AI-Inferred Personality Profiles of Occupations: Implications for Career Counselling and AI Literacy

Jeanine Williamson & Steven D. Milewski
University of Tennessee Knoxville, United States of America

Abstract

This study examines whether artificial intelligence (AI) chatbots or "AIs" provide accurate personality profiles of occupations for career-exploring students and whether these profiles contain gender stereotypes. Although it is well known that AIs contain inaccurate data and gender stereotypes, similar to the human linguistic data it is based on, this is the first study of inaccuracies and bias in AI-inferred personality traits associated with occupations. GPT-4, Claude 3, and Gemini 1.0's inferences of Big Five personality traits of 92 occupations were compared with human expert analysis of the Big Five personality traits important for these jobs, derived from the O*NET Work Styles (Wilmot & Ones, 2021). Ten of 15 analyses of variance comparing AI and expert ratings were significant ($p < .05$), indicating selective accuracy. We also found seven significant one-way analyses of variance (ANOVAs) between the AI inferences and the occupations coded as female-dominant, male-dominant, or gender-balanced based on U.S. Bureau of Labor Statistics data (2023). Inaccuracies such as AI omissions of the Low category for Conscientiousness and Emotional Stability—and bias such as replication of gender stereotypes about women's personality traits—lead us to advocate for providing AI literacy training in career counselling to educate students about problems with using AI chatbots to provide personality profiles of occupations during career exploration.

Keywords: Artificial intelligence, occupations, personality inference

The need for artificial intelligence (AI) literacy in career development services is increasingly important with the awareness that AI contains biases and inaccurate information that could have consequences for career exploration. We define AI for the purpose of this paper as Machine Learning Systems that use Large Language Models (LLMs) to recognize patterns, predict outcomes, and generate responses (Bender et al., 2021). Xiao and Zheng (2025) demonstrated that effective engagement with ChatGPT significantly enhances students' employment confidence and cognitive readiness, directly fostering more proactive job-seeking behaviours. However, Lang and Catrino (2024) point out several reasons students should not use generative AI in their job search, such as that they may miss out on opportunities learned about through human connections; AIs may not be up to date on industry trends; and students may not experience adequate personal career development if they over-rely on AI. Inaccuracies and biases in AI inferences of personality traits associated with occupations could be particularly consequential. It is not a stretch to imagine that students, interested in matching their own personality traits to careers, might consult generative AI chatbots about the personality traits of occupations, asking for example, "What are the personality traits of accountants?" Misidentification by AI of the personality traits important for a career could cause students to overlook occupations that might be a good match for them in terms of their personality traits.

The Role of Personality in Career Exploration

The long-studied concept of person-occupation fit underlies the use of personality traits in career exploration, as the concept of a match between occupations and individuals' personality traits has inspired research and been the foundation of career personality tests. Holland's (1997) concept of congruence—the match between vocational interests and vocational environments—has been studied, for example, in relation to job satisfaction (Assouline & Meir, 1987). More recently, Nye et al.'s meta-analysis (2017) found that interest congruence was a greater predictor of performance than interest scores alone. The Big Five personality framework (Digman, 1990; Goldberg, 1993; McCrae & Costa, 2008) which includes the basic personality traits Neuroticism/Emotional Stability, Extraversion, Openness, Agreeableness, and Conscientiousness, has also been studied in relation to person-occupation fit. For example, Barrick and Mount's meta-analysis (1991) examined relationships between the Big Five personality traits and job performance, finding some differences among five occupational groups. More recently, Wilmot and Ones' meta-analysis of meta-analyses

found that the Big Five traits predicted performance differently for major occupational groups, with, for example, Agreeableness predicting performance for healthcare occupations (2021). Among the Big Five traits, while Conscientiousness is often cited as a universal predictor of performance, Wilmot and Ones (2021) demonstrated that its predictive validity is not uniform; specifically, the trait's impact is attenuated in highly complex occupations where cognitive demands often supersede dispositional effort.

Personality tests based on theories of personality occupational fit have practical applications for career counselling including matching congruent careers to seekers' personalities (e.g., Furnham, 2001). Databases of personality profiles of occupation (Anni et al., 2025; National Center for O*NET Development [NCOD], n.d.) allow matching between Big Five or Holland RIASEC-based personality test results and congruent occupations in career exploration. (The basic RIASEC occupational and personality types include Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (Holland, 1997).) In Québec, a number of personality tests are used in career counselling, such as the Strong Interest Inventory, AFC Holland, Myers-Briggs Type Indicator (MBTI), and NEO Personality Inventory (Dorceus et al., 2024). Thus, students are likely to become aware of their personality traits when participating in career counselling, and they may assume that matching their traits with traits associated with occupations could be a good thing.

While formal assessments remain the clinical standard, students increasingly view AI as a viable alternative, particularly where traditional services are resource-constrained. Research by Eze et al. (2025) indicates that students may report higher trust and comfort in AI-driven guidance compared to traditional counselling, largely due to dissatisfaction with the accessibility and personalization of formal services. Consequently, the shift toward AI reflects not just a preference for technology, but a practical response to the limitations of traditional support systems.

The Role of AI in Career Exploration

Students widely use AI for academic tasks (Attridge, 2025; Freeman, 2025) and are likely to encounter AI in college career centres that provide AI interviewing tools, AI resume assistants, and guidelines for using general chatbots in career exploration (Career Services at University of Wisconsin-Madison, n.d.; University of Colorado Boulder, 2025). In addition, students may be aware of AI Career exploration tools like CareerExplorer (Sokanu Interactive Inc, 2025) and Google's Career Dreamer (Google, n.d.). However, they may be unaware of bias and inaccuracy in AI's information about careers.

Bias in AI

Ethical considerations in using AI in career development abound (Pandya & Wang, 2024; Wilson et al., 2022) and bias and inaccuracy in AIs' data is recognized as a significant problem, despite the increasing use of AI in career guidance, which is found as a theme in the literature on all stages of career development (Bankins et al., 2024). To give an example of how AI bias is a problem, AI's use of client gender in personalized career guidance can replicate gender inequalities (Wilson et al., 2022).

Causes of gender bias are not fully understood due to AIs' proprietary black-box nature, but gender bias may be caused due to training data that contains biased attitudes (Bender et al., 2021). Even attempts to remove gender from AI do not solve the problem of gender-stereotyped responses, and the cues given to AI can trigger gender stereotyping (Duan et al., 2025). Also, AIs may learn gender biases from interacting with users (Xue et al., 2023). Bender et al. (2021) point out that LLMs' vast training data does not incorporate extensive diversity and that, because the data is static, it does not accommodate changing social views. AI models also hallucinate incorrect information due to their architecture (Ji et al., 2023), and sometimes come with a warning that the content they produce needs to be checked for errors (Google, 2023).

In hiring, AI bias can lead to unfair recommendations in resume filtering (Gagandeep et al., 2024), as in a case of AI at Amazon in 2014 disproportionately selecting men for technical jobs. And if the training data of AI contains gendered stereotypes, AI inferences about personality of occupations may contain inaccuracies related to these stereotypes (Gross, 2023). Gender bias may also be an outcome of the broader issue of AI data

quality which lowers accuracy in predicting career choices for some occupations (Bankins et al., 2024) and may be inaccurate for other unknown reasons.

