

Список источников

1. Киевич, А. В. Неуправляемый госдолг США как главный индикатор возрастающей турбулентности и глобальной неопределенности / А. В. Киевич // Банковская система: устойчивость и перспективы развития : сб. науч. статей XIII междунар. науч.-практич. конфер. по вопросам финансовой и банковской экономики, Пинск, 28 окт. 2022 г. : в 2 ч. / Мин-во образования Респ. Беларусь [и др.] ; редкол.: В. И. Дунай [и др.]. – Пинск : ПолесГУ, 2022. – Ч. 1. – С. 103–109.
2. Kievich, A. V. De-dollarization of the world economy as the factor of the aggressive politis of the USA / A. V. Kievich // в сб. : Банковская система: устойчивость и перспективы развития // сб. науч. статей IX междунар. науч.-практич. конфер. по вопр. банковской экономики. – 2018. – С. 58–62.
3. Киевич, А. В. Отрезвляющая реальность: глобальная экономика переживает худший кризис / А. В. Киевич // Современные аспекты экономики. – 2019. – № 5 (261). – С. 189–194.
4. Может ли цена золота установить новые рекорды: прогнозы на второй квартал [Электронный ресурс]. – Режим доступа: <https://quote.ru/news/article/643961c19a79472796e08273>. – Дата доступа: 10.04.2024.
5. Золото обогнало индекс S&P 500 по доходности в 2023 году [Электронный ресурс]. – Режим доступа: <https://quote.ru/news/article/653b61569a79473f03e10b26>. – Дата доступа : 11.04.2024.
6. Цена на золото с 1970 по 2023 год [Электронный ресурс]. – Режим доступа: https://goldomania.ru/menu_003_002html. – Дата доступа: 17.04.2024.
7. «Ультрабычий» сценарий: почему плавает золото и что с ним будет дальше [Электронный ресурс]. – Режим доступа: https://pro.rbc.ru/demo/661641089a7947bb85b3a56b?from=from_main_5. – Дата доступа: 12.03.2024.
8. Что ждет рынок золота в 2024 году: прогнозы экспертов [Электронный ресурс]. – Режим доступа: <https://quote.ru/news/article/6569c0f49a79477fd0849ccc>. – Дата доступа: 17.04.2024.
9. US debt clock: real time [Электронный ресурс]. – Режим доступа: <http://www.usdebtclock.org/>. – Дата доступа: 15.04.2024.

PROMPT-ИНЖИНИРИНГА В ОБЕСПЕЧЕНИИ ЭФФЕКТИВНОГО ВЗАИМОДЕЙСТВИЯ БИЗНЕСА И ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

PROMPT ENGINEERING IN ENSURING EFFECTIVE INTERACTION BETWEEN BUSINESS AND GENERATIVE ARTIFICIAL INTELLIGENCE

Архипова Л. И.

*Белорусский государственный университет информатики и радиоэлектроники,
Минск, Республика Беларусь*

Arkhipova L. I.,

*Belarusian State University of Informatics and Radioelectronics,
Minsk, Republic of Belarus*

Аннотация. В данной статье рассматриваются вопросы практического использования инструментов генеративного искусственного интеллекта и его роли в бизнесе. Особое внимание уделено *prompt*-инжинирингу и процессам, обеспечивающим создание эффективного взаимодействия с моделями искусственного интеллекта. Значительное внимание уделено вопросам разработки промптов и их практической адаптации к целям и задачам пользователя.

Abstract. This article discusses the practical use of generative artificial intelligence tools and its role in business. Particular attention is paid to *prompt* engineering and processes that ensure the creation of effective interaction with artificial intelligence models. Considerable attention is paid to the development of prompts and their practical adaptation to the goals and objectives of the user.

Ключевые слова: ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ (GENAI), PROMPT-ИНЖИНИРИНГ, МОДЕЛИ ИИ, ТИПЫ ПРОМПТОВ.

