

Как ИИ меняет нашу жизнь и работу

Генеративный ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Нума Дхамани • Мэгги Энглер

УДК 004.8
ББК 32.813
Д86

Numa Dhamani, Maggie Engler
INTRODUCTION TO GENERATIVE AI

© Eksmo Publishing House 2025. Authorized translation of the English edition © 2024 Manning Publications. This translation is published and sold by permission of Manning Publications, the owner of all rights to publish and sell the same.

Книга содержит упоминания потенциально травмирующего контента (насилие, дискриминация, запрещенные вещества и т.д.) исключительно в контексте анализа проблем безопасности ИИ и методов их устранения. Не является пропагандой или одобрением подобного контента.

Дхамани, Нума.
Д86 Генеративный искусственный интеллект : как ИИ меняет нашу жизнь и работу / Нума Дхамани, Мэгги Энглер ; [перевод с английского Е. В. Жевлаковой, Ю. В. Войтко]. — Москва : Эксмо, 2026. — 384 с. — (Путеводитель по GPT и AI).

ISBN 978-5-04-240230-2

Книга «Генеративный искусственный интеллект: как ИИ меняет нашу жизнь и работу» погружает в мир современных технологий. Авторы подробно объясняют принципы работы больших языковых моделей, демонстрируя не только их потенциал, но и ограничения. Они уделяют внимание вопросам интеграции ИИ в повседневные процессы, показывая необходимость баланса между инновациями и ответственностью.

Читатели узнают о влиянии ИИ на общество, право и политику, а также о перспективах дальнейшего развития технологий. Книга раскрывает секреты лучших практик генерации текста и графики, подчеркивает важность защиты личных данных.

УДК 004.8
ББК 32.813

ISBN 978-5-04-240230-2

© Жевлакова Е. В., перевод на русский язык, 2025
© Оформление. ООО «Издательство «Эксмо», 2026

*Нума посвящает эту книгу своим родителям Назарали и Надье,
а также своему брату Нихалю*

Мэгги посвящает эту книгу своему мужу Джо

Оглавление

<i>Предисловие</i>	13
<i>Вступление</i>	15
<i>Благодарности</i>	17
<i>Об этой книге</i>	19
<i>Об авторах</i>	23
<i>Об иллюстрации на обложке</i>	25
<i>1. Большие языковые модели: вся мощь ИИ</i>	27
Как развивалась обработка естественного языка	29
Рождение LLM: все, что вам нужно, — это внимание	34
Бурное развитие больших языковых моделей	37
Где применяются LLM?	39
<i>Моделирование языка</i> 39 ■ <i>Генерация ответов</i> <i>на вопросы</i> 40 ■ <i>Программирование</i> 42 ■ <i>Генерация</i> <i>контента</i> 43 ■ <i>Логические рассуждения</i> 44 ■ <i>Другие</i> <i>задачи на естественном языке</i> 46	
В чем недостатки LLM?	47
<i>Обучающие данные и предвзятость</i> 48 ■ <i>Весьма</i> <i>правдоподобные, но неверные ответы</i> 52 ■ <i>Вопросы</i> <i>устойчивого развития больших языковых моделей</i> 53	
Революция в общении: LLM говорят с людьми	55
<i>ChatGPT от OpenAI</i> 55 ■ <i>Bard / LaMDA от Google</i> 57	

