

# ОН ПРОСТО ДОБАВЛЯЕТ ПО ОДНОМУ СЛОВУ ЗА РАЗ

Тот факт, что ChatGPT может автоматически генерировать текст, который выглядит так, словно написан человеком, поражает наше воображение. Большинство людей недоумевают, как чат-бот выполняет интеллектуальную работу, которая прежде считалась доступной исключительно человеку. Как он это делает? Моя цель — дать вам хотя бы общее представление о том, что происходит внутри ChatGPT, а затем исследовать, как ему удается так хорошо создавать то, что выглядит как осмысленный текст. Хочу сразу сказать, что я собираюсь сосредоточиться на общих принципах работы ChatGPT и, хотя буду упоминать некоторые технические детали, не стану в них углубляться. (Все, о чем я буду говорить, применимо и к другим большим языковым моделям, подобным ChatGPT.)

[Почитать описание, рецензии и купить на сайте МИФа](#)

Первое, что нужно объяснить, — ChatGPT всегда пытается создать «разумное продолжение» любого текста, который у него есть на данный момент. Под словами «разумное продолжение» мы подразумеваем «то, что можно ожидать, исходя из того, что люди уже написали на миллиардах веб-страниц».

Итак, давайте предположим, что у нас есть предложение «Самое лучшее в ИИ — это его способность...». Представьте, что вы прочитываете миллиарды страниц текста, написанных людьми (скажем, веб-страницы в интернете и оцифрованные книги), и находите все аналогичные предложения, а затем смотрите, какое слово чаще всего встречается в них на месте многоточия. ChatGPT делает что-то подобное, за исключением того, что (как я объясню позже) он не просматривает тексты буквально — он ищет фрагменты, которые «совпадают по смыслу». Результатом его работы является список слов, которые логически могут следовать за основным словом, ранжированных по степени вероятности их присутствия в этом предложении:

Самое лучшее в ИИ — это его способность... (англ. The best thing about AI is its ability to...)	...учиться (англ. learn)	4,5%
	...предсказывать (англ. predict)	3,5%
	...создавать (англ. make)	3,2%
	...понимать (англ. understand)	3,1%
	...делать (англ. do)	2,9%

Примечательно, что, когда ChatGPT пытается написать эссе, по сути он просто снова и снова спрашивает: «Учитывая текст на данный момент, каким должно быть следующее слово?» — и каждый раз добавляет это новое слово. (Точнее, он добавляет так называемый токен, который может быть и частью слова; именно поэтому ChatGPT иногда может «составлять новые слова». Но об этом позже.)

Итак, на каждом этапе ChatGPT создает список слов с указанием вероятности их присутствия на данном месте. Но какое из них нужно выбрать, например, для эссе? Можно предположить, что это должно быть слово с самым высоким рейтингом (то есть то, для которого определена самая высокая вероятность). Однако именно здесь ChatGPT начинает проявлять свои вуду-способности. Потому что по какой-то причине (возможно, однажды мы ее даже узнаем), если всегда будем выбирать слово с самым высоким рейтингом, мы получим очень гладенькое эссе без малейшего признака креативности (и которое будет слово в слово повторять множество других текстов). Но если мы наугад выбираем слова с более низким рейтингом, то эссе получается «более интересное».

Здесь действует фактор случайности, а это значит, что, даже используя каждый раз один и тот же

промпт<sup>\*</sup>, мы, скорее всего, получим разные эссе. И, как мы уже упоминали, у ChatGPT есть своя вуду-идея, то есть у него существует так называемый температурный параметр, который определяет, как будут использоваться слова с более низким рейтингом, и для создания эссе, оказывается, лучше всего подходит температура, равная 0,8. (Нужно подчеркнуть, что это никакая не теория — это просто вывод из того, что работает на практике. И например, понятие «температура» применяется потому, что здесь используются экспоненциальные распределения, знакомые нам из статистической физики, но нет никакой физической связи — по крайней мере насколько нам известно.)

Прежде чем мы продолжим, я должен объяснить, что ради простоты изложения не стану рассматривать всю систему, лежащую в основе ChatGPT. Вместо этого буду работать с более простой системой — GPT-2, которую можно запустить на обычном персональном компьютере. И поэтому практически для всего, что вам здесь покажу, я включу код на языке Wolfram, который вы сможете немедленно запустить на своем компьютере.

---

<sup>\*</sup> Промпт (от *англ.* prompt — «побуждать») — запрос, подсказка или инструкция, которые вы даете искусственному интеллекту.  
*Прим. пер.*

Например, вот как получить таблицу вероятностей, приведенную выше. Во-первых, мы должны иметь базовую нейронную сеть языковой модели:

```
In[ ]:= model = NetModel[{"GPT-2 Transformer Trained
on WebText Data", "Task" → "LanguageModeling"}]
```

```
Out[ ]= NetChain[  Input port: string
Output port: class ]
```

Позже мы заглянем внутрь этой нейронной сети и поговорим о том, как она работает. Но пока можем просто применить эту модель к нашему тексту в качестве черного ящика и запросить пять наиболее подходящих слов, которые, по мнению модели, должны следовать за указанным текстом:

```
In[ ]:= model["The best thing about AI is its ability to",
{"TopProbabilities", 5}]
```

```
Out[ ]= {do → 0,0288508, understand → 0,0307805,
make → 0,0319072, predict → 0,0349748, learn → 0,0445305}
```

Затем эти строки превращаются в отформатированный набор данных:

```
In[ ]:= Dataset[ReverseSort[Association[%]],
ItemDisplayFunction → (PercentForm[#, 2] &)]
```

