

УДК 004.032.26

5.2.2. Математические, статистические и инструментальные методы в экономике (физико-математические науки, экономические науки)

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ПРОМПТ-ИНЖИНИРИНГА ДЛЯ УПРАВЛЕНИЯ ПОВЕДЕНИЕМ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Попова Ольга Борисовна
Кандидат технических наук, доцент кафедры информационных систем и программирования
РИНЦ-SCIENCE INDEX SPIN-код: 1095-6863
popova_ob@mail.ru
ФГБОУ ВО «Кубанский государственный технологический университет», Краснодар, Россия

Князев Максим Владимирович
аспирант кафедры информационных систем и программирования
maksimka.knyazev.tm2002@yandex.ru
ФГБОУ ВО «Кубанский государственный технологический университет», Краснодар, Россия

Боярко Максим Владимирович
студент кафедры информационных систем и программирования
boyarko-m@mail.ru
ФГБОУ ВО «Кубанский государственный технологический университет», 350072, улица Московская, 2, Краснодар, Россия

Ведерников Георгий Валерьевич
студент кафедры информационных систем и программирования
george.vedernikov.01@mail.ru
ФГБОУ ВО «Кубанский государственный технологический университет», 350072, улица Московская, 2, Краснодар, Россия

В статье представлен систематический обзор и сравнительный анализ современных методов промпт-инжиниринга, направленных на повышение эффективности генеративных моделей искусственного интеллекта. Рассмотрены как классические подходы (Zero-Shot, Few-Shot, Chain of Thought), так и перспективные техники (Knowledge Generation, Directed Stimulus, Tree of Thoughts, ReAct). На основе детального сравнения по ключевым метрикам (точность, релевантность, ресурсозатратность, универсальность) выявлены оптимальные сценарии применения различных методов и определены их сильные и слабые стороны. Сформулированы требования к перспективному формализованному методу,

UDC 004.032.26

5.2.2. Mathematical, statistical and instrumental methods of economics (physical and mathematical sciences, economic sciences)

COMPARATIVE ANALYSIS OF INDUSTRIAL ENGINEERING METHODS FOR CONTROLLING THE BEHAVIOR OF GENERATIVE ARTIFICIAL INTELLIGENCE MODELS

Popova Olga Borisovna
Candidate of Technical Sciences, Associate Professor of the Department of Information Systems and Programming
RSCI-SCIENCE INDEX SPIN-code: 1095-6863
popova_ob@mail.ru
Kuban State Technological University, Krasnodar, Russia

Knyazev Maxim Vladimirovich
Graduate Student of the Department of Information Systems and Programming
maksimka.knyazev.tm2002@yandex.ru
Kuban State Technological University, Krasnodar, Russia

Boyarko Maxim Vladimirovich
student of the Department of Information Systems and Programming
nastya26112004@mail.ru
Kuban State Technological University, Krasnodar, Russia

Vedernikov Georgy Valerievich
student of the Department of Information Systems and Programming
nastya26112004@mail.ru
Kuban State Technological University, Krasnodar, Russia

The article presents a systematic review and comparative analysis of modern methods of prompt engineering aimed at improving the efficiency of generative AI models. Both classical approaches (Zero-Shot, Few-Shot, Chain of Thought) and promising techniques (Knowledge Generation, Directed Stimulus, Tree of Thoughts, ReAct) are considered. Based on a detailed comparison of key metrics (accuracy, relevance, resource consumption, and versatility), the optimal scenarios for using different methods are identified, and their strengths and weaknesses are determined. The requirements for a promising formalized method that integrates the advantages of the considered approaches are formulated

интегрирующему преимуществу рассмотренных подходов

Ключевые слова: ПРОМПТ-ИНЖИНИРИНГ, ГЕНЕРАТИВНЫЕ МОДЕЛИ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, УПРАВЛЕНИЕ ПОВЕДЕНИЕМ, СРАВНИТЕЛЬНЫЙ АНАЛИЗ, CHAIN OF THOUGHT, TREE OF THOUGHTS, KNOWLEDGE GENERATION, РЕЛЕВАНТНОСТЬ, ТОЧНОСТЬ ОТВЕТОВ

Keywords: PROMPT ENGINEERING, GENERATIVE MODELS, ARTIFICIAL INTELLIGENCE, BEHAVIOR MANAGEMENT, COMPARATIVE ANALYSIS, CHAIN OF THOUGHT, TREE OF THOUGHTS, KNOWLEDGE GENERATION, RELEVANCE, AND ACCURACY OF RESPONSES

<http://dx.doi.org/10.21515/1990-4665-214-026>

Введение. Современный этап развития искусственного интеллекта характеризуется стремительным ростом возможностей генеративных моделей, особенно больших языковых моделей. Это обусловило возникновение и активное развитие промпт-инжиниринга – дисциплины, посвященной оптимизации текстовых запросов для эффективного управления поведением моделей. Качество промпта напрямую определяет релевантность, точность и полезность ответа модели, что является критическим фактором для успешного внедрения AI-решений в бизнес-процессы и научную деятельность [1]. Несмотря на обилие эмпирических данных, область остро нуждается в систематизированных и формализованных методах, обеспечивающих воспроизводимость и масштабируемость подходов.

