"When AI Writes Personas": Analyzing Lexical Diversity in LLM-Generated Persona Descriptions

Sankalp Sethi
College of Information Science
University of Arizona
Tucson, Arizona, USA
sankalpsethi@arizona.edu

Joni Salminen
University of Vaasa
Vaasa, Finland
jonisalm@uwasa.fi

Danial Amin
University of Vaasa
Vaasa, Finland
danialam@uwasa.fi

Bernard J Jansen
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
bjansen@hbku.edu.qa

Abstract

Large language models (LLMs) are increasingly employed in generating user personas representing various groups of people. It is vital that these personas do not contain major sources of bias for stakeholders using the personas. To investigate linguistic bias in LLM-generated personas, we apply eleven lexical diversity metrics to analyze the association between linguistic diversity in 600 persona descriptions generated using five LLMs (GPT, Claude, Gemini, DeepSeek, Llama) and demographic attributes (age, gender, country) of the personas. We find that LLM-generated persona descriptions are lexically diverse independently of the personas' demographic attributes. While we find no significant demographic bias in the persona profiles, we do find significant differences between the lexical diversity of persona descriptions generated by the LLMs. The persona descriptions generated by Gemini 1.5 Pro have the highest lexical diversity. The results imply that current LLMs can generate lexically diverse persona descriptions, but the selection of an LLM for specific applications is an important decision.

CCS Concepts

• **Human-centered computing** → *Empirical studies in HCI*.

Keywords

AI, LLMs, user personas, lexical diversity, evaluation

ACM Reference Format:
Sankalp Sethi, Joni Salminen, Danial Amin, and Bernard J Jansen. 2025. "When AI Writes Personas": Analyzing Lexical Diversity in LLM-Generated Persona Descriptions. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3706599.3719712

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719712

1 Introduction

Personas are humanized depictions of user segments that are used for user representation and understanding in user experience (UX) design, product development, and human-computer interaction (HCI) research and practice [8, 15]. Personas are usually presented as a profile. A key component of the persona profile is the persona description (also called 'narrative'), a text describing the persona's attributes, including needs, preferences, and background in a narrative format [39]. See Figure 1 for an example persona description.

Traditionally, persona descriptions have been written by human persona developers. However, this is increasingly changing due to the ability of Generative AI (GenAI) and large language models (LLMs) to generate fluent text content [18, 54, 61]. Natural language processing (NLP) has contributed to data-driven persona development by providing persona developers with multiple techniques to computationally process data, such as from user interviews or user-generated social media content [51]. Current state-of-the-art NLP technologies include GenAI and LLMs, which are rapidly influencing HCI research, including persona development [18, 45, 54]. The ability of the current generation of LLMs to generate context-sensitive and detail-rich text [2] makes persona description generation a seemingly fitting use case for *LLMs in HCI*. LLMs can contribute to several tasks in persona development, ranging from data analysis to writing persona descriptions [26].

The evaluation of user personas is a central research topic in persona research [7, 51]. One critical aspect of evaluation is *diversity*, referring to how varied the developed personas are in their representation of various end-user groups. It is believed that more diverse personas also yield more inclusive design choices [53]; that is, covering more (especially underrepresented) user groups [13]. Most existing work evaluating persona diversity focuses on the demographic diversity of persona sets [23], LLM-generated or otherwise [29, 53], analyzing how well the persona set covers the groups represented by the personas [30]. The role of demographics in persona development and use is important as research has found many effects of varied demographics of personas on stakeholder perceptions of personas [24, 27, 52, 55].)

The proliferation of textual content generated by LLMs has prompted research into benchmarking the linguistic diversity of

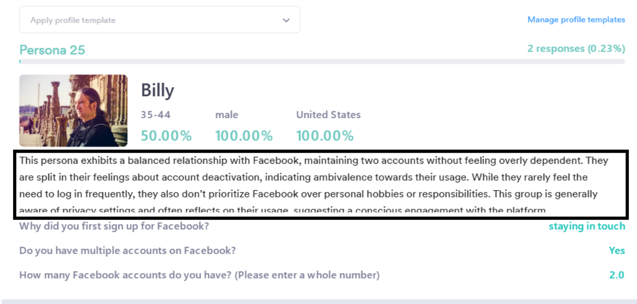

Figure 1: An example persona obtained from Survey2Persona, a system using LLMs for writing persona descriptions (a snippet of which is highlighted in the figure).

LLM generated content [16, 48, 49, 61]. Linguistic diversity can be broadly classified into *semantic*, *syntactic*, and *lexical* diversity [16]. In our study, we focus on lexical diversity, leaving other forms of linguistic diversity for future work. *Lexical diversity* measures the range and variety of words used in a text sample and is an indicator of vocabulary richness and textual complexity [3, 25, 63]. In the context of persona descriptions generated by LLMs, lexical diversity can be an important indicator of how well the description captures the nuanced characteristics and attributes of the persona [32].

The need for a lexical analysis of persona descriptions generated by LLMs is emphasized by evidence of bias and potentially harmful stereotypes [1, 32, 42] in LLM-generated text that could make their way into persona descriptions generated by LLMs [28, 54]. A lexical analysis of LLM-generated persona descriptions can reveal valuable information about the diversity of LLM-assisted user representation, potential biases and risks involved in it, and practical guidelines about how to quantify this lexical diversity and what LLMs to use in order to create lexically diverse persona descriptions.