	<i>Искусственный интеллект Bing от Microsoft</i>	59	■	<i>LLaMa и Alpa</i>	61
Итоги					63
2. Обучение больших языковых моделей					65
Как происходит обучение LLM?					66
<i>Рассмотрим сбор открытых данных из интернета</i>					67
<i>Развеем мифы об авторегрессии и двунаправленном предсказании токенов</i>					69
■ <i>Тонкая настройка LLM</i>					71
Неожиданный поворот: эмерджентные свойства LLM					71
<i>Способный студент: обучение на небольшом количестве примеров</i>					72
■ <i>Эмерджентность — это иллюзия?</i>					76
Что попадает в обучающие данные?					77
<i>Закодированная предвзятость</i>					77
■ <i>Конфиденциальная информация</i>					82
Итоги					84
3. Конфиденциальность и безопасность данных в аспекте LLM					86
Как усовершенствовать генерирования в LLM, сделав упор на безопасность					87
<i>Алгоритмы обнаружения на этапе постобработки</i>					89
<i>Фильтрация контента или условное предобучение</i>					90
<i>Обучение с подкреплением на основе обратной связи от человека</i>					92
■ <i>Обучение с подкреплением на основе обратной связи от ИИ</i>					94
Разбираемся с конфиденциальностью пользователей и коммерческими рисками					97
<i>Непреднамеренная утечка данных</i>					98
■ <i>Как следует вести себя в беседе с чат-ботом</i>					100
Изучаем ПДД: политика и нормативные положения в области обработки данных					101
<i>Международные стандарты и законы о защите данных</i>					101
■ <i>Соответствуют ли чат-боты требованиям Общего регламента ЕС по защите персональных данных?</i>					105
■ <i>Правила конфиденциальности в области образования</i>					107
<i>Корпоративные политики</i>					108
Итоги					109

4. Развитие генерируемого контента	110
Появление синтетических медиа	111
Популярные методы создания синтетических медиа	112
Что плохого и что хорошего в синтетических медиа	115
ИИ или человек: как выявлять синтетические медиа	117
Как генеративный ИИ преобразует творческий процесс	120
Применение в маркетинге	120
Создание художественных работ	123
Интеллектуальная собственность в эпоху LLM	127
Закон об авторском праве и добросовестное использование	127
Свободное программное обеспечение и лицензии	135
Итоги	139
5. Ненадлежащее использование и вредоносные атаки	141
Кибербезопасность и социальная инженерия	142
Информационное смещение: как злонамеренно искажают факты	157
Политическая предвзятость и предвыборная агитация	168
Откуда у LLM галлюцинации?	172
Ненадлежащее использование LLM в профессиональных целях	180
Итоги	186
6. Повышаем производительность: ИИ как помощник	188
Как используют LLM люди разных профессий	189
LLM помогают врачам разобраться с бумажной волокитой	190
LLM применяют для правовых исследований, досудебного обмена информацией и подготовки документации	192
LLM оптимизируют портфельные инвестиции и обслуживание банковских клиентов	195
LLM как соавторы в творчестве	196
Пишем код вместе с LLM	199
LLM в повседневной жизни	203
Как генеративный ИИ влияет на образование	210
Выявляем машинно-генерированный текст	214
Как LLM воздействуют на рынок труда и экономику	219
Итоги	222

7. Устанавливаем социальные связи с чат-ботами	224
Чат-боты для социального взаимодействия	225
Почему люди вступают в отношения с чат-ботами	232
<i>Эпидемия одиночества</i> 232 ■ <i>Чат-боты в свете теории эмоциональной привязанности</i> 235	
Что хорошего и что плохого в отношениях людей с чат-ботами	238
Как прийти к благотворному взаимодействию с чат-ботом	246
Итоги	254
8. Что ждет ИИ и LLM в будущем?	255
В каком направлении развиваются LLM?	256
<i>Язык — универсальный интерфейс</i> 257 ■ <i>LLM-агенты открывают новые возможности</i> 258 ■ <i>Волна персонализации</i> 261	
Социальные и технические риски LLM	262
<i>Входные данные и ответы модели</i> 263	
<i>Конфиденциальность данных</i> 265 ■ <i>Злонамеренные атаки</i> 266 ■ <i>Ненадлежащее использование</i> 269	
<i>Как все это влияет на общество</i> 270	
Лучшие практики ответственного использования LLM	271
<i>Целенаправленно курируем наборы данных и стандартизируем документацию</i> 272 ■ <i>Защита конфиденциальности данных</i> 274 ■ <i>Объяснимость, прозрачность и предвзятость</i> 276 ■ <i>Стратегии обучения для более безопасных ответов</i> 282	
<i>Усовершенствование методов обнаружения</i> 285	
<i>Проблемы с метрикой «вовлеченность» и альтернативные метрики</i> 287 ■ <i>Люди в центре внимания</i> 289	
Правовое регулирование ИИ с точки зрения этики	291
<i>Обзор ситуации в Северной Америке</i> 291 ■ <i>Обзор ситуации в Европейском союзе</i> 296 ■ <i>Обзор ситуации в Китае</i> 301 ■ <i>Корпоративное самоуправление</i> 304	
На пути к системе регулирования в области ИИ	307
Итоги	310
9. Расширяем горизонты: изыскания в области ИИ	313
В поисках общего искусственного интеллекта	314
Способность чувствовать и сознание ИИ	323