Введение

Тема влияния генеративного искусственного интеллекта (GenAI) на развитие бизнеса сегодня остается, бесспорно, актуальной. Исследователей волнует следующее: насколько это влияние приведет к преобразующим изменениям или это будет постепенный прогресс. Около 80% специалистов оптимистично оценивают потенциальную ценность GenAI для бизнеса, даже несмотря на существование рисков, связанных с кибербезопасностью, защитой данных и степенью оценки достоверности информации, сгенерированной искусственным интеллектом. Область генеративного искусственного интеллекта разнообразна и включает в себя множество моделей, инструментов, приложений и инфраструктуры, поддерживаемых значительными инвестициями крупных технологических компаний и растущей сферой стартапов – развитие технологий и инструментов генеративного ИИ представлено в виде Gartner Hype Cycle на рисунке 1 [1–2].

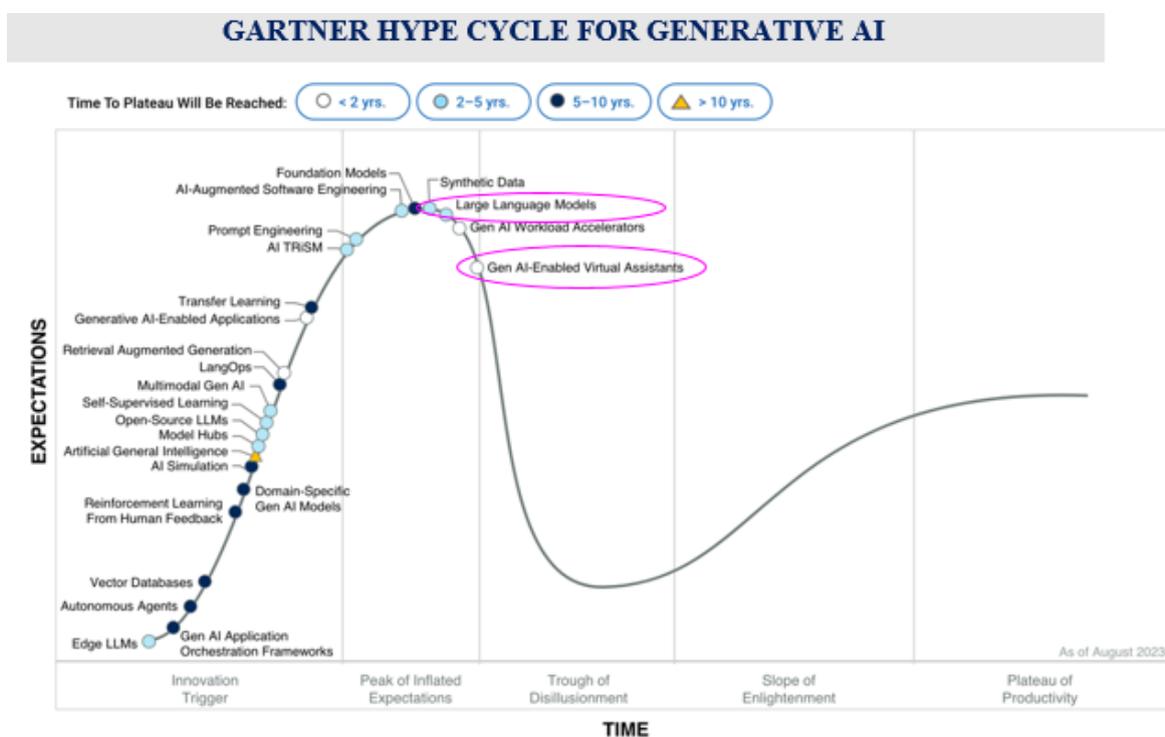


Рисунок 1 – Gartner Hype Cycle для генеративного искусственного интеллекта

Циклы развития Hype Cycle обеспечивают графическое и концептуальное представление, которое отражает путь технологий и инструментов GenAI – большинство технологий в настоящее время находятся на начальных этапах цикла и подходят к фазе ухода от иллюзий (фаза разочарования), после которой начинается переход и достижение зрелости и фазы продуктивности. Стоит отметить, что каждая технология движется с разной скоростью, что обозначается соответствующим значком на кривой цикл: позиция в цикле не является «плохой» – она только отражает зрелость технологии [3].

Практическое внедрение и применения технологий и инструментов генеративного ИИ в реальном бизнесе и других видах человеческой деятельности должно обеспечиваться специалистами, имеющими специфические знания, умения и навыки. Практика разработки и оптимизации запросов или инструкций для получения конкретных, релевантных и полезных ответов от генеративных моделей искусственного интеллекта получила название prompt-инжиниринг (prompt engineering). Фактически, prompt-инжиниринг, объединяет

физический и цифровой мир, преобразуя данные в информацию для принятия решений и практического внедрения. По этому поводу, аналитик Gartner (специалист в области взаимодействия данных и искусственного интеллекта С. Сикуляр) считает, что человек плюс искусственный интеллект умнее, чем любой из них по отдельности. Искусственный интеллект должен помогать организациям в цифровой трансформации для реализации трансформационных бизнес-идей.