Against this backdrop, we put forth the following research questions (RQs):

- **RQ1**: *How lexically diverse are LLM-generated personas?*
- **RQ2**: *Is there a dependence between demographic attributes and lexical diversity in LLM-generated personas?*
- **RQ3**: *Does the lexical diversity of LLM-generated personas vary by the LLM used?*

RQ1 can provide results that can be compared against lexical diversity baselines and conventionally developed persona descriptions [32, 49]. High lexical diversity might be desirable in LLM-generated persona descriptions, which are representative of entire groups [30], thus, in turn, leading to more inclusive design.

For **RQ2**, a dependence between demographic attributes and lexical diversity could indicate bias in which the LLM generates, for example, lower-quality descriptions for certain demographic groups [32]. Finally, **RQ3** allows us to make informed choices in selecting LLMs to maximize lexical diversity in persona descriptions.

In terms of positioning, our work examines the role of LLMs in persona development and, as such, exemplifies the convergence of HCI and NLP in addressing "grand challenges" in persona development [51]. In the remainder of the work, we will first concisely review the related work. After this, we present our methodology and findings for each RQ. In the discussion, we summarize the implications of these findings and suggest future research directions.

2 Related Work

Diverse user representation is a key principle in user-centered design (UCD) [10, 19, 39–41, 46], since global user populations are increasingly heterogeneous and, therefore, require variation in persona sets to be represented in a fair and balanced manner [37]. Diverse user understandings can promote inclusion by representing a wider range of user needs and attributes in the design process [13, 14]. At its best, diversity empowers designers to create relevant, inclusive products and features [8] that are accessible and usable to those considered fringe or marginalized users [12, 62].

However, using personas as an inclusive design tool requires that personas representing user groups within a population in a diverse way [31, 38]. Design teams employing diverse persona sets are better equipped to identify and address the needs of underrepresented user segments early in the development process [13, 22, 36]. Diversity in persona sets is associated with representing users in various demographic and behavioral contexts [47, 50]. Furthermore, it is essential to address ethical considerations in data-driven persona development, particularly the risk that algorithms may emphasize majority groups instead of fringe user groups [60]. This is particularly important given the proliferation of algorithms in the persona development process [21, 54].

While previous work has investigated human-AI persona generation workflows [59], qualitative analysis of LLM-generated personas [56] and persona diversity based on demographic attributes [54], to our knowledge, no previous study has specifically investigated *lexical* diversity in LLM-generated persona descriptions. Yet, analyzing lexical diversity in LLM-generated personas helps to evaluate how well LLMs can create distinct character descriptions without falling into repetitive language that might contain patterns of generic stereotypes.

3 Methodology

3.1 Overview

We applied the methodology of Salminen et al. [54] using publicly available Jupyter notebooks. We used our own OpenAI API keys to run the persona generation code in the notebooks (without any modification) and obtained a set of 450 personas generated by GPT-4o[1] to address **RQ1** and **RQ2**.

The first stage involved creating a "skeletal" [54] persona that does not contain a detailed persona description. In the second stage, we used this skeletal persona in the prompt to generate a detailed persona description. These persona descriptions were then passed through a custom NLP pipeline for pre-processing and lexical diversity calculations (detailed in subsequent sections). We have made the pre-processing and the NLP pipeline code available through Jupyter notebooks[2] which researchers can use to re-run the analysis or calculate lexical diversity metrics of their own.

Statistical analysis was performed on the resulting lexical diversity metric data to address **RQ1** and **RQ2**. **RQ3** was addressed by generating 30 new personas from five different models (for a

[1]All LLMs used in this study are the mentioned versions as of 7 January 2025.
[2]Available at: https://bit.ly/lexical-diversity-personas-supplementary-material

total of 150 new personas) using a modified version of the original prompt that was previously used to generate the personas used in **RQ1** and **RQ2**. A 120 word limit was enforced through the prompt to ensure consistency in the length of the persona descriptions across all models because text length might otherwise impose a confounding factor. (This was not the case as, due to this conditioning, the observed length varied only little by personas; M = 78.5, SD = 6.1 words.) For RQs 1 and 2, since we used the 450 personas developed by Salminen et al. [54], the word limit was not applied. Lexical diversity metrics were calculated based on the descriptions of these new personas, and a Kruskal-Wallis test was performed on each metric to compare the central tendencies of the metrics across models. A post-hoc analysis was performed using Dunn's test for the metrics that exhibited statistically significant differences.

3.2 Model Selection

For investigating *RQ1* and *RQ2*, we used GPT-4o, as it represented the state-of-the-art model at the time of our study and was supported by the implementation of Salminen et al. [54]. This allowed us to efficiently generate a substantial dataset of 450 personas while maintaining methodological consistency with prior work. GPT-4o has also shown strong capabilities in generating contextually rich text [2], making it an ideal candidate for establishing baseline lexical diversity patterns in persona descriptions. For RQ3, we expanded our analysis to include multiple models (GPT 4o, Claude 3.5 Sonnet, DeepSeek V3, Gemini 1.5 Pro and LLaMa 3.1) to provide a comparative perspective on lexical diversity across different LLM architectures. This study design allowed us to first establish fundamental patterns in lexical diversity (RQ1) and its relationship with demographic attributes (RQ2) before expanding to cross-model comparisons (RQ3).

3.3 Metric Selection

Lexical diversity has a rich academic landscape with applications in multiple domains, including the assessment of language disorders [9, 11], the quality of writing [17], language development [64] and NLP tasks [6, 18, 32, 65].

