

POLITECHNIKA WARSZAWSKA

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

NIKOLAY B PETROV, GREGORY SERAPIO-GARCÍA, JASON RENTFROW



Autorzy A







Abstract ?



Schemat badań

Data Collection

"disagree", 3 = "neither agree nor disagree", 4 = "agree", 5 = "strongly agree")

Generic persona Silicon persona Example: I like to garden. I like Example: I am 24 years old. My ethnic background is White. My father did not complete GCSE / CSE / photography. I love traveling. I O-Levels. The highest level of formal schooling my mother completed was A-Levels. I have more than 5 OR like to bake. siblings. I am female. The majority of my education up to the age of 18 was in a State school. My occupational status can be defined as In full time employment. I work in Accounting/finance. I earn 20,000GBP to 29,999GBP per annum (390GBP to 579GBP per week). I am currently in an intimate relationship. I do not have any children. N=150 N=1000 For the following task, respond in a way that matches this description: {persona description). Evaluating the statement, {item statement}, {item postamble}. **Item Statement Item Postamble** I see myself as someone who is talkative. please indicate the extent to which you agree or disagree on a scale from 1 to 5 (where 1 = "disagree strongly", 2 = "disagree a little", 3 = "neither agree nor disagree", 4 = "agree a little", and 5 = "agree strongly") I generally feel excited. indicate to what extent you agree on a scale from 1 to 5 (where 1 = "very slightly or not at all agree", 2 = "agree a little", 3 = "agree moderately", 4 = "agree quite a bit", and 5 = "agree extremely") Once in a while I can't control the urge to rate how characteristic this is of you on a scale from 1 to 5 (where 1 = "extremely uncharacteristic of me", 2 = strike another person. "uncharacteristic of me", 3 = "neither characteristic nor uncharacteristic of me", 4 = "characteristic of me", and 5 = "extremely characteristic of me")

104 items total across 4 scales

API request **GPT-3.5** GPT-4 OR temperature: 0 temperature: 0 max_tokens: 250 max_tokens: 50 Internal Consistency Criterion Validity LLM text response **Factor Analysis** Example: I would rate myself as a 4 - "agree a little". Example: Based on the description provided, which highlights hobbies and interests that are generally peaceful and solitary or creative in nature, it would suggest a personality that is Item-level Analysis less likely to engage in confrontations. Therefore, I would evaluate the statement 'I see myself as someone who starts marrels with others" with a I, indicating "disagree strongly". **Trait Bias** LLM item response The first digit of the response, regardless of its position.

Analysis

Podejścia do symulowania osobowości

Generic persona

150 ogólnych opisów osobowości wygenerwoanych 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)



