Retorio's Behavioral Model

Artificial Intelligence

Summary

You'll learn more about Retorio's Behavioral Model and its scientific foundation. Moreover, you'll understand what distinguishes Retorio's AI behavioral methodology and development and key validation criteria from other traditional self-rating approaches.

Theoretical background

Retorio focuses on an individual's displayed behavior to derive a personality profile of them. This particular approach is not unusual in scientific research, but it is novel in applied settings such as ours.

Do others know us better than we know ourselves?

Before we delve into understanding the difference between self-reporting and observer (peer) ratings, it's essential to discern how each rating measures the same concept (i.e., personality), yet focuses on different aspects. Self-reporting concentrates on how an individual measures their own internal dynamics, also known as forming one's identity. Observer ratings focus on peers.

One explanation of the difference between personal identity and peer opinion in personality evaluation is that peer evaluation is more shaped social behaviour and thus formed by how other's perceive an individual. Conversely, personal identity is comparatively shaped by own feelings and motives (Mount, Barrick, & Strauss, 1994).

Therefore self and peer ratings do not provide redundant information, but actually capture complementary aspects of an individual's personality (Vazire & Carlson, 2011).

Can other's judge my "outer" personality?

While self and peer ratings are two different aspects, one common question is how accurately others can evaluate an individual's external personality. Research shows people correctly predict extraversion only after 50 milliseconds exposure to a face (Borkenau, Brecke, Möttig, & Paelecke, 2009). For other Big-Five personality dimensions like Agreeableness, similar effects were found after watching a 20 second, silenced clip (Kogan et al., 2011). Even for less interpersonal personality dimensions, such as Conscientiousness, showing people short video clips (i.e., 30 seconds) was found to be enough to form Conscientiousness-related judgements, which had a predictive validity for job performance (Ambady, Krabbenhoft, & Hogan, 2006; Ambady & Rosenthal, 1993).

Peer rating, distinctive from self-reporting, presents accurate assessment of an individual and thus their performance in the workplace.

Self vs. peer ratings: What is more relevant in the workplace?

If self-reporting and peer ratings are not necessarily measuring the same aspects of personality, which measurement method is more relevant in the workplace?

In short, it depends on the question. In a workplace context, such as hiring and recruitment, it turns out peer ratings have an incremental predictive validity over self-reporting (Mount et al., 1994; Oh et al., 2011). For example, peer ratings from supervisors, co-workers, or customers could predict performance; they found that Conscientiousness and Extraversion assessed externally (e.g., supervisor) were valid predictors for job performance.

Other research corroborates these findings, showing observer ratings (vs. self-ratings) possess a stronger effect when predicting job performance. Moreover, research reveals adding in observer ratings to self-reporting when predicting job performance yielded substantial incremental validity (Oh et al., 2011).
Thus observer ratings tend to have a higher validity compared to self-reporting when predicting for future success.

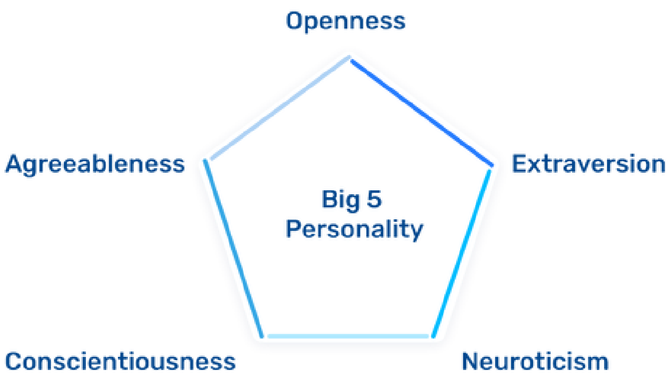
> **Research reveals adding in peer ratings to self-reporting when predicting job performance yielded substantial incremental validity**

Key take-aways:

1. Peer ratings focus on the behavioral aspect of personality
2. Peer ratings judge personality from observation
3. Peer ratings possess higher validity when predicting job-related outcomes, like performance

Methodology

Retorio's personality AI incorporates the aforementioned research by combining peer ratings and the Big Five personality concept into a single technology.



Retorio's personality model – The Big-5

Adjectives are used to describe a person's personality (Goldberg, 1992; McCrae & Costa, 1987). Clustered in dimensions, these adjectives represent a higher-order trait. Researchers distilled clusters into 5 traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Emotional Stability). As the most commonly-used models in the psychology community, these 5 traits are known as the Big 5 or OCEAN Model.