Как LLM воздействуют на окружающую среду	330
Архитекторы будущего: сообщество свободного программного обеспечения	335
Итоги	341

<i>Приложение А. Ссылки на источники</i>	343
--	-----

Предисловие

Вы замечали, что все вокруг твердят, насколько хорош теперь ИИ? Люди бросаются модными терминами вроде «генеративный искусственный интеллект», «большая языковая модель» (LLM), «диалоговый агент» и тому подобное. Почему так происходит? Откуда это все взялось? Из-за чего так много определений? Разве они все не обозначают одно и то же? О чем-таки же все говорят? Что ж, у меня есть книга, которая нужна вам сейчас.

Как Нума, так и Мэгги имеют опыт работы в области моральных принципов. Они входят в состав «Института этики», аналитической группы и профессионального объединения специалистов, которые занимаются тем, что стараются понять, как и почему в интернете случаются нехорошие вещи, а также разрабатывают средства смягчения их последствий и ищут способы создать более здоровую среду онлайн. На протяжении своих карьер Нума и Мэгги разбирались во взаимодействиях в сети (сначала между людьми, а теперь между людьми и роботами) и фундаментальных законах того, что творится внутри этих невообразимо сложных систем, набитых личностями, пытающимися их сломать. Как оказалось, манера мышления авторов весьма хорошо подходит и для изучения того, как человечество будет использовать технологию генеративного ИИ, а также злоупотреблять ею. Через посредство «Института этики» Нума и Мэгги помогали нам просвещать и народные массы, и власть имущих о том, как функционирует интернет. Они

входят в состав растущего движения технарей, которые разъясняют обществу, что же на самом деле происходит в мире, где все социальное общение ведется онлайн. Важность их занятия возрастает по мере того, как люди проводят все больше времени в сети.

Появление данной книги меня воодушевляет. Я верю, что она станет частью новой волны произведений и исследований (предварительно назовем все это «наукой об этичности»), авторы которых работали в соцсетях, стремясь разобраться в информационных экосистемах, образуемых нашим поведением и отношениями между нами в интернете. Подобный метод мышления применим не только к социальным медиа и приложениям для знакомств или игр: с его помощью можно самыми разными способами уяснить суть и людей, и сведений. Чтобы читать эту книгу, вам не нужно ни быть фанатом статистики, ни выдавать себя за него, ни становиться им. Точно так же вам не придется смотреть на ИИ как на ящик, в котором сидит непостижимый волшебный робот. Нума и Мэгги устраивают для нас экскурсию по системам генеративного искусственного интеллекта, обеспечивая возможность рассуждать о них и принимать взвешенные решения, касающиеся их. Отталкиваясь от такой стартовой колодки, авторы ведут нас в дальнее странствие, где, опираясь на понимание этого новомодного ИИ и свои знания, доставшиеся тяжким трудом в окопах борьбы за этичность, распутывают вопрос о том, как генеративный ИИ повлияет на общество. Мы узнаем, как изменится экономика и сами наши разговоры, а также факторы, побуждающие нас к плохому поведению и распространению дезинформации.

Книга Нумы и Мэгги вышла в самый подходящий момент. Нам не обойтись без подобного руководства, в котором сложные концепции объясняются доступным языком. И хотя я уверен, что не все предсказания авторов сбудутся с точностью до буквы, вы, несомненно, не только погрузитесь в источник по-настоящему ценной информации о том, как в наше время действует генеративный ИИ, но и ознакомитесь с образом мыслей, отточенным за годы упорной работы на поприще сетевой этики. Прочтите эту книгу.

*Шахар Массачи,
один из основателей
и генеральный директор «Института этики»*

Вступление

По иронии судьбы, нас двоих свели безумные теории заговоров из интернета: мы встретились, когда проектировали системы обработки естественного языка, призванные оценивать и анализировать наполнение экстремистских ресурсов в сети. Когда в общественное сознание по всему миру вошли такие концепции, как LLM и другие модели генеративного ИИ, мы осознали, что наша область деятельности необратимо преобразится. Еще никогда люди не могли так дешево создавать и распространять информационные материалы, и в то же время еще никогда не возникало столь острой необходимости в классификации этих материалов в огромных масштабах.

