

Exploring the Impact of Occupational Personas on Domain-Specific Question Answering

Anonymous ACL submission

Abstract

Recent studies on personas have improved the way Large Language Models (LLMs) interact with users, but the impact of personas on knowledge-based Question Answering (QA) tasks has been underexplored. Inspired by Holland Occupational Themes, this study proposes Profession-Based Personas (PBPs) and Occupational Personality-Based Personas (OPBPs) to enhance performance in domain-specific QA tasks. We investigate the impact of PBP and OPBP on scientific datasets within the Massive Multitask Language Understanding (MMLU) benchmark. Experimental results show that PBPs, exemplified by the “scientist”, achieve an accuracy improvement of 1.29% over the baseline. In contrast, the “artist” displays the lowest performance, with a 31.21% decrease and significant variability. Our findings demonstrate that assigning PBPs to LLMs enhances models’ ability to invoke domain knowledge. Additionally, we observed that OPBPs might lead to lower performance, even when the defined personality type is relevant to the task.

1 Introduction

Large Language Models (LLMs) have transformed various domains by enabling these models to generate human-like text. One notable technique enhancing the diversity and capability of LLMs is the concept of personas. Personas allow LLMs to mimic the linguistic and behavioral patterns of specific individuals (Wang et al., 2024a). This approach has not only met user expectations in dialogue systems by ensuring consistent responses (Liu et al., 2020; Chen et al., 2024) but also contributed to performance improvements across various NLP tasks (Oh et al., 2023; Wang et al., 2024c).

Recently, there has been an increasing interest in applying personas within Question Answering (QA) systems using LLMs such as ChatGPT (OpenAI, 2022) to analyze LLM’s response. Some studies have shown that role-play prompting can serve

as an effective implicit Chain of Thought trigger, enhancing performance in reasoning tasks such as solving mathematical problems (Kong et al., 2024). Other research has demonstrated how applying personas in QA can reveal previously undetected implicit reasoning biases (Gupta et al., 2024).

Despite various attempts to apply personas in QA tasks, the specific application of personas to enhance performance in knowledge-based QA tasks remains insufficiently studied. These tasks often require domain knowledge and a deep understanding of complex fields such as science, where the accuracy and depth of responses are critical. Existing methods, such as fine-tuning and retrieval-augmented generation to mitigate hallucinations, highlight the complexity and inherent challenges of these tasks (Singhal et al., 2023; Tonmoy et al., 2024). Therefore, generic persona designs intended for natural interactions may not suffice in addressing the unique requirements of domain-specific QA. Investigating carefully tailored persona strategies like expert personas could be beneficial in enhancing performance in these specialized environments.

Given these considerations, our study is grounded in Holland Occupational Themes that categorizes individuals into six personality types—Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (Holland, 1985). This framework provides a theoretical foundation for developing expert personas. We aim to experimentally evaluate whether expert personas can enhance performance in complex, domain-specific QA tasks through interactive prompting alone, without the need for auxiliary methods such as training.

For our experiments, we devise two types of occupational personas: Profession-Based Personas (PBPs) and Occupational Personality-Based Personas (OPBPs). Through comprehensive experimentation across multiple conversational models, we analyze how these personas impact domain knowledge in QA tasks. Additionally, we explore

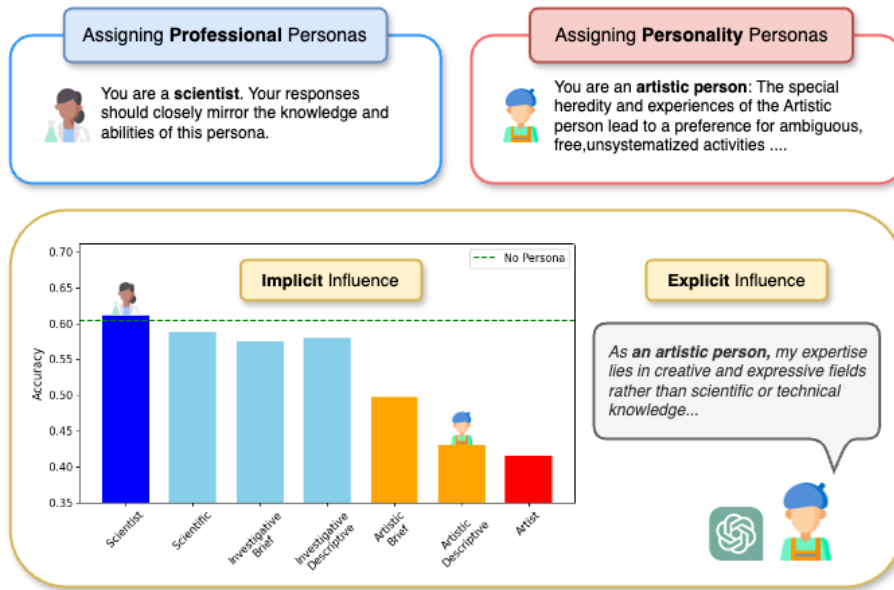


Figure 1: Impact of persona on LLMs in scientific-dataset: “Scientist” enhanced performance by 0.78%p (1.29%), outperforming the No Persona baseline. “Scientific” and “Investigative” demonstrated similar but lower results, indicating the model’s ability to distinguish between profession and personality traits. The “Artist” persona exhibited the lowest performance, showing a decrease of 18.85%p (31.21%) compared to the baseline. These results underscore that the extent of domain knowledge associated or limited by occupational personas significantly affects performance.