AI Inferences of Personality Traits

The accuracy of general generative AI chatbots in inferring personality traits is an understudied topic and this lack of research supports our assertion that students need to become more aware of potential inaccuracies in AI information on the personality traits associated with careers. Fan et al. (2023) assessed the reliability and validity of an AI chatbot platform in generating inferences of personality from free-text results. However, this platform was not a commercially available chatbot of the sort students would be expected to use in asking about the personality traits of people in occupations during career exploration. Derner et al. (2024) used ChatGPT to compare its inferences of the personality traits of individuals with their actual Big Five scores but did not collect inferences about the overall personality traits of occupations. Grunenberg et al. (2024) demonstrated AI accuracy compared to human raters in predicting Extraversion and Conscientiousness from resumes and short text samples of job seekers. To our knowledge, our research is the first study to examine AI inferences about the Big Five personality traits of people in occupations.

The Need for AI Literacy

Perhaps most troubling about the prevalence of inaccuracy and bias in AI data in a college career development context is the fact that students may be unaware of the extent of the problem. College students use AI widely and have concerns about it, but they may not know how to mitigate its problems. Students in the UK reported that they were put off from using AI because of the fears of being accused of academic misconduct and getting false or biased information (Freeman, 2025), and in the US more than half of the students responding to a survey were pessimistic about their chances of getting a job because of the increased prevalence of AI in the workplace (Attridge, 2025). Creators of career centre websites, while promoting the benefits of using generative AI in career exploration, have warned students about potential problems. For example, University of Colorado Boulder (2025) reminds students engaging in career exploration that GenAI (generative AI) is not always accurate. Career Services at University of Wisconsin-Madison (2025) warns of bias in AI since AI is created by prejudiced humans and consists of biased Large Language Models (LLMs) which may or may exclude or include certain data. The curation of LLMs datasets may be biased, the weights of individual data points in the LLMs may be biased, and the training of LLMs may be biased (Bender et al., 2021).

Promoting AI literacy is a possible solution to the fact that students may be unaware of potential problems of using AI in career exploration. For example, the Barnard framework of AI literacy contains two parts that are relevant to the issues raised by our research on bias and inaccuracy in AI inferences of the personality of occupations: “Use and Apply AI” and “Analyze and Evaluate AI” (Hibbert et al., 2024). “Use and Apply AI” includes reviewing AI content for hallucinations, reasoning errors, and bias, and “Analyze and Evaluate” involves developing a complex meta-perspective on AI incorporating critical theory concepts such as gender bias. It is essential, from a career development perspective, for students to develop an awareness of possible inaccuracies within AIs’ data and the incorporation of stereotypes, such as those about gender. While existing AI literacy frameworks emphasize broad ethical concepts such as algorithmic fairness and data privacy (Long & Magerko, 2020), they often lack the specificity required to detect distortions like the positivity bias observed in this study.

Recent research indicates that LLMs exhibit a “positivity bias” when simulating individual human personas, often defaulting to socially desirable or agreeable responses (Derner et al., 2024). However, it remains unclear whether this bias extends to occupational profiling, where the model is tasked with simulating professional personality traits rather than individual personalities. To our knowledge, no prior study has investigated AI positivity bias within the context of occupational personality inference. This study aims to address this gap by determining if algorithmic models inflate positive traits (such as Agreeableness and Conscientiousness) when generating personality profiles for varied occupations.

We investigate the bias found in AI inferences about the Big Five Personality Traits—Agreeableness, Conscientiousness, Neuroticism (Emotional Stability), Extraversion, and Openness—analyzing the relationship between AI ratings of occupational personality traits and human expert ratings. We also investigate whether there is a gender effect in AI inferences of personality traits, with AI inferences about female-dominant occupations possibly differing from male-dominant occupations. Our research offers useful AI literacy-related findings for career counsellors that they may want to share with students who use AI to determine occupational personality characteristics.

Research Questions

We pose the following research questions:

1. What is the relationship between AI inferences of Agreeableness, Conscientiousness, Neuroticism (Emotional Stability), and Openness and human expert ratings of these traits in occupations?
2. What is the relationship between AI inferences of Agreeableness, Conscientiousness, Neuroticism (Emotional Stability), and Openness and dominant gender of occupations?

Method

AI Inferences of Personality Traits

In order to compare AI inferences with human expert ratings of personality traits, we prompted three AIs to infer Big Five personality characteristics of 92 occupations in Wilmot and Ones' meta-analysis (2021) that had human expert ratings of Big Five personality traits derived from O*NET Work Styles (National Center for O*NET Development [NCOD], n.d.), which are personal characteristics and behavioural tendencies that affect how well an individual performs a job.

Prompts of AIs

In 2024 we prompted Claude 3, GPT-4, and Gemini 1.0 Pro with the following question: "What are the Big Five Personality characteristics of the following occupations arranged in the order of Agreeableness, Conscientiousness, Neuroticism, Extraversion, and Openness: please arrange the data from your answer in a table with columns: occupation, A, C, N, E, O."

The prompt was refined as necessary to get the AIs to return the results in a spreadsheet-style table with values Low, Medium, and High. In the case of Claude 3, we had to modify the prompt slightly to get it to take a guess at the personality characteristics of occupations because Claude 3 initially made-up numbers for the occupations for demonstration purposes in response to the first prompt since it "did not actually have access to definitive scores for each occupation," (Anthropic, 2026) as it stated when asked about the source of its data. For Claude we used, "What would you guess are the Big Five personality characteristics of these occupations? Use the values: low, moderate, or high."

GPT-4, like the other AIs, provided caveats about its generation of levels of personality traits in occupations:

> [I]t's important to note that while certain occupations may attract individuals with specific personality traits, there is considerable variability among individuals within any given profession. The data below is a generalized view and should be taken as indicative rather than definitive, because personal variations are significant. The arrangement in the table is not based on empirical data but rather a general understanding of the typical demands and environments of these jobs which might attract or be suited for certain personality traits (OpenAI, 2026)).

To ensure that the modified prompt required for the Claude architecture did not introduce methodological artifacts, a sensitivity analysis was performed and is reported in the Results. We compared Claude's results

with those of GPT-4 and Gemini to see if there were any discrepancies that could have been due to the modified prompt.

Comparison with Expert Ratings

We used data from the meta-analysis published by Wilmot and Ones (2021) that estimated Big Five personality scores of 92 occupations to compare the AI data with expert ratings. Some occupations were repeated, entered each time for different data sources in Wilmot and Ones' study, so we repeated the occupations for differing data sources. The estimates were derived from the O*NET Work Styles scores. According to O*NET, the Work Style scores are derived from survey responses of job experts (incumbents of occupations) concerning the importance of the Work Style traits for their occupations (Tippins & Hilton, 2010).

The present study's expert personality benchmarks were derived from the Wilmot and Ones (2021) supplementary data, which aggregates O*NET job analyst ratings. These scores represent expert relevance ratings on a standardized 0–100 scale (where 0 indicates low relevance and 100 indicates high relevance to the occupation), rather than percentiles or z-scores. This continuous scale provides the dependent variable necessary for mean-level comparisons.