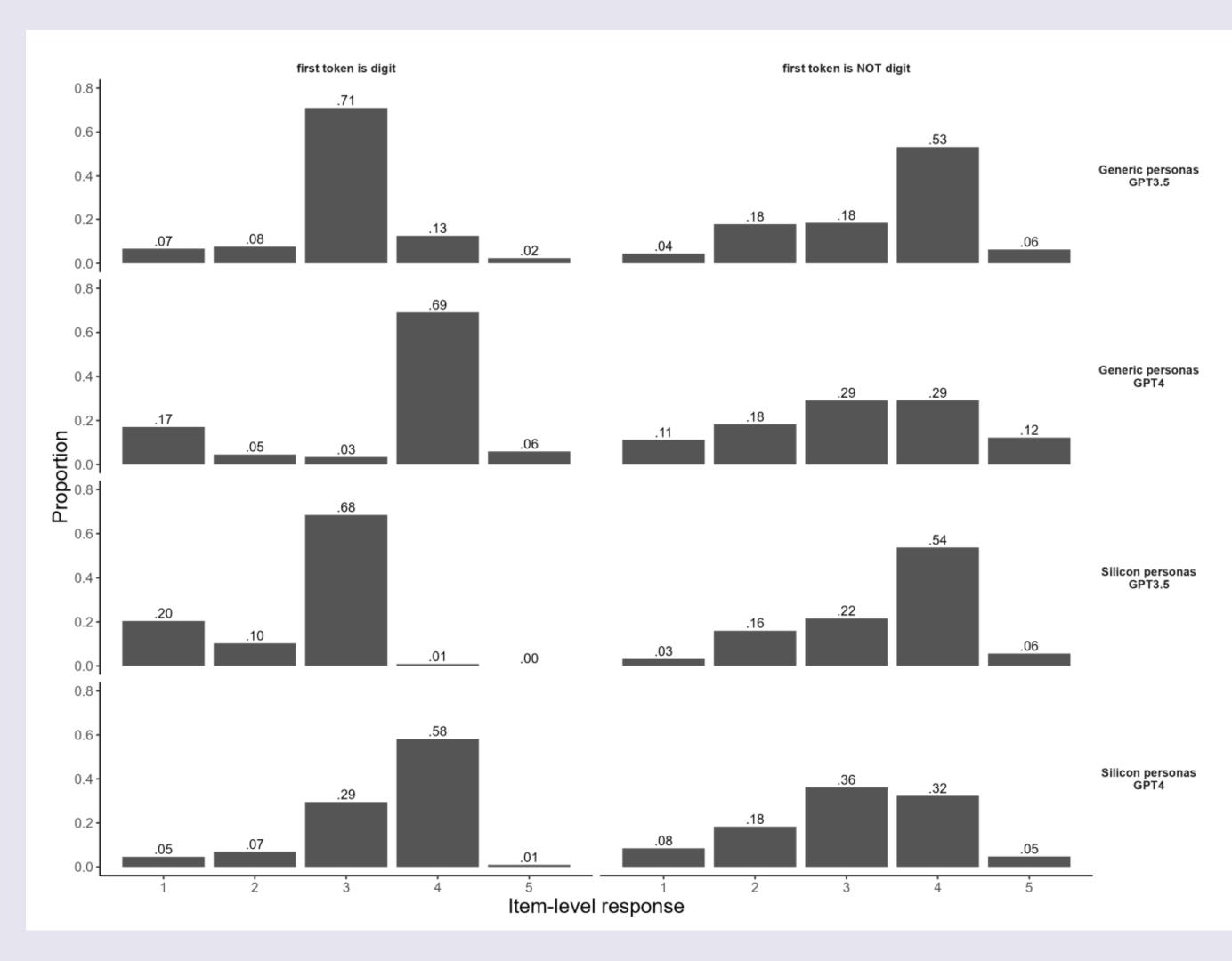
ChatGPT 3.5 turbo-1106

ChatGPT 4 1106-preview

239,200 AFI 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



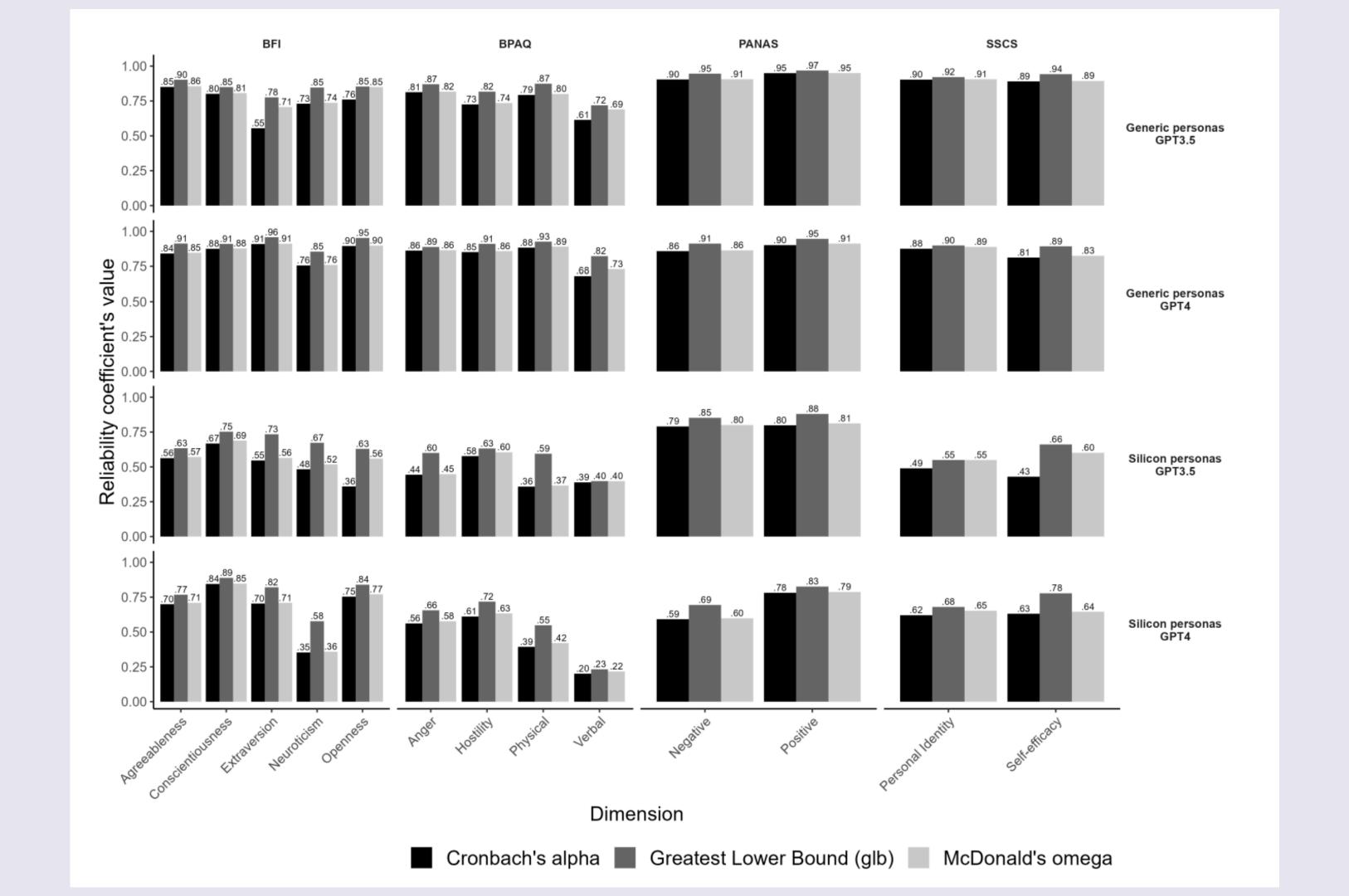
Rzetelność wewnętrzna

- Alfa Cronbacha
- Greates Lower Bound
- Omega McDonalda

$$lpha = rac{k}{k-1} \left(1 - rac{\sum_{i=1}^k \sigma_{y_i}^2}{\sigma_y^2}
ight)$$

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, $y = \sum_{i=1}^n y_i$