В мировой и российской научной практике уже сложился комплекс эмпирических методов управления поведением генеративных моделей, основанных на вероятностной природе языковых моделей [2, 4]. К традиционным, устоявшимся техникам принято относить нулевой контекст (Zero-Shot), несколько примеров (Few-Shot), назначение роли (Role Assignment), цепочку размышлений (Chain of Thought), самосогласованность (Self-Consistency) и расширение контекста (Context Expansion). Несмотря на свою распространенность, эти подходы зачастую не в полной мере учитывают архитектурные особенности трансформеров [3] и не гарантируют стабильно высокого качества ответов при решении

<http://ej.kubagro.ru/2025/10/pdf/26.pdf>

специализированных задач, что стимулирует поиск более совершенных методик.

Среди перспективных направлений выделяется метод генерации знаний (Knowledge Generation Prompting), основанный на принципе явной активации релевантных знаний модели перед финальным ответом. Этот подход предполагает двухэтапное взаимодействие: сначала модель генерирует фактологическую базу по теме, а затем синтезирует итоговый ответ на ее основе. Технология доказала свою эффективность в снижении вероятности галлюцинаций за счет принудительной активации нужных ассоциативных связей, демонстрируя на экспериментах повышение точности на 15-20% базовым техникам.

Другим значимым направлением является метод направленного стимулирования (Directed Stimulus Prompting). Его суть заключается в инъекции в промпт обязательных семантических маркеров, которые детерминируют содержание генерируемого ответа. В отличие от назначения роли, где задается общий контекст, Directed Stimulus предполагает точечное внедрение конкретных концептов, тезисов и аспектов, подлежащих обязательному отражению в выводе. Это делает метод особенно ценным для задач структурированной аналитики, где требуется гарантированное покрытие всех аспектов проблемы.

Для работы со слабоструктурированными проблемами, где исходные параметры не могут быть четко формализованы изначально, высокую эффективность показывает итеративный диалоговый метод (Iterative Prompting). Он реализует принцип адаптивного уточнения запроса через последовательность взаимодействий с моделью. Каждая следующая итерация строится на анализе результатов предыдущих, что позволяет осуществлять постепенную декомпозицию сложных задач и уточнение семантических нюансов.

Эволюционным развитием техники цепочки размышлений стал метод древовидного мышления (Tree of Thoughts), который реализует нелинейный подход к поиску решений. В отличие от последовательного CoT, этот метод предполагает генерацию множества альтернативных путей рассуждения с последующей оценкой их перспективности через механизм самооценки модели. Наиболее релевантные ветви развиваются дальше, что позволяет модели осуществлять параллельный поиск решения с элементами, а не последовательный перебор.

Методология ReAct (ReAct Prompting) интегрирует Reasoning (рассуждение) и Action (действие) в единый цикл взаимодействия. Модель не только осуществляет логические рассуждения, но и генерирует запросы к внешним источникам информации для восполнения дефицита знаний. Это позволяет преодолевать ограничения, связанные с тренировочными данными модели, и обеспечивать актуальность и достоверность генерируемых ответов, что особенно востребовано в областях с высокой динамикой обновления знаний.

Сравнив основные техники управления поведением генеративных моделей и проведя анализ их характеристик, была подготовлена сравнительная таблица (Таблица 1).

Проведенный сравнительный анализ наглядно демонстрирует существенное превосходство современных методов над классическими подходами. Если традиционные техники (Zero-Shot, Few-Shot, Role Assignment) обеспечивают точность в диапазоне 25-75%, то современные методы (Knowledge Generation, Directed Stimulus, Tree of Thoughts) уверенно достигают показателей 75-97%. Это сопровождается увеличением стабильности ответов, прецизионным контролем их содержания и значительным, на 30-60%, снижением подверженности галлюцинациям.

Эффективность различных методов варьируется в зависимости от категории задач. Для научных исследований Knowledge Generation превосходит Chain of Thought по фактологической точности на 15-20%, обеспечивая системную проверку достоверности контента. В бизнес-аналитике Directed Stimulus демонстрирует преимущество перед Role Assignment в контроле структуры ответа, гарантируя покрытие всех критических аспектов. Для стратегического планирования Tree of Thoughts обеспечивает на 25-30% более качественные решения compared to Context Expansion за счет исследования альтернативных сценариев.