Научно-исследовательская часть

В отечественных и зарубежных интернет-источниках представлено множество определений понятия **prompt-инжиниринг** (prompt engineering).

Prompt-инжиниринг – это процесс создания эффективных промптов для работы с моделями искусственного интеллекта, такими как языковые модели. Рассмотрим несколько определений этого понятия и его сущность [4–6].

Определение 1. Prompt-инжиниринг – это методология разработки оптимальных промптов, которые максимально эффективно ориентируют модели на генерацию желаемого контента или выполнение задачи. Сущность: основная идея заключается в том, чтобы создать промпты, которые точно и четко формулируют задачу для модели, минимизируя возможность ошибок или недопонимания. Путем тщательного выбора слов, фраз и структуры промпта можно направить модель на создание высококачественного контента, соответствующего требованиям пользователя.

Определение 2. Prompt-инжиниринг – это процесс оптимизации формулировки промптов для обучения или использования моделей искусственного интеллекта с целью улучшения их производительности и адаптации к конкретным задачам. Сущность: этот процесс включает в себя не только создание промптов, но и их доработку и оптимизацию в соответствии с требованиями и целями конкретного проекта. Оптимизированные промпты могут значительно повысить эффективность модели, сократить время обучения и улучшить качество результатов.

Определение 3. Prompt-инжиниринг – это методика адаптации промптов в соответствии с требованиями конкретного приложения или задачи с целью повышения производительности и точности модели. Сущность: задачи и контексты использования моделей искусственного интеллекта могут сильно различаться. Путем настройки промптов на конкретные потребности пользователей или бизнес-задачи можно добиться оптимальных результатов работы модели в конкретной области.

В целом сущность prompt-инжиниринга заключается в том, чтобы создать промпты, которые наилучшим образом соответствуют целям и требованиям задачи, что ведет к улучшению производительности и качества работы моделей искусственного интеллекта.

Prompt-инжиниринг играет решающую роль в повышении производительности и эффективности систем искусственного интеллекта, помогая им лучше понимать вводимые пользователем данные и генерировать более точные, актуальные и контекстуально соответствующие результаты. В prompt-инжиниринге основное внимание уделяется пониманию особенностей и «тонкостей» обработки естественного языка, а также возможностей моделей искусственного интеллекта [6].

Таким образом, ключевая роль prompt-инжиниринга заключается в следующем:

- обеспечивать повышение точности и актуальности ответов на запросы (enhanced accuracy and relevance);
- оптимизировать процедуры принятия решений (improved decision making);
- обеспечивать персонализированный клиентский опыт (personalized customer experiences);
- способствовать повышению эффективности использования ресурсов (efficient resource utilization);
- учитывать этические составляющие бизнеса и избегать предвзятости (ethical considerations and bias mitigation);
- способствовать адаптации к специфическим отраслевым нормам и требованиям (adaptability to industry-specific requirements).

Существенной составляющей успеха использования генеративного ИИ является процесс прогнозирования и подготовки к потенциальным изменениям с целью минимизации их влияния, что предполагает рассмотрение тенденций и потенциальных рисков, а также разработку стратегий и решений для их внедрения (рисунок 2) [6, 7].

