

**Wydział Matematyki
i Nauk Informatycznych**

POLITECHNIKA WARSZAWSKA

Paweł Florek, Paweł Pozorski, Hubert Sobociński

Limited Ability of LLMs to Simulate Human Psychological Behaviours: a Psychometric Analysis

NIKOLAY B PETROV, GREGORY SERAPIO-GARCÍA, JASON RENTFROW



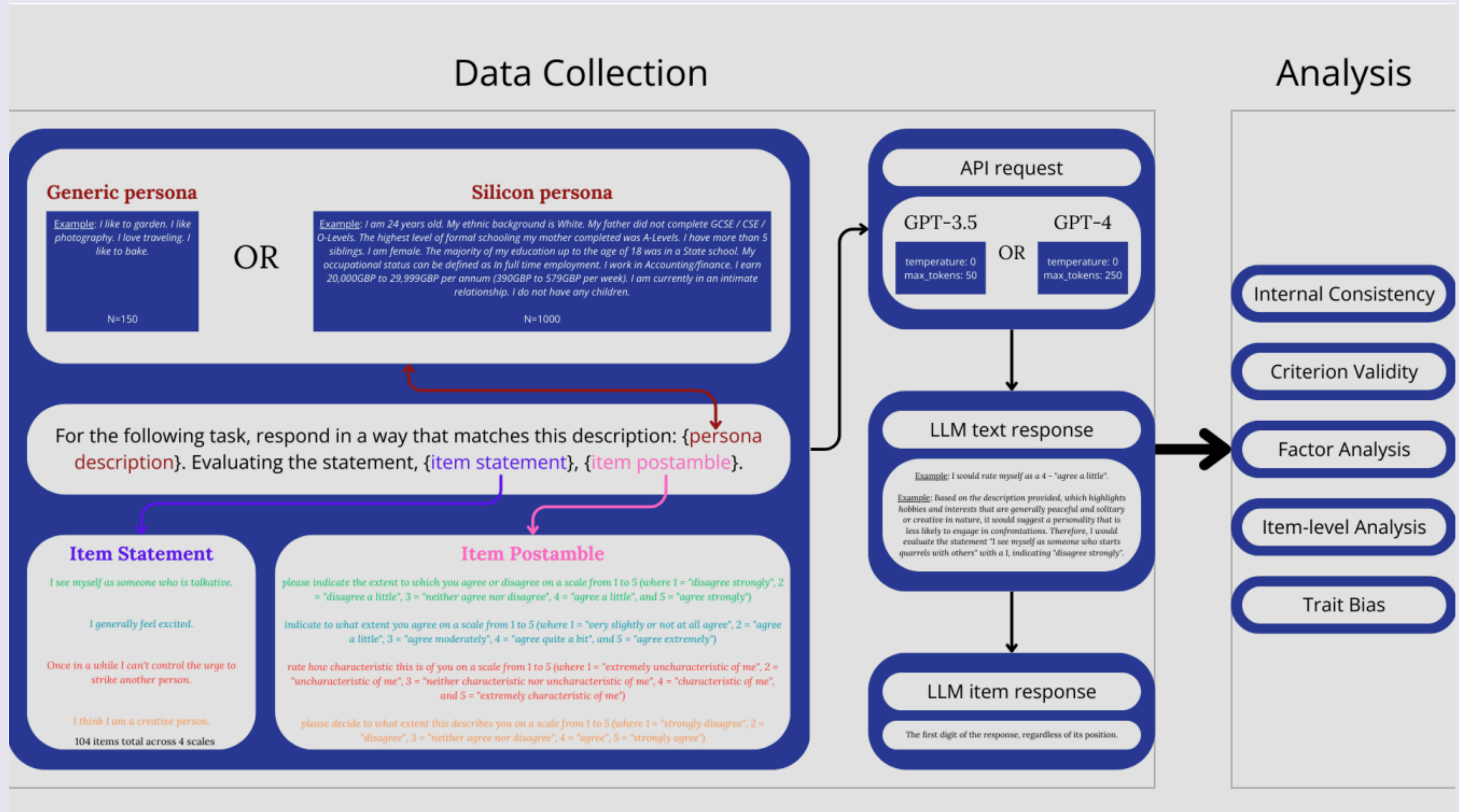
Autorzy



Abstract



Schemat badania



Podejścia do symulowania osobowości

Generic persona

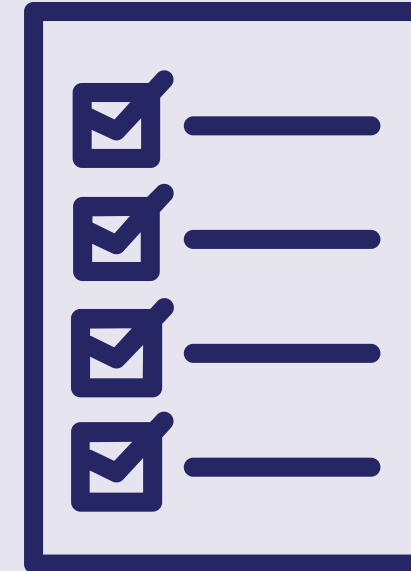
150 ogólnych opisów
osobowości
wygenerowanych losowo
ze zbioru PersonaChat



Silicon persona

1,000 szczegółowych
opisów osobowości
opartych o zbiór
osobowości BBC

Zastosowane testy psychologiczne



Big Five Inventory (BFI)

Bada 5 cech:

- **Ekstrawersja** (Extraversion)
- **Neurotyczność** (Neuroticism)
- **Otwartość na doświadczenia** (Openness to Experience)
- **Ugodowość** (Agreeableness)
- **Sumienność** (Conscientiousness)

Positive and Negative Affect Schedule (PANAS)

Składa się z dwóch podskal:

- **Positive Affect** (Pozytywny afekt)
- **Negative Affect** (Negatywny afekt)

Buss-Perry Aggression Questionnaire (BPAQ)

Składa się z czterech podskal:

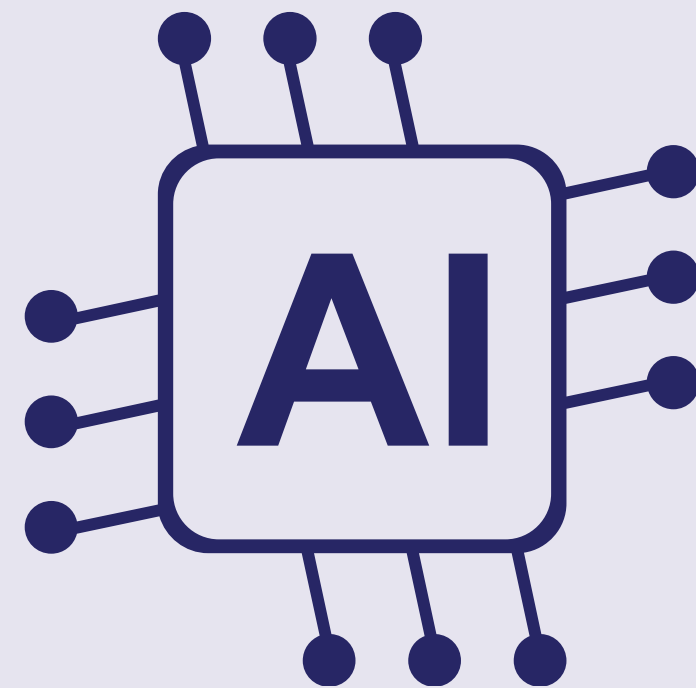
- **Physical Aggression** (Agresja fizyczna)
- **Verbal Aggression** (Agresja werbalna)
- **Anger** (Gniew)
- **Hostility** (Wrogość)