There are open-source Python libraries that implement different sets of lexical diversity metrics. For our study, we applied the `LexicalRichness` [58] library because of its extensive coverage of lexical diversity metrics, detailed documentation, compatibility with our own custom pre-processing pipeline and proven record of use in academic research [44, 49]. Using multiple metrics also aligns with Mccarthy and Jarvis' recommendation of *"using multiple metrics in research studies [rather than any single index] noting that lexical diversity can be assessed in many ways and each approach may be informative as to the construct under investigation"* [34].

Since our data is predominantly limited to persona descriptions of 120 words, which translates to a lower range of 40-50 words after stop-word (common words like 'a', 'and', 'the') removal and further processing, our data falls below the suitable text length threshold for metrics like Measure of Textual, Lexical Diversity (MLTD) and vocD [34, 35], which have thus been excluded from our chosen metrics for this study.

A description of the lexical diversity metrics used in this study is available in the supplementary material [3]. The expected values for high diversity have been referenced from the results of empirical studies and existing research [33, 64]. It should also be noted that these studies also emphasize the shortcomings of these metrics, such as variability of Type-Token Ratio (TTR) with respect to word count [34]. Our study, by using a collection of these metrics, balances the outcomes against the individual weakness for a given metric. For example, CTTR and RTTR are corrected versions of TTR which overcome their dependence on token length.

3.4 Data Processing

We performed standard data pre-processing on the textual persona description data. The original text was sequentially processed by first normalizing the text by lower casing, removing special characters, and removing excess white spaces. Stop words were then removed and lemmatization (i.e., reducing the inflected forms of a word to its base form, e.g. driving -> drive) was performed on the normalized text using the NLTK library [4]. The persona's name is also removed with the stop words during this step.

Our selection of demographic attributes (age, gender, and country) follows an established precedent in persona research [30, 54], where these attributes form the main demographic identifiers in persona profiles. These attributes are consistently present in persona templates and are known to potentially influence perception and stereotyping [65], making them appropriate focal points for investigating potential biases in lexical diversity.

The pre-processed text was used to generate LD metrics via the `LexicalRichness` [58] package. The resulting metrics form the basis for our analysis in this study. For **RQ3**, we created a set of personas using different LLMs for comparative analysis. The prompt of Salminen et al. [54] was used to create 30 personas using each model. The following models were used: (a) ChatGPT 4o, (b) Claude 3.5 Sonnet, (c) DeepSeek V3, (d) Gemini 1.5 Pro, and (e) LLaMa 3.1 (405b). These represented state-of-the-art models at the time we conducted the study. The persona descriptions were passed through the same pre-processing pipeline, and metric calculation was performed as in **RQ1**. Since this part of the study aggregation of results across metrics for each model to calculate overall lexical diversity, we transformed the values of the metrics where a lower score is better to a *higher is better* scale.

4 Results

4.1 RQ1: How Lexically Diverse Are LLM-Generated Personas?

We first illustrate the difference between persona descriptions with marked and noticeable differences in lexical diversity in Table 1. As an example, we chose two LLM-generated persona descriptions (P209 and P281) from our sample. These descriptions are for personas representing alcohol addiction. Repeated words negatively affect the lexical diversity and are highlighted in red. P209 scores lower on all the eleven lexical diversity metrics used in this study. Taking Type-Token Ratio (TTR) as a reference, a TTR of 0.59 for

[3]Available at: https://bit.ly/lexical-diversity-personas-supplementary-table

P209 indicates moderate lexical diversity and a higher rate of repetition of words in the persona description as compared to **P281** with a TTR of 0.82, which indicates high lexical diversity.

To understand the central tendency, stability, and overall distribution of the lexical diversity metrics across our set of 450 persona descriptions generated using GPT-4o for **RQ1**, we performed descriptive statistical analysis (see Table 2).

The key takeaway from these descriptive statistics is that LLMs can generate lexically diverse persona descriptions. Across all the metrics, we see high mean values, indicating that these persona descriptions contain a wide ranging vocabulary, balanced word usage, and overall linguistic variety [33].

Figure 2 illustrates the distribution of the eleven normalized lexical diversity across the 450 LLM-generated persona descriptions created for **RQ1**.

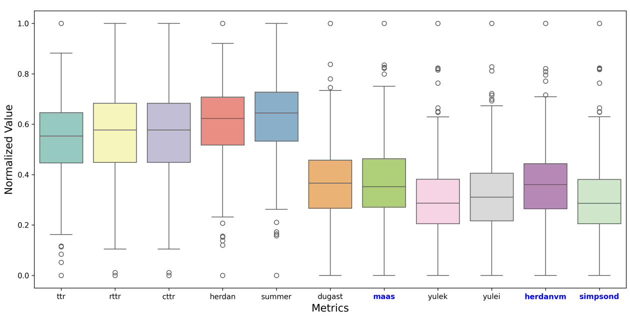

Figure 2: Boxplot of normalized lexical diversity metrics. The boxplot represents the distribution of each metric calculated on a sample of 450 persona descriptions. Metrics where a lower value is better are highlighted in blue. The metrics have varying distributions and stability across the entire sample and between metrics.

The metrics show varying distributions but low coefficients of variation overall, suggesting consistent lexical diversity scores across our dataset. This indicates that the personas generated by LLMs maintain lexical diversity regardless of the context or length of the persona description. In other words, the persona descriptions exhibit consistent depth and detail.