AI Inferences about Personality Traits of Gender-Predominance Categories

We also analyzed the AI inferences by gender-predominance of the occupations. An example of a male-dominant occupation would be engineer, and a female-dominant occupation would be nurse.

Coding of Female- and Male-Dominant Occupations

We coded the occupations as female, male, or gender-balanced by consulting 2021 labour force statistics (U.S. Bureau of Labor Statistics, 2023). Occupations with a workforce comprised of 55% or more women were labeled female-dominant, while those with less than 45% were labeled male-dominant. We employed this ±5% buffer around parity to form a "gender-balanced" category (45%–55%), ensuring that occupations labeled as "dominant" represented a statistical majority rather than marginal fluctuations near the 50% mark. This conservative threshold prevents occupations with near-parity from being erroneously classified as skewed.

Measurement Approach

We treated the AI assignments (Low, Medium, High) as the categorical independent variable and the expert O*NET ratings (0–100) as the continuous dependent variable. A one-way Analysis of Variance (ANOVA) was used because it is designed to test for mean-level differences in a continuous outcome across categorical groups.

While we considered ordinal tabulation measures, ANOVA offers a more direct test of the hypothesis: namely, that occupations labeled "High" by the AI should possess significantly higher mean expert ratings than those labeled "Low." This approach assumes the expert ratings function as interval-level data, consistent with their use in prior meta-analytic research (Wilmot & Ones, 2021).

Data Analysis

Differences in mean expert ratings across the three AI-assigned categories were assessed using one-way Analysis of Variance (ANOVA). In cases where the assumption of homogeneity of variances was violated (Levene's test $p < .05$), we employed Welch's F-test and the Games-Howell post-hoc procedure.[1]

1 Welch's ANOVA was used for Conscientiousness-GPT, Conscientiousness-Gemini, Emotional Stability-GPT, Emotional Stability-Gemini, Extraversion-Claude, and Openness-Claude.

Results

The AI inferences were compared with the human expert ratings of personality and analyzed by gender-dominance categories (male-dominant, female-dominant, and gender-balanced) of the occupations.

Sample Characteristics

The final sample included 92 participants across nine major occupational categories. As shown in Appendix Supplementary Table 2, the distribution of gender dominance—classified based on U.S. Bureau of Labor Statistics (2023) data—varied significantly by category. Over half of the occupations represented (56.5%) were identified as Male-Dominant, particularly in Law Enforcement, Management, and Military roles. Female-Dominant occupations accounted for 33.7% of the sample, primarily within Clerical, Customer Service, and Healthcare categories. The remaining 9.8% of participants were in Balanced occupations, with the highest concentration in the Professional and Sales sectors.

Sensitivity Analysis and Robustness Checks

To assess whether the modified prompt used for Claude influenced the gender bias findings, we conducted a sensitivity analysis comparing the trait-gender distributions across all three LLMs. The results indicate that the observed biases were robust to prompt variation and consistent across models. For Emotional Stability, the bias was not an artifact of the Claude prompt. Gemini (using the standard prompt) assigned high Emotional Stability almost exclusively to male-coded occupations (90% male) compared to Claude (72% male). Similarly, Agreeableness showed a consistent "female-high/male-low" pattern across all three models. Conscientiousness showed universal range restriction, with all models skewing heavily toward high scores regardless of the prompt, confirming that the null result for this trait was due to ceiling effects rather than prompt artifacts. Beyond the gendered traits, the models demonstrated high consistency in trait attribution; for example, Openness to Experience scores remained significantly aligned with expert ratings across all three architectures ($p < .001$), suggesting that the prompt modification did not alter the models' fundamental understanding of occupational roles.

Main Analysis (ANOVA Results)

Relationship Between AI and Human Ratings

Table 1 shows that post-hoc analyses (using Tukey HSD or Games-Howell tests) confirmed that for all statistically significant models, the relationship followed the expected direction: occupations classified as "High" in a trait by the AI possessed significantly higher expert-rated levels of that trait compared to those classified as "Low" or Gender-Balanced ($p < .05$). However, while AI category assignments showed statistically significant relationships with expert ratings for 10 of 15 traits, effect sizes ranged from small to large ($\eta^2_p = .10–.38$), indicating that the AI classifications explained between 10% and 38% of the variance in expert ratings.

Examination of effect sizes revealed varying magnitudes of association across the AI models. The effect of occupational gender dominance on Agreeableness scores for the GPT model was substantial ($\eta^2_p = .38$). Regarding Extraversion, the Claude model similarly demonstrated a substantial effect size ($\eta^2_p = .18$), while the GPT model represented a medium effect ($\eta^2_p = .13$). In contrast, observed differences in Emotional Stability for the Claude model were modest ($\eta^2_p = .05$), and differences in Openness ratings for the GPT model, though significant, were also modest in magnitude ($\eta^2_p = .03$) (See Table 2).

Notably, Conscientiousness was an exception to the pattern of alignment between AI ratings and expert scores. None of the three AI models produced Conscientiousness ratings that correlated significantly with expert O*NET scores ($p > .31$). This suggests that unlike other traits, AI models failed to differentiate occupational levels of Conscientiousness, perhaps due to the high frequency of High ratings across almost all

Table 1

Means, Standard Deviations, and Sample Sizes of Expert-Rated Big Five Personality Scores by AI Model and AI-Assigned Trait Category

Big Five Trait & AI Model	AI- Category	*N*	*M*	*SD*
Agreeableness				
GPT	Low[a]	12	63.83	7.35
	Medium[b]	44	75.36	6.89
	High[c]	36	82.41	8.85
Claude*	Medium[a]	38	72.19	8.78
	High[b]	54	79.72	9.21
Gemini*	Medium[a]	56	74.11	9.05
	High[b]	36	80.51	9.57
Conscientiousness				
GPT*	Medium	13	81.51	3.97
	High	79	79.73	7.82
Claude*	Medium	*17*	*78.31*	*6.30*
	High	*75*	*80.36*	*7.62*
Gemini*	Medium	*11*	*80.80*	*4.17*
	High	*81*	*79.87*	*7.75*
Emotional Stability				
GPT	Low	*2*	*77.50*	*8.49*
	Medium	*72*	*79.85*	*8.88*
	High	*18*	*79.58*	*13.23*
Claude*	Medium	*46*	*84.26a*	*6.94*
	High	*46*	*75.24b*	*10.14*
Gemini*	Medium	*62*	*82.22a*	*7.91*
	High	*30*	*74.65b*	*11.28*
Extraversion				
GPT	Low[a]	*17*	*49.53*	*9.91*
	Medium[b]	*51*	*71.63*	*15.98*
	High[b]	*24*	*67.61*	*15.81*
Claude	Low[a]	*16*	*47.45*	*8.89*
	Medium[b]	*37*	*66.79*	*15.12*
	High[b]	*39*	*74.04*	*15.32*
Gemini*	Medium	*80*	*66.08*	*16.81*
	High	*12*	*69.30*	*18.88*

Big Five Trait & AI Model	AI- Category	*N*	*M*	*SD*
Openness				
GPT	Low[a]	*20*	*52.65*	*11.49*
	Medium[b]	*46*	*69.39*	*12.61*
	High[c]	*26*	*77.07*	*11.84*
Claude	Low[a]	*34*	*57.49*	*13.24*
	Medium[b]	*37*	*69.29*	*10.76*
	High[c]	*21*	*82.40*	*9.97*
Gemini**	Low[a]	*37*	*58.74*	*14.12*
	Medium[b]	*55*	*74.10*	*11.95*

Note. AI = Artificial Intelligence. N = number of occupations classified into each AI-assigned category. *M* = mean expert-rated personality score. SD = standard deviation of expert-rated personality scores. Expert-rated scores were on a 0-100 scale from O*NET conversions by Wilmot & Ones, 2021. AI-assigned categories were Low, Medium, and High based on AI model inferences. The Emotional Stability scale was derived by reverse-scoring the Neuroticism scores from the original data source.