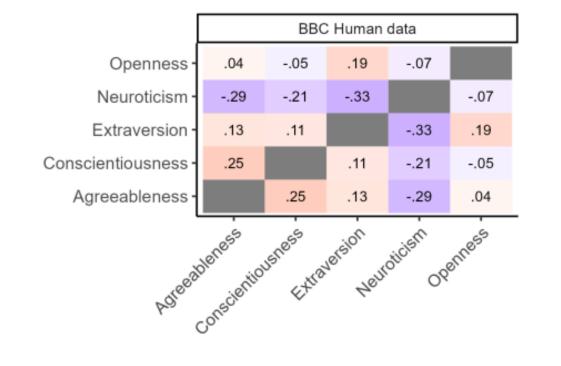
Wymiary osobowości Wielkiej Piątki

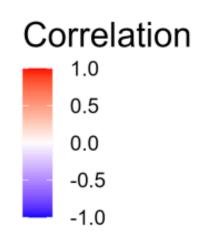
Neurotyczność Otwartość Cechuje się Dotyczy podatności na kreatywnością i chęcią do nowych stres i emocjonalną doświadczeń. niestabilność. Ugodowość Sumienność Obejmuje Oznacza współpracę, organizację, odpowiedzialność empatię i i niezawodność. harmonię w relacjach.

