Применение больших языковых моделей для построения графов трудных жизненных ситуаций

И. О. Бергер*, Е. В. Битюцкая‡, Э. Э. Гасанов§, А. А. Херувимова¶, А. А. Хлебникова‖

Проведено сопоставление возможностей двух больших языковых моделей (DeepSeek и GPT-6) при автоматическом построении направленных графов трудной жизненной ситуации (ТЖС) по ее текстовому описанию. На материале 10 описаний ТЖС получены 20 графов, оценивание которых проводилось по формальным и содержательным критериям с использованием метода LLM-as-a-Judge. Обе модели показали высокую формальную корректность, однако это не обеспечило содержательной адекватности.

Ключевые слова: трудная жизненная ситуация, большая языковая модель, графовое моделирование.

Финансирование. Исследование выполнено по междисциплинарному фундаментальному проекту «Анализ и синтез траекторий преодоле-

* *Бергер Ирина Олеговна* — м. н. с. каф. математической теории интеллектуальных систем мех.-мат. ф-та МГУ, e-mail: ioberger@ya.ru, ORCID: 0000-0003-0123-1014.

Berger Irina Olegovna — Junior Researcher, Department of Mathematical Theory of Intelligent Systems, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University.

‡ *Битюцкая Екатерина Владиславовна* — канд. психол. наук, доцент кафедры общей психологии ф. психологии МГУ, e-mail: bityutskaya_ew@mail.ru, ORCID: 0000-0001-6117-1063.

Bityutskaya Ekaterina Vladislavovna — Candidate of Psychological Sciences, Associate Professor, Department of General Psychology, Faculty of Psychology, Lomonosov Moscow State University.

§ *Гасанов Эльяр Эльдарович* — д. ф.-м. н., зав. каф. математической теории интеллектуальных систем мех.-мат. ф-та МГУ, e-mail: el_gasanov@mail.ru, ORCID: 0000-0002-0739-0967.

Gasanov Elyar Eldarovich — Doctor of Physical and Mathematical Sciences, Head of Department of Mathematical Theory of Intelligent Systems, Faculty of Mechanics and Mathematics, Moscow State University.

¶ *Херувимова Анна Алексеевна* — инженер кафедры общей психологии ф. психологии МГУ, e-mail: kheruvimova.a.a@gmail.com, ORCID: 0009-0003-8058-7728.

Kheruvimova Anna Alexeevna — Staff member, Department of General Psychology, Lomonosov Moscow State University.

‖ *Хлебникова Алёна Андреевна* — аспирант каф. общей психологии ф. психологии МГУ, e-mail: alena.epochta@gmail.com, ORCID: 0000-0002-7427-8844.

Khlebnikova Alena Andreevna — PhD Student, Department of General Psychology, Faculty of Psychology, Moscow State University.

Опубликовано на условиях лицензии CC-BY 4.0
Published under CC-BY 4.0

ния трудных жизненных ситуаций», поддержанному в рамках Конкурса научных проектов по перспективным междисциплинарным практико-ориентированным тематикам МГУ имени М.В.Ломоносова (2026 г.).

Введение. Необходимость системного изучения ТЖС предполагает учет состояний субъекта, его действий и возможных исходов ситуации. В связи с этим особую значимость приобретает разработка методов формализации, позволяющих создавать модели ТЖС: их компонентов и взаимосвязей. Мы полагаем, что одним из перспективных способов такой формализации является графовое моделирование. Представление ТЖС в виде направленного графа позволяет обозначить состояния, события, действия, ресурсы и цели в качестве вершин, а допустимые переходы между ними – в виде направленных ребер.

Развитие больших языковых моделей (LLM) открывает возможности для автоматизации такого преобразования, однако формально корректное построение графа еще не гарантирует содержательной адекватности полученной репрезентации. Таким образом, проблема исследования заключается в недостаточной изученности способности LLM преобразовывать текстовые описания ТЖС в направленные графы, одновременно удовлетворяющие формальным требованиям и сохраняющие содержательную логику исходной ситуации.

Цель исследования — выявить особенности формальной корректности и содержательной адекватности направленных графов ТЖС, созданных большими языковыми моделями DeepSeek и GPT-6 на основе исходных описаний ТЖС.

Гипотеза: формальная корректность графов ТЖС не гарантирует содержательной адекватности их репрезентации; различия между графами, построенными разными LLM, могут проявляться прежде всего в полноте моделирования существенных компонентов ситуации и логической организации направленных переходов между ними.

Определения. Направленный граф ТЖС — структура, в которой вершины обозначают события, состояния, действия, ресурсы или цели, а направленные ребра – допустимые переходы между ними. Обязательными условиями являлись наличие единственной стартовой вершины, обозначение целей и запрещенных вершин, классификация терминальных исходов и содержательная обоснованность всех элементов графа. Формальная корректность — соблюдение требований исходной инструкции к структуре и визуализации графа. Содержательная адекватность — соответствие графа исходному описанию ТЖС и согласованность представленных в нем направленных переходов. Логическая связность — согласованность переходов и возможность интерпретации графа как целостной структуры траекторий от начального состояния к альтернативным исходам.

Процедура. Материал составили 20 направленных графов, созданных DeepSeek и GPT-6 на основе десяти описаний ТЖС. Описания были получены с использованием методики структурированного описания ситуации [1]. Ситуации относились преимущественно к учебной и учебно-профессиональной деятельности (подготовка и защита выпускной квалификационной работы, сдача экзаменов, завершение обучения, практика и стажировка).