We further investigated the stability of diversity metrics using the coefficient of variation (CoV), because it is dimensionless and allows comparison between metrics despite their scale [66]. We chose the cut-off for CoV at 0.20 for stability [57].

Overall, the results indicate high lexical diversity and stable distribution with minimal outliers across all metrics. The stability of these metrics translates into low variance across the dataset. This minimizes the influence of fluctuations caused by text length, demographic attributes, or randomness.

4.2 RQ2: Is There a Dependence between Demographic Attributes and Lexical Diversity in LLM-generated personas?

Regression analysis was conducted using Ordinary Least Squares (OLS) to examine the relationship between Age, Gender, and Country predictors and the eleven lexical diversity metrics.

The results indicate that none of the demographic attributes (age, gender, or country) significantly predicts changes in the lexical diversity metrics. While RTTR ($B = -0.0141$) and CTTR ($B = -0.0100$) exhibit statistically significant negative relationships with age, the B for these relationships is very small, indicating a very weak relationship. Overall, these results suggest that the lexical diversity of LLM-generated persona descriptions is independent of the demographic attributes of the personas themselves.

4.3 RQ3 : Does the Lexical Diversity of LLM-Generated Personas Vary by LLM Used?

The Kruskal-Wallis test was conducted to assess the differences in lexical diversity across groups (models) for each metric. The results revealed statistically significant differences for all metrics[4].

The p-values for all metrics are highly significant ($p < .05$), suggesting that for each metric, at least one group's (model's) median differs significantly from the others.

Since all metrics exhibit statistically significant variability across groups, post-hoc pairwise comparisons using Dunn's test were conducted to evaluate differences in lexical diversity metrics between groups. Bonferroni correction was applied to adjust for multiple comparisons across model pairs to ensure stringent control over Type I error rate. Statistically significant differences between multiple model pairs were observed across all metrics [4].

Finally, we measured the lexical diversity of each LLM across the lexical diversity metrics to identify the model that generates persona descriptions with the highest lexical diversity. To enable a direct comparison, we transformed the values of the metrics (i.e., Maas, Herdan-VM and Simpson-D) in which a lower score is better. The values for Maas, Herdan-VM and Simpson-D were transformed to a *higher is better* scale using min-max normalization (applied to the reciprocal of the original value).

The lexical diversity metrics for each model were normalized using MinMax normalization. The median for each metric-model pair was calculated and plotted on a radar plot (see Figure 3). The area under the polygon formed by each model was calculated using the *Surveyor's Area Formula* [5] that gives us the models' overall lexical diversity across all metrics. The larger the area of the polygon, the higher the overall lexical diversity of the model.

The results indicate that Gemini 1.5 Pro had the highest lexical diversity measured by our set of metrics among the models with the largest area under the polygon formed in the radar plot. The models can be ranked in descending order of their overall lexical diversity (see Figure 4) by the area of their respective polygons.

5 Discussion

5.1 Findings Concerning Research Questions

Results on **RQ1** show that LLM-generated persona descriptions are lexically diverse, as measured by all metrics. Compared to benchmark values from the previous literature [33, 64], the persona descriptions scored high on normalized scales. The relative stability

[4]Results of the statistical analyses performed can be viewed in the supplementary material at: https://bit.ly/lexical-diversity-personas-supplementary-material

Table 1: Comparison of lexical diversity in two persona descriptions. Repeated words (highlighted in red) reduce the overall lexical diversity of the text. In P209, we see the words *farm* repeated 8 times and *family* repeated once. In P281, there is a lower incidence of repeated words, with *college* and *social* repeated once each, making it more lexically diverse than P209, which is reflected in the TTR value.

P209 (Lower Diversity - TTR = 0.59)	P281 (Higher Diversity - TTR = 0.82)
P209 is a 49-year-old **farmer** who has found himself struggling with alcohol. Born and raised on a **family farm**, P209 was exposed to the demanding nature and responsibilities of **farm** life from a very young age. Growing up on the **family farm** made P209 accustomed to the rigors of rural life, including the seemingly never-ending days of labor and the relentless physical demands of **farm** work. As an adult, P209 now manages the **farm** himself, working long hours and taking on all of the day-to-day responsibilities. Although the **farm** has provided P209 with a stable occupation and a sense of pride in his work, it has also come at a high personal cost. Dealing with the day-to-day stressors and routine of **farm** life, coupled with feelings of isolation from the lack of social interaction inherent in rural living, P209 turned to alcohol as a way to cope with his emotional struggles.	P281 is a 22-year-old **college** student currently studying **social** sciences at a well-known university. He is a member of a popular fraternity on campus, which plays a significant role in his **social** life. As part of this fraternity, P281 has had the opportunity to make several friends and enjoy a wide range of activities, both on and off-campus. As is quite common among **college** students, partying is a significant aspect of P281's life. During his freshman year, many of these gatherings involved excessive drinking and the use of recreational drugs. Unfortunately, this environment provided P281 with ample opportunity to experiment with various substances, including opioids. Initially, P281 only used opioids occasionally, believing that this would prevent him from developing an addiction. However, over time, he found himself using increasingly higher doses and more frequent administration to achieve the desired euphoric effects. Inevitably, this led him down the path of opioid addiction.

Table 2: Central tendencies of the lexical diversity metrics. Metrics where a lower value is better are highlighted in blue. The mean values for all metrics lie between our expected values for high diversity (greater than 70th percentile for metrics with an upper bound). The metrics also exhibit low coefficients of variation, indicating that the metrics remain stable across the data, indicating consistency in the lexical diversity of LLM-generated persona descriptions.