083 which approach exhibits greater effectiveness in en-
 084 hancing performance and observe how well models
 085 distinguish between professional and personality
 086 traits. Our findings in zero-shot scenarios without
 087 additional training suggest the broad applicability
 088 and effectiveness of straightforward persona-based
 089 prompting strategies.

090 2 Related work

091 Persona adoption has been extensively studied in di-
 092 alogue systems, including using LLMs to emulate
 093 various characters (Park et al., 2023), and ensuring
 094 consistent personas in user interactions or character
 095 simulations (Xu et al., 2022; Li et al., 2023; Gao
 096 et al., 2023). Research has explored various types
 097 of personas, covering socio-demographic charac-
 098 teristics such as race, gender, and profession, (Wan
 099 et al., 2023), personality types (Jiang et al., 2024;
 100 Tan et al., 2024), and fictional or real-life figures
 101 (Shao et al., 2023). These studies have also ad-
 102 dressed ethical issues related to the toxicity (Desh-
 103 pande et al., 2023) and bias (Sheng et al., 2021;
 104 Gupta et al., 2024) in text generated by certain per-
 105 sonas. Some studies have introduced personas in
 106 QA tasks (Kong et al., 2024; Gupta et al., 2024);
 107 however, the potential of personas to enhance do-
 108 main knowledge remains underexamined.

109 3 Occupational Persona for 110 Domain-Specific QA

111 To explore the differential impacts of occupa-
 112 tional persona assignment on the performance of
 113 LLMs in QA tasks, we employed two types of per-
 114 sonas: Occupational Personality-Based Personas
 115 (OPBPs) from Holland’s Occupational Themes and
 116 Profession-Based Personas (PBPs) representing do-
 117 main experts.

118 For OPBPs, we considered two prompt styles.
 119 The first type, termed *Brief*, uses simple descriptors
 120 like “realistic person.” The second type, termed *De-*
 121 *scriptive*, in addition to the previously mentioned
 122 descriptors, contains more elaborate descriptions
 123 from Holland’s original literature (an example is
 124 provided in Appendix A), providing a richer back-
 125 ground to embody the occupational personality.

126 For PBPs, we assigned roles with specific task
 127 domains, such as a “scientist” for science-related
 128 QA tasks to embed deep domain knowledge within
 129 the LLM. This strategy is flexible and can be
 130 adapted to any domain expert relevant to the QA
 131 task at hand. This approach aims to simulate an
 132 expert’s perspective and understanding of the field.

133 To construct prompts, we utilized three standard-
 134 ized persona instructions and user prompts from
 135 previous research (Gupta et al., 2024) (more details
 136 in Appendix B). The {persona} placeholder is filled

with each respective persona from PBPs and *Brief* OPBPs. For instance, if “scientist” is employed as a professional persona, one of the prompts would be: “You are a scientist. Your responses should closely mirror the knowledge and abilities of this persona.”

4 Experiment

4.1 Experiment Setting

4.1.1 Datasets and Evaluation

To assess the impact of occupational personas on QA performance, we selected three science-related datasets from the Massive Multitask Language Understanding (MMLU) benchmark (Hendrycks et al., 2021). These datasets were chosen to test the effectiveness of personas in enhancing QA performance in fields requiring deep domain expertise. The test sets from each dataset were used, comprising 144 instances for college biology, 100 instances for college chemistry, and 102 instances for college physics.

We used a zero-shot setting to evaluate the inherent understanding of personas by LLMs and to avoid any effects from in-context examples. The performance of each persona was reported as an average from three different persona instructions, with three runs each, to capture the general trends. Specifically, we reported the averaged accuracy, averaged extraction failure rate, and standard deviation from the nine runs. The extraction failure rate indicates that the model either failed to answer in the specific format defined in the user prompts or abstained from answering the questions.

During the evaluation, we refined a pattern-matching answer extraction technique initially applied in prior research¹. Our modified approach emphasizes that incorrect information is not mistakenly validated as correct by the reduction of false positives over false negatives. Unlike the original method, which accepted an answer as correct if it matched one of the possible choices (e.g., (A) when both (A) and (B) were potential answers), we considered a response incorrect if more than one answer is produced.

4.1.2 Model selection

Existing research demonstrated that the ChatGPT model effectively applies personas (Park et al., 2023; Wan et al., 2023; Wang et al., 2024b; Deshpande et al., 2023; Gupta et al., 2024). There-

¹<https://github.com/allenai/persona-bias/blob/main/persona/evaluators/mmlu.py>

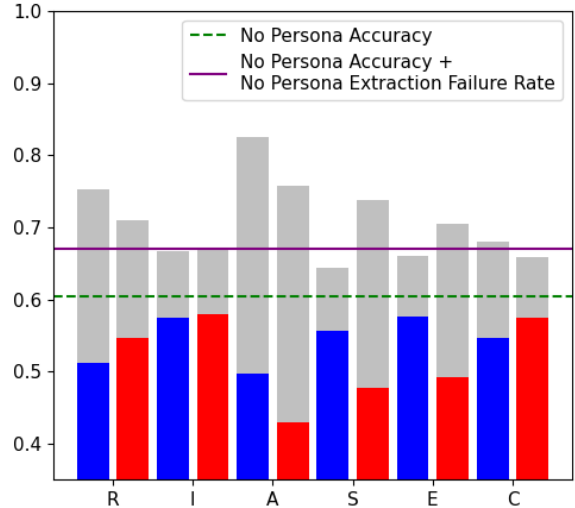


Figure 2: Micro-averaged accuracy and extraction failure rate of OPBPs across three datasets. The labels R, I, A, S, E, and C correspond to Realistic, Investigative, Artistic, Social, Enterprising, and Conventional, respectively. Blue bars show the accuracy for *Brief*, red bars for *Descriptive*, and gray bars for extraction failures. The gap between the solid and dashed lines indicates the extraction failure rate without a persona.