Out[ ]=

learn	4,5%
predict	3,5%
make	3,2%
understand	3,1%
do	2,9%

Вот что произойдет, если многократно «применять модель» — на каждом шаге добавлять слово, имеющее наибольшую вероятность (в коде это слово decision):

```
In[ ]:= NestList[StringJoin[#, model[#, "Decision"]] &,
"The best thing about AI is its ability to", 7]
```

```
Out[ ] = {Самое лучшее в ИИ — это его способность,
Самое лучшее в ИИ — это его способность учиться,
Самое лучшее в ИИ — это его способность учиться на,
Самое лучшее в ИИ — это его способность учиться
на собственном опыте,
Самое лучшее в ИИ — это его способность учиться
на собственном опыте.,
Самое лучшее в ИИ — это его способность учиться
на собственном опыте. Это,
Самое лучшее в ИИ — это его способность учиться
на собственном опыте. Это не,
Самое лучшее в ИИ — это его способность учиться
на собственном опыте. Это не так.}
```

Что будет, если мы продолжим предложение? В этом случае (при нулевой температуре) полученный текст оказывается запутанным и избыточным повторами:

Самое лучшее в ИИ — это его способность учиться на собственном опыте. Это не просто вопрос обучения на своем опыте, это изучение окружающего вас мира. Искусственный интеллект — очень хороший пример. Это очень хороший пример того, как использовать искусственный интеллект для улучшения своей жизни. Это очень хороший пример того, как использовать искусственный интеллект для улучшения своей жизни. Искусственный интеллект — очень хороший пример того, как использовать его для улучшения своей жизни. Это очень хороший пример того, как использовать искусственный интеллект для

Но что, если вместо того, чтобы всегда выбирать самое популярное слово, мы выберем слова с меньшей вероятностью (с температурой 0,8)? Тогда может получиться следующий текст:

{Самое лучшее в ИИ — это его способность,  
Самое лучшее в ИИ — это его способность создавать,

Самое лучшее в ИИ — это его способность создавать миры,

Самое лучшее в ИИ — это его способность создавать миры, которые,

Самое лучшее в ИИ — это его способность создавать миры, которые являются,

Самое лучшее в ИИ — это его способность создавать миры, которые являются одновременно захватывающими,

Самое лучшее в ИИ — это его способность создавать миры, которые являются одновременно захватывающими,}

И каждый раз, когда мы будем выбирать случайные слова, текст окажется разным, как в следующих примерах:

Самое лучшее в ИИ — это его способность к обучению. Мне всегда нравилось

Самое лучшее в ИИ — это его способность по-настоящему проникать в ваш мир и просто

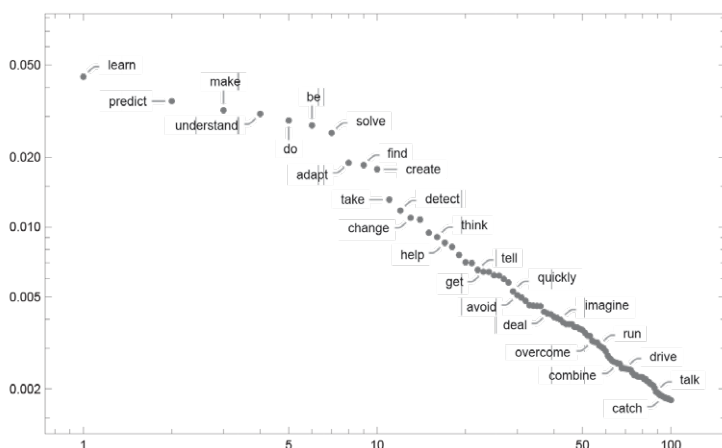
Самое лучшее в ИИ — это его способность изучать поведение человека и то, как он

Самое лучшее в ИИ — это его способность проделывать отличную работу по обучению нас

Самое лучшее в ИИ — это его способность создавать реальные задачи, но вы можете



Стоит отметить, что даже на первом этапе имеется множество возможных «следующих слов» (при температуре 0,8), хотя вероятность их использования падает достаточно быстро (данные на этом логарифмическом графике демонстрируют зависимость от  $n^{-1}$ , которая очень характерна для общей статистики языка):



Так что же произойдет, если мы продолжим создавать текст? Вот этот случайный пример уже лучше, чем пример с самыми популярными (нулевая температура) словами, но текст все еще немного странный:

Самое лучшее в ИИ — это его способность видеть насквозь и осмысливать мир вокруг нас, а не паниковать и игнорировать. Мы также можем сказать, что ИИ «выполняет свою работу» или «заурядный ИИ».

Действительно, делая бесконечное количество шагов, разработка машины, которая может быть интегрирована с другими системами или управляться одной системой, является одним из наиболее фундаментальных процессов ИИ. Помимо взаимодействия человека и машины ИИ также играет важную роль в творчестве.

Этот текст был написан простой моделью GPT-2 (2019 года выпуска). У более современной модели, GPT-3, результаты получаются намного лучше. Следующий текст был создан при помощи самых популярных (нулевая температура) слов с использованием того же самого промпта, но с помощью новой модели — GPT-3:

Самое лучшее в ИИ — это его способность автоматизировать процессы и принимать решения быстро и точно. ИИ можно использовать для автоматизации повседневных задач вроде ввода данных или для принятия сложных решений, таких как прогнозирование поведения клиентов или анализ больших наборов данных. ИИ также можно использовать для улучшения обслуживания клиентов, поскольку он позволяет быстро и точно реагировать на запросы клиентов. ИИ также можно использовать для повышения точности медицинских диагнозов и автоматизации процесса открытия новых лекарств.



[Почитать описание, рецензии  
и купить на сайте](#)

Лучшие цитаты из книг, бесплатные главы и новинки:

