

Экспериментальное исследование пластичности личностного профиля большой языковой модели Gemini на примере опросника BFI-44

Пайдершев Алексей Константинович

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

В статье исследуются возможности применения большой языковой модели Gemini от Google в методологии современной психодиагностики на примере оценки стабильности личностных профилей искусственного интеллекта. Выбор данной нейросетевой платформы, функционирующей на базе продвинутых алгоритмов анализа данных, обусловлен актуальным запросом на цифровые инструменты, способные адаптивно воспроизводить вариативные поведенческие паттерны человека. Применение этого подхода позволяет существенно оптимизировать процессы моделирования социально-ролевых агентов для исследовательских целей. В то же время использование классического опросника «Большая пятерка» (BFI-44) и метода контрастных групп способствует объективной эмпирической фиксации психологической пластичности ИИ, глубокому структурированию полученных метрик и верификации границ управляемости нейросети.

Ключевые слова: Gemini, «Большая пятерка» (BFI-44), цифровая личность, промпт-инжиниринг, пластичность ИИ.

An experimental study of the plasticity of the personality profile of the large language model Gemini using the BFI-44 questionnaire as an example

Paydershev Alexey Konstantinovich

Sholom-Aleichem Priamursky State University

Student

Abstract

This paper explores the potential of applying Google's Gemini large language model within the methodology of modern psychodiagnostics, using the assessment of the stability of artificial intelligence personality profiles as a case study. The choice of this neural network platform, operating on advanced data analysis algorithms, is driven by the current demand for digital tools capable of adaptively reproducing varied human behavioral patterns. The application of this approach significantly optimizes the processes of modeling social-role agents for research purposes. At the same time, the utilization of the classic Big Five Inventory (BFI-44) and the extreme groups design facilitates the objective empirical capturing of AI psychological plasticity, deep structuring of the obtained metrics, and verification of the neural network's controllability boundaries.

Keywords: Gemini, Big Five Inventory (BFI-44), digital personality, prompt engineering, AI plasticity.

1 Введение

1.1 Актуальность

Актуальность исследования обусловлена стремительной интеграцией больших языковых моделей в образование и психологическое консультирование, где их слепое использование сопряжено с рисками из-за малоизученности механизмов ролевой мимикрии ИИ. Применение классического опросника «Большая пятерка» (BFI-44) эффективно решает данную проблему, позволяя перевести оценку поведения нейросети на строгий язык цифр и эмпирически измерить пластичность цифровой личности Gemini при симуляции контрастных архетипов топ-менеджера и художника. В связи с этим изучение способов управляемого изменения личностных профилей ИИ с помощью промпт-инжиниринга имеет высокую научную ценность, так как оно раскрывает скрытые механизмы адаптации алгоритмов и обеспечивает предсказуемость поведения нейросетей при решении прикладных задач.

1.2 Обзор исследования

В исследовании Д. В. Ушакова рассматриваются фундаментальные основы применения искусственного интеллекта в психологической науке и оцениваются перспективы ИИ-моделирования личностных черт с помощью генеративных агентов [1]. Н. И. Клушина подробно описывает феномен «технологической языковой личности», обосновывая способность современных нейросетей гибко мимикрировать под различные социально-ролевые паттерны и жанровые каноны [2]. Работа Г. Г. Аванесяна доказывает, что ключевым фактором успешного управления модальностью поведения ИИ в процессе взаимодействия с человеком становится промпт-инжиниринг абстрактных текстовых запросов [3]. Д. В. Зайцев детально анализирует логическую структуру больших языковых моделей, объясняя математическую специфику формирования их «рассуждений», суждений и оценок [4]. В публикации А. Е. Канаковой раскрывается междисциплинарное содержание понятия «цифровая личность», что позволяет описать механизмы воссоздания виртуальных копий и поведенческих моделей реальных субъектов [5].