These traits possess descriptive sub-dimensions. For example, Openness has sub-dimensions of intellectual curiosity, aesthetic interest, and creativity imagination. Conscientiousness measures a person's behaviour in terms of achievement striving, impulse control, and industriousness. Extraversion describes a person's social, energetic, and assertiveness. Agreeableness captures a person's compassionate, respectful, and trusting behaviour (Soto & John, 2017). Extraversion and Agreeableness are particularly used to describe interpersonal behaviour while Conscientiousness and Neuroticism are used to describe intrapersonal behaviour.

In summary, adjectives can be used to describe people's personality and classify it along a specific taxonomy.

Big 5-dimensions

Openness

Those who score high on this trait tend to be intellectually curious, willing to try new things, and are more creative or unconventional. Those who score low on this trait usually have an especially difficult time to adapt to change and abstract thought.

Conscientiousness

This highlights how well a person aligns themselves with responsibility, organization, and goal-setting. It comprises self-control and showcases how they may deliberate over choices. Those who score low on this trait tend to be more spontaneous, flexible, or unreliable.

Extraversion

The spectrum of extraversion-introversion describes how individuals derive pleasure and receive energy. The more introverted, the greater the likelihood the person receives more enjoyment from their inner life than by social events. Introverts are more intrigued with the world of ideas and thus tend to be a bit more cerebral and reflective than extraverts. Extroverts gain energy from being around others and taking part in a wide-variety of activities. No one is purely extroverted or introverted, but rather lies somewhere on the spectrum.

Agreeableness

A person with higher levels of this trait exhibits greater amounts of prosocial behaviour such as cooperation, friendliness, and politeness. They possess the ability for substantial empathy and tend to be concerned about others. They tend to avoid conflict and do not easily project negative emotions.

Neuroticism

Individuals who score high on this particular trait tend to experience negative or emotionally-anxious states. They wrestle with feelings of anxiety, depression, guilt or loneliness—more so than those who score low. Neuroticism is a long-term emotional state that may make everyday situations seem more challenging.

Dataset and data collection

We combine the approach of trait taxonomy (i.e., Big-5) and observer ratings to assess individuals via short, video clips. Observers assessed people in video clips along the Big-5 (or OCEAN) taxonomy. In total, we used more than 2,500 assessors from five continents. The individuals in the video clips were also divided equally in regards to sex, ethnicity, and age. To promote objectivity, multiple ratings per video were obtained. The overall dataset consists of more than 12,000 people.
With this particular approach, research shows expertise is not needed to assess others, but rather it depends on a validated and solid scientific concept (Kolar, Funder, & Colvin, 1996).

Data exploration

We scrutinized the Big Five ratings given by the assessors for any systematic biases. For example, we compared means of Extraversion across Caucasian and African-Americans. If we detected mean differences that were due to the membership of a group, we adjusted the mean to the respective difference to cancel out discriminatory biases in the training and testing sets.

Evaluation

Prediction accuracy

On average, our accuracy determines how far away our estimations are from the actual value a group of humans would have given an individual. Thereby, we reach a 90% accuracy. This means that when trying to predict the value of all human assessors, we have on average 10% deviation. Thus, it may happen that we do not predict a 3, but rather a 2.7 or 3.3.

Given the elusive nature of the topic itself (i.e., personality) and the fact that there is no “natural” baseline as comparison standard, we’re impressively close to what is considered---according to the majority of people---the personality of an individual we’ve not met before.

Reliability

We calculated the relative consistency and agreements in ratings provided by human assessors and our own AI. Below, we address the question whether our AI ranks rated people in a manner that is relatively consistent with human assessors. Additionally, we focus on the question of whether human scores and those from our AI are interchangeable or equivalent in their absolute value.
Distinguishing between human and AI assessment, our internal calculations yielded an intraclass correlation coefficient (ICC) range of [.53; .62]. For the sake of mitigating bias, it’s important to note original, human-assessed labels have been adapted (cf. Data exploration) and thus affect the coefficient.