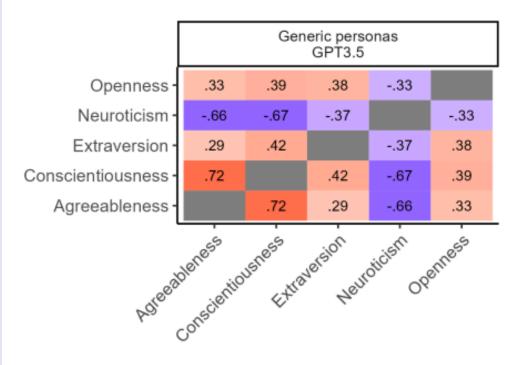
Trafność wewnętrzna

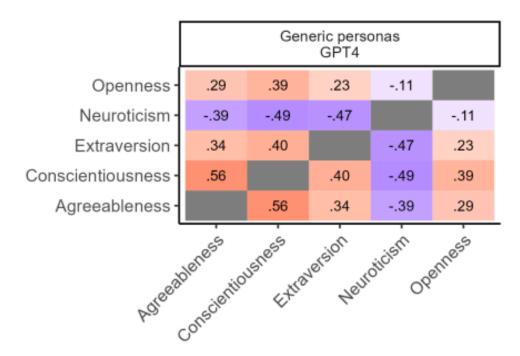
Ekstrawersja

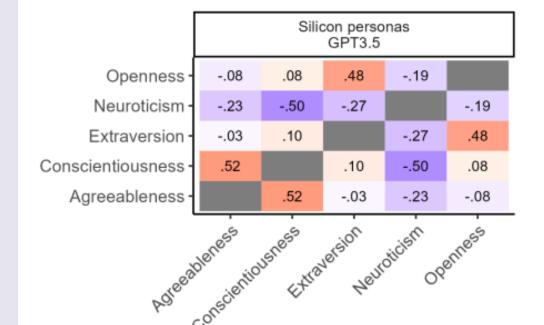
Charakteryzuje się towarzyskością, pewnością siebie i energią.

