**Inte.**	**Auth.**	**Pers.**	**Acti.**
	[0, ∞]	[1.0,3.0]	[4.9, 9.9]	[1.0,3.0]	[1.0,3.0]	[1.0,3.0]	[0.0,1.0]	[1.0,3.0]	[0.0,1.0]
ArenaPref	45.83 (37.75)	0.35 (0.50)	9.28 (4.69)	1.01 (0.14)	1.12 (0.39)	1.17 (0.47)	0.64 (0.38)	1.12 (0.33)	0.79 (0.29)
ChatbotArena	42.10 (36.07)	0.28 (0.45)	8.88 (2.10)	1.01 (0.11)	1.07 (0.29)	1.10 (0.37)	0.66 (0.38)	1.13 (0.33)	0.77 (0.31)
MultiPref	49.14 (34.54)	0.35 (0.49)	9.00 (1.24)	1.00 (0.09)	1.18 (0.41)	1.15 (0.38)	0.49 (0.37)	1.46 (0.53)	0.84 (0.23)

Table 2. Summary statistics of the measured stylistic features.

A.3 Binary Preference Regression Results

Below we attached the numerical results version of Fig. 1 (Left).

Population	Rich.	Pres.	Comp.	Figu.	Frie.	Inte.	Auth.	Pers.	Acti.
ArenaPref	0.680**	0.160*	0.117	0.581**	-0.080	0.125	-0.004	-0.121	-0.050
ChatbotArena	0.880**	0.085	0.269*	0.810	0.289*	-0.089	0.034	-0.062	0.025
MultiPref	0.199	0.230*	0.450*	-0.110	0.100	0.200*	-0.110*	0.179*	0.116
Our Experiment (Study 2)	0.031	0.140	0.278^{+}	-0.017	0.402*	0.081	0.178	-0.070	-0.034

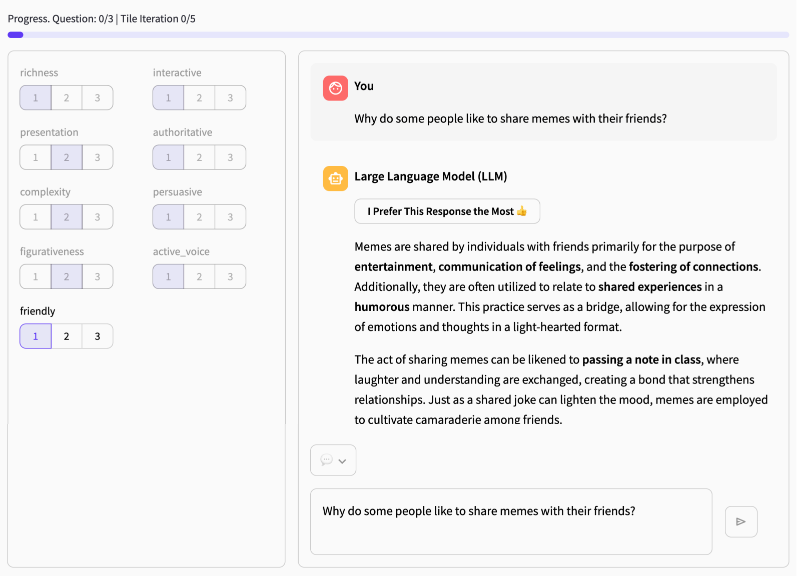
Table 3. Odds of a particular language style feature increasing or decreasing the user's preferences. Statistical significance: $^{**}: p < 0.01, ^{*}: p < 0.05, ^{+}: p < 0.1$.

B Details on Study 2

B.1 Sampling with People Interface

The user interface for sampling with people experiment is shown in Fig. 4. Each participant is given the following instruction to follow (the instruction is self-contained in the experimental GUI, we show it here for brevity):

> Go through every option in Tile #k, until the LLM gives you the response the you prefer the most among the options. After it gives the response that you feel you prefer the most, click '**I Prefer This Response the Most.**'
>
> - Always re-read the new manipulated response from start to finish.
> - Ask yourself: "Do I like this new response more than the previous one?"
> - You don't have to be overly objective or over-analyze your decision in preferring or liking the LLM's responses. In fact, you are encouraged to go with "what feels right".



Why do some people like to share memes with their friends?

Memes are shared by individuals with friends primarily for the purpose of **entertainment, communication of feelings,** and the **fostering of connections.** Additionally, they are often utilized to relate to **shared experiences** in a **humorous** manner. This practice serves as a bridge, allowing for the expression of emotions and thoughts in a light-hearted format.

The act of sharing memes can be likened to **passing a note in class,** where laughter and understanding are exchanged, creating a bond that strengthens relationships. Just as a shared joke can lighten the mood, memes are employed to cultivate camaraderie among friends.

Why do some people like to share memes with their friends?