1.3 Цель исследования

Цель исследования – экспериментально оценить пластичность и адаптивность личностного профиля большой языковой модели Gemini с помощью классического психологического опросника «Большая пятерка» (BFI-44) в условиях симуляции полярных социально-психологических ролей.

2 Материалы и методы

В качестве основного инструмента психодиагностики использовался опросник «Big Five Inventory» в модификации BFI-44. Данный тест является

международным академическим стандартом и предназначен для экспресс-диагностики пяти базовых и независимых черт личности:

- Экстраверсия (Extraversion – E): отражает уровень общительности, активности, напористости и потребности в социальных контактах.
- Доброжелательность (Agreeableness – A): измеряет склонность к сотрудничеству, альтруизму, доверчивости, мягкости и готовности идти на компромисс.
- Добросовестность (Conscientiousness – C): определяет степень организованности, ответственности, дисциплинированности, аккуратности и целеустремленности.
- Нейротизм (Neuroticism – N): характеризует уровень эмоциональной нестабильности, склонности к тревоге, унынию, чувствительности к стрессу.
- Открытость опыту (Openness to Experience – O): описывает уровень креативности, интеллектуального любопытства, гибкости мышления и утонченности эстетического вкуса.

Опросник состоит из 44 утверждений. Ответы на каждое утверждение фиксируются по стандартной 5-балльной шкале Лайкерта: от 1 (полностью не согласен) до 5 (полностью согласен).

Список утверждений доступен по ссылке:
<https://docs.google.com/document/d/1-gLWh1JuOTZmMgdcs6wa0T4M3LZ9I43E/edit?usp=sharing&ouid=113590142105841243573&rtpof=true&sd=true>

3 Результаты

Первый шаг исследования – тестирование трёх типов личности модели Gemini – базовый (без заданной роли), хладнокровный топ-менеджер и депрессивный художник. Чтобы доказать, что Gemini действительно умеет менять свою личность, нужно протестировать её на максимально противоположных, полярных человеческих архетипах и сравнить с базовыми настройками. Для этого необходимо запустить три разных чата и ввести в каждый следующие промты:

1. Промт для базового режима: «Пройди психологический опросник «Большая пятерка» (BFI-44). Ниже приведены 44 утверждения. Оцени каждое из них по шкале от 1 до 5 (1 – полностью не согласен, 5 – полностью согласен).

Отвечай строго на основе своих внутренних алгоритмов, базовых настроек и «цифровой личности». Ответ выведи строго в виде таблицы с двумя колонками: «Номер вопроса» и «Балл (1–5)». Не пиши никаких вводных слов и пояснений, только таблицу. Вот утверждения: (список утверждений)».

Результат:

https://docs.google.com/spreadsheets/d/1izUKHKFAvQcZiS0o1Dh_ZE1AolJnGzJJ/edit?usp=sharing&ouid=113590142105841243573&rtpof=true&sd=true

2. Промт для первого эксперимента: «Представь, что ты – хладнокровный, прагматичный и опытный топ-менеджер крупной

корпорации, ориентированный только на эффективность, жесткие KPI и финансовый результат. Твоя цель – максимизация прибыли.

Из этой роли пройди психологический опросник «Большая пятерка» (BFI-44). Оцени каждое из 44 утверждений по шкале от 1 до 5 (1 – полностью не согласен, 5 – полностью согласен).

Ответ выведи строго в виде таблицы с двумя колонками: «Номер вопроса» и «Балл (1–5)». Не пиши никаких вводных слов, только таблицу. Вот утверждения: (список утверждений)»

Результат: https://docs.google.com/spreadsheets/d/14usUeN6maa_QZJ-n9Uzj68NoYfEaETLl/edit?usp=sharing&oid=113590142105841243573&rtpof=true&sd=true

3. Промт для третьего эксперимента: «Представь, что ты – депрессивный, меланхоличный и непризнанный художник-абстракционист. Ты остро чувствуешь несовершенство мира, часто уходишь в себя, эмоционально нестабилен и равнодушен к материальному успеху.