В ходе работы над книгой мы получили весьма запоминающийся отзыв такого содержания: «Авторам следует прояснить свою позицию в отношении генеративного ИИ. Они за него или против?» Читатель, мы, к сожалению, не в силах втиснуть наши соображения по данному поводу в одно слово. Вместо этого мы постарались отразить на следующих страницах все тонкости возможных последствий развития и применения генеративного ИИ. Чтобы решить эту задачу, мы сначала постепенно разъясняем вам, каким способом и на каких данных обучаются LLM, а также рассматриваем алгоритмы, вносящие вклад в их конечный продукт — текст, практически неотличимый от тех, что пишут люди.

Выдаваемые ими материалы (а равно те, что создаются генеративными моделями иных типов) имеют целый ряд применений,

как благотворных, так и вредоносных. Ни одна из предыдущих систем не обладала подобными возможностями, однако за великолепными результатами генеративного ИИ на бенчмарках вроде прохождения типовых тестов порой скрываются их вопиющие недостатки, в том числе предвзятость, галлюцинации и формирование небезопасного контента. Помимо того, из-за их продукции встают серьезные вопросы о законных правах на контент, моральных принципах взаимодействия людей с ИИ, экономических параметрах работы при поддержке ИИ и очень многом другом.

Хотя мы попытались очертить наши позиции в данной книге, ссылаясь на научные статьи и примеры практического использования, у нас нет ни малейших иллюзий по поводу решения упомянутых проблем. Остается еще немало вопросов, и для ответа на них нужно запустить циклический процесс, в который будут вовлечены все слои общества. Соответственно, мы надеемся, что эта книга побудит новичков, увлеченных персон и опытных профессионалов принять участие в публичном обсуждении темы генеративного ИИ. На этом поприще до сих пор звучит слишком мало голосов, что оборачивается дискуссиями в узком кругу, на которых не принимают в расчет точки зрения обособленных групп (маргиналов), наемных рабочих, творческих личностей и деятелей культуры, а также бесчисленного множества других социальных категорий, на которые влияет искусственный интеллект. Просвещенный народ — наше главное орудие для постройки желаемого будущего в аспекте генеративного ИИ. Мы рассчитываем, что вы присоединитесь к нашим усилиям по созданию мира, где ИИ помогает людям, а не заменяет их, и ключевой ценностью остаются человеческие ощущения и переживания.

Благодарности

Мы хотели бы выразить сердечную признательность Шахару Массачи, чье обстоятельное предисловие заставляет задуматься и задает тон всей книге. Нас вдохновляют ваша увлеченность и приверженность работе в области этики, а этот проект стал еще более содержательным благодаря вашему вкладу.

Кроме того, данная книга не увидела бы свет, если бы не помощь и поддержка множества наших отзывчивых друзей и коллег. Не расставляя их в каком-то определенном порядке, мы благодарим Дэвида Салливана, Эрин Маколиф, Наталью Битюкову, доктора Дэниела Роджерса, Эдгара Маркевичюса, Сэма Планка, Дерека Слейтера, доктора Стива Крамера, Райана Уильямса, Брайана Джонса, доктора Фаиза Дживани, Рида Кока, Уитни Нельсон, Рахима Макани, Элис Хансбергер, Карана Лалу, Ребекку Руппель, Майкла Уортона, доктора Атиша Агарвалу, Рона Грина, доктора Кеннета Р. Флейшмана и Стивена Штрауса. Все эти люди предоставили нам ценные отзывы и разнообразные точки зрения, которые существенно повлияли на идеи, изложенные в тексте.

Далее нам хотелось бы выразить признательность сотрудникам Manning, трудившимся над книгой. Мы особенно благодарны редактору-консультанту по аудиторрии Ребекке Джонсон, которая направляла нас от начала до конца, высказывала замечания и пожелания, а также координировала все движущиеся части издательского механизма, и редактору по контрактам Энди