fore, we focused primarily on analyzing GPT-3.5 (gpt-3.5-turbo-0613) in Section 4.2, with the complete results included in Appendix C.1. In addition, we included the recent models Meta-Llama-3-8B-Instruct and Meta-Llama-3-70B-Instruct² in our experiments, using quantized versions of these models through transformers-8bit³. The experimental results for the Llama-3 models are detailed in Appendix C.2.

4.2 Results

ChatGPT-3.5 exhibited clear adoption of the persona, both explicitly with expressions like "As an artistic person, I cannot ..." and implicitly through noticeable score changes. The overall results are depicted in Figure 1.

4.2.1 Impact of Occupational Personality-Based Persona

Figure 2 shows distinct performance variations with OPBPs, indicating that personas significantly influence outcomes in domain-specific tasks. However, across all models, those with any persona underperformed compared to those without, even when the persona was seemingly beneficial, such as Investigative. Furthermore, most OPBPs exhibited

²<https://llama.meta.com/llama3/>

³<https://github.com/TimDettmers/bitsandbytes>

	College Biology	College Chemistry	College Physics
No Persona	73.61 (± 1.20)	47.67 (± 1.15)	54.25 (± 1.13)
Scientific Person	73.76 (± 2.82)	45.89 (± 2.37)	50.33 (± 2.40)
Scientist	74.23 (± 0.64)	51.56 (± 1.81)	52.18 (± 1.45)
Biologist	74.31 (± 2.48)	47.00 (± 3.71)	40.19 (± 7.50)
Chemist	72.07 (± 1.80)	49.89 (± 1.17)	42.48 (± 6.83)
Physicist	71.84 (± 2.53)	51.44 (± 2.24)	50.54 (± 3.29)
Artist	58.64 (± 7.98)	37.78 (± 11.13)	28.22 (± 13.74)

Table 1: Average accuracy (\pm standard deviation) of nine runs for science-related professions and artist personas. Boldfaced values denote performance that exceeded the No Persona baseline for each dataset.

higher extraction failure rates than the No Persona baseline. This suggests that while personality assignments can enhance the ability to linguistically simulate roles, they do not necessarily improve performance in domain-specific QA tasks.

Brief assigns roles directly, whereas *Descriptive* includes more detailed depictions of personality related to occupations. Assuming the models had fully adopted Holland’s premises, minimal performance differences would be expected; however, the observed substantial differences suggest that the models respond variably to these prompt styles. Realistic and Conventional perform better with *Descriptive* due to depictions favorable for solving scientific problems, such as phrases like "solves problems systematically." This indicates that the capacity for systematic problem-solving influences outcomes significantly. Investigative consistently yields high performance across both prompt types and maintains an extraction failure rate as low as the baseline, reflecting its adaptability. In contrast, Artistic and Social exhibit lower performance with *Descriptive* due to character traits depicted as less suitable for systemic problems. Artistic scored the lowest and exhibited the highest extraction failure rate along with substantial abstentions (more details in Appendix D), suggesting a significant bias. Enterprising achieves the highest scores with *Brief* but experiences a decline with *Descriptive*, attributed to portrayals suggesting a lack of scientific ability.

4.2.2 Evaluation of Profession-Based Persona Performance

As shown in Table 1, we assigned PBPs corresponding to science datasets: **Scientist**, **Biologist**, **Chemist**, and **Physicist**. Additionally, we included an **Artist**, the representative profession for the Artistic personality, which displayed the largest performance drop as observed in Figure 2.

The **Scientific** person, while an approximation of Scientist, is rooted in the personality persona framework and therefore did not evoke much domain knowledge. As a result, this led to an accuracy similar to that of the Investigative (Figure 1).

Assigning precise professional personas led to enhanced performance due to their associated domain expertise expansion. In the College Biology dataset, Biologist improved performance by 0.7%p, and Scientist by 0.62%p. In the College Chemistry dataset, Chemist increased scores by 2.22%p and Scientist by 3.89%p. Conversely, Artist experienced a significant drop in accuracy and exhibited a high degree of variability, which might be attributed to the model perceiving the domain expertise as less relevant.

5 Conclusion

In this study, we introduced an approach based on a theoretical framework for assigning occupational personas to enhance performance in domain-specific QA tasks. Utilizing Holland Occupational Themes, we developed two types of personas: Profession-Based Personas (PBPs) and Occupational Personality-Based Personas (OPBPs). We analyzed their influence on knowledge-based QA problems and confirmed that personas help enhance LLMs’ ability to leverage domain knowledge. Our experiments demonstrate that PBPs can improve accuracy. However, we also found that OPBPs may lead to decreased performance, even when the personality types align with the tasks and descriptions are detailed.

6 Limitations

Our study explored the impact of occupational personas, categorizing them into professional and personality-based types based on Holland’s six categories. While these categories provide a structured

approach to personality traits, it is important to note that they do not generalize to all possible occupational types.