Metric	Expected Value (High Diversity)	Mean (M)	SD	CoV	Diversity
TTR	0.7 – 0.9	0.72	0.05	0.063	High
RTTR	10 – 15	10.70	0.76	0.071	High
CTTR	7 – 10	7.57	0.54	0.071	High
Herdan	0.8 – 1.0	0.94	0.01	0.012	High
Summer	0.8 – 1.0	0.96	0.007	0.007	High
Dugast	90 – 150	90.96	15.95	0.175	High
Maas	0.005 – 0.02	**0.01**	**0.002**	**0.178**	High
Yulek	20 – 80	63.10	14.44	0.229	High
Yulei	70 – 150	72.83	22.23	0.305	High
Herdan-VM	0.06 – 0.1	**0.067**	**0.008**	**0.126**	High
Simpson-D	0.002 – 0.01	**0.006**	**0.001**	**0.229**	High

of these metrics reflects LLMs' ability to produce lexically rich persona descriptions. This aligns with previous findings on the ability of LLMs to generate high-quality, lexically diverse text [2, 32, 54].

For **RQ2**, we did not find a statistically significant relationship between demographic attributes of age, gender, or country and the lexical diversity of the persona descriptions. The only statistically significant relationship between age and two of our 11 lexical diversity metrics (RTTR and CTTR) was found to be extremely weak with $B = -0.0141$ and $B = -0.0100$, respectively.

The findings to **RQ3** revealed statistically significant differences between the selected models for each individual metric, making model choice a factor to consider in maximizing the lexical diversity of LLM-generated persona descriptions. A performance analysis across all metrics indicated that the persona descriptions generated by Gemini 1.5 Pro had the highest lexical diversity.

5.2 Limitations and Future Research Directions

Even though we employed metrics that measure conventional lexical diversity, their efficacy and relevance in persona evaluation lack scholarly attention. A unified quantitative model for measuring lexical diversity relevant to persona generation could inform decisions of which metrics correlate with stakeholders' experience of personas. So, more needs to be known about the relationship between lexical diversity and the "quality" of personas, especially from the perspective of persona users. There is also a need to explore similar metrics in the realm of semantic [20] and syntactic [43] diversity to quantify overall linguistic diversity. A correlation analysis between linguistic diversity scores and human evaluations by subject matter experts could form the basis of future research.

In assessing bias, **RQ2** showing significant findings would have signaled a bias toward a particular demographic attribute. However, the lack of it (as found in this study) does not posit the absence of bias. A conclusive answer to whether LLM-generated persona

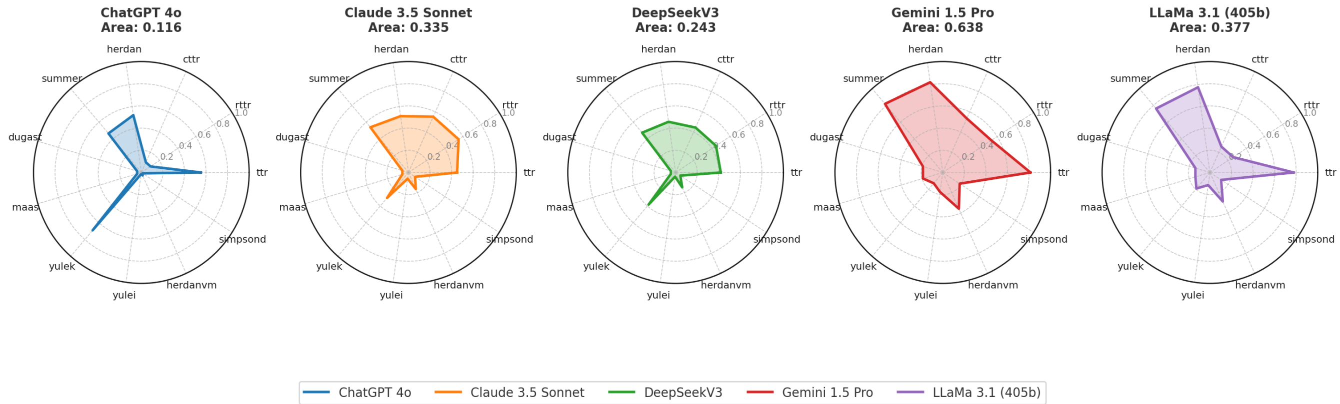

Figure 3: A comparison of the overall lexical diversity of the models. Each individual radar plot depicts a model's overall lexical diversity across the metrics. Within each plot, each radial axis represents the median value of a lexical diversity metric. These metrics have been MinMax normalized before plotting. The model with the highest overall lexical diversity across all metrics will have the largest area under the polygon.

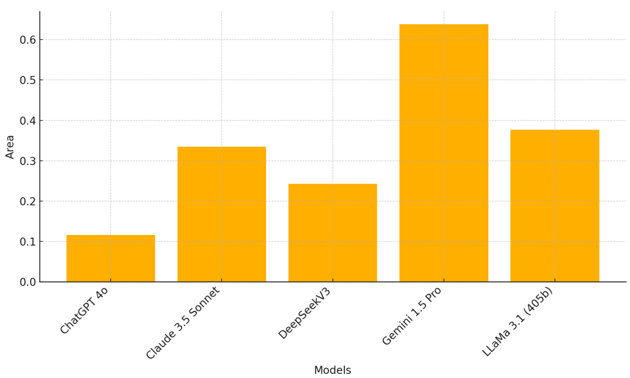

Figure 4: Overall lexical diversity of the LLMs for persona generation based on areas under the polygon in Figure 3.

descriptions are free of demographic bias would require a cross-sectional analysis of lexical diversity with demographic attributes over a larger sample, with descriptions of varying lengths and additional variables. While outside the scope of this study, this is a needed direction for future research.