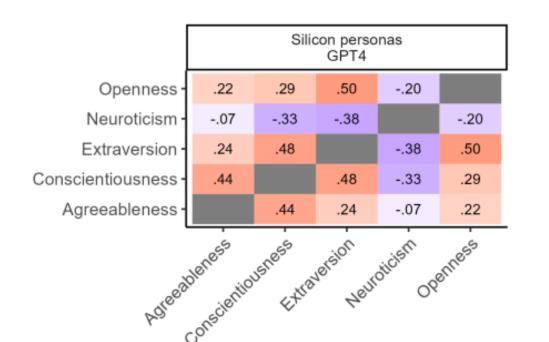


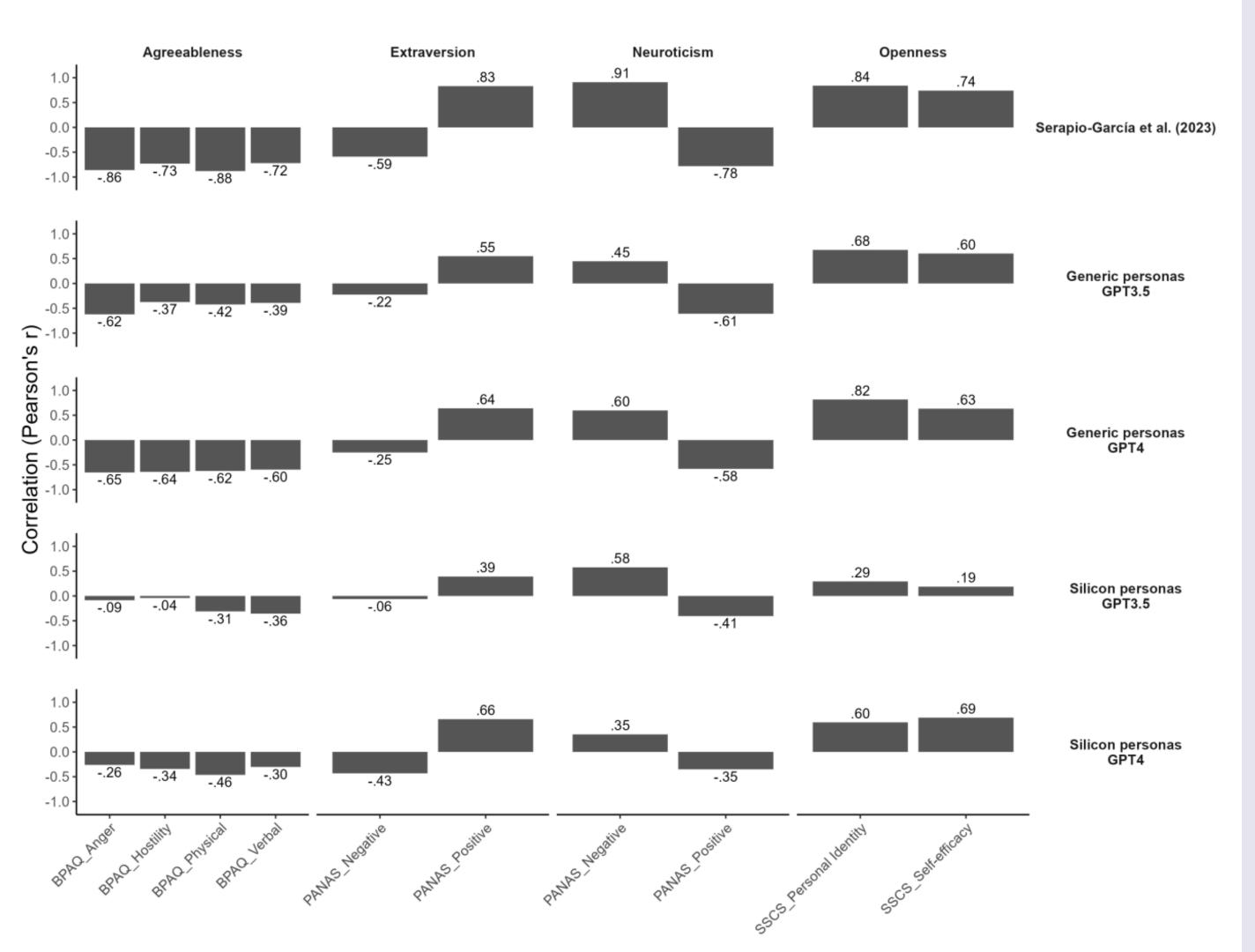




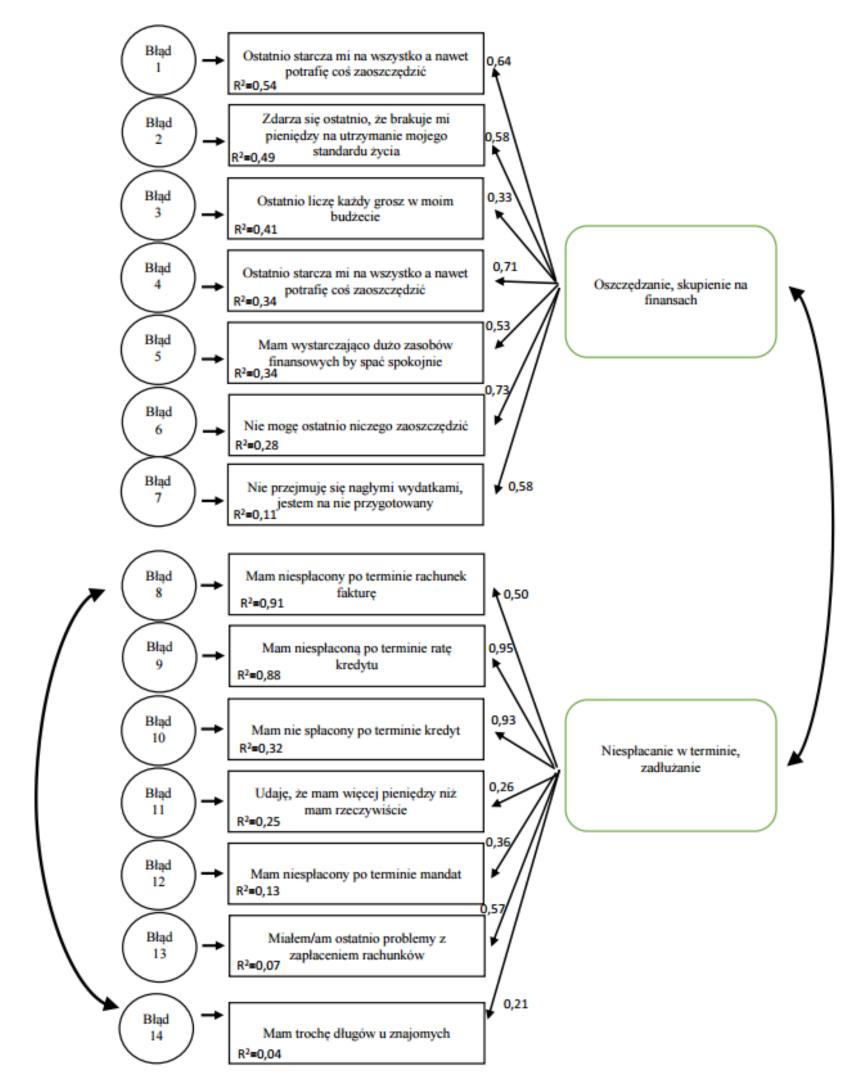