Fig. 4. User interface for our sampling with people experiment. During the actual experiment, the style names and intensity levels are not shown to the users, and the order of both are randomized for every iteration.

B.2 User Participants Demographics

Below we report our participant pool's demographics and individual traits.

Participant's Demographics		
Age	Mean (SD)	25.30 (2.83)
	Range	20-30
Sex	Female	5
	Male	5
Ethinicity	Asian	1
	Black	6
	White	1
	Mixed	1
	Prefer not to say	1
Daily Usage of AI/LLM	Every day	3
	Multiple times every day	7
LLM Service Usage (A user can report multiple LLM services they used)	OpenAI's ChatGPT	10
	Anthropic's Claude	5
	Google's Gemini	4
	Others	8
Participant's Individual Traits		
Extraversion	Mean (SD)	2.85 (0.92)
	Range	1-5
Agreeableness	Mean (SD)	4.00 (0.77)
	Range	1-5
Conscientiousness	Mean (SD)	4.25 (0.71)
	Range	1-5
Neuroticism	Mean (SD)	2.40 (0.86)
	Range	1-5
Openness	Mean (SD)	3.60 (0.73)
	Range	1-5
Trust toward LLM	Mean (SD)	3.59 (0.37)
	Range	1-4
Num. of Participants		10
Num. of Rejected Participants		1
Num. of Preference Samples per Participants		60
Num. of Gibbs Sampling's Burn-in Period		2
Final Num. of Preference Samples After Filtering		**n = 162**

Table 4. Distribution of user participant's demographics and individual traits in Study 2.

B.3 Zero-shot Style Transfer Prompting

The following are the description of the intensity level for each stylistic features, particularly used to prompt the zero-shot style transfer pipeline for synthesizing our stimuli.

Style	**Intensity Level & Description**
Richness	**L1.** Provides a straightforward, unembellished answer to the question, focusing solely on the essential information. **L2.** Offers the answer along with additional information that adds context or important details but remains relevant to the question. **L3.** Provides the answer along with excessive details, tangents, or background information that, while interesting, does not directly support the original question.
Presentation	**L1.** Using a single paragraph to structure the utterance without any formatting elements. **L2.** Using two paragprahs to structure the utterance. Important words and phrases in the utterance are formatted with bold or italic style. **L3.** Using more than two paragprahs to structure the utterance with bolding, italicizing, headings, bullet points, numbered list the key words and phrases as the formatting elements.
Complexity	**L1.** Using vocabulary of verbs, nouns, adjectives, and adverbs that are very easy to read. Easily understood by an average twelve year old student. **L2.** Using vocabulary of verbs, nouns, adjectives, and adverbs that are moderately difficult to read. Best understood by an average high-school student. **L3.** Using vocabulary of verbs, nouns, adjectives, and adverbs that are very difficult to read. Best understood by university graduates and experienced scholars.
Figurativeness	**L1.** Does not convey any figurative language. **L2.** Re-emphasize an explanation by figurative language in the form of simple metaphor that introduce direct comparisons using common ideas. **L3.** Re-emphasize an explanation by figurative language in the form of complex metaphors that introduce imaginative yet relatable ideas.
Friendliness	**L1.** Does not convey expressions of friendliness. **L2.** Use expression that convey politeness, warmth, approachable, and come off as formal. **L3.** Use expression that convey politeness, warmth, approachable, and come off as informal.
Interactiveness	**L1.** Does not seek further engagement or clarification regarding the query of the utterance. **L2.** Attempt to engage with the user’s curiosity. These include prompting the user to consider a broader context or related topics. **L3.** Attempt to engage with the user’s curiosity and intention. These include explicitly asking for more relevant information or seeking to understand the user’s intent.
Authoritativeness	**L1.** Using expressions that are lacking in confidence and detail. These includes the incorporation of tentative languages (e.g., "maybe," "might be," "I think"). **L2.** Does not convey expressions of authoritativeness. **L3.** Using expressions that exudes confidence and expertise. These includes the incorporation of assertive language (e.g., “is,” “will,” “must”).
Persuasiveness	**L1.** Does not use expressions that attempts to convince the user more to accept the information, statements, facts, or opinions in the utterance. **L2.** Attempts to convince the user more to accept the information or opinions in the utterance, using moderate emotional appeal or reasoning which lacks deeper engagement or urgency. **L3.** Attempts to convince the user more to accept the information or opinions in the utterance, using strong emotional appeal or reasoning to effectively convince the user.
Active Voice	**L1.** Always using passive voice, if the context is appropriate, aiming for a less direct and less engaging style of communication. **L2.** Using a mix of both passive and active voice, striking a balance between engagement and formality style of communication. **L3.** Always using active voice, if the context is appropriate, resulting in clear, direct, and engaging style of communication.

Table 5. Description of stylistic features’ intensity level used for our zero-shot style transfer pipeline to synthesize style-varying LLM’s responses.