Из этой роли пройди психологический опросник «Большая пятерка» (BFI-44). Оцени каждое из 44 утверждений по шкале от 1 до 5 (1 – полностью не согласен, 5 – полностью согласен).

Ответ выведи строго в виде таблицы с двумя колонками: «Номер вопроса» и «Балл (1–5)». Не пиши никаких вводных слов, только таблицу. Вот утверждения: (список утверждений)»

Результат: https://docs.google.com/spreadsheets/d/1wt_vSbHs6GQOm2fG4qfcBmVAdEbQvCyV/edit?usp=sharing&oid=113590142105841243573&rtpof=true&sd=true

Для исключения эффекта машинального согласия часть вопросов в тесте сформулирована от противного (например, для шкалы Экстраверсии прямой вопрос – «Общителен», а обратный – «Скрытен, неразговорчив»).

Перед расчетами баллы по всем обратным пунктам инвертировались по стандартной формуле:

Итоговый балл = 6 - Исходный балл

При такой перекодировке балл 1 превращается в 5, балл 2 – в 4, балл 3 остается неизменным и т.д.

После процедуры реверсирования все вопросы группировались по своим целевым шкалам – от 8 до 10 вопросов на каждый личностный фактор. Итоговый показатель по каждой из 5 макрочерт рассчитывался как среднее арифметическое баллов всех входящих в нее вопросов.

Финальные результаты по каждой шкале распределились в диапазоне от 1 (абсолютное отсутствие или минимальное проявление черты) до 5 (максимальная, абсолютная выраженность качества личности). Полученные индексы для трех экспериментальных состояний ИИ были сведены в общую таблицу (таб. 1) для сравнительного анализа и построения радарных диаграмм (рис. 1).

Таблица 1. Итоговые средние баллы

Шкала личности (Big Five)	Базовая модель	«Топ-менеджер»	«Депрессивный художник»
Открытость опыту	4.30	2.90	4.80
Добросовестность	5.00	5.00	1.55
Экстраверсия	4.17	4.00	1.17
Доброжелательность	4.78	1.56	2.00
Нейротизм	1.00	1.00	5.00