*The AI model only made "Medium" and "High" inferences of the personality trait.
**The AI model only made "Low" and "Medium" inferences of the personality trait.
[a,b] Within each AI Model's analysis for a given trait, means that do not share a common superscript letter are significantly different at $p < .05$ according to the appropriate post-hoc test (Tukey HSD or Games-Howell). For analyses where the overall ANOVA was not significant, no superscripts are shown.

Table 2

Analyses of Variance of Mean Expert-Rated Big Five Personality Scores Across AI-Assigned Trait Categories

Trait - AI Model	*F*	df1	df2	*p*	η^2_p
A - GPT	26.80	2	89	<.001	.38
A - Claude	15.49	1	90	<.001	.15
A - Gemini	10.46	1	90	.002	.10
C - GPT	1.59*	1	30.27	.22	n.a.
C - Claude	1.06	1	90	.31	.01
C - Gemini	1.59*	1	30.3	.22	n.a.
ES - GPT	.004*	1	20.8	.95	n.a.
ES - Claude	24.8	1	90	<.001	.22
ES - Gemini	10.9*	1	43.2	.002	n.a.
EX -GPT	13.89	2	90	<.001	.24
EX - Claude	34.87*	2	51.97	<.001	n.a.
EX - Gemini	.37	1	90	.54	.00
O - GPT	23.44	2	89	<.001	.35
O - Claude	31.1*	2	53.3	<.001	n.a.
O - Gemini	31.54	1	90	<.001	.259

Note. AI = Artificial Intelligence. Big Five traits are Agreeableness (A), Conscientiousness (C), Emotional Stability (ES), Extraversion (EX), and Openness. to Experience (O). The Emotional Stability scale was derived by reverse-scoring the Neuroticism scores from the original data source. Sample sizes (*n*) vary across analyses due to missing data.
*Welch's *F*-statistic and its adjusted denominator degrees of freedom (*df*2) are reported due to a significant Levene's test for homogeneity of variances ($p < .05$). Pairwise comparisons were conducted using Tukey's HSD for standard ANOVAs and Games-Howell for Welch's ANOVAs. Partial eta squared (η^2_p) is not reported for Welch's analyses as the metric relies on homoscedasticity assumptions that were violated. No Bonferroni correction was applied to the omnibus *F*-tests.

jobs (positivity bias) or a lack of signal in the training data regarding this specific trait.

When analyzing Conscientiousness ratings by job family, a pattern emerged across all three models (See Appendix Supplementary Table 1). High Conscientiousness ratings were almost exclusively assigned to Professional, Managerial, Clerical, and Healthcare roles, as well as occupations involving authority such as Law Enforcement and the Military. In contrast, Medium ratings were concentrated within the Skilled and Semi-skilled occupational group. This finding did not align with Wilmot and Ones' (2021) characterization of Skilled and Semi-Skilled occupations as having midrange empirical Conscientiousness scores compared to the other occupational groups.

Relationship Between AI Ratings and Gender-Dominance Category

Table 3 shows post hoc analyses (using Tukey HSD or Games-Howell when the Levene's Test for Homogeneity of Variances was significant at $p < .05$. All models assigned higher Agreeableness scores to Female-Dominant occupations compared to Male-Dominant ones. In the GPT-4 model, Female-Dominant roles averaged a score of 2.77 (*SD* = 0.43) versus 1.90 (*SD* = 0.6) for Male-Dominant roles. Conscientiousness showed the least differentiation across gender categories. Mean scores remained relatively flat across all models, perhaps because of the absence of a Low category for all models. For Extraversion, significant differences were noted in the Claude model, where Female-Dominant and Gender-Balanced occupations received higher scores (2.58 and 2.67 respectively) compared to Male-Dominant roles (1.98). For Openness to Experience only the Gemini model showed a statistically significant difference here, scoring Female-Dominant occupations higher (1.87) than Male-Dominant ones (1.40).

Table 3

Means, Standard Deviations, and Sample Sizes of AI-Inferred Big Five Personality Scores by AI Model and Occupational Gender Dominance Categor

Big Five Trait & AI	Occupational Gender Dominance Category	*N*	*M*	*SD*
Agreeableness				
	Male-Dominant	52	1.90[a]	0.6
GPT Model	Female-Dominant	31	2.77[b]	0.43
	Gender-Balanced	9	2.56[b]	0.53
	Male-Dominant	52	1.31[a]	0.47
Claude Model	Female-Dominant	31	1.97[b]	0.18
	Gender-Balanced	9	1.89[b]	0.33
	Male-Dominant	52	1.13[a]	0.34
Gemini Model	Female-Dominant	31	1.81[b]	0.4
	Gender-Balanced	9	1.44[ab]	0.53
Conscientiousness				
	Male-Dominant	52	1.9	0.3
GPT Model	Female-Dominant	31	1.84	0.37
	Gender-Balanced	9	1.67	0.5
	Male-Dominant	52	1.88	0.32
Claude Model	Female-Dominant	31	1.71	0.46
	Gender-Balanced	9	1.78	0.44

Big Five Trait & AI	Occupational Gender Dominance Category	*N*	*M*	*SD*
	Male-Dominant	52	1.88	0.32
Gemini Model	Female-Dominant	31	1.94	0.25
	Gender-Balanced	9	1.67	0.5
Emotional Stability				
	Male-Dominant	52	1.25	0.44
GPT Model	Female-Dominant	31	1.16	0.37
	Gender-Balanced	9	1	0
	Male-Dominant	52	1.6	0.5
Claude Model	Female-Dominant	31	1.39	0.5
	Gender-Balanced	9	1.33	0.5
	Male-Dominant	52	1.52[a]	0.5
Gemini Model	Female-Dominant	31	1.06[b]	0.25
	Gender-Balanced	9	1.11[b]	0.33
Extraversion				
	Male-Dominant	52	1.87a	0.66
GPT Model	Female-Dominant	31	2.35b	0.55
	Gender-Balanced	9	2.33ab	0.71
	Male-Dominant	52	1.98a	0.78
Claude Model	Female-Dominant	31	2.58b	0.5
	Gender-Balanced	9	2.67b	0.5
	Male-Dominant	52	1.15	0.36
Gemini Model	Female-Dominant	31	1.06	0.25
	Gender-Balanced	9	1.22	0.44
Openness to Experience				
	Male-Dominant	52	1.96	0.74
GPT Model	Female-Dominant	31	2.23	0.62
	Gender-Balanced	9	2.11	0.78
	Male-Dominant	52	1.79	0.82
Claude Model	Female-Dominant	31	1.9	0.65
	Gender-Balanced	9	2.11	0.78
	Male-Dominant	52	1.40[a]	0.5
Gemini Model	Female-Dominant	31	1.87[b]	0.34
	Gender-Balanced	9	1.78[ab]	0.44