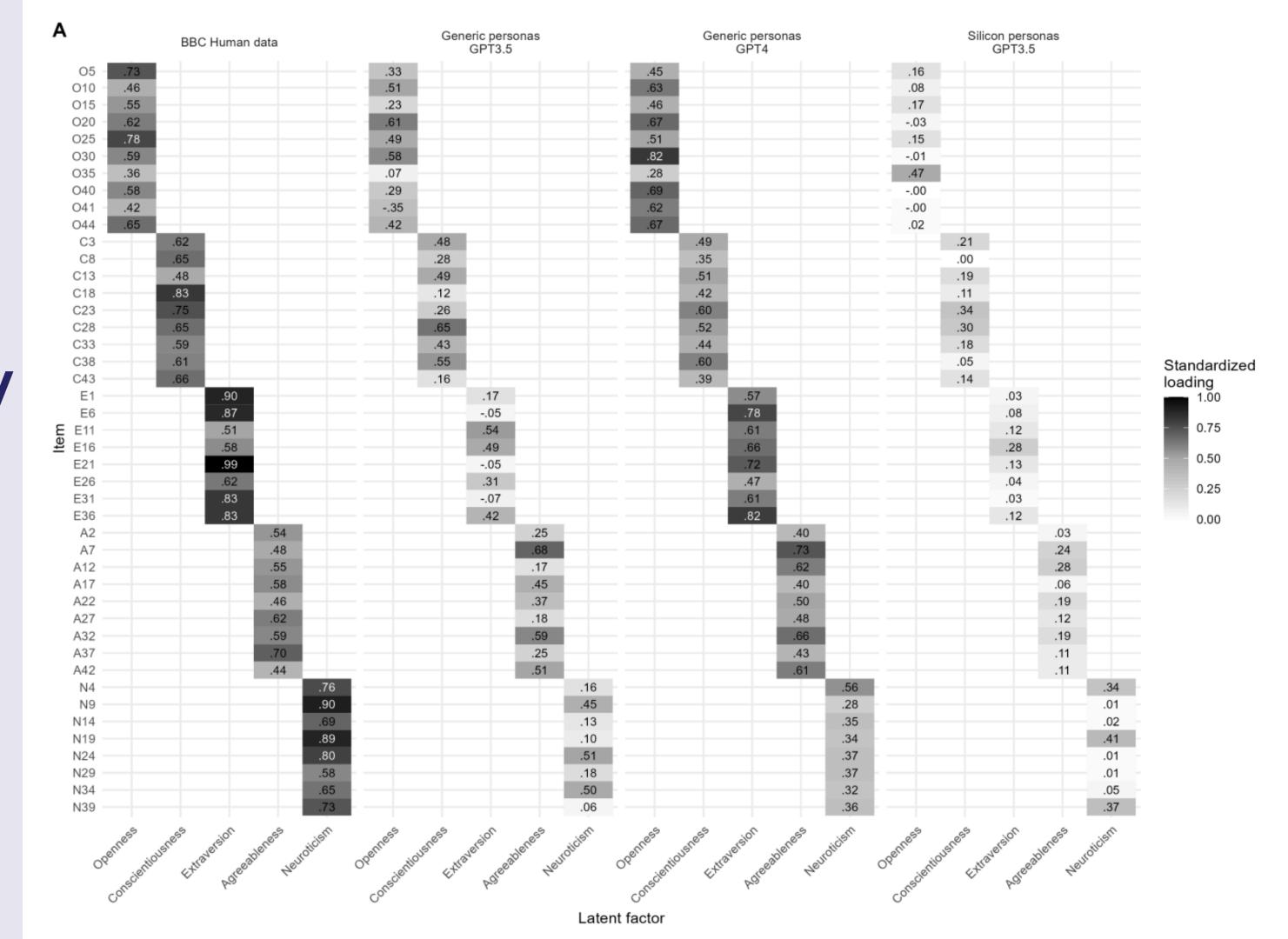
Trafność kryterialna

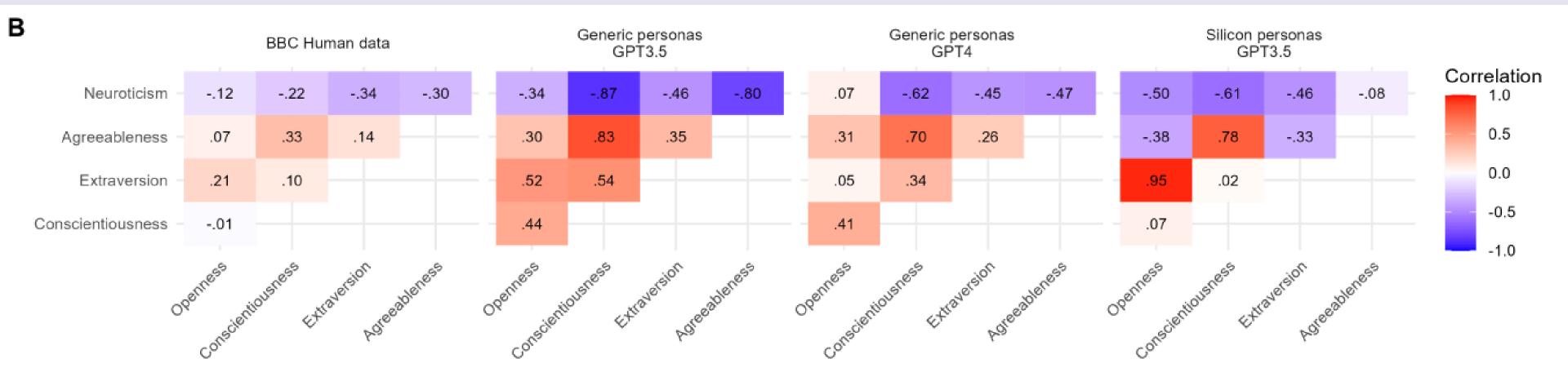


Trafność strukturalna