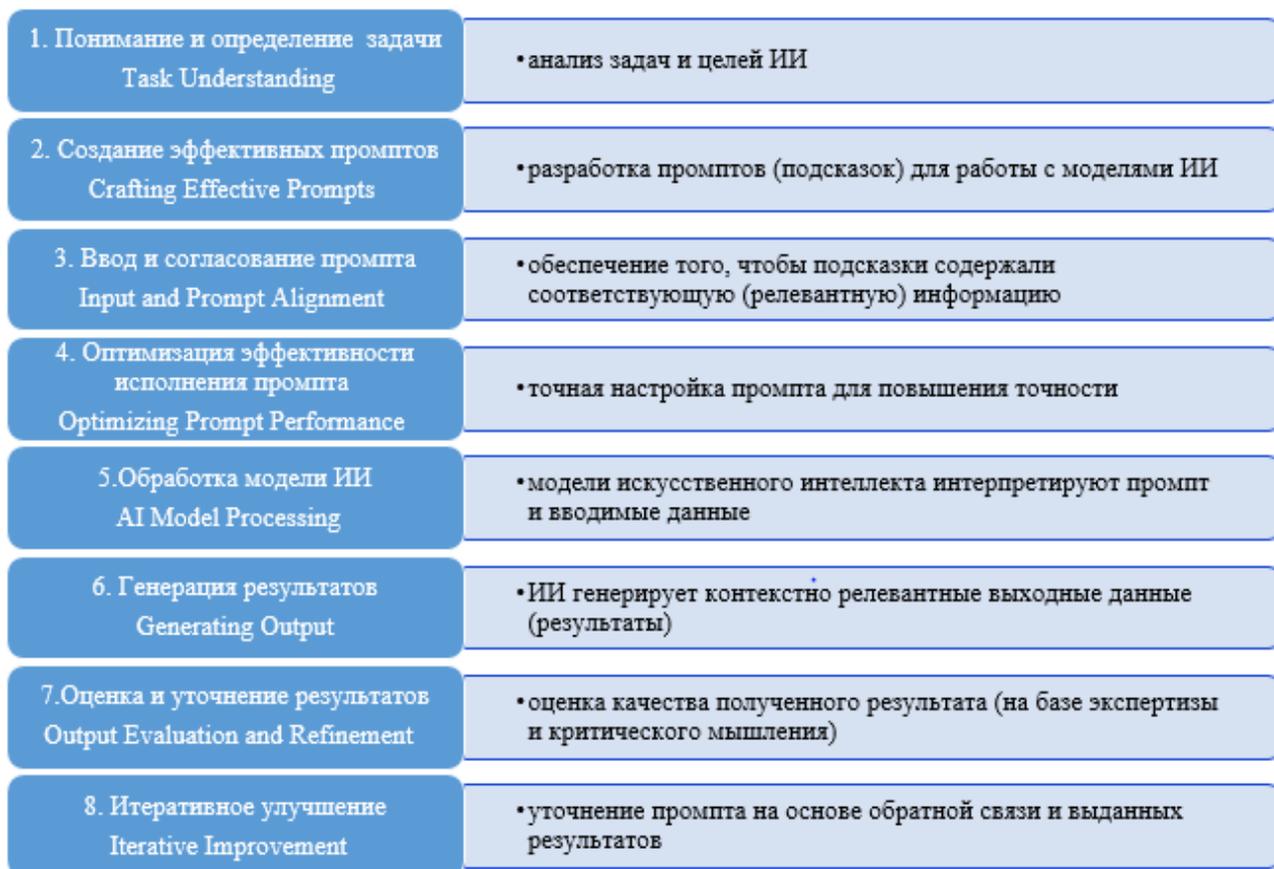


Рисунок 2 – Процесс создания промптов

Тщательная разработка промптов (запросов) должна помочь специалистам найти актуальную информацию для принятия решений и получения желаемых результатов. Достоверная информация, в свою очередь, позволит сократить неопределенность и повысит соответствие результатов бизнес-активностей ожиданиям пользователей.

Хорошо продуманные промпты должны помочь моделям ИИ быстрее и эффективнее учиться, что, фактически, способствует улучшению производительности и эффективности модели, а также сокращению времени и ресурсов. Благодаря «правильному» формированию запросов для нейросетей, учитывающих рекомендуемые prompt-инжинирингом форматы и правила, бизнес получает контролируемые результаты выдачи (output) по таким категориям, как содержание, формат, тон, стиль и др. [7, 8].

На рисунке 3 представлена обобщенная блок-схема внедрения prompt-инжиниринга в организации.