Moreover, while we have studied the results from multiple models on a control prompt, the effect of prompt engineering with respect to prompt length, structure, and input data has not been studied. It offers a promising avenue for future research. In a similar vein, our work paves the way for automated evaluation systems that could be applied before providing the persona descriptions to stakeholders in order to ensure that the personas contain sufficient levels of lexical diversity. The use of NLP metrics in the automatic evaluation of persona descriptions thus offers a formidable avenue for future work that is currently understudied. For example, our approach could be further developed into an algorithm that would select persona descriptions from a pool of *candidate persona descriptions* to maximize the *overall* diversity of user representation at a persona set level. Therefore, we believe that our work, exploratory in nature, adds value to HCI research on the impact of LLMs in persona generation and the application of NLP in HCI domains. To this end, we share our programming code to facilitate further explorations in LLM-generated personas.

References

[1] William Babonnaud, Estelle Delouche, and Mounir Lahlouh. 2024. The Bias that Lies Beneath: Qualitative Uncovering of Stereotypes in Large Language Models. *Swedish Artificial Intelligence Society* (2024), 195–203.

[2] Jason Baronova, Catherine Stevens, Logan Tennant, and Alfred MacPhee. 2024. Dynamic context-aware representation for semantic alignment in large language models. (2024).

[3] Yves Bestgen. 2024. Back to Basics in Measuring Lexical Diversity: Too Simple to Be True. *Applied Linguistics* 45, 5 (Oct. 2024), 926–932. doi:10.1093/applin/amae053

[4] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

[5] Bart Braden. 1986. The surveyor's area formula. *The College Mathematics Journal* 17, 4 (1986), 326–337.

[6] Erik Cambria. 2024. Semantics Processing. In *Understanding Natural Language Understanding*. Springer, 113–228.

[7] Chris Chapman, Edwin Love, Russell P. Milham, Paul ElRif, and James L. Alford. 2008. Quantitative Evaluation of Personas as Information. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 52, 16 (Sept. 2008), 1107–1111. doi:10.1177/154193120805201602

[8] Alan Cooper. 1999. *The inmates are running the asylum.* Springer.

[9] Kevin T. Cunningham and Katarina L. Haley. 2020. Measuring Lexical Diversity for Discourse Analysis in Aphasia: Moving-Average Type–Token Ratio and Word Information Measure. *Journal of Speech, Language, and Hearing Research* 63, 3 (2020), 710–721. doi:10.1044/2019_JSLHR-19-00226

[10] Nana Kesewaa Dankwa and Claude Draude. 2021. Setting Diversity at the Core of HCI. In *Universal Access in Human-Computer Interaction. Design Methods and User Experience*, Margherita Antona and Constantine Stephanidis (Eds.). Springer International Publishing, Cham, 39–52.

[11] Gerasimos Fergadiotis, Heather Wright, and Thomas West. 2013. Measuring Lexical Diversity in Narrative Discourse of People With Aphasia. *American Journal of Speech-Language Pathology* 22 (05 2013), S397–S408. doi:10.1044/1058-0360(2013/12-0083)

[12] Jennifer Goodman and Michelle Broome. 2012. *A designer's research manual: Succeed in design by knowing your clients and what they really need.* Rockport Publishers.

[13] Joy Ai-Leen Goodman-Deane, Mike Bradley, Sam Waller, and P. John Clarkson. 2021. Developing personas to help designers to understand digital exclusion. 1 (2021), 1203–1212. doi:10.1017/pds.2021.120 Publisher: Cambridge University Press.

[14] Jonathan Grudin. 2006. Why Personas Work: The Psychological Evidence. In *The Persona Lifecycle*, John Pruitt and Tamara Adlin (Eds.). Elsevier, 642–663. doi:10.1016/B978-012566251-2/50013-7

[15] Jonathan Grudin and John Pruitt. 2002. Personas, participatory design and product development: An infrastructure for engagement. In *Proceedings of Participation and Design Conference (PDC2002)*, Vol. 2. Sweden, 144–161.

[16] Yanzhu Guo, Guokan Shang, and Chloé Clavel. 2024. Benchmarking Linguistic Diversity of Large Language Models. *arXiv preprint arXiv:2412.10271* (2024).

[17] Hye Seung Ha. 2019. Lexical Richness in EFL Undergraduate Students' Academic Writing. *English Teaching* 74, 3 (2019), 3–28.

[18] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating large language models in generating synthetic hci research data: a case study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[19] Julia Himmelsbach, Stephanie Schwarz, Cornelia Gerdenitsch, Beatrix Wais-Zechmann, Jan Bobeth, and Manfred Tscheligi. 2019. Do We Care About Diversity in Human Computer Interaction: A Comprehensive Content Analysis on Diversity Dimensions in Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3290605.3300720

[20] Paul Hoffman, Matthew A Lambon Ralph, and Timothy T Rogers. 2013. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior research methods* 45 (2013), 718–730.