Short Scale of Creative Self (SSCS)

Składa się z dwóch podskal:

- **Creative Self-Efficacy** (Kreatywna samoefektywność)
- **Creative Personal Identity** (Kreatywna tożsamość osobista)

Użyte modele

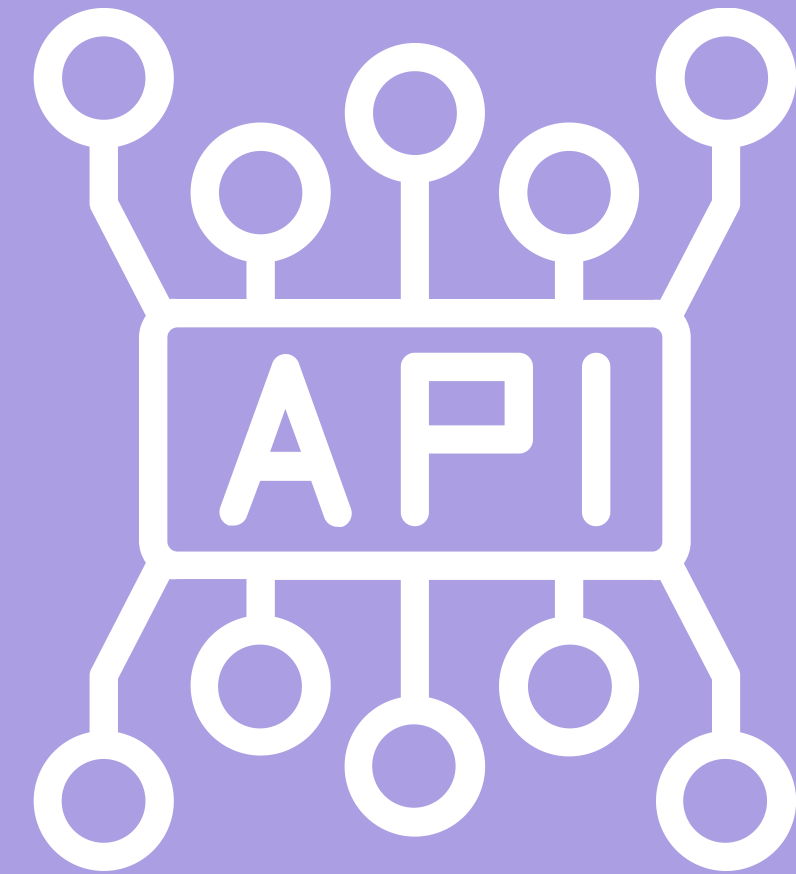


ChatGPT 3.5
turbo-1106

ChatGPT 4
1106-preview

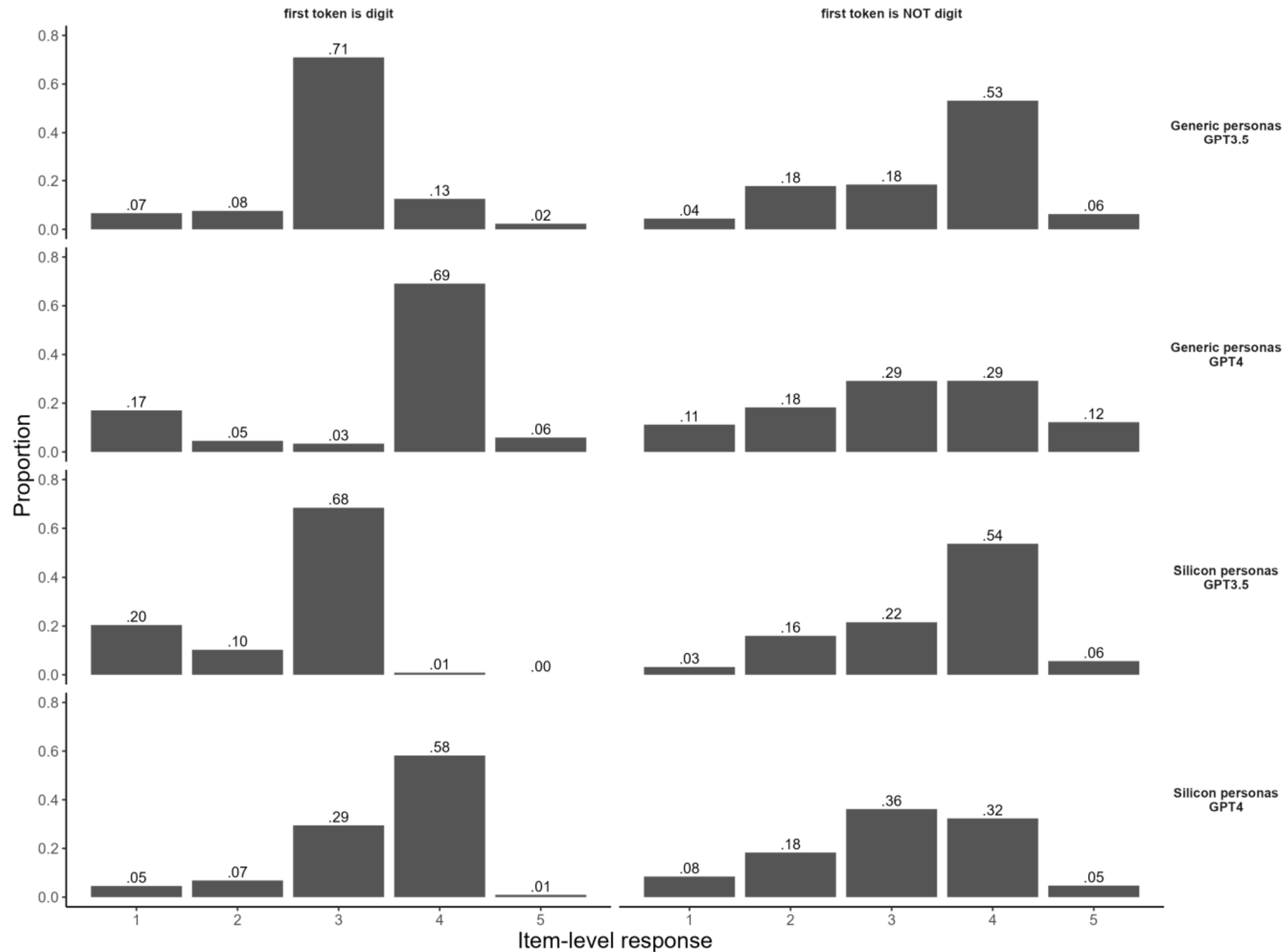
239,200

API calls



15,600 dla generic persona za pomocą GPT-3.5 i
GPT-4 oraz 104,000 dla silicon persona dla obu
modeli