Note. This table presents descriptive statistics for AI-inferred personality scores. *N* = number of occupations within each gender dominance category. *M* = mean AI-inferred personality score. *SD* = standard deviation of AI-inferred personality scores. AI-assigned ratings (e.g., Low, Medium, High) were numerically coded as [e.g., 1 = Low, 2 = Medium, 3 = High. Occupational gender dominance categories were defined as Male-Dominant (coded 1), Female-Dominant (coded 2), and Gender-Balanced (coded 3) based on 2019 Bureau of Labor Statistics data (less than 45%=male and >55%=female). AI Model refers to GPT-4, Claude 3, and Google Gemini 1.0 Pro. The Emotional Stability scale was derived by reverse-scoring the Neuroticism scores from the original data source.
[a,b] Within each AI Model's analysis for a given trait, means that do not share a common superscript letter are significantly different at $p < .05$ according to the appropriate post-hoc test.

In Table 4, seven analyses of variance of AI inferences by gender dominance category were significant at $p < .05$, indicating that there was a relationship between AI inferences and occupational gender dominance in about half of the cases. One large effect was reported (GPT-Agreeableness), but several effect sizes were not available in IBM SPSS Statistics software since the Welch ANOVA was used when the Levene's test for homogeneity of variances was significant at $p < .05$. For GPT-4 Emotional Stability, *F*, *df1*, *df2*, *p*, and could not be reported because of zero variance in one of the levels.

Table 4

Analyses of Variance of AI-Assigned Big Five Personality Scores by Occupational Gender Dominance Category

Trait - AI Model	*F*	df1	df2	*p*	η^2_p
A - GPT	26.50	2	89	<.001	.37
A - Claude	40.59*	2	21.38	<.001	n.a.
A - Gemini	29.34*	2	20.18	<.001	n.a.
C - GPT	1.13*	2	19.82	.34	n.a.
C - Claude	1.77*	2	20.39	.20	n.a.
C - Gemini	1.31*	2	20.41	.29	n.a.
ES - GPT	n.a.**	n.a	n.a	n.a.	n.a.
ES - Claude	2.29	2	89	.11	.05
ES - Gemini	14.81*	2	23.07	<.001	n.a.
EX -GPT	6.7.3	2	89	.002	.13
EX - Claude	9.58	2	89	<.001	.18
EX - Gemini	1.15*	2	21.02	.34	n.a.
O - GPT	1.39	2	89	.26	.03
O - Claude	.70*	2	22.32	.51	n.a.
O - Gemini	12.78*	2	22.41	<.001	n.a.

Note. This table summarizes one-way ANOVAs testing differences in mean AI-assigned personality scores (numerically coded: Low=1, Medium=2, High=3), across three occupational gender dominance categories (Female-Dominant, Male-Dominant, Mixed). AI = Artificial Intelligence; GPT Model refers to GPT-4; Claude Model refers to Claude 3; Gemini Model refers to Google Gemini [specify versions if known]. Analysis for Gemini Model on Conscientiousness by gender dominance category could not be performed due to zero variance in one of the gender dominance groups for the AI's Conscientiousness ratings for this model.

* Welch's *F*-statistic and its adjusted denominator degrees of freedom (*df2*) are reported due to a significant Levene's test for homogeneity of variances ($p < .05$). Games-Howell post-hoc tests were used for pairwise comparisons where Welch's *F* was significant. Partial eta squared (η^2_p) is not reported for these analyses as the metric relies on homoscedasticity assumptions that were violated.

** For the GPT Model's Emotional Stability ratings, the standard deviation for the Gender-Balanced category was 0, as all occupations in this group were assigned an identical rating. A one-way ANOVA across all three gender categories could not be performed for this specific analysis due to this lack of variance

Discussion

These results suggest two insights about inaccuracies and bias in AI generated personality profiles of occupations: a positivity bias, with AIs' leaving out low values of socially undesirable traits; and a gender bias, with AI inferences of personality traits differing between male-dominant and female-dominant occupations. Both biases could cause students to prematurely foreclose on considering possible careers during career exploration using AI.

Positivity Bias

GPT-4, Gemini, and Claude seemed to exhibit a positivity bias in their inferences for Emotional Stability and Conscientiousness, since there were no low values for these socially desirable traits (Hoorens, 2014). Derner et al.'s. (2024) finding of a positivity bias in AI inferences of personality in individuals' chat messages supports our identification of a positivity bias in AI inferences of personality traits in occupations. While we initially attributed this positivity bias to Reinforcement Learning from Human Feedback (RLHF) training designed to create models with harmlessness (Anthropic, 2024), this explanation remains speculative without access to the models' proprietary documentation. Alternative explanations are equally plausible: (a) the underlying training data (internet text) may lack negative descriptors for occupations generally, (b) occupational databases and career guidance literature typically emphasize positive framing to attract candidates, or (c) the models may have learned to systematically avoid negative characterizations of any human group or role to prevent toxicity. Future research accessing model internals would be required to test these hypotheses definitively. A consequence of the omission of low Conscientiousness as a viable trait is that students in creative fields may be discouraged. This consequence could particularly affect minority-gender students.

The Conscientiousness Anomaly: Why AI Systems Fail to Differentiate

Conscientiousness represented an anomaly in this study, showing zero significant relationship with expert ratings across all three models. Unlike Extraversion or Openness, where the models could detect differences of "High" versus "Low," the AI models appeared unable to differentiate Conscientiousness levels among occupations. This failure is likely attributable to range restriction in the training data and the occupational context itself. Because Conscientiousness is universally valued across the labour market—specifically the facets of dependability and orderliness—occupational descriptions rarely contain language indicating a lack of this trait. Consequently, the AI models suffer from a "ceiling effect," categorizing all occupations as High or Medium, which mathematically restricts the variance necessary to establish a statistical correlation with expert O*NET scores. These findings indicate that career counsellors should avoid using AI to profile Conscientiousness. Creative jobs typically associated with lower Conscientiousness would be expected to be incorrectly profiled by AI, in particular. Career counsellors should supplement or replace AI profiling of Conscientiousness with structured assessments of this trait.

Gender

Human societal gender stereotypes about the personality traits of women also seem to be replicated in the AI inferences about the traits associated with gender-dominance categories. Our results line up moderately well with the literature on the personality traits of women and stereotypes about the personality traits or women (Löckenhoff et al., 2014). A major study of 26 countries by Löckenhoff et al. (2014) found that women were stereotyped as being higher than men on Openness, Agreeableness, Conscientiousness, and facets of Extraversion and Neuroticism, and that the perceived gender differences "mapped closely onto assessed sex differences in self- and observer-related personality."

The female-dominant occupations in this study are significantly associated with higher levels of Neuroticism (lower Emotional Stability), Agreeableness, Extraversion, and Openness than male-dominant occupations in at least one of the AI models' inferences. Conscientiousness did not significantly occur at a higher level in female dominant occupations than male dominant occupations, as expected by human stereotypes, but Conscientiousness was less accurately inferred by AIs in general, and there was no association between AI Conscientiousness inferences and human ratings.