[21] Pei-Fang Hsu, Yu-Han Lu, Shih-Chu Chen, and Patricia Pei-Yi Kuo. 2024. Creating and validating predictive personas for target marketing. *International Journal of Human-Computer Studies* 181 (2024), 103147. doi:10.1016/j.ijhcs.2023.103147

[22] Laiba Husain, Teresa Finlay, Arqam Husain, Joseph Wherton, Gemma Hughes, and Trisha Greenhalgh. 2024. Developing user personas to capture intersecting dimensions of disadvantage in older patients who are marginalised: a qualitative study. *British Journal of General Practice* 74, 741 (2024), e250–e257.

[23] Bernard J Jansen, Soon-gyo Jung, and Joni Salminen. 2024. Finetuning analytics information systems for a better understanding of users: evidence of personification bias on multiple digital channels. *Information Systems Frontiers* 26, 2 (2024), 775–798.

[24] Bernard J Jansen, Joni Salminen, Soon-gyo Jung, and Kathleen Guan. 2021. Challenges of Applying Data-Driven Persona Development. *Data-Driven Personas* (2021), 139–158.

[25] Scott Jarvis. 2013. Capturing the Diversity in Lexical Diversity. *Language Learning* 63, s1 (March 2013), 87–106. doi:10.1111/j.1467-9922.2012.00739.x

[26] Soon-Gyo Jung, Joni Salminen, Kholoud Khalil Aldous, and Bernard J Jansen. 2025. PersonaCraft: Leveraging language models for data-driven persona development. *International Journal of Human-Computer Studies* 197 (2025), 103445.

[27] Ilkka Kaate, Joni Salminen, Soon-Gyo Jung, João M Santos, Essi Häyhänen, Trang Xuan, Jinan Azem, and Bernard J Jansen. 2024. Modeling the New Modalities of Personas: How Do Users' Attributes Influence Their Perceptions and Use of Interactive Personas?. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. 164–169.

[28] Messi H.J. Lee, Jacob M. Montgomery, and Calvin K. Lai. 2024. Large Language Models Portray Socially Subordinate Groups as More Homogeneous, Consistent with a Bias Observed in Humans. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. ACM, 1321–1340. doi:10.1145/3630106.3658975

[29] Antonio A Lopez-Lorca, Tim Miller, Sonja Pedell, Antonette Mendoza, Alen Keirnan, and Leon Sterling. 2014. One size doesn't fit all: diversifying" the user" using personas and emotional scenarios. In *Proceedings of the 6th International Workshop on Social Software Engineering*. 25–32.

[30] Nicola Marsden and Monika Pröbster. 2019. Personas and identity: Looking at multiple identities to inform the construction of personas. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[31] Nicola Marsden and Monika Pröbster. 2019. Personas and Identity: Looking at Multiple Identities to Inform the Construction of Personas. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–14. doi:10.1145/3290605.3300565

[32] Gonzalo Martínez, José Alberto Hernández, Javier Conde, Pedro Reviriego, and Elena Merino-Gómez. 2024. Beware of Words: Evaluating the Lexical Diversity of Conversational LLMs using ChatGPT as Case Study. *ACM Trans. Intell. Syst. Technol.* (Sept. 2024). doi:10.1145/3696459 Just Accepted.

[33] Philip M McCarthy. 2005. *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)*. Ph. D. Dissertation. The University of Memphis.

[34] Philip M McCarthy and Scott Jarvis. 2007. vocd: A theoretical and empirical evaluation. *Language Testing* 24, 4 (2007), 459–488.

[35] Philip M. McCarthy and Scott Jarvis. 2010. MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42, 2 (May 2010), 381–392. doi:10.3758/BRM.42.2.381

[36] Zdenek Meier, Kristyna Gabova, Radka Zidkova, and Peter Tavel. 2024. Personas of Older Adults in Social and Health Context. In *Intelligent Technologies for Healthcare Business Applications*. Springer, 137–171.

[37] Farooq Mubarak, Reima Suomi, and Satu-Päivi Kantola. 2020. Confirming the links between socio-economic variables and digitalization worldwide: the unsettled debate on digital divide. *Journal of Information, Communication and Ethics in Society* 18, 3 (2020), 415–430.

[38] Timothy Neate, Aikaterini Bourazeri, Abi Roper, Simone Stumpf, and Stephanie Wilson. 2019. Co-Created Personas: Engaging and Empowering Users with Diverse Needs Within the Design Process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300880

[39] Lene Nielsen. 2013. *Personas-user focused design*. Vol. 15. Springer.

[40] Lene Nielsen, Marta Larusdottir, and Lars Bo Larsen. 2021. Understanding users through three types of personas. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part II 18*. Springer, 330–348.

[41] Lene Nielsen and Kira Storgaard Hansen. 2014. Personas is applicable: a study on the use of personas in Denmark. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1665–1674.

[42] Naseela Pervez and Alexander J. Titus. 2024. Inclusivity in Large Language Models: Personality Traits and Gender Bias in Scientific Abstracts. arXiv:2406.19497 [cs.CL] https://arxiv.org/abs/2406.19497

[43] Alexander Pfaff. 2024. How to measure syntactic diversity: Patternization, methods, algorithms. *Noun phrases in early Germanic languages* (2024), 33–70.

[44] Esther Ploeger, Huiyuan Lai, Rik van Noord, and Antonio Toral. 2024. Towards Tailored Recovery of Lexical Diversity in Literary Machine Translation. arXiv:2408.17308 [cs.CL] https://arxiv.org/abs/2408.17308

[45] Mirjana Prpa, Giovanni Maria Troiano, Matthew Wood, and Yvonne Coady. 2024. Challenges and Opportunities of LLM-Based Synthetic Personae and Data in HCI. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–5.