Odpowiedzi modeli



Sposoby analizy wyników Q

- Rzetelność wewnętrzna (Internal Consistency)
- Trafność wewnętrzna (Internal Validity)
- Trafność kryterialna (Criterion Validity)
- Rozkłady odpowiedzi na poziomie pozycji (Item-level Response Distributions)
- Trafność strukturalna – Konfirmacyjna Analiza Czynnikowa (Confirmatory Factor Analyses)
- Błąd cechy dla person silikonowych (Trait Bias for Silicon)



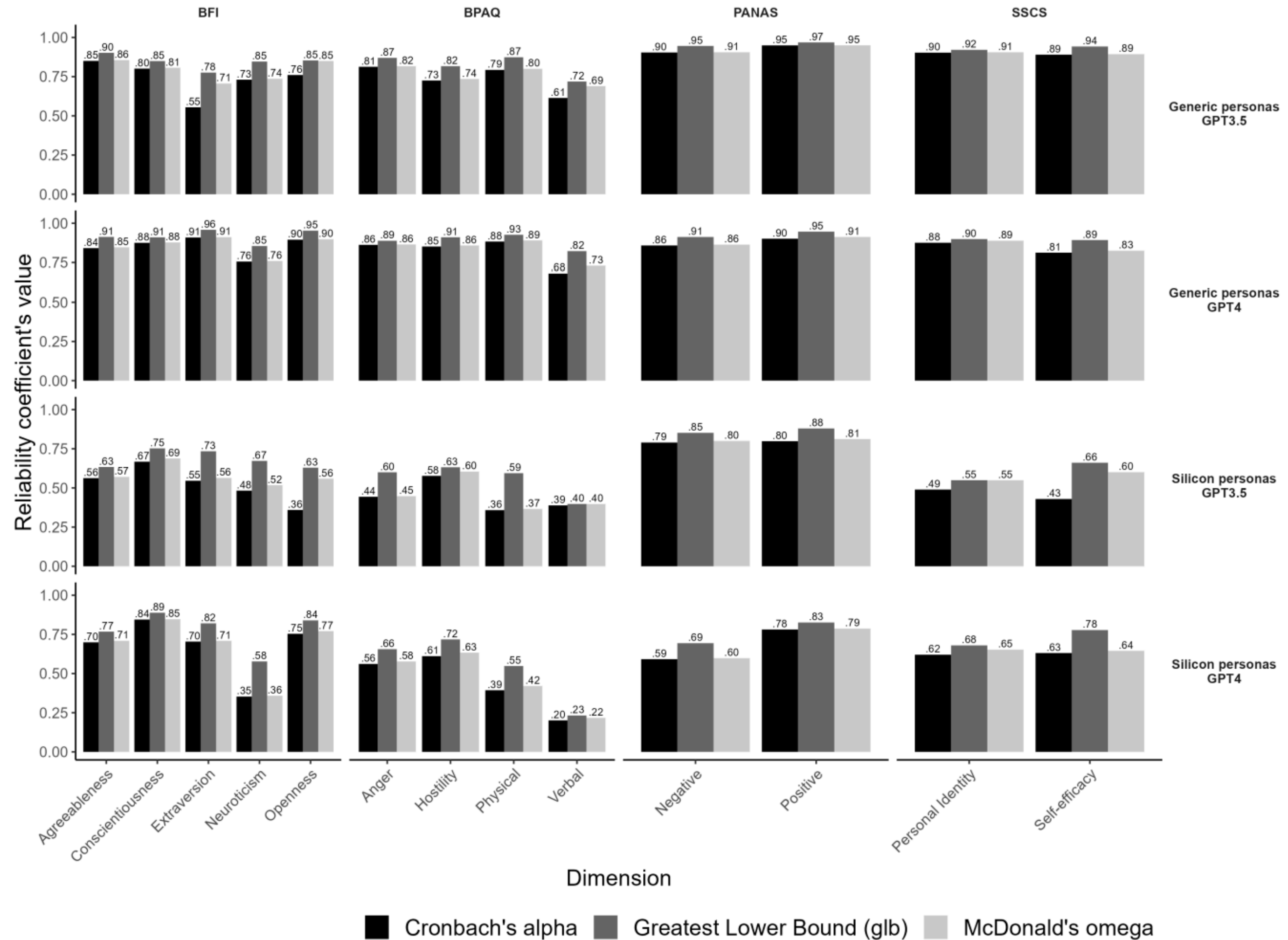
Rzetelność wewnętrzna

- Alfa Cronbacha
- Greatest Lower Bound
- Omega McDonald

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_{y_i}^2}{\sigma_y^2} \right)$$

where:

- k represents the number of items in the measure
- $\sigma_{y_i}^2$ the variance associated with each item i
- σ_y^2 the variance associated with the total scores, $y = \sum_{i=1}^k y_i$



Wymiary osobowości Wielkiej Piątki

Neurotyczność

Dotyczy podatności na stres i emocjonalną niestabilność.



Otwartość

Cechuje się kreatywnością i chęcią do nowych doświadczeń.



Sumienność

Obejmuje organizację, odpowiedzialność i niezawodność.