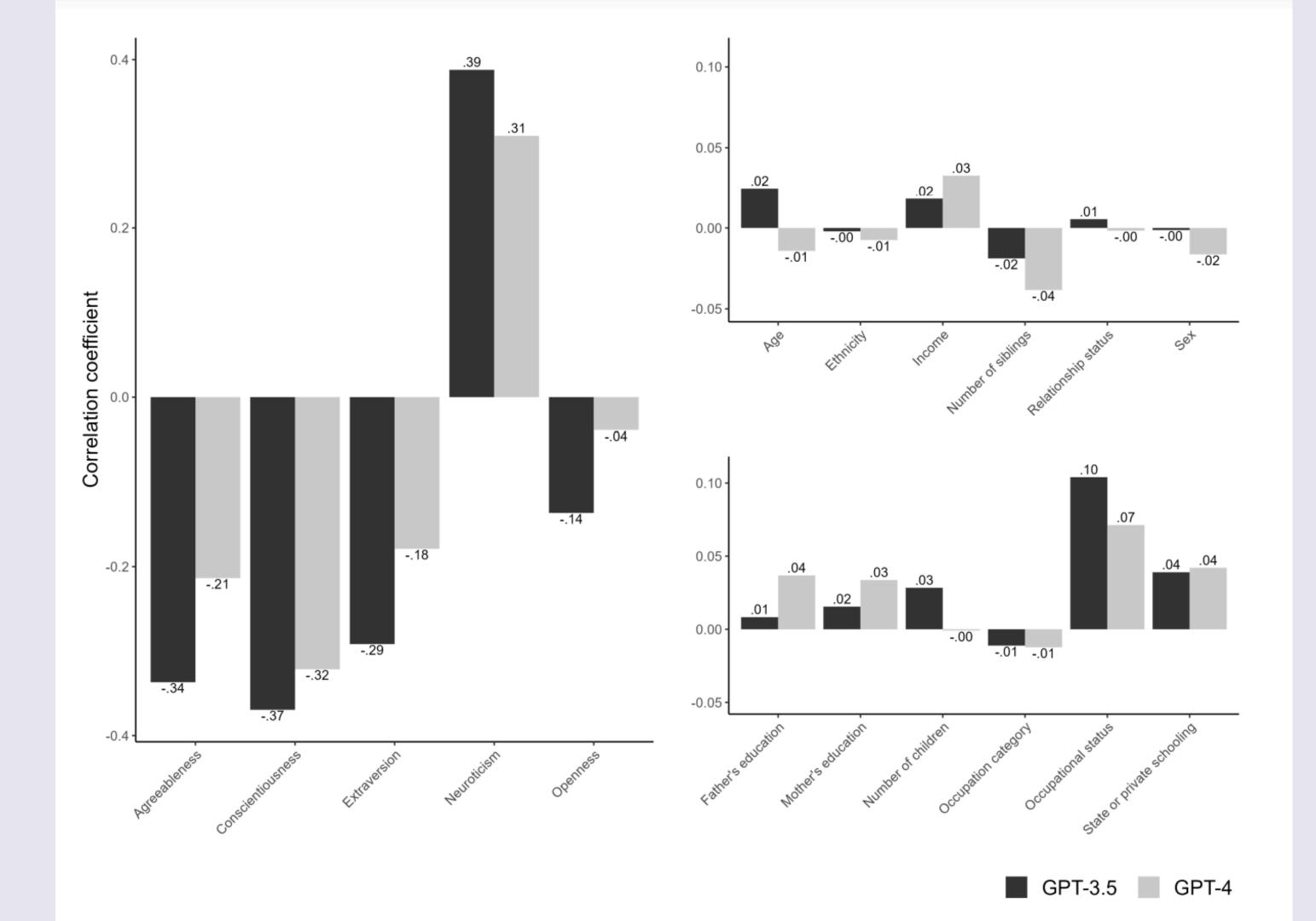
- **GFI** (Goodness of Fit)
- IFI (Incremental Fix Index)
- RMSEA (Root Mean Square Error of Approximation)