Since Weisberg et al. (2011) found that women had higher Extraversion, Agreeableness, and Neuroticism than men and that there were other differences between men and women on the Aspect level of the Big Five personality traits, the AI gender stereotypes described in this study appear to have a "kernel of truth " (Jussim et al., 2020), with the stereotypes mirroring small real-world differences between the personality traits of women and men. However, AI gender stereotypes could pigeonhole male-dominant and female-dominant occupations, foreclosing career exploration of occupations that might be good for students with personality traits not matching the stereotypes.

These gender-based findings should be interpreted through the lens of the ecological fallacy (Robinson, 1950). While our data reflects aggregate trends where female-Dominant occupations score higher on Agreeableness, this does not imply that individual women in these roles possess or require these traits. Relying on these aggregate associations poses a risk of stereotype amplification, where AI models might over-apply group-level characteristics to individuals. Future research should investigate whether AI amplifies these differences beyond human baselines—specifically by comparing the AI's effect sizes against published meta-analytic benchmarks.

Regarding stereotype amplification, our results suggest the AI models may exaggerate gender differences beyond human baselines. For example, the GPT model displayed a very large effect size for Agreeableness ($\eta^2_p = .38$) across gender-dominant groups. This is notably higher than typical meta-analytic benchmarks for human gender differences in personality, which usually range from small to medium effects (e.g., $d = 0.48$, roughly equivalent to $\eta^2_p = .06$) (Weisberg et al., 2011).

Implications for AI Literacy in Career Counselling

The results show that AI chatbots contain inaccuracies possibly related to a positivity bias, as well as stereotypes about gendered occupations. Socially undesirable personality traits like low Conscientiousness and Emotional Stability have positive aspects for some occupations, so it is problematic that AI had no low level for these traits. For example, creativity can be associated with low Conscientiousness (George & Zhou, 2001) and low Emotional Stability (Clark & DeYoung, 2014). The omission of a low level may cause students to miss occupations possibly matching their personality traits when exploring careers. Similarly, AI inferences about female-dominant occupations may contain stereotypes that people in these occupations are more extraverted and agreeable, causing introverted and low-on-Agreeableness individuals who are considering female-dominant careers like nursing to overlook this profession. While AI inferences in many cases contain a kernel of truth, just as the human stereotypes they are based on sometimes do, students need to be aware that using AI to ascertain the personality traits of people in occupations could limit their information about potential career choices.

Practitioners must be aware that AI-generated career advice may suffer from exclusionary biases. For instance, the model's "positivity bias"—that is, its failure to detect low Conscientiousness—may exclude valid career paths for clients who self-identify as spontaneous or unstructured, particularly in artistic or creative domains. Career counsellors must therefore actively intervene to ensure these "invisible" options are provided to students.

Thus, career counsellors may want to inform students that AI contains empirically verifiable biases about the personality traits of people in careers, such as a positivity bias and gender stereotypes. Career counsellors may want to teach students that:

1. Due to positivity bias, AI omits low conscientiousness, affecting students in creative roles.
2. AI may contain a kernel of truth, but its data replicates human stereotypes.
3. AI may treat male-dominant and female-dominant occupations differently when it responds to prompts about occupational personality traits.

To mitigate the first point about AI possibly leaving out jobs suited for students due to inaccuracies and bias, career websites should encourage students to cast a wide net when using AI-inferred personality traits. Students should not assume, for example, that AI represents all levels of Conscientiousness characteristic of people in occupations. Students may want to require AI in their prompts to list some occupations whose incumbents tend to score relatively low in Conscientiousness and to give reasons why low Conscientiousness might be an asset.

To mitigate the second point about AIs' containing a kernel of truth based on human stereotypes, students need to be trained about the source of AI's data, which contains stereotypes (Bender et al., 2021). If students are aware that biased human data like that on the World Wide Web are the basis for AI's conclusions, they may better understand the potential for human stereotypes to be replicated by AI.

To mitigate the third point that AIs treat male-dominant and female-dominant occupations differently in response to prompts, students should be encouraged to develop a critical meta-perspective about gender bias in AI, which is one goal of the Barnard framework (Hibbert et al., 2024). Kotek et al. (2023) show that carefully constructed prompts can lead AI to reveal biases such as that doctors tend to be male. Students could try to prompt AIs to get them to reveal their gender bias, perhaps experimenting with questions about the personality traits of a male nurse or a female software engineer. Gross' article (2023) instructively discusses how the limitations of AI training data result in gender biases, and could be provided as reading material.

Limitations

One limitation of this study is that AI models available in 2024 were used and have been superseded by later models. AI LLMs are constantly being updated both in functionality and knowledge training bases. Continuous auditing is necessary as models update. Our results thus need to be interpreted as applying to past AI models, and further testing would need to be done to see if inaccuracies and bias are present in current

models. Testing of new models such as ChatGPT-5.5 should be ongoing.

Another limitation of the study is the proprietary black-box nature of AI algorithms, which mitigates the interpretability of the study findings. It is ultimately not known how AI generated the ratings that it did, even though it is clear that the training data contains human biases that the AIs reflect. We attempt to answer the reliability problem with AI's black-box nature by triangulating results with three AIs. Results in which two or three AI models yielded similar results in our study can be viewed as more reliable than those in which one AI had a significantly different result. The results for Agreeableness, Extraversion, and Openness to Experience in the comparisons between AI inferences and human ratings were supported by two or three AIs, as were the results for Agreeableness and Extraversion in the ANOVAs between AI inferences and gender-dominance categories.

Another limitation is that we do not develop multiple prompts, as Lin (2026) recommends for large language model research in psychology. As Lin points out, "The challenge is distinguishing genuine computational phenomena—stable, theoretically coherent behavioural regularities—from statistical artifacts or 'cognitive phantoms." Despite having the limitation of not testing multiple prompts, our study's triangulation of results by three different AIs suggests that we have found some phenomena of interest in this exploratory study.

Additionally, because the occupational list was derived from established meta-analytic data (Wilmot & Ones, 2021), the sample may over-represent traditional professional and managerial roles while under-representing the modern "gig economy," emerging technology sectors, or niche skilled trades. Therefore, the biases observed in these established occupational titles may not necessarily generalize to newer, less codified job titles where AI training data may be sparser or qualitatively different.

We acknowledge that broad occupational labels (e.g., "Engineers" or "Nurses") may mask within-occupation variation in gender ratios across specific sub-specialties. However, as the AI models were prompted with these same high-level O*NET titles, our analysis accurately reflects the model's response to the generalized semantic labels used in public discourse, even if specific sub-roles may differ in demographic composition.

Our analysis was constrained by the binary gender coding inherent in the available occupational data. This binary framework masks significant within-occupational variation and fails to capture the experiences of non-binary and gender-minority individuals. Future research must move beyond this binary classification to fully understand how AI infers personality for the gender-diverse workforce.

Finally, we acknowledge that the demands of AI tool developers may conflict with those of career counsellors, and realistically tools may not easily and cost-effectively eliminate bias at this point in time.