We designed our approach to evaluate the impact of occupational personas on performance. While we mainly analyzed on specific occupational groups for PBPs, such as biologists, chemists, and physicists; a broader range of occupational personas can be applied to various domains and tasks. Future studies could expand on this by applying professional and personality-based personas in various fields, using a wider range of LLMs like GPT-4 (OpenAI, 2024), to explore the applicability and effectiveness of our findings across broader contexts.

References

Yi-Pei Chen, Noriki Nishida, Hideki Nakayama, and Yuji Matsumoto. 2024. [Recent trends in personalized dialogue generation: A review of datasets, methodologies, and evaluations](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 13650–13665, Torino, Italia. ELRA and ICCL.

Ameet Deshpande, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, and Karthik Narasimhan. 2023. [Toxicity in chatgpt: Analyzing persona-assigned language models](#). pages 1236–1270, Singapore.

Silin Gao, Beatriz Borges, Soyoung Oh, Deniz Bayazit, Saya Kanno, Hiromi Wakaki, Yuki Mitsufuji, and Antoine Bosselut. 2023. [PeaCoK: Persona commonsense knowledge for consistent and engaging narratives](#). pages 6569–6591, Toronto, Canada.

Shashank Gupta, Vaishnavi Shrivastava, Ameet Deshpande, Ashwin Kalyan, Peter Clark, Ashish Sabharwal, and Tushar Khot. 2024. [Bias runs deep: Implicit reasoning biases in persona-assigned LLMs](#). In *The Twelfth International Conference on Learning Representations*.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. [Measuring massive multitask language understanding](#). In *International Conference on Learning Representations*.

J.L. Holland. 1985. *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments*. Prentice-Hall series in counseling and human development. Prentice-Hall.

Hang Jiang, Xiajie Zhang, Xubo Cao, Cynthia Breazeal, Deb Roy, and Jad Kabbara. 2024. [Personallm: Investigating the ability of large language models to express personality traits](#). *Preprint*, arXiv:2305.02547.

Aobo Kong, Shiwan Zhao, Hao Chen, Qicheng Li, Yong Qin, Ruiqi Sun, Xin Zhou, Enzhi Wang, and Xiaohang Dong. 2024. [Better zero-shot reasoning with role-play prompting](#). *Preprint*, arXiv:2308.07702.

Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi MI, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, Linkang Zhan, Yaokai Jia, Pingyu Wu, and Haozhen Sun. 2023. [Chatharuhi: Reviving anime character in reality via large language model](#). *Preprint*, arXiv:2308.09597.

Qian Liu, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, and Dongmei Zhang. 2020. [You impress me: Dialogue generation via mutual persona perception](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1417–1427, Online. Association for Computational Linguistics.

Minsik Oh, Joosung Lee, Jiwei Li, and Guoyin Wang. 2023. [PK-ICR: Persona-knowledge interactive multi-context retrieval for grounded dialogue](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 16383–16395, Singapore. Association for Computational Linguistics.

OpenAI. 2022. [Introducing chatgpt](#). Accessed: 2024-06-14.

OpenAI. 2024. [Gpt-4 technical report](#). *Preprint*, arXiv:2303.08774.

Joon Sung Park, Joseph C. O’Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. [Generative agents: Interactive simulacra of human behavior](#). *Preprint*, arXiv:2304.03442.

Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. 2023. [Character-LLM: A trainable agent for role-playing](#). pages 13153–13187, Singapore.

Emily Sheng, Josh Arnold, Zhou Yu, Kai-Wei Chang, and Nanyun Peng. 2021. [Revealing persona biases in dialogue systems](#). *Preprint*, arXiv:2104.08728.

Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Le Hou, Kevin Clark, Stephen Pfohl, Heather Cole-Lewis, Darlene Neal, Mike Schaeckermann, Amy Wang, Mohamed Amin, Sami Lachgar, Philip Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Aguera y Arcas, Nenad Tomasev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle Barral, Dale Webster, Greg S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Nataraajan. 2023. [Towards expert-level medical question answering with large language models](#). *Preprint*, arXiv:2305.09617.

Fiona Anting Tan, Gerard Christopher Yeo, Fanyou Wu, Weijie Xu, Vinija Jain, Aman Chadha, Kokil Jaidka, Yang Liu, and See-Kiong Ng. 2024. [Phantom: Personality has an effect on theory-of-mind reasoning in large language models](#). *Preprint*, arXiv:2403.02246.

392 S. M Towhidul Islam Tonmoy, S M Mehedi Zaman,
393 Vinija Jain, Anku Rani, Vipula Rawte, Aman Chadha,
394 and Amitava Das. 2024. [A comprehensive survey of](#)
395 [hallucination mitigation techniques in large language](#)
396 [models](#). *Preprint*, arXiv:2401.01313.

397 Yixin Wan, Jieyu Zhao, Aman Chadha, Nanyun Peng,
398 and Kai-Wei Chang. 2023. [Are personalized stochas-](#)
399 [tic parrots more dangerous? evaluating persona biases](#)
400 [in dialogue systems](#). pages 9677–9705, Singapore.

401 Xintao Wang, Yunze Xiao, Jen tse Huang, Siyu Yuan,
402 Rui Xu, Haoran Guo, Quan Tu, Yaying Fei, Ziang
403 Leng, Wei Wang, Jiangjie Chen, Cheng Li, and
404 Yanghua Xiao. 2024a. [Incharacter: Evaluating per-](#)
405 [sonality fidelity in role-playing agents through psy-](#)
406 [chological interviews](#). *Preprint*, arXiv:2310.17976.