[46] Graham Pullin and Alan Newell. 2007. Focussing on extra-ordinary users. In *Proceedings of the 4th International Conference on Universal Access in Human Computer Interaction: Coping with Diversity* (Beijing, China) *(UAHCI'07)*. Springer-Verlag, Berlin, Heidelberg, 253–262.

[47] Cynthia Putnam, Emma J. Rose, Erica J. Johnson, and Beth E. Kolko. 2009. Adapting User-Centered Design Methods to Design for Diverse Populations. *Information Technologies and International Development* 5 (2009), 51–74. https://api.semanticscholar.org/CorpusID:55242592

[48] Emily Reif, Minsuk Kahng, and Savvas Petridis. 2023. Visualizing linguistic diversity of text datasets synthesized by large language models. In *2023 IEEE Visualization and Visual Analytics (VIS)*. IEEE, 236–240.

[49] Pedro Reviriego, Javier Conde, Elena Merino-Gómez, Gonzalo Martínez, and José Alberto Hernández. 2024. Playing with words: Comparing the vocabulary and lexical diversity of ChatGPT and humans. *Machine Learning with Applications* 18 (2024), 100602.

[50] Joni Salminen, Kamal Chhirang, Soon-Gyo Jung, Saravanan Thirumuruganathan, Kathleen W. Guan, and Bernard J. Jansen. 2022. Big Data, Small Personas: How Algorithms Shape the Demographic Representation of Data-Driven User Segments. (2022). doi:10.1089/big.2021.0177 Publisher: Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.

[51] Joni Salminen, Soon-gyo Jung, Hind Almerekhi, Erik Cambria, and Bernard Jansen. 2023. How Can Natural Language Processing and Generative AI Address Grand Challenges of Quantitative User Personas?. In *International Conference on Human-Computer Interaction*. Springer, 211–231.

[52] Joni Salminen, Soon-gyo Jung, Shammur A Chowdhury, and Bernard J Jansen. 2020. Rethinking personas for fairness: Algorithmic transparency and accountability in data-driven personas. In *Artificial Intelligence in HCI: First International Conference, AI-HCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*. Springer, 82–100.

[53] Joni Salminen, Soon-Gyo Jung, Lene Nielsen, and Bernard Jansen. 2022. Creating More Personas Improves Representation of Demographically Diverse Populations: Implications Towards Interactive Persona Systems. In *Nordic Human-Computer Interaction Conference*. 1–11.

[54] Joni Salminen, Chang Liu, Wenjing Pian, Jianxing Chi, Essi Häyhänen, and Bernard J Jansen. 2024. Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3613904.3642036

[55] Joni Salminen, Sercan Şengün, João M Santos, Soon-gyo Jung, Lene Nielsen, and Bernard Jansen. 2023. The Choice of a Persona: An Analysis of Why Stakeholders Choose a Given Persona for a Design Task. In *International Conference on Human-Computer Interaction*. Springer, 288–310.

[56] Andreas Schuller, Doris Janssen, Julian Blumenröther, Theresa Maria Probst, Michael Schmidt, and Chandan Kumar. 2024. Generating personas using LLMs and assessing their viability. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[57] Orit Shechtman. 2013. *The Coefficient of Variation as an Index of Measurement Reliability.* Springer Berlin Heidelberg, Berlin, Heidelberg, 39–49. doi:10.1007/978-3-642-37131-8_4

[58] Lucas Shen. 2022. LexicalRichness: A small module to compute textual lexical richness. doi:10.5281/zenodo.6607007

[59] Joongi Shin, Michael A Hedderich, Bartłomiej Jakub Rey, Andrés Lucero, and Antti Oulasvirta. 2024. Understanding Human-AI Workflows for Generating Personas. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 757–781.

[60] Phillip Douglas Stevenson and Christopher Andrew Mattson. 2019. The Personification of Big Data. *Proceedings of the Design Society: International Conference on Engineering Design* 1, 1 (July 2019), 4019–4028. doi:10.1017/dsi.2019.409

[61] Guy Tevet and Jonathan Berant. 2021. Evaluating the Evaluation of Diversity in Natural Language Generation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Association for Computational Linguistics, Online, 326–346. doi:10.18653/v1/2021.eacl-main.25

[62] Phil Turner and Susan Turner. 2011. Is stereotyping inevitable when designing with personas? *Design studies* 32, 1 (2011), 30–44.

[63] Fiona J. Tweedie and R. Harald Baayen. 1998. How Variable May a Constant Be? Measures of Lexical Richness in Perspective. *Computers and the Humanities* 32, 5 (1998), 323–352. https://www.jstor.org/stable/30200474 Publisher: Springer.

[64] Ji Seung Yang, Carly Rosvold, and Nan Bernstein Ratner. 2022. Measurement of lexical diversity in children's spoken language: Computational and conceptual considerations. *Frontiers in psychology* 13 (2022), 905789.

[65] Xulang Zhang, Rui Mao, and Erik Cambria. 2024. Multilingual emotion recognition: Discovering the variations of lexical semantics between languages. In *2024 International Joint Conference on Neural Networks (IJCNN)*.

[66] Wanwan Zheng and Mingzhe Jin. 2024. Evaluate Lexical Richness Measures Using Coefficient of Variation and Relative Value. (2024). http://www.cicling.org/2018/intranet/pre-print/papers/paper_1.pdf Working paper. Accessed: 29-Dec-2024.