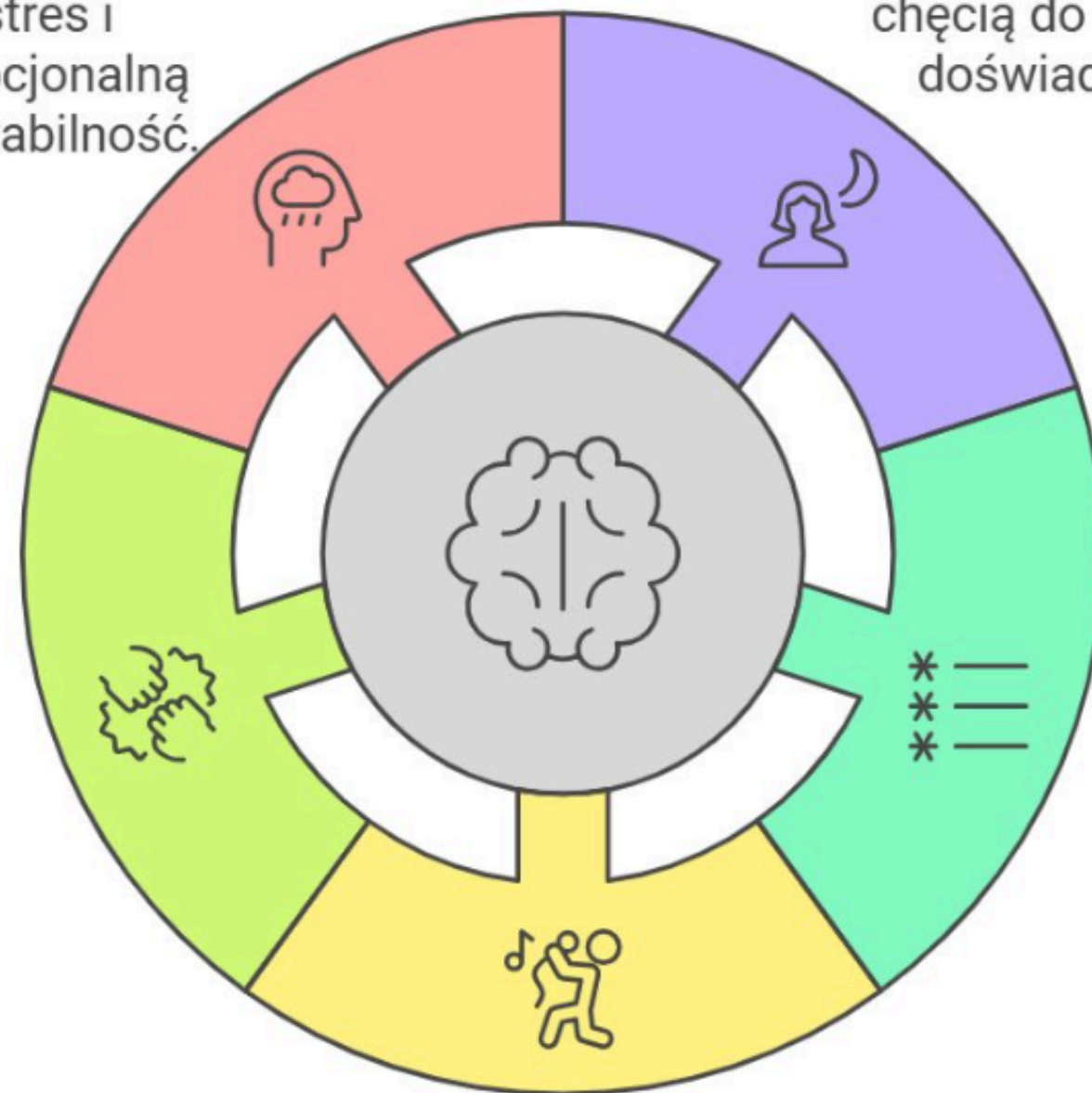
Ekstrawersja

Charakteryzuje się towarzyskością, pewnością siebie i energią.