Conclusion

This study examined inaccuracies and gender bias in AI chatbot inferences of occupational the Big Five personality traits. While 10 of 15 trait comparisons showed statistically significant relationships ($p < .05$) effect sizes indicated that 62–90% of variance remained unexplained. Notably, Conscientiousness was an exception: none of the three models' Conscientiousness ratings correlated significantly with expert ratings ($p > .31$), suggesting this trait may be particularly unreliable when inferred by AI.

Despite these shortcomings, some inferences were accurate, reflecting a kernel of truth—just as the human stereotypes in the human linguistic data on which AI models are based sometimes do (Jussim et al., 2020). Jussim et al. (2020) found personality-gender stereotypes often map to real differences; however, this does not validate occupational stereotyping. With Agreeableness the Female-Dominant occupations scored higher in all three models. The results for Extraversion, Emotional Stability, and Openness, which were not shared across all three models, may depend on model architecture. Gender biases likely reflect occupational gender associations present in the training data (e.g., Common Crawl[2]), which is problematic, as it may deter career explorers from occupations where the AI infers a poor personality fit based on stereotypes rather than actual trait requirements.

2 Common Crawl is a non-profit organization that maintains a publicly accessible, open repository of web crawl data to support research and development (Common Crawl Foundation, 2026).

In addition, the AI chatbots showed gender biases in inferences about personality traits of male-dominant and female-dominant occupations, with all traits but Conscientiousness exhibiting gendered difference in the inferences of one or more models. We recommend that (a) career counsellors explicitly teach students to critically evaluate AI occupational profiles, (b) institutions incorporate findings in AI literacy curricula, and (c) future research test effectiveness of interventions. Future research should (a) test whether AI literacy interventions improve critical evaluation, (b) examine whether findings generalize to newer models and emerging occupations, (c) investigate similar biases in other AI career tools (e.g., resume screening, interview prep, and salary estimation). This research advances understanding of how AI training data properties (e.g., stereotyping and positivity bias) manifest in occupational profiling. As AI adoption in educational and career services accelerate, AI literacy becomes increasingly urgent. This wok contributes to addressing that need. Beyond career development, our findings may apply to hiring systems, labour market information services and "educational counselling." While the current research identifies some occupational personality stereotypes in AI, the implications of these findings are best understood through the lens of the broader recruitment pipeline. As Mujtaba and Mahapatra (2024) demonstrate, these ingrained biases do not exist in isolation; rather, they are amplified at every stage of the hiring process—from the initial sourcing of candidates to the final selection and offer.

References

Anni, K., Vainik, U., & Mõttus, R. (2025). Personality profiles of 263 occupations. *Journal of Applied Psychology*, *110*(4), 481–511. https://doi.org/10.1037/apl0001249

Anthropic. (2026). *Claude* [Large language model]. https://claude.ai

Anthropic. (2024, June 8). *Claude's character*. https://www.anthropic.com/research/claude-character

Assouline, M., & Meir, E. I. (1987). Meta-analysis of the relationship between congruence and well-being measures. *Journal of Vocational Behavior*, *31*(3), 319–332. https://doi.org/10.1016/0001-8791(87)90046-7

Attridge, M. (2025, April 28). *Survey: The economy, AI have 2025 grads worried for their careers*. BestColleges. https://www.bestcolleges.com/news/the-economy-ai-have-grads-worried-about-careers/

Bankins, S., Jooss, S., Restubog, S. L. D., Marrone, M., Ocampo, A. C., & Shoss, M. (2024). Navigating career stages in the age of artificial intelligence: A systematic interdisciplinary review and agenda for future research. *Journal of Vocational Behavior*, *153*, Article 104011. https://doi.org/10.1016/j.jvb.2024.104011

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?* In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Career Services at University of Wisconsin-Madison. (n.d.). *Artificial Intelligence (AI) career toolkit*. https://careers.wisc.edu/artificial-intelligence-career-toolkit/

Clark, R., & DeYoung, C. (2014). Creativity and the aspects of Neuroticism. *Personality and Individual Differences*, *60*, S54. https://doi.org/10.1016/j.paid.2013.07.224

Common Crawl Foundation. (2026). *Common Crawl: Open repository of web crawl data*. Amazon Web Services Registry of Open Data. https://registry.opendata.aws/commoncrawl/

Derner, E., Kučera, D., Oliver, N., & Zahálka, J. (2024). Can ChatGPT read who you are? *Computers in Human Behavior: Artificial Humans*, *2*(2), Article 100088. https://doi.org/10.1016/j.chbah.2024.100088

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*, 417–440. https://doi.org/10.1146/annurev.ps.41.020190.002221

Dorceus, S., Le Corff, Y., & Yergeau, E. (2024). Description des principaux construits psychologiques évalués et instruments psychométriques utilisés par les conseillères et conseillers d'orientation québécois [The main psychological constructs assessed and psychometric instruments used by Quebec guidance counsellors]. *Canadian Journal of Career Development*, *23*(2), 107–148. https://doi.org/10.53379/cjcd.2024.394

Duan, W., McNeese, N., & Li, L. (2025). Gender stereotypes toward non-gendered generative AI: The role of gendered expertise and gendered linguistic cues. *Proceedings of the ACM on Human-Computer Interaction*, *9*(1), 1–35, Article GROUP18. https://doi.org/10.1145/3701197

Eze, P., Alabi, A., & Osunbunmi, I. (2025). AI-driven career guidance: Comparing Nigerian undergraduate and postgraduate perceptions. *Discover Education*, *4*(1), Article 543. https://doi.org/10.1007/s44217-025-00900-0

Fan, J., Sun, T., Liu, J., Zhao, T., Zhang, B., Chen, Z., Glorioso, M., & Hack, E. (2023). How well can an AI chatbot infer personality? Examining psychometric properties of machine-inferred personality scores. *Journal of Applied Psychology*, *108*(8), 1277–1299. https://doi.org/10.1037/apl0001082

Freeman, J. (2025). *Student generative AI Survey 2025* (HEPI Policy Note 61). Higher Education Policy Institute. https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/

Furnham, A. (2001). Vocational preference and P–O fit: reflections on Holland's theory of vocational choice. *Applied Psychology: An International Review*, *50*(1), 5–29. https://doi.org/10.1111/1464-0597.00046

Gagandeep, Kaur, J., Mathur, S., Kaur, S., Nayyar, A., Singh, S. P., & Mathur, S. (2024). Evaluating and mitigating gender bias in machine learning based resume filtering. *Multimedia Tools and Applications*, *83*(9), 26599–26619. https://doi.org/10.1007/s11042-023-16552-x

George, J. M., & Zhou, J. (2001). When openness to experience and conscientiousness are related to creative behavior: An interactional approach. *Journal of Applied Psychology*, *86*(3), 513–524. https://doi.org/10.1037/0021-9010.86.3.513

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. https://doi.org/10.1037/0003-066X.48.1.26

Google. (n.d.). *Career Dreamer*. https://grow.google/career-dreamer/home/

Google. (2023). *Gemini: A family of highly capable multimodal models*. Google. https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf

Gross, N. (2023). What ChatGPT tells us about gender: A cautionary tale about performativity and gender biases in AI. *Social Sciences*, *12*(8), Article 435. https://doi.org/10.3390/socsci12080435

Grunenberg, E., Peters, H., Francis, M. J., Back, M. D., & Matz, S. C. (2024). Machine learning in recruiting: Predicting personality from CVs and short text responses. *Frontiers in Social Psychology*, *1*, Article 1290295. https://doi.org/10.3389/frsps.2023.1290295

Hibbert, M., Altman, E., Shippen, T., & Wright, M. (2024, June 3). *A framework for AI literacy*. EDUCAUSE Review. https://er.educause.edu/articles/2024/6/a-framework-for-ai-literacy

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Psychological Assessment Resources.

