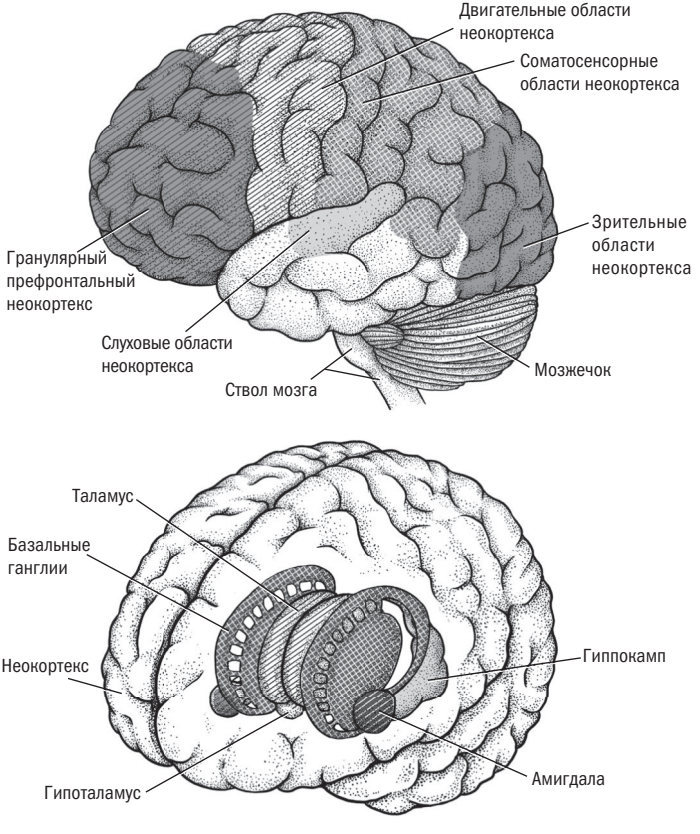
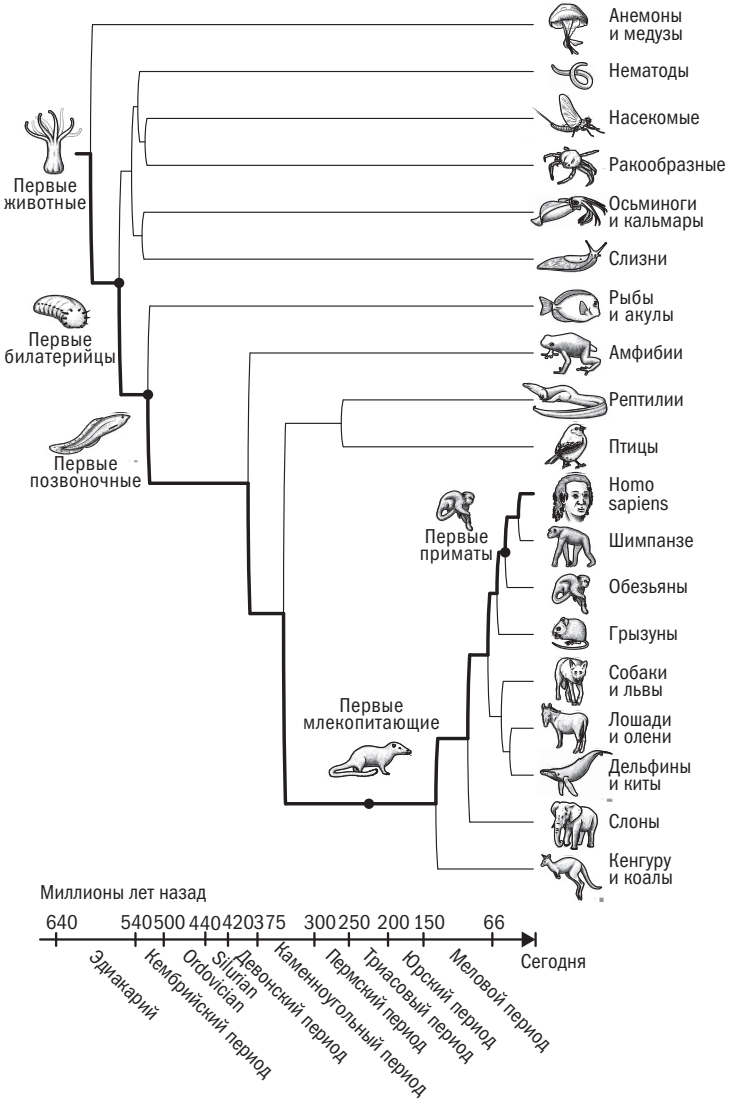
Оглавление

Всем, кто задумывается о том, что такое Вселенная — от огромного до самого малого — и обо всём, что между тем и другим.