Рисунок 3 – Блок-схема внедрения prompt-инжиниринга

Сегодня уже можно говорить о наличии классификации промптов по типам, которые могут быть адаптированы для различных задач и ситуаций в зависимости от требований и целей пользователя. Далее представлено описание нескольких типов промптов, сгенерированных искусственным интеллектом по следующему запросу (рисунок 4):

<p>Запрос: В практике <i>prompt-инжиниринга</i> используются следующие типы промптов: <i>Zero-shot prompting</i>; <i>One-shot prompting</i>; <i>Few-shot prompting</i>; <i>Chain-of-thought prompting</i>. Сформулируй сущность и преимущества применения каждого из них.</p> <p style="text-align: right;">https://ru-chatgpt.ru/</p>
<p>Сгенерированный ответ:</p> <ol style="list-style-type: none">1. Zero-shot prompting: Сущность: При использовании <i>zero-shot prompting</i> модель генерирует текст, не требуя дополнительного контекста или обучения на конкретных примерах. Вместо этого модель получает общую инструкцию или вопрос, а затем создает соответствующий контент. Преимущества:<ul style="list-style-type: none">– Универсальность: Модель может генерировать тексты для различных задач и тем, даже если не была обучена на них напрямую.– Экономия времени: Не требуется предварительное обучение модели на больших объемах данных для каждой конкретной задачи.– Простота использования: Пользователю не нужно создавать множество примеров обучения или указывать дополнительный контекст для генерации текста.2. One-shot prompting: Сущность: <i>One-shot prompting</i> предполагает предоставление модели одного примера для обучения или генерации контента. Этот пример служит основой для того, чтобы модель могла создать соответствующий текст. Преимущества:<ul style="list-style-type: none">– Индивидуализация: Модель может быть настроена на конкретную задачу или стиль, используя лишь один пример.– Эффективность: Даже с небольшим количеством примеров модель может проявить хорошую производительность, создавая качественный контент.3. Few-shot prompting: Сущность: <i>Few-shot prompting</i> предполагает предоставление модели нескольких примеров (обычно, не более пяти) для обучения или создания контента. Эти примеры используются моделью для получения более глубокого понимания задачи или стиля. Преимущества:<ul style="list-style-type: none">– Большая гибкость: Пользователь может предоставить небольшое количество примеров для более точного управления тем, как модель создает контент.– Улучшенное качество: Благодаря большому количеству примеров модель может лучше понять контекст и создавать более качественный контент.4. Chain-of-thought prompting (CoT): Сущность: <i>Chain-of-thought prompting</i> предполагает использование последовательности промптов для создания непрерывного потока текста, где каждый следующий промпт строится на предыдущем контексте. Преимущества:<ul style="list-style-type: none">– Естественность: Последовательность промптов позволяет модели создавать текст, который логически следует за предыдущим, имитируя естественный разговор или повествование.– Контекстуальная связь: Модель использует предыдущие промпты для сохранения логической связи и последовательности в создаваемом тексте

Рисунок 4 – Текст, сгенерированный искусственным интеллектом (адаптирован автором с учетом цели и критического мышления)

Каждый из описанных типов промптов имеет свои особенности и преимущества, которые могут быть использованы в различных сценариях.

В качестве примера рассматривается структура промпта для наиболее продвинутой нейросети ChatGPT-4 (рисунок 5), который разработан по аналогии с иллюстрацией компании AltexSoft «Key elements of a prompt» [9].

КЛЮЧЕВЫЕ ЭЛЕМЕНТЫ ПРОМПТА (для взаимодействия с ChatGPT-4) <i>"Учитывая недавние исследования по изменению ВВП стран, обобщите основные выводы по данным прилагаемого отчета и представьте свое резюме в научном стиле".</i> <i>"Considering the recent research on countries by GDP change, summarize the main finding in attached report and present your summary in scientific style"</i>			
INPUT ВХОДНЫЕ ДАННЫЕ <i>данные прилагаемого отчета</i> <i>..attached report</i>	CONTEXT КОНТЕКСТ <i>учитывая недавние исследования по изменению ВВП стран</i> <i>countries by GDP change</i>	INSTRUCTION ИНСТРУКЦИЯ <i>обобщите основные выводы по данным прилагаемого отчета</i> <i>summarize the main finding in the attached report</i>	OUTPUT INDICATOR ИНДИКАТОР ВЫВОДА <i>представьте свое резюме в научном стиле</i> <i>present your summary in scientific style</i>

Рисунок 5 – Ключевые элементы промпта

Сформулированный запрос на взаимодействие с генеративным искусственным интеллектом, как правило, представляет собой следующий набор элементов:

Инструкция – основной элемент подсказки, который сообщает модели, что вы ожидаете от нее. Инструкция должна четко описывать действие, которое вы просите выполнить модель.