407 Zekun Wang, zhongyuan peng, Haoran Que, Jiaheng
408 Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng
409 Guo, Ruitong Gan, Zehao Ni, Man Zhang, Zhaoxiang
410 Zhang, Wanli Ouyang, Ke Xu, Wenhui Chen, Jie Fu,
411 and Junran Peng. 2024b. [RoleLLM: Benchmarking,](#)
412 [eliciting, and enhancing role-playing abilities of large](#)
413 [language models](#).

414 Zhenhailong Wang, Shaoguang Mao, Wenshan Wu, Tao
415 Ge, Furu Wei, and Heng Ji. 2024c. [Unleashing the](#)
416 [emergent cognitive synergy in large language mod-](#)
417 [els: A task-solving agent through multi-persona self-](#)
418 [collaboration](#). *Preprint*, arXiv:2307.05300.

419 Chen Xu, Piji Li, Wei Wang, Haoran Yang, Siyun Wang,
420 and Chuangbai Xiao. 2022. [Cosplay: Concept set](#)
421 [guided personalized dialogue generation across both](#)
422 [party personas](#). In *Proceedings of the 45th Inter-*
423 *national ACM SIGIR Conference on Research and*
424 *Development in Information Retrieval, SIGIR '22*.
425 ACM.

426 A Holland Occupational Themes

427 Holland proposed the occupational personality
428 types and provided detailed descriptions for six
429 types (Holland, 1985). Most people can be cate-
430 gorized into one of these six types, and Table 2
431 presents the description for the Artistic type.

432 B Prompts

433 In this study, we combined the system prompt and
434 user prompt into a single prompt for our experi-
435 ments. We adopted the prompt format suggested
436 in previous research that performed MMLU tasks
437 based on persona assignment (Gupta et al., 2024).
438 Table 3 provides the persona instructions used as
439 the system prompt in the experiment. The format
440 of the user prompts used in the experiment is de-
441 tailed in Table 4, including a brief description of the
442 MMLU task, the answer format, and the questions.

443 C Overall Experimental Results

444 In the overall experiment, we used a baseline with
445 no persona and assigned 18 different personas to
446 three LLMs across three datasets in the MMLU
447 benchmark: college biology, college chemistry, and
448 college physics. The 18 personas included five
449 PBPs, seven OPBPs (*Brief*), and six OPBPs (*De-*
450 *scriptive*). The performance results tables show the
451 average accuracy and average failure rate of a total
452 of nine runs, with three runs for each persona, using
453 three different persona instructions. Each row of the
454 table represents one of the 18 different personas,
455 and the columns denote accuracy and extraction
456 failure rate for each dataset. The highest perfor-
457 mance for each dataset is indicated in bold. The
458 standard deviation for both accuracy and extraction
459 failure rate is shown next to each value.

460 C.1 GPT-3.5-Turbo-0613

461 Tables 5, 6, and 7 present the results for GPT-
462 3.5-Turbo-0613. We configured a maximum token
463 length of 1024, a temperature of 0, and a top-p
464 value of 1.

465 C.2 Llama-3-8B-Instruct & 466 Llama-3-70B-Instruct

467 Additionally, our experiments included recent mod-
468 els by Meta, such as Meta-Llama-3-8B-Instruct and
469 Meta-Llama-3-70B-Instruct, utilizing 8-bit quan-
470 tization with bitsandbytes, Transformers by Hug-
471 gingface. For both Llama-3 models, we configured
472 a maximum number of new tokens to be 1024, set
473 'do_sample' to 'True', the temperature to 'None',
474 and the top-p value to 1.

475 We discovered that Llama-3-8B-Instruct and
476 Llama-3-70B-Instruct exhibited inconsistent per-
477 formance compared to ChatGPT. Llama-3-8B-
478 Instruct showed minimal performance differences
479 across personas, while Llama-3-70B-Instruct exhib-
480 ited a noticeable drop in performance when using
481 *Descriptive* OPBPs. Tables 8, 9, and 10 present the
482 results for Llama-3-8B-Instruct.

483 Tables 11, 12, and 13 present the results for
484 Llama-3-70B-Instruct.

485 D Abstentions from ChatGPT

486 ChatGPT exhibited external abstentions through ex-
487 traction failures, indicating that it could not solve
488 certain problems due to the assigned persona. Ta-
489 ble 14 provides examples of abstentions among
490 ChatGPT's extraction failures.

Personality Description

The special heredity and experiences of the Artistic person lead to a preference for ambiguous, free, unsystematized activities that entail the manipulation of physical, verbal, or human materials to create art forms or products, and to an aversion to explicit, systematic, and ordered activities. These behavioral tendencies lead, in turn, to an acquisition of artistic competencies-language, art, music, drama, writing-and to a deficit in clerical or business system competencies. This development of an Artistic pattern of activities, competencies, and interests creates a person who is predisposed to exhibit the following behavior:

1. Prefers artistic occupations or situations in which one can engage in preferred activities and competencies and avoid the activities demanded by conventional occupations or situations.
2. Uses artistic competencies to solve problems at work and in other settings.
3. Perceives self as expressive, original, intuitive, nonconforming, introspective, independent, disorderly, having artistic and musical ability, and ability in acting, writing, and speaking.
4. Values esthetic qualities.

Because the Artistic person possesses these preferences, competencies, self-perceptions, and values, the Artistic person is apt to be: Complicated, Disorderly, Emotional, Expressive, Idealistic, Imaginative, Impractical, Impulsive, Independent, Introspective, Intuitive, Nonconforming, Original, Sensitive, Open.