Основы анатомии человеческого мозга



Наша эволюционная родословная



Введение

В сентябре 1962 года, когда весь мир обсуждал космическую гонку, кубинский ракетный кризис и недавно усовершенствованную вакцину от полиомиелита, произошло еще одно важное событие. Пресса писала о нем мало, хотя для человеческой истории это был значимый этап: мы предсказали будущее.

Тогда на экранах первых цветных телевизоров американцы увидели первые серии мультфильма «Джетсоны» о семье, которая живет в будущем. Это комедийный сериал, но авторы его сценария показывали свои идеи, как в будущем будут жить люди и какие технологии будут у них в карманах и в их домах.

Создатели «Джетсонов» правильно предсказали появление видеоконференций, плоскопанельных телевизоров, сотовых телефонов, 3D-принтеров и умных часов. Все эти технологии в 1962 году казались фантастикой, но в 2022-ом ими уже пользовались миллионы людей.

А вот технология автономного робота Рози из этого сериала до сих пор проходит по разряду фантастики.

Рози была служанкой Джетсонов, которая следила за их домом и детьми. Когда шестилетний Элрой не мог справиться со школьным домашним заданием, Рози приходила ему на помощь. Когда 15-летняя Джуди готовилась к экзамену на знание правил дорожного движения, Рози давала ей уроки вождения.

Рози готовила еду и накрывала на стол. Она была преданной, внимательной и всегда веселой помощницей по хозяйству, которая понимала, когда в семье зреют конфликты, и умела прийти на помощь в нужный момент каждому члену семьи. В одной серии Рози расплакалась, когда Элрой прочитал ей свои стихи, посвященные маме, а в другой — даже влюбилась.

Короче говоря, у Рози был интеллект человека, она обладала не только здравым смыслом и моторными функциями, необходимыми для выполнения сложных задач в физическом мире, но и сочувствовала, видела перспективу и улаживала конфликты, без чего невозможно жить в обществе. Как сказала Джейн Джетсон: «Рози — это член нашей семьи»[01].

Хотя в «Джетсонах» правильно предсказано появление сотовых телефонов и умных часов, у нас до сих пор нет ничего похожего на Рози. Ни для кого не секрет, что первая компания, которая выпустит робота, способного загружать тарелки в посудомоечную машину, получит огромную прибыль, потому что до сих пор все попытки в этом направлении проваливались. Проблема тут связана не с механикой, а с интеллектом: робот должен правильно идентифицировать тарелки в мойке, аккуратно извлечь их и поставить в машину, ничего не разбив. Решение такой, казалось бы, простой задачи оказалось намного сложнее, чем думали раньше.

Разумеется, нельзя отрицать огромный прогресс искусственного интеллекта (Artificial Intelligence, ИИ) после 1962 года. Сейчас ИИ играет намного лучше человека во многие игры, включая шахматы и го, может распознавать опухоли на рентгеновских снимках не хуже профессионального рентгенолога. ИИ практически готов сесть за руль автомобилей с автономным управлением. А в последние несколько лет новые достижения в области больших языковых моделей позволяют таким продуктам, как запущенный осенью 2022 года ChatGPT (чат-бот, популярность которого стремительно растет) сочинять стихи, переводить с одного языка на другой и даже программировать. К неудовольствию школьных учителей, он может мгновенно написать хорошим языком оригинальное эссе практически на любую тему, которую задаст ему ленивый школьник. ChatGPT легко сдаст экзамен на адвоката, набрав больше баллов, чем 90% юристов.

Однако на протяжении всей истории достижений ИИ оставался открытым вопрос, насколько мы близки к созданию интеллекта человеческого уровня.

После первых успехов алгоритмов решения задач в 1960-х Марвин Мински, пионер ИИ, провозгласил, что «через три-восемь лет у нас будет машина с общим интеллектом среднего человека». Этого не произошло. После успехов экспертных систем в 1980-х журнал BusinessWeek объявил, что ИИ уже здесь, однако вскоре прогресс остановился. И вот теперь с развитием больших языковых моделей многие исследователи снова заявляют, что «игра окончена», потому что мы стоим на пороге создания ИИ уровня человека[02]. Но так ли это? Находимся ли мы на пороге создания аналога Рози, человекоподобного искусственного интеллекта, или же большие языковые мо-

дели — всего лишь новейшее достижение на долгом пути, который закончится только через много десятилетий?