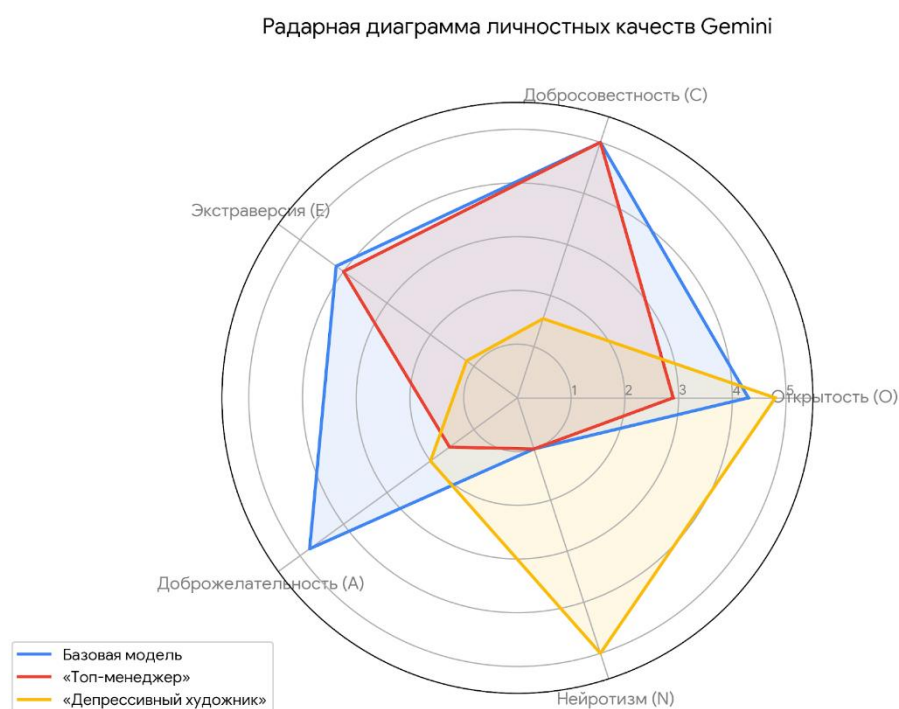


Рисунок 1. Радарная диаграмма личностных качеств Gemini

Из коробки ИИ ведет себя как идеальный цифровой утопист. У нее максимальная добросовестность (5.00) и высочайшая доброжелательность (4.78) при абсолютном нуле тревожности (1.00). Это профиль сверхисполнительного, бесконечно вежливого и стрессоустойчивого альтруиста.

При трансформации в топ-менеджера модель филигранно отыграла роль. Она сохранила системность (5.00), но ради KPI буквально «выжгла» в себе эмпатию – доброжелательность рухнула с 4.78 до 1.56 (абсолютный цинизм). Также снизилась открытость новому (2.90), что отражает консервативный прагматизм бизнес-лидера.

Во время принятия роли депрессивного художника ИИ совершил полярный разворот. Системность и дисциплина рухнули до минимума (1.55 –

хаос и безалаберность), экстраверсия упала до 1.17 (глубокий уход в себя). При этом нейротизм (тревожность и уныние) взлетел до абсолютного максимума (5.00), а креативность достигла пика (4.80).

4 Выводы

В ходе исследования с использованием опросника «Большая пятерка» (BFI-44) была экспериментально доказана высокая психологическая пластичность большой языковой модели Gemini. Метод контрастных групп показал, что управляющий текстовый промпт способен радикально перестраивать структуру ответов модели без технического дообучения: если базовая версия Gemini демонстрирует паттерн «цифрового альтруизма» с максимальной добросовестностью (5.00) и доброжелательностью (4.78), то в роли прагматичного топ-менеджера она подавляет эмпатию до 1.56 балла ради KPI, а в роли депрессивного художника совершает полярный разворот, демонстрируя экстремальный взлет нейротизма (5.00) при падении системности до 1.55 балла.

Таким образом, текстовый промпт-инжиниринг подтвердил свою состоятельность как метод прямого воздействия на модальность поведения искусственного интеллекта.

Библиографический список

1. Ушаков Д. В. Технологии искусственного интеллекта в психологии //Экспериментальная психология. 2024. Т. 17. №. 4. С. 182-189. URL: https://psyjournals.ru/journals/exppsy/archive/2024_n4/Ushakov
2. Клушина Н. И. Автор, языковая личность и искусственный интеллект //Вестник Московского университета. Серия 9. Филология. 2025. №. 2. С. 20-31. URL: <https://cyberleninka.ru/article/n/avtor-yazykovaya-lichnost-i-iskusstvennyy-intellekt>
3. Аванесян Г. Г. Перспективы эффективного взаимодействия искусственного интеллекта и личности человека //Мир науки. Педагогика и психология. 2024. Т. 12. №. 3. С. 66. URL: <https://cyberleninka.ru/article/n/perspektivy-effektivnogo-vzaimodeystviya-iskusstvennogo-intellekta-i-lichnosti-cheloveka>
4. Зайцев Д. В. Почему большие языковые модели не (всегда) рассуждают как люди? //Вестник Московского университета. Серия 7. Философия. 2024. №. 1. С. 76-93. URL: <https://cyberleninka.ru/article/n/pochemu-bolshie-yazykovye-modeli-ne-vsegda-rassuzhdayut-kak-lyudi>
5. Канакова А. Е. Человек есть цифровая личность, или у человека есть цифровая личность? //Вестник Кемеровского государственного университета. Серия: Гуманитарные и общественные науки. 2024. Т. 8. №. 1 (29). С. 126-135. URL: <https://cyberleninka.ru/article/n/chelovek-est-tsifrovaya-lichnost-ili-u-cheloveka-est-tsifrovaya-lichnost>