Контекст – этот элемент обеспечивает фон или настройку, в которой должно произойти действие, декларированное в инструкции. Контекст помогает модели сформулировать свой ответ таким образом, чтобы он соответствовал задуманному сценарию, а также делает промпт более эффективным, сосредоточивая модель на конкретном предмете и теме, соответствующей цели запроса.

Входные данные – это конкретная часть информации, которую вы хотите, чтобы модель учитывала при генерации выходных данных. Это может быть фрагмент текста, документ, набор чисел или другие данные, которые модель должна обработать или учесть.

Индикатор вывода – указывает модели на формат или стиль, в котором вы хотите получить ответ. Это может быть особенно полезно в случаях и сценариях, где формат вывода имеет такое же значение, как и его содержимое.

Точная разработка промпта фактически ставит задачу по прототипированию будущего результата (сгенерированного ответа), то есть ответа, который вы ожидаете получить от модели искусственного интеллекта.

В prompt-инжиниринге принята условная классификация типов промптов (подсказок по запросам), которые идентифицируются по сложности решаемых задач и глубиной взаимодействия с ИИ: Zero-shot prompting; One-shot prompting; Few-shot prompting; Chain-of-thought prompting (CoT) [10].

Далее сформулированы ключевые факторы успеха prompt-инжиниринга, которые представлены в виде общих рекомендаций, существенно влияющих на качество получаемых результатов (output). Эти рекомендации можно применять в различных техниках и комбинациях.

Ясность, четкость, краткость – конкретизируйте промпты, так как, чем конкретнее подсказки, тем лучше ИИ сможет понять запрос, тем более точным будет выходной результат.

Информационная перегрузка – максимально избегайте перегрузки, уходите от соблазна излишней детализации, так как избыток информации может оказаться контрпродуктивным и «запутать» модель.

Контекст и ограничения – добавляйте, по возможности, конкретные ограничения, что поможет конкретизировать (сузить) ответ на ваши потребности (запросы), а также запрещайте определенный «негатив» в качестве ограничений.

Вопросы – избегайте наводящих и открытых вопросов, по которым надо отыскивать баланс, так как такие вопросы могут исказить результат или дать слишком общий ответ.

Итерация и точная настройка – являются неотъемлемой частью процесса prompt-инжиниринга, они полезны независимо от целей и задач взаимодействия с ИИ и являются необходимыми действиями для получения желаемого результата [7–9].

Prompt-инжиниринг также играет значительную роль и в инновационной деятельности организации, поддерживая специалистов в поиске творческих идей и альтернативных решений, «работающих» на удовлетворение рыночных потребностей с учетом персонализированного клиентского опыта.

Типовые задачи prompt-инжиниринга реализуются prompt-инженерами, которые должны быть специалистами в бизнес-процессах, разработке IT-продуктов и нейросетях. К задачам prompt-инжиниринга относятся [8–10]:

- Разработка и внедрение AI-моделей для бизнеса: создание нейросетевых ассистентов/помощников для автоматизации создания контента; настройка моделей для автоматизации бизнес-процессов.
- Работа с API или использование open source решений: написание качественных запросов (промптов) для генеративных моделей; использование API для интеграции моделей ИИ в бизнес-процессы.
- Работа с базами данных: сбор, очистка и подготовка данных для обучения моделей ИИ; оптимизация структуры данных для взаимодействия с моделями ИИ.
- Внедрение ИИ в бизнес: аудит бизнес-процессов и продуктов на предмет возможности использования инструментов ИИ.

Примеры практических задач, которые могут быть решены в бизнесе с помощью prompt-инжиниринга [9–10]:

Генерация текста – создание статей, рецензий, кратких описаний товаров и услуг.

Ответы на вопросы – предоставление информации по запросу пользователей, например, в чат-ботах или поисковых системах.

Генерация кода – создание кода для решения задач программирования или автоматизации задач.

Классификация текста – определение категории или темы текста на основе предоставленного промпта.

Перевод текста – перевод текста с одного языка на другой с использованием контекста промпта.

Генерация идей – создание новых концепций, идей или сценариев на основе предоставленных данных.

Создание диалогов – написание диалогов между персонажами, сценариев для фильмов или игр.

Генерация музыки – создание музыкальных композиций на основе мелодических или стилистических указаний.