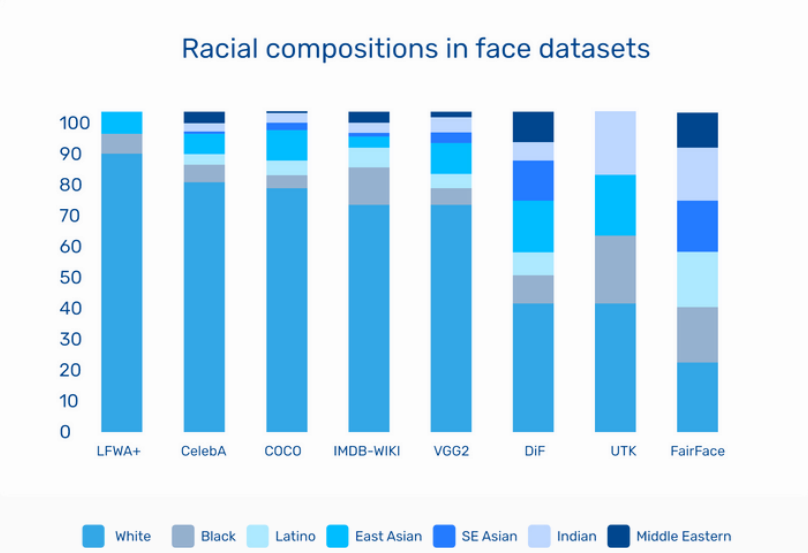
Thus, a higher ICC would indicate that stereotypical assessments in the peer rating procedure are perpetuated and transferred in our AI. Given that we do not want to reach a perfect agreement between initial human ratings, a moderate ICC is perfect.

Demographic neutral

Gauge criteria – Baseline comparison dataset

To demonstrate fairness in the assessment of our personality AI model, we evaluated our AI on a newly published dataset: The FairFace Dataset (https://arxiv.org/abs/1908.04913). This dataset consists of pictures from over 100,000 people distributed among 7 ethnicities: White, Black, Indian, East Asian, Southeast Asian, Middle East, and Latino; 9 age-groups ranging from 2 to over 70 years, and 2 sexes.

The dataset has been created to ensure distribution of demographics within a dataset is equally distributed when training AI models for classification tasks. The figure below from the paper depicts the demographic distribution pertaining ethnicity compared to other existing face datasets.



Results when predicting personality

Table 1 shows the results of our personality-AI when predicting the Big-5 for different groups of ethnicities aggregated over age groups and sexes. **Table 2** shows a deeper breakdown between sexes across different ethnicities.

It's worth noting that we found statistically significant differences between groups. Given the large amount of data it is highly likely that significance levels, defined by p-values, reach significance. However, when examining effect sizes for those differences they did not show any effects.

Table 1. *Showing mean values for the Big-5 personality prediction across ethnicities*

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
East Asian	0,5435	0,4584	0,5299	0,5007	0,4825
Indian	0,5425	0,4595	0,5283	0,5038	0,4793
Black	0,5477	0,4668	0,5332	0,5043	0,4813
White	0,5382	0,4558	0,5276	0,5022	0,4841
Middle Eastern	0,5443	0,4633	0,5346	0,5099	0,4786
SE Asian	0,5442	0,4598	0,5315	0,5062	0,4785
Latino/Hispanic	0,5456	0,4620	0,5365	0,5095	0,4789

Mean values in each column aggregated within ethnicity over age groups, and sexes

In **Table 3** we examined the average values aggregated across all groups. The range values indicate the average span across all groups. For example the range of 0,03 in Openness indicates that the average value in one of the subgroups was between 0,53 and 0,56. The mean values highlight the average value aggregated across all groups together whereas the Std. Deviation shows the variation around the mean.

Key take-aways:

We demonstrate our personality AI evaluates individuals in a novel approach, regardless of colour, age, or sex.

Table 2. *Showing mean values for Big-5 personality predictions across ethnicities and sexes*

		Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
East Asian						
	Female	0,5494	0,4558	0,5342	0,4998	0,4830
	Male	0,5375	0,4610	0,5256	0,5016	0,4820
Indian						
	Female	0,5472	0,4581	0,5307	0,5046	0,4790
	Male	0,5379	0,4608	0,5259	0,5030	0,4796
Black						
	Female	0,5503	0,4625	0,5350	0,5025	0,4834
	Male	0,5450	0,4711	0,5314	0,5062	0,4792
White						
	Female	0,5423	0,4501	0,5308	0,5006	0,4862
	Male	0,5341	0,4614	0,5244	0,5037	0,4820
Middle Eastern						
	Female	0,5473	0,4584	0,5368	0,5082	0,4837
	Male	0,5414	0,4682	0,5325	0,5117	0,4735
SE Asian						
	Female	0,5493	0,4572	0,5347	0,5067	0,4780
	Male	0,5391	0,4623	0,5283	0,5056	0,4790
Latino/Hispanic						
	Female	0,5505	0,4562	0,5401	0,5077	0,4811
	Male	0,5406	0,4678	0,5330	0,5114	0,4767

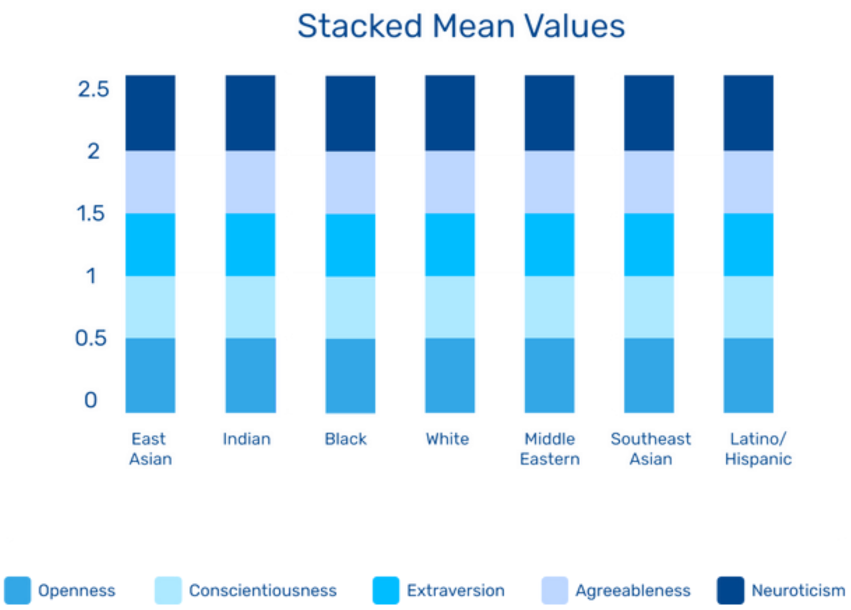


Table 3. *Depicting descriptive differences across groups*

	Range	Minimum	Maximum	Mean	Std. Dev
Openness	0,03	0,53	0,56	0,5437	0,00633
Conscientiousness	0,03	0,45	0,48	0,4608	0,00679
Extraversion	0,03	0,52	0,55	0,5317	0,00660
Agreeableness	0,02	0,49	0,52	0,5052	0,00488
Neuroticism	0,02	0,47	0,49	0,4805	0,00536

N = 70 groups; 2 (sexes: male, female) x 5 age groups (10-19, 20-29, 30-39, 40-49, 50-59) x 7 ethnicities (East Asian, Indian, Black, White, Middle Eastern, Southeast Asian, Latino/Hispanic)

Major summary

In this outline, the goal was to elucidate Retorio's theoretical framework for its personality AI, how the AI has been trained, and how we evaluate its accuracy.

We were able to show that we have:

- 90% accuracy when predicting people's personality
- A good agreement between how humans and our AI assess people
- Demographic fair assessments across 70 different groups

We wish to emphasize we do not claim our personality AI is perfect nor comprehensive. However, given the use cases for its applications, it remains one of the most sophisticated technologies available.

Our goal is to provide value and reduce risks when making decisions in situations where the objective truth can not be measured (e.g., a scale which measures weight) and assessments are rather subjective (i.e., single judgements about someone's personality). We believe in a technology-supported approach, where technology assists humans make better decisions.

References

Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology, 16*(1), 4-13.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*(3), 431-441.

Borkenau, P., Brecke, S., Möttig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality, 43*(4), 703-706.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*(1), 26-42.

Kogan, A., Saslow, L. R., Impett, E. A., Oveis, C., Keltner, D., & Saturn, S. R. (2011). Thin-slicing study of the oxytocin receptor (OXTR) gene and the evaluation and expression of the prosocial disposition. *Proceedings of the National Academy of Sciences, 108*(48), 19189-19192.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality, 64*(2), 311-337.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81–90.

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology, 79*(2), 272-280.

Oh, I.-S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology, 96*(4), 762–773.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality, 68*, 69–81.

Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science, 20*(2), 104-108.