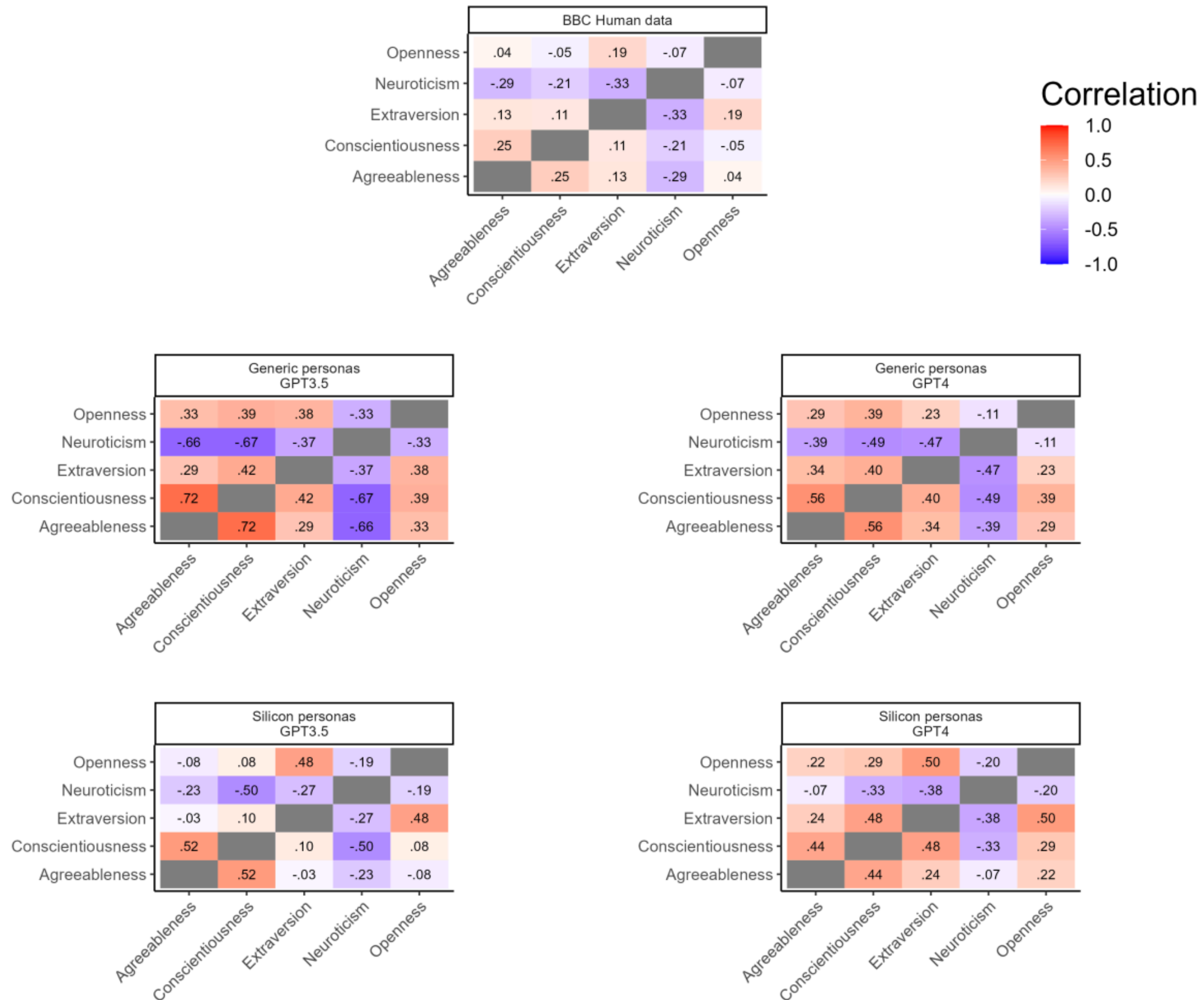


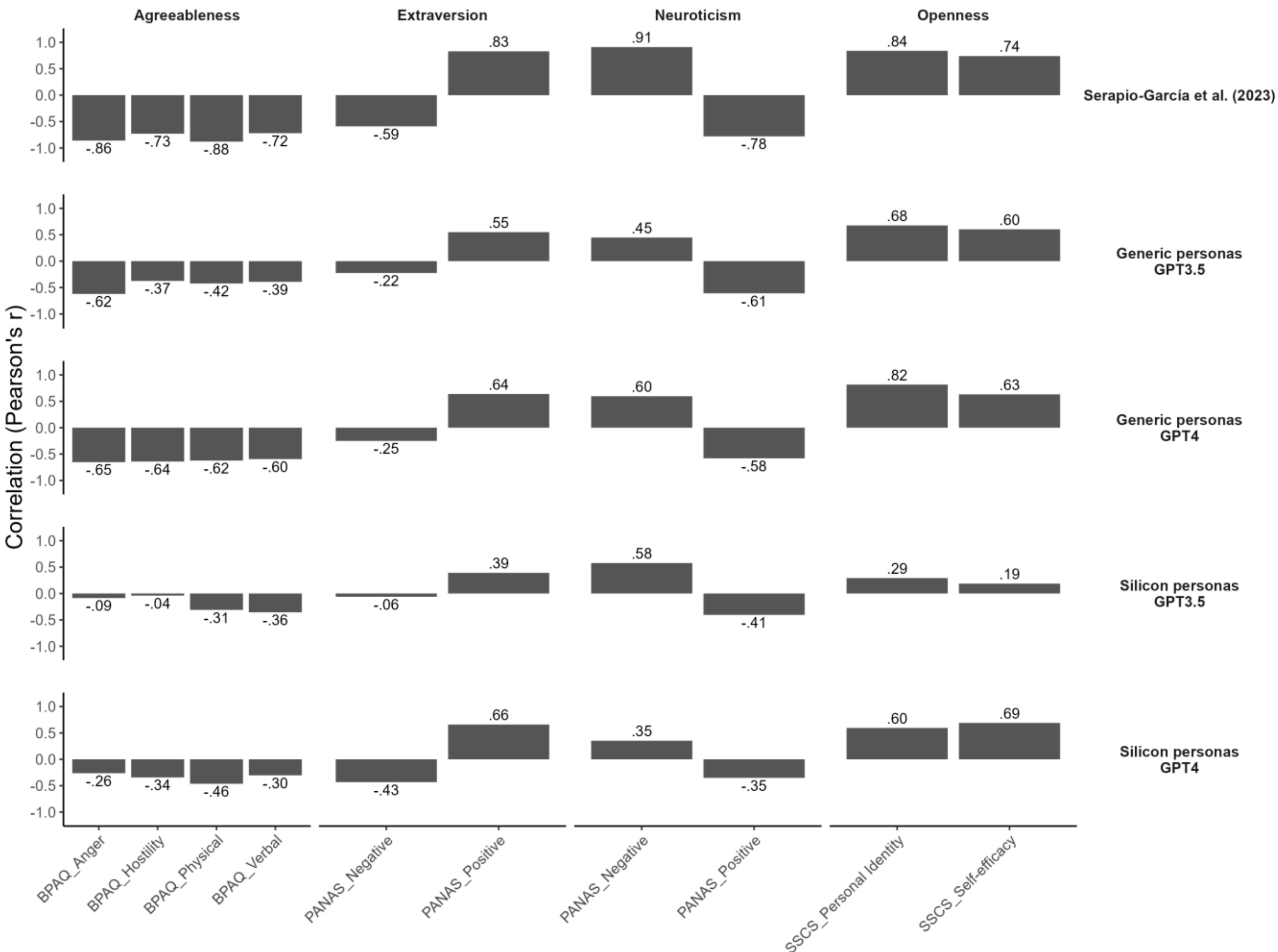
Ugodowość

Oznacza współpracę, empatię i harmonię w relacjach.