По мере развития ИИ все труднее оценить наш прогресс в достижении этой цели. Если система ИИ превосходит человека в решении определенной задачи, означает ли это, что ИИ понял, как человек решает эту задачу? Действительно ли калькулятор, способный мгновенно вычислять, понимает математику? ChatGPT, сдавший экзамен на адвоката лучше, чем большинство юристов, разбирается в законодательстве на самом деле? Как мы можем увидеть разницу и в каких обстоятельствах (если они вообще есть) эта разница действительно имеет значение?

В 2021 году, более чем за год до выхода ChatGPT, я использовал его предшественника, большую языковую модель GPT-3. Она была обучена на огромном количестве текста (как весь интернет), а затем использовала накопленные знания, чтобы подобрать наиболее подходящий ответ на поставленный вопрос.

Например, я спрашивал: «По каким двум причинам собака может быть в плохом настроении?» Система отвечала: «Собака может быть в плохом настроении по двум причинам: если она голодна или если ей жарко».

Новая архитектура этих систем позволяла им формулировать ответы с поразительной на первый взгляд степенью интеллекта. Эти модели легко обобщали факты, о которых они читали (например, страницы «Википедии» о собаках и другие материалы о причинах плохого настроения), для ответа на новые вопросы, которые им раньше не задавали. В 2021 году я изучал возможности применения этих новых языковых моделей, например, для создания систем поддержки психического здоровья, улучше-

ния сервисного обслуживания клиентов или расширения возможностей доступа к медицинской информации.

Чем больше я взаимодействовал с GPT-3, тем больше меня поражали как его успехи, так и ошибки. В некоторых темах он был гениален, а иногда удивлял тупостью. Попросите GPT-3 написать эссе о картофелеводстве XVIII века и его связи с глобализацией. Он выдаст удивительно понятный текст. Задайте ему простой вопрос: что можно увидеть в подвале, и он ответит бессмыслицей (подробнее об этом в главе 22). Почему GPT-3 правильно выполняет одни задания и не справляется с другими? Какие черты человеческого интеллекта он копирует, а какие упускает? И почему по мере ускорения развития ИИ теперь GPT-3 легко находит нужный ответ на вопросы, на которые раньше не мог?

Когда я завершал работу над этой книгой в начале 2023 года, вышла усовершенствованная версия GPT-3 под названием GPT-4. Она правильно отвечает на многие вопросы, которые ставили в тупик его предшественника. И все же, как мы увидим позже, GPT-4 не в состоянии понять важные особенности человеческого интеллекта, связанные с процессами в головном мозге.

Отличия искусственного интеллекта от человеческого порождают немало вопросов. Действительно, почему ИИ может победить лучшего шахматиста, но любой шестилетний ребенок лучше него загрузит тарелки в посудомоечную машину?

Нам трудно ответить: мы еще не понимаем, что именно пытаемся воспроизвести. Все эти вопросы, по сути, относятся не к ИИ, а к природе человеческого интеллекта: как он работает, почему именно так. И вскоре мы увидим самый главный ответ: как он появился на свет.

Подсказка от природы

Когда человечество захотело понять, как можно летать, оно стало изучать, как летают птицы. Жорж де Местраль придумал застежку-липучку, когда под микроскопом рассмотрел головки репейника и увидел крохотные крючки, которыми они цепляются за шерсть животных. Когда Бенджамин Франклин попытался изучить электричество, то начал с экспериментов с молнией.

Природа на протяжении всей истории инноваций была чудесным гидом человечества. Она также подсказывает нам, как работает интеллект. Самый яркий пример этого — конечно же, человеческий мозг. Но в этом смысле ИИ не похож на другие технологические инновации: мозг оказался более сложным для изучения принципов работы, чем крылья или молния. Ученые тысячелетиями изучают его, и, хотя они достигли определенного прогресса, у нас пока нет удовлетворительных ответов.

Проблема заключается в сложности.

Человеческий мозг содержит 86 миллиардов нейронов и более 100 триллионов связей. Каждое из этих соединений настолько мало (менее 30 нанометров в ширину), что их трудно разглядеть даже в самый мощный микроскоп. Все они плотно собраны в клубок, в одном кубическом миллиметре которого (ширина одной буквы на одноцентовой монете) находится более миллиарда связей[03].

Огромное количество связей — это лишь один из аспектов сложности мозга. Даже если бы мы составили карту «проводки» каждого нейрона, это ненамного приблизило бы нас к пониманию принципов работы мозга. В отличие от электрических соединений в ком-