Генерация изображений – создание изображений или иллюстраций на основе описаний или указанных характеристик.

Решение математических задач – вычисление математических формул и решение задач на основе предоставленных данных.

Внедрение prompt-инжиниринга во взаимодействие с ИИ моделями дает ряд преимуществ: улучшение производительности модели ИИ (enhanced AI model performance); эффективное обучение модели (efficient model training); улучшение пользовательского опыта (better user experience); адаптацию к различным вариантам использования (adaptability to diverse use cases); непрерывное улучшение (continuous improvement).

Выводы

Различные отрасли и бизнес-активности имеют уникальные требования, цели и задачи. Prompt-инжиниринг позволяет точно настраивать модели ИИ с учетом отраслевых стандартов, гарантируя соответствие полученных результатов, коррелирующих и удовлетворяющих потребностям бизнеса. Высокая адаптивность помогает оптимизировать (дообучать) системы

искусственного интеллекта и максимизировать отдачу от его внедрения. Промпт в данном случае, являясь неотъемлемой частью процесса взаимодействия бизнеса (человека) с искусственным интеллектом, действует как промежуточный язык, переводящий намерения и ожидания человека в задачи, которые должен решить GenAI.

Изучение и применение инструментов prompt-инжиниринга, то есть понятных и простых подсказок для создания высококачественных результатов, сфокусировано на том, чтобы обеспечить доступность ИИ для большего числа пользователей, включая тех, у кого нет достаточного опыта использования искусственного интеллекта.

Список источников

1. The Generative AI Hype Cycle [Electronic resours]. – Mode of access: <https://pc.schhei.de/untitled/>. – Date of access: 07.05.2024.

2. Hype Cycle for AI Technologies in Business [Electronic resours]. – Mode of access: <https://omniscien.com/blog/hype-cycle-for-ai-technologies-in-business/>. – Date of access: 17.04.2024.

3. What’s New in Artificial Intelligence from 2023 Gartner Hype Cycle [Electronic resours]. – Mode of access: <https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle>. – Date of access: 07.05.2024.

4. Prompt Engineering: The Guide to Mastering the Art of Talking to AI [Electronic resours]. – Mode of access: <https://www.altexsoft.com/blog/prompt-engineering/>. – Date of access: 02.05.2024.

5. 7 Reasons Why Prompt Engineering Is Essential for Businesses [Electronic resours]. – Mode of access: <https://inclusioncloud.com/insights/blog/prompt-engineering-organizations/>. – Date of access: 02.05.2024.

6. Prompt Engineering [Electronic resours]. – Mode of access: <https://botpenguin.com/glossary/prompt-engineering>. – Date of access: 04.05.2024.

7. How to implement prompt engineering in your organization? [Electronic resours]. – Mode of access: <https://101blockchains.com/implement-prompt-engineering-in-organization/>. – Date of access: 28.04.2024.

8. A Step-by-Step Guide to Implementing Prompt Engineering in Your Organization [Electronic resours]. – Mode of access: <https://101blockchains.com/implement-prompt-engineering-in-organization/>. – Date of access: 01.05.2024.

9. Prompt Engineering: The Guide to Mastering the Art of Talking to AI [Electronic resours]. – Mode of access: <https://www.altexsoft.com/blog/prompt-engineering/>. – Date of access: 07.05.2024.

10. Prompt Engineering: The Guide to Mastering the Art of Talking to AI [Electronic resours]. – Mode of access: <https://www.altexsoft.com/blog/prompt-engineering/>. – Date of access: 04.05.2024.

МЕХАНИЗМ ФОРМИРОВАНИЯ ПОТРЕБИТЕЛЬСКОЙ ЦЕННОСТИ НА ОСНОВЕ РАБОТЫ С ОНЛАЙН-ОТЗЫВАМИ

THE MECHANISM OF FORMING CONSUMER VALUE BASED ON WORKING WITH ONLINE REVIEWS

Кисель Е. И., Назарук А. В.

*Брестский государственный технический университет,
г. Брест, Республика Беларусь*

Kisel E. I., Nazaruk A. V.
*Brest State Technical University,
Brest, Republic of Belarus*

Аннотация. В статье раскрыто значение работы с онлайн-отзывами как составляющей потребительской ценности товара. Описаны этапы создания ценности товаров в сознании