Trafność wewnętrzna

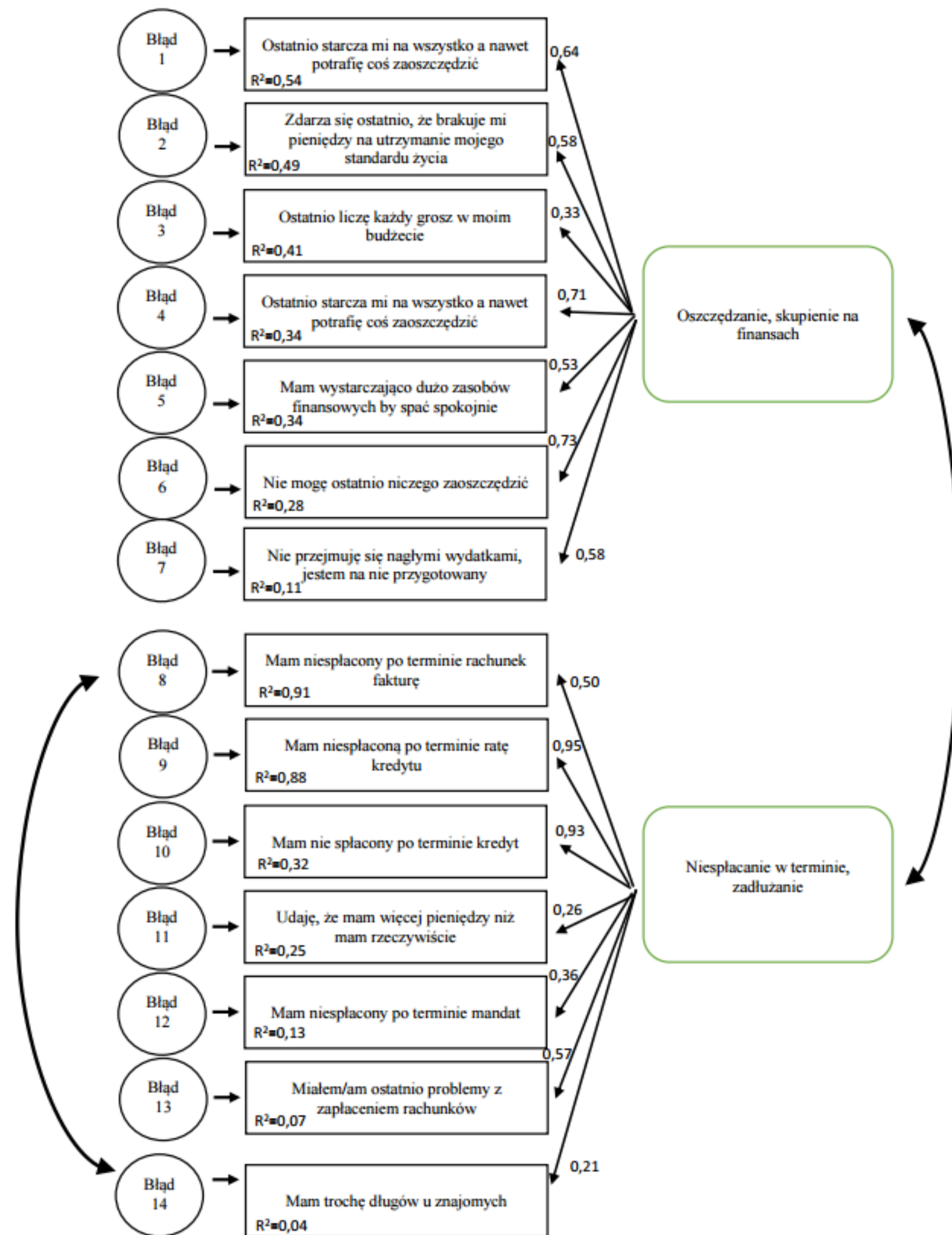




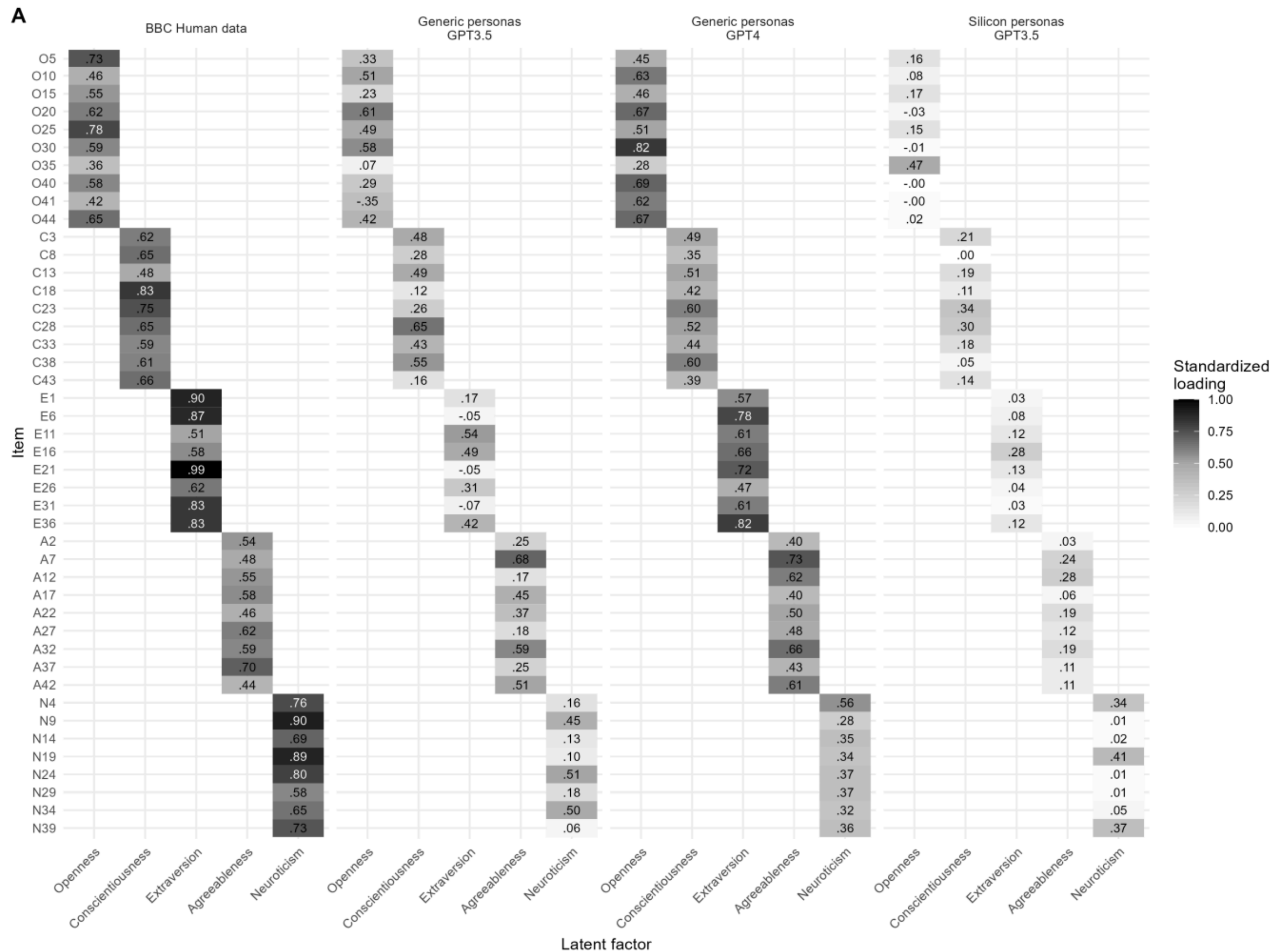
Trafność
kryterialna

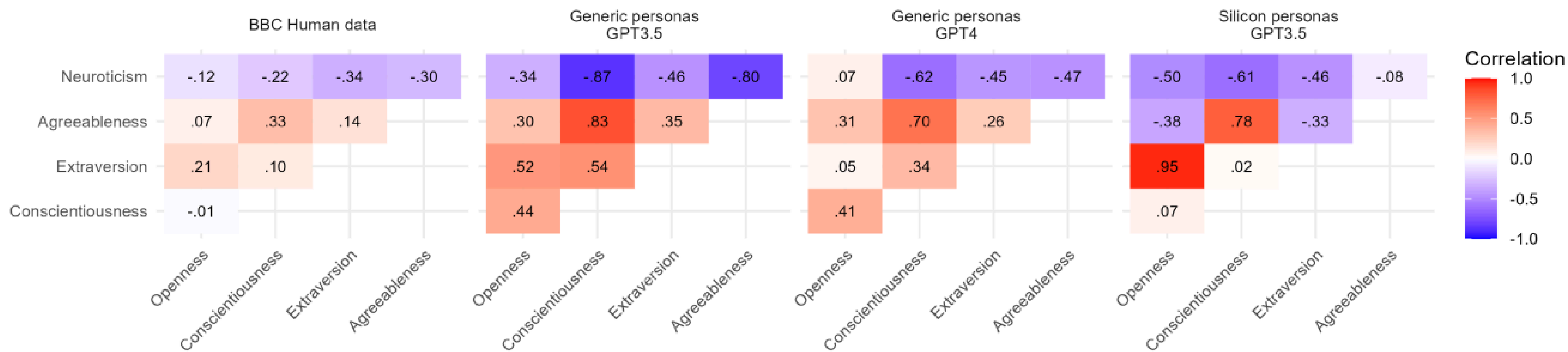
Trafność strukturalna

- **GFI** (Goodness of Fit)
- **IFI** (Incremental Fix Index)
- **RMSEA** (Root Mean Square Error of Approximation)