Несмотря на более высокие временные затраты современных методов, их совокупная эффективность (качество решения/затраты) превосходит классические подходы. Ярким примером является Iterative Prompting, который при решении сложных задач требует на 40-60% меньше итераций по сравнению с последовательным применением Few-Shot. С точки зрения массового внедрения в бизнес-процессы наиболее перспективными представляются Directed Stimulus и Knowledge Generation, так как они обеспечивают оптимальный баланс между качеством и ресурсозатратностью. В то же время, для стратегических задач с высокой стоимостью ошибки целесообразно применение более требовательных, но и более мощных методов, таких как Tree of Thoughts и ReAct.

На основе проведенного анализа становятся очевидны контуры будущего формализованного метода управления поведением генеративных моделей. Такой метод должен интегрировать ключевые преимущества современных техник: борьбу с галлюцинациями от Knowledge Generation и прецизионный контроль содержания от Directed Stimulus. Его архитектура должна быть адаптивной, сочетая быстроту выполнения, характерную для Role Assignment, и глубину анализа, присущую Tree of Thoughts [5]. Критически важными требованиями являются оптимизация ресурсопотребления до уровня, не превышающего 4000-8000 токенов, и

обеспечение универсальности применения в различных предметных областях. Разработка и внедрение подобного комплексного подхода позволят вывести взаимодействие с искусственным интеллектом на новый уровень надежности и эффективности.

Таблица 1 – Сравнительная таблица эффективности техник управления поведением генеративных моделей.

Техника	Точность / Релевантность	Затраты времени	Экономия токенов	Универсальность	Сложность реализации	Когнитивная нагрузка на пользователя	Лучшие сценарии применения
Zero-Shot	Низкая (25-40%)	Минимальные	Высокая	Универсальная	Очень низкая	Низкая	Простые фактологические запросы
Few-Shot	Средняя (50-70%)	Средние	Низкая	Ограниченная	Низкая	Средняя	Задачи с четкими паттернами
Role Assignment	Средняя (60-75%)	Низкие	Средняя	Средняя	Низкая	Низкая	Специализированные консультации
Chain of Thought	Высокая (70-85%)	Средние	Низкая	Средняя	Средняя	Высокая	Логические и математические задачи
Self-Consistency	Очень высокая (80-90%)	Высокие	Низкая	Ограниченная	Высокая	Средняя	Критически важные решения
Context Expansion	Высокая (75-85%)	Высокие	Низкая	Средняя	Средняя	Высокая	Комплексные исследовательские задачи
Knowledge Generation	Очень высокая (80-92%)	Средние	Низкая	Средняя	Средняя	Средняя	Научные и медицинские вопросы
Directed Stimulus	Высокая (75-88%)	Низкие/Средние	Средняя	Широкая	Низкая	Высокая	Бизнес-аналитика и отчетность
Iterative Prompting	Наивысшая (90-95%)	Очень высокие	Низкая	Широкая	Переменная	Очень высокая	Творческие и исследовательские задачи
Tree of Thoughts	Наивысшая (92-97%)	Экстремально высокие	Очень низкая	Специализированная	Очень высокая	Высокая	Стратегическое планирование
ReAct Prompting	Очень высокая (85-93%)	Высокие	Низкая	Средняя	Высокая	Средняя	Задачи с актуальными данными

Список литературы

1. Ведомости. Экономический эффект от внедрения AI-решений в Сбере превысил 1 трлн рублей за 5 лет. 2024.
2. Brown T. B. et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems. 2020.
3. Vaswani A. et al. Attention is All You Need // Advances in Neural Information Processing Systems. 2017.
4. Косников С. Н., Золкин А. Л. Основы анализа данных и интеллектуальные системы. – М.: Лань, 2025.
5. Алексейчук А. С. Введение в нейронные сети: модели, методы и программные средства. – М.: МАИ, 2023.

References

1. Vedomosti. Jekonomicheskij jeffekt ot vnedrenija AI-reshenij v Sbere prevysil 1 trln rublej za 5 let. 2024.
2. Brown T. B. et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems. 2020.
3. Vaswani A. et al. Attention is All You Need // Advances in Neural Information Processing Systems. 2017.
4. Kosnikov S. N., Zolkin A. L. Osnovy analiza dannyh i intellektual'nye sistemy. – M.: Lan', 2025.
5. Aleksejchuk A. S. Vvedenie v nejronnye seti: modeli, metody i programmnye sredstva. – M.: MAI, 2023.