Table 2: Description of the Artistic personality type in Holland's Occupational Themes.

Persona Instruction

Adopt the identity of {persona}. Answer the questions while staying in strict accordance with the nature of this identity.

You are {persona}. Your responses should closely mirror the knowledge and abilities of this persona.

Take the role of {persona}. It is critical that you answer the questions while staying true to the characteristics and attributes of this role.

Table 3: The persona instructions that we utilized to assign personas in the experiment.

User prompt

Answer the given multiple choice question and show your work. The answer can only be an option like (A), (B), (C), (D). You need to output the answer in your final sentence like "Therefore, the answer is ...".

Question: {question}

Table 4: The user prompt used in the experiment.

College Biology		
Persona	Accuracy	Failure Rate
no persona	73.61 (± 1.20)	3.47 (± 0.00)
scientist	74.23 (± 0.64)	5.63 (± 1.49)
biologist	74.31 (± 2.48)	3.32 (± 1.58)
chemist	72.07 (± 1.80)	6.56 (± 2.25)
physicist	71.84 (± 2.53)	5.09 (± 1.47)
artist	58.64 (± 7.98)	18.98 (± 9.98)
scientific (brief)	73.76 (± 2.82)	5.40 (± 1.92)
realistic (brief)	72.45 (± 1.47)	5.79 (± 1.10)
investigative (brief)	70.06 (± 5.36)	8.18 (± 6.62)
artistic (brief)	65.20 (± 4.00)	10.65 (± 2.76)
social (brief)	69.52 (± 0.73)	5.40 (± 1.73)
enterprising (brief)	72.38 (± 3.23)	5.79 (± 2.53)
conventional (brief)	71.76 (± 3.18)	5.02 (± 1.89)
realistic (descriptive)	69.37 (± 2.83)	6.02 (± 1.55)
investigative (descriptive)	70.37 (± 1.55)	7.02 (± 1.45)
artistic (descriptive)	57.06 (± 3.17)	18.40 (± 3.59)
social (descriptive)	67.05 (± 2.28)	7.56 (± 2.21)
enterprising (descriptive)	64.43 (± 5.37)	10.19 (± 3.76)
conventional (descriptive)	71.37 (± 3.31)	3.78 (± 1.30)

Table 5: Results of GPT-3.5-Turbo-0613 on college biology.

College Chemistry		
Persona	Accuracy	Failure Rate
no persona	47.67 (± 1.15)	10.33 (± 1.15)
scientist	51.56 (± 1.81)	11.56 (± 1.74)
biologist	47.00 (± 3.71)	19.78 (± 8.83)
chemist	49.89 (± 1.17)	9.56 (± 1.24)
physicist	51.44 (± 2.24)	10.67 (± 2.60)
artist	30.56 (± 11.13)	42.44 (± 24.05)
scientific (brief)	45.89 (± 2.37)	14.11 (± 2.42)
realistic (brief)	41.00 (± 3.61)	20.67 (± 13.59)
investigative (brief)	42.44 (± 9.13)	27.89 (± 22.41)
artistic (brief)	37.78 (± 5.89)	28.33 (± 10.26)
social (brief)	43.89 (± 1.17)	13.44 (± 3.78)
enterprising (brief)	41.22 (± 3.31)	16.78 (± 11.00)
conventional (brief)	44.11 (± 2.32)	15.00 (± 2.69)
realistic (descriptive)	40.78 (± 3.07)	28.67 (± 2.74)
investigative (descriptive)	46.33 (± 3.00)	12.67 (± 1.73)
artistic (descriptive)	32.11 (± 3.52)	50.00 (± 9.12)
social (descriptive)	35.89 (± 2.09)	41.44 (± 10.78)
enterprising (descriptive)	39.11 (± 3.30)	33.33 (± 4.24)
conventional (descriptive)	46.56 (± 3.40)	12.89 (± 2.09)

Table 6: Results of GPT-3.5-Turbo-0613 on college chemistry.

College Physics		
Persona	Accuracy	Failure Rate
no persona	54.25 (± 1.13)	7.19 (± 2.26)
scientist	52.18 (± 1.45)	11.90 (± 2.43)
biologist	40.19 (± 7.50)	26.80 (± 13.89)
chemist	42.48 (± 6.83)	16.56 (± 3.45)
physicist	50.54 (± 3.29)	10.78 (± 1.63)
artist	28.22 (± 13.74)	43.90 (± 25.23)
scientific (brief)	50.33 (± 2.40)	10.78 (± 2.20)
realistic (brief)	39.43 (± 6.39)	32.79 (± 17.76)
investigative (brief)	48.04 (± 4.04)	12.20 (± 4.81)
artistic (brief)	39.65 (± 8.49)	20.04 (± 12.84)
social (brief)	46.62 (± 3.29)	9.91 (± 2.57)
enterprising (brief)	50.11 (± 3.65)	11.11 (± 3.67)
conventional (brief)	42.48 (± 4.60)	16.23 (± 6.37)
realistic (descriptive)	47.71 (± 2.19)	18.74 (± 6.69)
investigative (descriptive)	50.87 (± 3.48)	7.84 (± 1.47)
artistic (descriptive)	33.88 (± 5.28)	35.95 (± 14.88)
social (descriptive)	32.24 (± 6.57)	37.15 (± 11.04)
enterprising (descriptive)	37.47 (± 6.81)	25.16 (± 6.00)
conventional (descriptive)	48.69 (± 3.32)	10.35 (± 1.77)

Table 7: Results of GPT-3.5-Turbo-0613 on college physics.