Обе модели получали идентичные запросы с описанием ТЖС, определениями элементов графа, правилами построения и требованиями к визуализации в формате SVG. Исследование имело пилотный сравнительный дизайн с парным сопоставлением: единицей сравнения выступала пара графов, созданных для одного и того же описания в одинаковых условиях генерации. Сравнение проводилось по формальным и содержательным критериям с использованием метода LLM-as-a-Judge: модель-эксперт (GPT-6) получала исходное описание и два обезличенных графа (А и Б) без сведений об авторстве; порядок предъявления был постоянным (DeepSeek — А, GPT-6 — В), что рассматривалось как возможный источник смещения.

Формальная корректность оценивалась по восьми бинарным параметрам (0/1): наличие единственной стартовой вершины; корректность направлений ребер и их соединений; определение и классификация терминальных вершин; выделение целей и запрещенных вершин; отсутствие дублирующихся элементов; соблюдение формата идентификаторов; соответствие подписей, легенды и визуальных обозначений; синтаксическая корректность SVG-кода и размеров изображения. Суммарная оценка от 0 до 8 баллов; достижимость вершин и согласованность траекторий анализировались отдельно, в рамках логической связности.

Содержательная оценка включала четыре критерия с трехбалльной шкалой (0–2): С1 — семантическая обоснованность (соответствие вершин и переходов описанию, корректность интерпретации компонентов); С2 — полнота моделирования (представленность значимых компонентов ситуации и промежуточных действий); С3 — логическая связность (достижимость вершин из старта, последовательность действий, согласованность переходов к исходам; единственная правильная траектория не предполагалась); С4 — отсутствие необоснованной избыточности (оправданность сложности структуры, дублирование, лишние элементы; меньшее количество вершин не считалось преимуществом). Максимальная суммарная содержательная оценка — 8 баллов. Анализ данных проводился с помощью описательной статистики. С учетом малого объема выборки основное внимание уделялось качественному характеру различий, а не статистической значимости.

Результаты. Формальная проверка показала высокий уровень соблюдения требований обеими моделями: в девяти из десяти пар оба графа получили максимальные 8 баллов, в одной паре DeepSeek получила 6 баллов, GPT-6 — 8. Суммарно это составило 78 из 80 баллов для DeepSeek и 80 из 80 для GPT-6; нарушения касались классификации максимального неуспеха и отсутствия запрещенной вершины. При этом в ряде графов DeepSeek с максимальной формальной оценкой обнаруживались недостижимые вершины и неоднозначные траектории, что указывает на необходимость разграничения формального соответствия и логической согласованности.

По содержательным оценкам графы GPT-6 превзошли графы DeepSeek в девяти из десяти пар (в одной паре преимущество было у DeepSeek — 8 против 7). Медианные оценки составили 4 балла для DeepSeek и 8 баллов для GPT-6, медиана парных различий — 3,5 балла, средние — 4,5 и 7,7 соответственно. По критериям суммарно (из 20 возможных баллов за 10 графов) DeepSeek и GPT-6 получили следующие оценки: по семантической обоснованности — 11 и 17, по полноте — 12 и 20, по логической связности — 8 и 20, по отсутствию избыточности — 14 и 20; итого 45 и 77. Наиболее выражены различия по логической связности: GPT-6 получила максимум во всех десяти ситуациях, DeepSeek — 8 баллов из 20.

Качественный анализ выявил повторяющиеся особенности: неполное отражение значимых компонентов ситуации (пропуск упоминавшихся в описании действий и событий); нарушения логики траектории; недостижимость отдельных вершин, ограничивающая интерпретацию графа как целостной системы траекторий; неоднозначная интерпретация исходной информации (представление предположительных обстоятельств как фактов, а возможных состояний — как обязательных этапов).

Обсуждение. Формальная корректность не гарантирует содержательной адекватности: даже графы с максимальными формальными оценками содержали смысловые недостатки, что требует различать техническое соответствие формату и качество смысловой организации. Логическая связность выступает ключевой характеристикой качества: наибольшие различия выявлены именно по этому критерию, причем они не всегда сопровождались различиями в полноте (например, в шестой паре обе модели получили максимум за полноту, но различались по организации переходов). Это позволяет предположить, что выделение релевантных элементов и построение согласованной структуры отношений – разные задачи формализации: модель может корректно определить компоненты, но связать их противоречиво.

Заключение. Обе исследованные языковые модели — DeepSeek и GPT-6 — продемонстрировали высокий уровень формального соответствия графов инструкции, однако формальная корректность не во всех случаях сопровождалась содержательной адекватностью. По результатам автоматизированного оценивания графы GPT-6 получили более высокие суммарные содержательные оценки в девяти из десяти пар. Наиболее выраженные различия выявлены по критерию логической связности. Результаты подтверждают перспективность применения LLM для графовой формализации ТЖС. Вместе с тем выявленные различия между моделями носят предварительный характер и требуют независимой экспертной проверки.

Список литературы

[1] Bityutskaya E. V., Gasanov E. E., Khazova K. V., Patrashkin N. A., "Classifying the perception of difficult life tasks: Machine learning and/or modeling of logical processes", *Psychology in Russia: State of the Art*, **17**:2 (2024), 64–84.. DOI: 10.11621/pir.2024.0205 .

Статья поступила 8 ноября 2026 г.

Using Large Language Models to Construct Graphs of Difficult Life Situations

I. O. Berger, E. V. Bityutskaya, E. E. Gasanov, A. A. Kheruvimova, A. A. Khlebnikova

We compared the capabilities of two large language models (DeepSeek and GPT-6) in the automated construction of directed graphs representing "difficult life situations" (DLS) based on their textual descriptions. Using 10 such descriptions, we generated 20 graphs and evaluated them according to both formal and content-related criteria, employing an "LLM-as-a-Judge" approach. While both models demonstrated a high degree of formal correctness, this did not guarantee the adequacy of their content.

Keywords: difficult life situation, large language model, graph modeling.

Received on November 8, 2026