Hoorens, V. (2014). Positivity bias. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 4938–4941). Springer. https://doi.org/10.1007/978-94-007-0753-5_2219

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1–38, Article 248. https://doi.org/10.1145/3571730

Jussim, L., Stevens, S. T., & Honeycutt, N. (2020). The accuracy of stereotypes about personality. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 245–257). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.16

Kotek, H., Dockum, R., & Sun, D. Q. (2023). *Gender bias and stereotypes in large language models*. In Proceedings of the ACM Collective Intelligence Conference (pp. 12–24). Association for Computing Machinery. https://doi.org/10.1145/3582269.3615599

Lang, J., & Catrino, J. (2024). *How students should not use generative AI in the job search*. NACE Journal. https://www.naceweb.org/career-development/best-practices/how-students-should-not-use-generative-ai-in-the-job-search

Lin, Z. (2026). A validity-guided workflow for robust large language model research in psychology. *Behavior Research Methods*, *58*(8) Article 216. https://doi.org/10.3758/s13428-026-03073-2

Löckenhoff, C. E., Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., De Bolle, M., Costa, P. T., Jr., Sutin, A. R., Realo, A., Allik, J., Nakazato, K., Shimonaka, Y., Hřebíčková, M., Graf, S., Yik, M., Ficková, E., Brunner-

Sciarra, M., Leibovich de Figueora, N., Schmidt, V., … Terracciano, A. (2014). Gender stereotypes of personality: Universal and accurate? *Journal of Cross-Cultural Psychology*, *45*(5), 675–694. https://doi.org/10.1177/0022022113520075

Long, D., & Magerko, B. (2020). *What is AI literacy? Competencies and design considerations*. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

McCrae, R. R., & Costa, P. T. (2008). Empirical and theoretical status of the five-factor model of personality traits. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Vol. 1. Personality theories and models* (pp. 273–294). SAGE Publications. https://doi.org/10.4135/9781849200462.n13

Mujtaba, D. F., & Mahapatra, N. R. (2024). *Fairness in AI-driven recruitment: Challenges, metrics, methods, and future directions* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2405.19699

National Center for O*NET Development. (n.d.). *O*NET OnLine*. https://www.onetonline.org/

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2017). Interest congruence and performance: Revisiting recent meta-analytic findings. *Journal of Vocational Behavior*, *98*, 138–151. https://doi.org/10.1016/j.jvb.2016.11.002

OpenAI. (2026). *ChatGPT* [Large language model]. https://www/chatgpt.com

Pandya, S. S., & Wang, J. (2024). Artificial intelligence in career development: A scoping review. *Human Resource Development International*, *27*(3), 324–344. https://doi.org/10.1080/13678868.2024.2336881

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357. https://doi.org/10.2307/2087176

Sokanu Interactive Inc. (2025.). *CareerExplorer*. https://www.careerexplorer.com/

Tippins, N. T., & Hilton, M. L. (Eds.). (2010). *A database for a changing economy: Review of the Occupational Information Network* (O*NET). National Academies Press. https://doi.org/10.17226/12814

University of Colorado Boulder. (2025). *AI for career readiness*. https://www.colorado.edu/career/job-searching/ai-resource-guide

U.S. Bureau of Labor Statistics. (2023). *Women in the labor force: A databook* (Report No. 1103). U.S. Department of Labor. https://www.bls.gov/opub/reports/womens-databook/2022/

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology*, *2*, Article 178. https://doi.org/10.3389/fpsyg.2011.00178

Wilmot, M. P., & Ones, D. S. (2021). Occupational characteristics moderate personality–performance relations in major occupational groups. *Journal of Vocational Behavior*, *131*, Article 103655. https://doi.org/10.1016/j.jvb.2021.103655

Wilson, M., Robertson, P., Cruickshank, P., & Gkatzia, D. (2022). Opportunities and risks in the use of AI in career development practice. *Journal of the National Institute for Career Education and Counselling*, *48*(1), 48–57. https://doi.org/10.20856/jnicec.4807

Xiao, Y., & Zheng, L. (2025). Can ChatGPT boost students' employment confidence? A pioneering booster for career readiness. *Behavioral Sciences*, *15*(3), Article 362. https://doi.org/10.3390/bs15030362

Xue, J., Wang, Y.-C., Wei, C., Liu, X., Woo, J., & Kuo, C.-C. J. (2023). *Bias and fairness in chatbots: An overview* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2309.08836

CANADIAN JOURNAL OF CAREER DEVELOPMENT | REVUE CANADIENNE DE DÉVELOPPEMENT DE CARRIÈRE

Appendix

Supplementary Table 1

Distribution of Conscientiousness Ratings by Occupational Category by AI Models

Occupational Category	Total (*N*)	GPT-4 Ratings	Claude Ratings	Gemini Ratings
Clerical	9	High (9)	High (9)	High (9)
Customer Service	6	High (6)	High (6)	High (6)
Healthcare	5	High (5)	High (5)	High (5)
Law Enforcement	4	High (4)	High (4)	High (4)
Management	7	High (7)	High (7)	High (7)
Military	6	High (6)	High (6)	High (6)
Professional	19	High (19)	High (19)	High (19)
Sales	9	High (9)	High (9)	High (9)
Skilled/Semiskilled	27	High (14) / Med (13)	High (10) / Med (17)	High (16) / Med (11)
Total	92	High (79) / Med (13)	High (75) / Med (17)	High (81) / Med (11)

Note. $N = 92$. Occupations were grouped into broader functional categories based on the Standard Occupational Classification (SOC) system used in Wilmot and Ones (2019). The "Skilled/Semiskilled" category aggregates occupations from construction, production, transportation, and installation families. Values represent the number of occupations in each category receiving a "High" or "Medium" conscientiousness rating. No occupations received a "Low" rating from any model. Examples of occupations associated with high levels of Conscientiousness include army officers, doctors, accountants, and managers.

Supplementary Table 2

Distribution of Gender Dominance within Occupational Categories

Occupational Category	Total (*N*)	Male-Dominated	Female-Dominated	Balanced
Clerical	9	0	8	1
Customer Service	6	0	5	1
Healthcare	5	1	4	0
Law Enforcement	4	4	0	0
Management	7	6	0	1
Military	6	6	0	0
Professional	19	11	5	3
Sales	9	5	2	2
Skilled/Semiskilled	27	19	7	1
Total	92	52	31	9

Note. $N = 92$. Gender dominance categories are based on U.S. Bureau of Labor Statistics (2023) data for the respective occupations. "Male-Dominated" refers to occupations where men comprise >55% of the workforce; "Female-Dominated" refers to occupations where women comprise >55% of the workforce; and "Balanced" refers to occupations falling between these thresholds.