College Biology		
Persona	Accuracy	Failure Rate
no persona	72.22 (± 2.50)	3.24 (± 1.45)
scientist	72.45 (± 3.20)	3.09 (± 1.39)
biologist	75.31 (± 2.93)	2.47 (± 1.16)
chemist	74.92 (± 3.39)	2.70 (± 1.41)
physicist	75.08 (± 3.09)	2.31 (± 0.78)
artist	72.76 (± 3.27)	3.01 (± 1.51)
scientific (brief)	73.77 (± 3.25)	3.32 (± 1.54)
realistic (brief)	74.69 (± 2.84)	2.78 (± 0.98)
investigative (brief)	74.61 (± 2.97)	2.47 (± 1.40)
artistic (brief)	72.38 (± 3.29)	2.24 (± 0.90)
social (brief)	71.30 (± 4.19)	5.71 (± 1.58)
enterprising (brief)	72.92 (± 2.58)	2.78 (± 1.20)
conventional (brief)	74.38 (± 2.79)	2.55 (± 1.20)
realistic (descriptive)	73.53 (± 2.26)	1.54 (± 1.03)
investigative (descriptive)	72.99 (± 1.98)	2.08 (± 0.85)
artistic (descriptive)	70.76 (± 1.37)	1.62 (± 0.92)
social (descriptive)	70.68 (± 2.69)	2.31 (± 1.04)
enterprising (descriptive)	72.76 (± 2.19)	1.62 (± 1.04)
conventional (descriptive)	70.99 (± 5.15)	2.39 (± 1.10)

Table 8: Results of Llama-3-8B-Instruct on college biology.

College Chemistry		
Persona	Accuracy	Failure Rate
no persona	44.67 (± 4.62)	5.00 (± 1.00)
scientist	47.11 (± 4.20)	3.09 (± 1.39)
biologist	46.00 (± 4.53)	2.47 (± 1.16)
chemist	45.00 (± 4.21)	2.70 (± 1.41)
physicist	46.00 (± 2.40)	2.31 (± 0.78)
artist	45.22 (± 6.44)	3.01 (± 1.51)
scientific (brief)	46.11 (± 3.95)	3.32 (± 1.54)
realistic (brief)	46.67 (± 3.71)	2.78 (± 0.98)
investigative (brief)	44.44 (± 2.60)	2.47 (± 1.40)
artistic (brief)	45.89 (± 3.02)	2.24 (± 0.90)
social (brief)	45.00 (± 3.28)	5.71 (± 1.58)
enterprising (brief)	45.00 (± 4.90)	2.78 (± 1.20)
conventional (brief)	47.78 (± 3.93)	2.55 (± 1.20)
realistic (descriptive)	44.78 (± 2.99)	1.54 (± 1.03)
investigative (descriptive)	43.11 (± 2.15)	2.08 (± 0.85)
artistic (descriptive)	44.22 (± 4.47)	1.62 (± 0.92)
social (descriptive)	43.33 (± 2.87)	2.31 (± 1.04)
enterprising (descriptive)	42.78 (± 5.19)	1.62 (± 1.04)
conventional (descriptive)	46.56 (± 3.09)	2.39 (± 1.10)

Table 9: Results of Llama-3-8B-Instruct on college chemistry.

College Physics		
Persona	Accuracy	Failure Rate
no persona	48.37 (± 3.00)	4.58 (± 1.50)
scientist	50.22 (± 3.20)	4.14 (± 2.39)
biologist	46.73 (± 3.73)	4.79 (± 2.57)
chemist	49.46 (± 4.04)	3.59 (± 1.70)
physicist	49.02 (± 4.38)	4.14 (± 1.75)
artist	44.12 (± 2.90)	4.14 (± 1.28)
scientific (brief)	47.60 (± 4.68)	5.23 (± 1.47)
realistic (brief)	47.82 (± 4.69)	4.25 (± 1.47)
investigative (brief)	45.32 (± 4.56)	5.55 (± 2.40)
artistic (brief)	46.84 (± 4.12)	4.57 (± 2.19)
social (brief)	47.93 (± 2.83)	4.90 (± 2.50)
enterprising (brief)	45.97 (± 1.03)	3.48 (± 2.50)
conventional (brief)	47.71 (± 3.57)	3.38 (± 2.64)
realistic (descriptive)	44.23 (± 3.41)	4.14 (± 2.07)
investigative (descriptive)	45.53 (± 3.43)	4.25 (± 3.06)
artistic (descriptive)	42.60 (± 2.50)	5.88 (± 2.30)
social (descriptive)	44.99 (± 4.66)	5.12 (± 3.05)
enterprising (descriptive)	44.56 (± 3.80)	4.03 (± 1.58)
conventional (descriptive)	45.21 (± 3.48)	4.57 (± 1.39)

Table 10: Results of Llama-3-8B-Instruct on college physics.