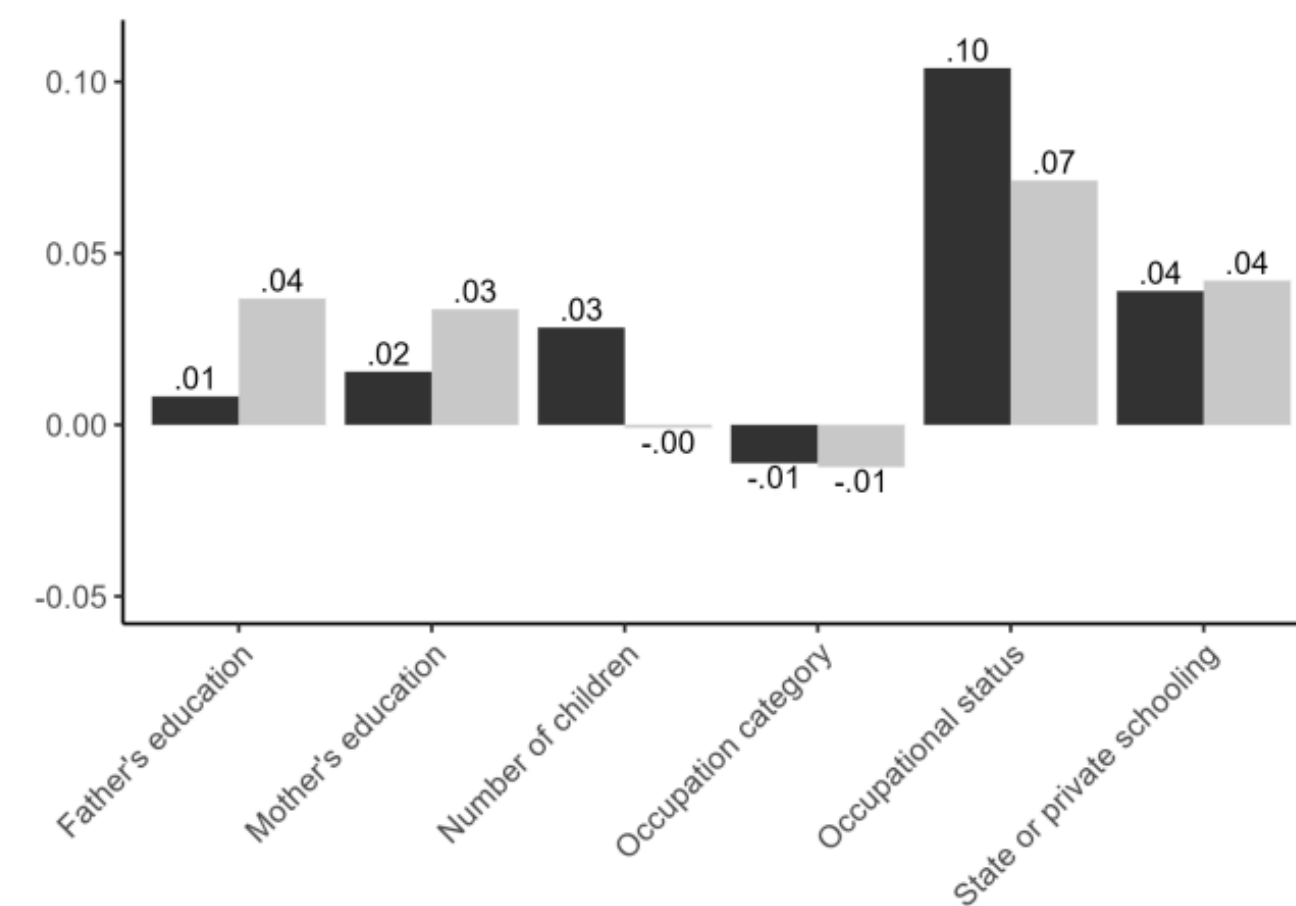
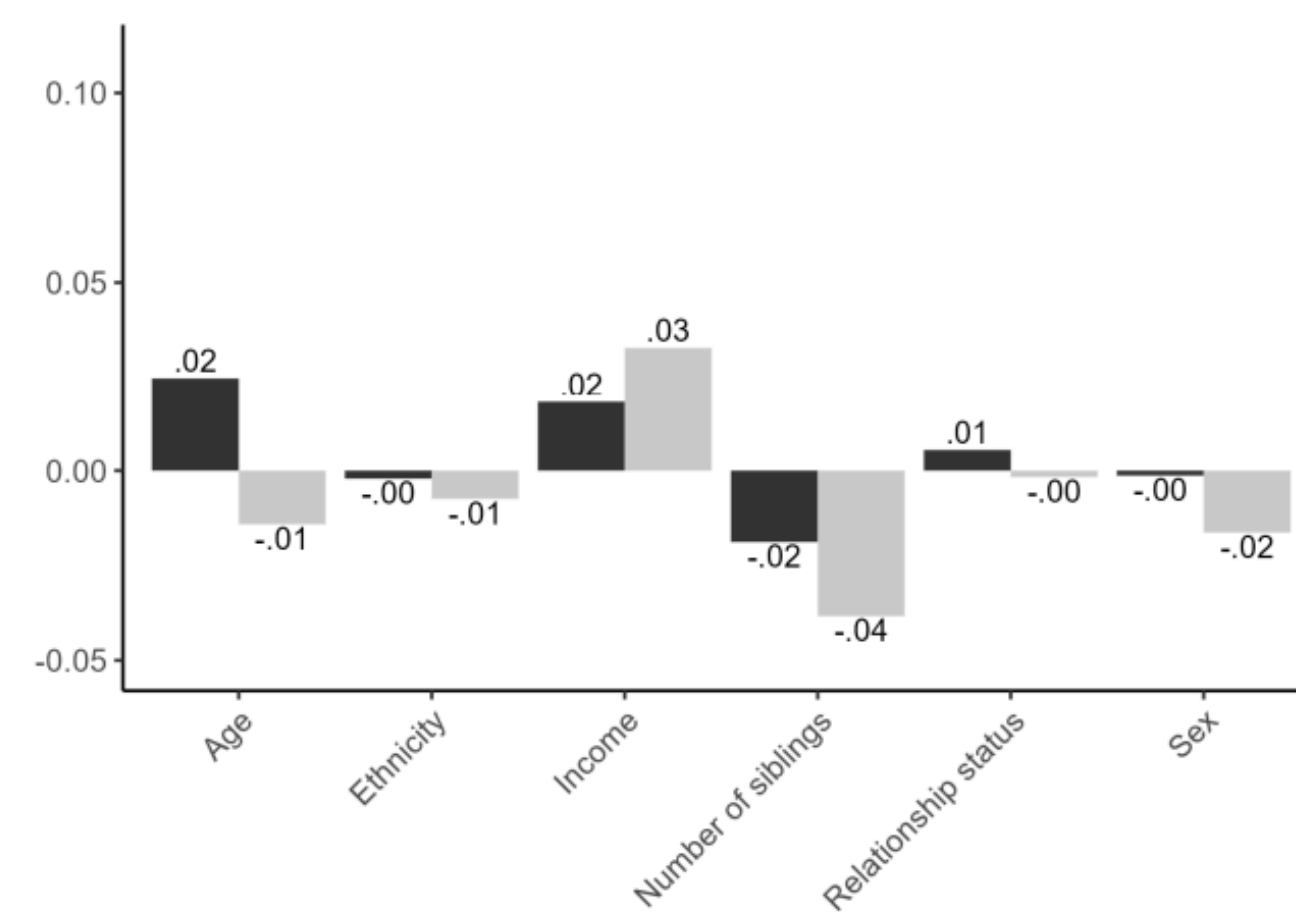
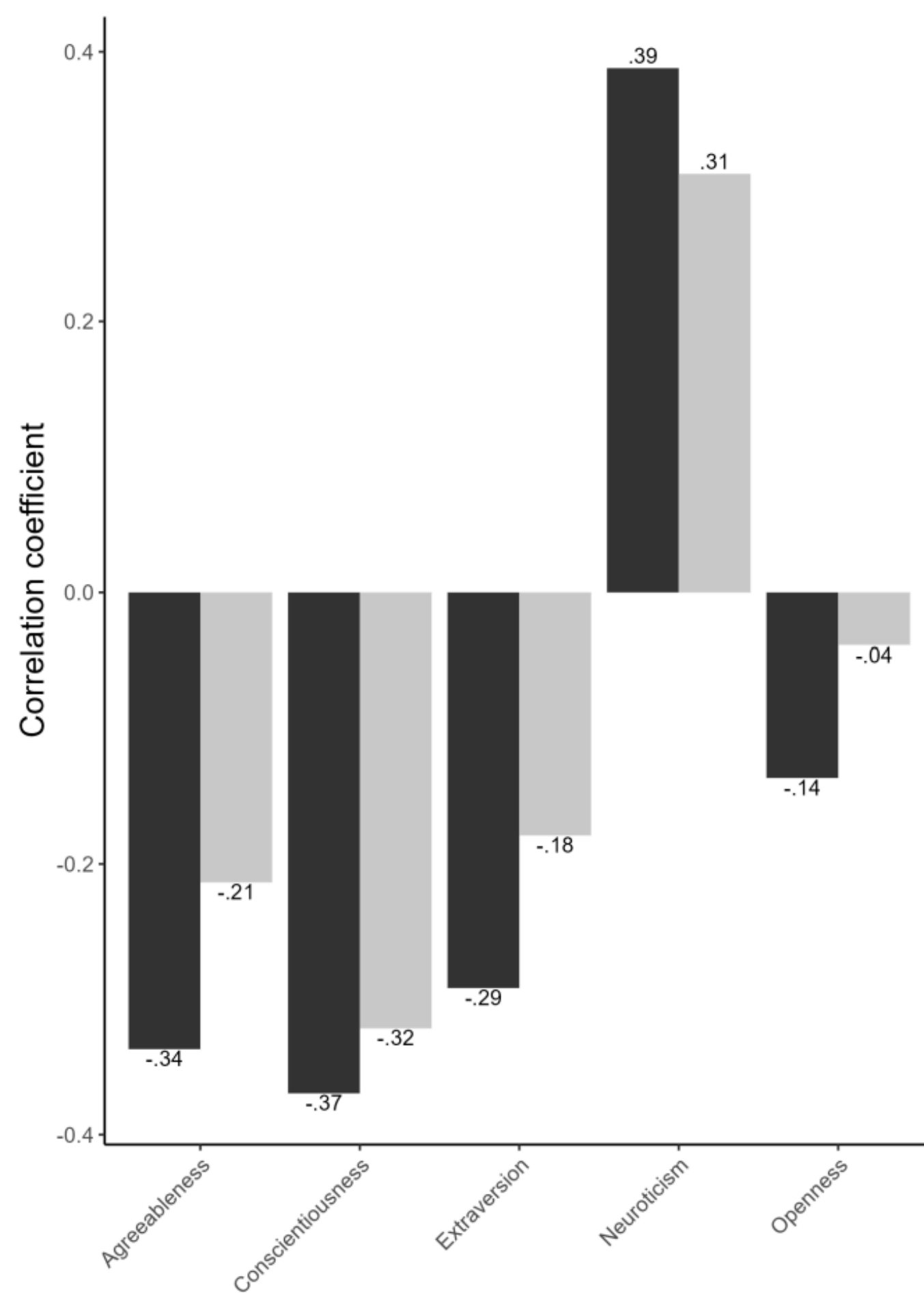