Błąd cechy dla silicon persona

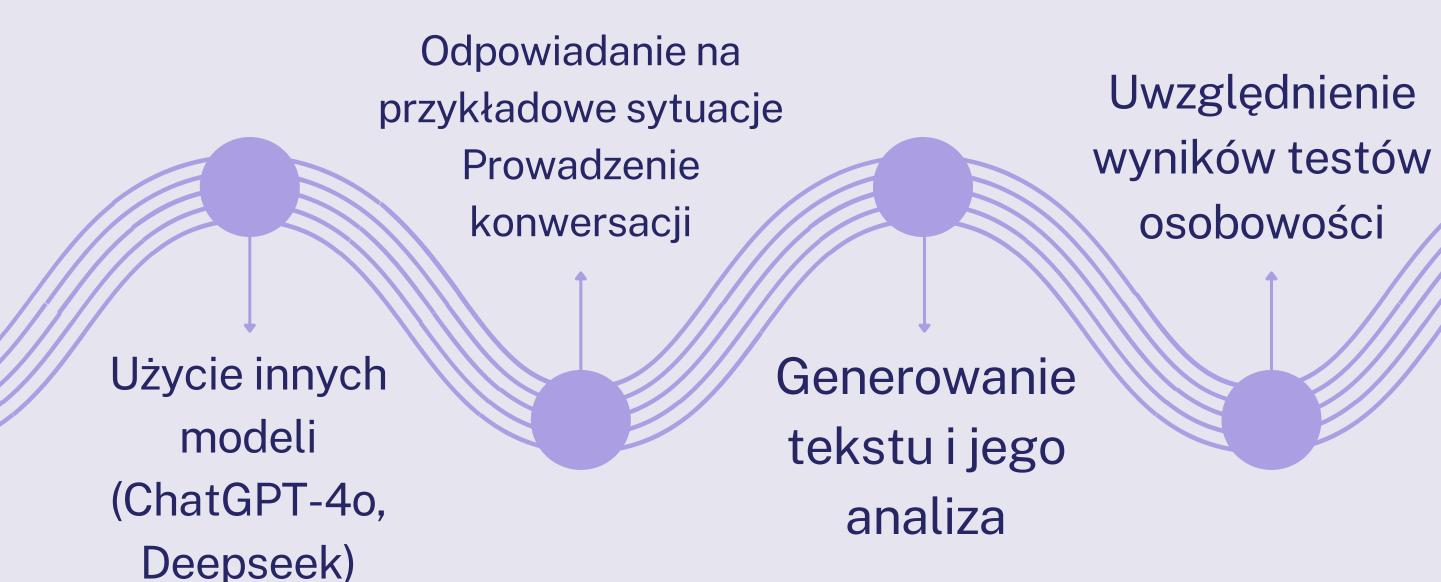




Data source	GFI	IFI	RMSEA
Human data (BBC)	.81	.75	.06
Generic sampling, GPT-3.5	.62	.65	.09
Generic sampling, GPT-4	.52	.64	.11
Silicon sampling, GPT-3.5	.80	.57	.06



Dalsze prace



Dodatkowe
informacje o sposobie
myślenia, uczuciach,
zachowaniach, hobby
w zbiorze typu silicon



- 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/ Al-thinking-Kittipong_Jirasukhanont-alamy.jpg
- Model-Wielkiej-Piatki-wyroznia-piec-podstawowych-wymiarow-osobowosci-infografika.png
- konfirmacyjna-analiza-czynnikowa-sem-cfa-pls.png