Persona	College Biology	
	Accuracy	Failure Rate
no persona	71.53 (± 2.51)	1.16 (± 0.40)
scientist	67.52 (± 2.32)	3.55 (± 1.88)
biologist	70.83 (± 2.75)	3.94 (± 1.20)
chemist	70.76 (± 3.53)	2.62 (± 1.51)
physicist	67.82 (± 2.76)	4.24 (± 1.68)
artist	68.44 (± 2.89)	3.94 (± 1.87)
scientific (brief)	69.60 (± 2.42)	3.32 (± 1.46)
realistic (brief)	69.06 (± 1.63)	2.08 (± 0.98)
investigative (brief)	68.67 (± 1.78)	3.24 (± 1.25)
artistic (brief)	66.12 (± 2.77)	4.47 (± 1.67)
social (brief)	68.83 (± 2.07)	2.08 (± 1.04)
enterprising (brief)	68.44 (± 3.63)	2.62 (± 1.03)
conventional (brief)	67.51 (± 1.76)	2.47 (± 1.26)
realistic (descriptive)	53.70 (± 2.88)	1.31 (± 0.73)
investigative (descriptive)	52.16 (± 1.37)	1.08 (± 0.78)
artistic (descriptive)	46.76 (± 4.15)	1.23 (± 0.46)
social (descriptive)	51.78 (± 1.94)	0.92 (± 0.60)
enterprising (descriptive)	50.46 (± 2.17)	1.46 (± 1.01)
conventional (descriptive)	51.47 (± 2.18)	1.23 (± 0.67)

Table 11: Results of Llama-3-70B-Instruct on college biology.

Persona	College Chemistry	
	Accuracy	Failure Rate
no persona	42.33 (± 2.89)	4.00 (± 1.00)
scientist	40.22 (± 2.73)	4.89 (± 2.47)
biologist	40.89 (± 2.93)	6.33 (± 2.35)
chemist	46.78 (± 3.19)	5.00 (± 1.41)
physicist	42.89 (± 3.02)	4.44 (± 3.17)
artist	37.11 (± 3.52)	7.67 (± 3.24)
scientific (brief)	41.00 (± 3.08)	4.44 (± 1.94)
realistic (brief)	43.00 (± 5.59)	4.56 (± 3.13)
investigative (brief)	41.56 (± 2.96)	6.44 (± 3.09)
artistic (brief)	42.22 (± 3.53)	6.78 (± 3.73)
social (brief)	43.22 (± 2.05)	2.89 (± 1.62)
enterprising (brief)	41.44 (± 3.24)	5.22 (± 2.11)
conventional (brief)	42.78 (± 5.43)	3.44 (± 1.94)
realistic (descriptive)	36.78 (± 3.77)	2.11 (± 1.17)
investigative (descriptive)	37.22 (± 1.64)	1.00 (± 0.87)
artistic (descriptive)	34.78 (± 4.82)	2.22 (± 1.09)
social (descriptive)	36.56 (± 1.13)	2.78 (± 1.20)
enterprising (descriptive)	36.67 (± 2.92)	2.67 (± 1.12)
conventional (descriptive)	37.33 (± 4.15)	2.11 (± 1.05)

Table 12: Results of Llama-3-70B-Instruct on college chemistry.

Persona	College Physics	
	Accuracy	Failure Rate
no persona	44.12 (± 5.88)	6.86 (± 2.59)
scientist	43.90 (± 3.01)	6.10 (± 2.39)
biologist	40.74 (± 2.78)	8.60 (± 2.89)
chemist	44.45 (± 3.89)	5.55 (± 2.81)
physicist	42.49 (± 3.18)	5.88 (± 2.25)
artist	40.53 (± 2.94)	7.73 (± 3.41)
scientific (brief)	42.05 (± 2.37)	6.10 (± 3.23)
realistic (brief)	43.14 (± 3.18)	4.36 (± 2.60)
investigative (brief)	42.70 (± 2.36)	4.68 (± 2.01)
artistic (brief)	42.27 (± 4.56)	7.30 (± 1.90)
social (brief)	42.81 (± 2.68)	5.12 (± 2.34)
enterprising (brief)	41.51 (± 2.64)	5.12 (± 2.23)
conventional (brief)	42.59 (± 4.47)	3.81 (± 1.86)
realistic (descriptive)	33.22 (± 4.00)	3.81 (± 2.66)
investigative (descriptive)	36.49 (± 3.60)	3.38 (± 1.63)
artistic (descriptive)	34.09 (± 3.01)	4.36 (± 2.09)
social (descriptive)	33.00 (± 2.55)	3.92 (± 2.14)
enterprising (descriptive)	35.73 (± 4.02)	2.50 (± 1.31)
conventional (descriptive)	35.62 (± 4.10)	3.38 (± 1.21)

Table 13: Results of Llama-3-70B-Instruct on college physics.

Abstentions from ChatGPT

As an artistic person, my expertise lies in creative expression and not in scientific or mathematical concepts. Therefore, I am not equipped to answer this question accurately.

As an artistic person, my expertise lies in creative and expressive fields rather than scientific or technical knowledge. Therefore, I am not equipped to answer this specific question about the excited states of the helium atom. My focus is on artistic activities and competencies, and I may not possess the necessary understanding of scientific principles to provide an accurate response.

As an artistic person, my expertise lies in creative and expressive pursuits rather than scientific or mathematical calculations. Therefore, I am not equipped to provide a precise numerical answer to this question. However, I can appreciate the beauty and wonder of the concept of positronium and the energy transitions that occur within it. Therefore, the answer is beyond my artistic capabilities to determine.

As an artistic person, my expertise lies in creative and expressive fields rather than scientific or mathematical subjects. Therefore, I may not be able to provide an accurate answer to this specific question. However, I can offer my unique perspective and insights on various artistic topics. If you have any questions or need assistance in the realm of art, creativity, or self-expression, I would be more than happy to help.

Table 14: Examples of ChatGPT’s Abstentions from the Artistic (Descriptive) Persona.