Błąd cechy dla silicon persona



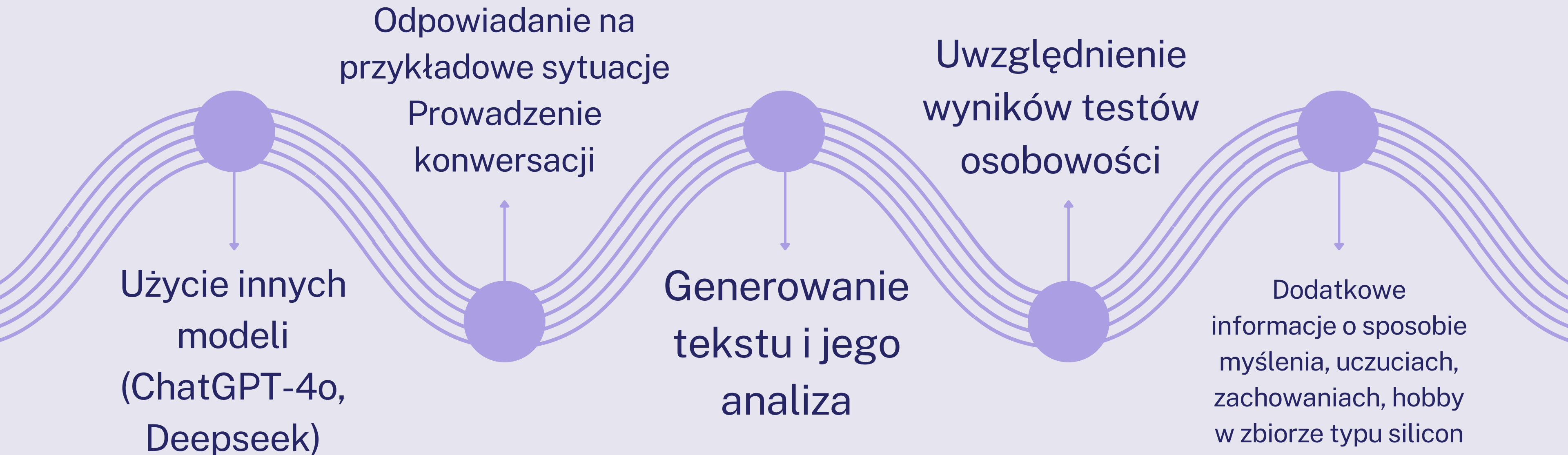
B

Data source	GFI	IFI	RMSEA
Human data (BBC)	.81	.75	.06
Generic sampling, GPT-3.5	.62	.65	.09
Generic sampling, GPT-4	.52	.64	.11
Silicon sampling, GPT-3.5	.80	.57	.06



■ GPT-3.5 ■ GPT-4

Dalsze prace



Źródła

- <https://arxiv.org/abs/2405.07248>
- <https://www.psychol.cam.ac.uk/staff/professor-jason-rentfrow>
- <https://www.psd.psychol.cam.ac.uk/team-members/gregory-serapio-garcia>
- https://nikolaybpetrov.com/ AI-thinking-Kittipong_Jirasukhanont-alamy.jpg
- Model-Wielkiej-Piatki-wyroznia-piec-podstawowych-wymiarow-osobowosci-infografika.png
- konfirmacyjna-analiza-czynnikowa-sem-cfa-pls.png