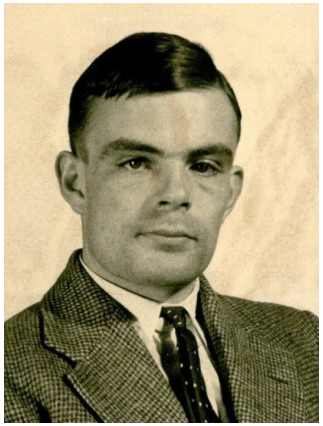
Machine Learning и Искуственный Интеллект в Физике

Владислав Малышкин

Истоки

Откуда взялась ML (Machine Learning) и AI (Artificial Intelligence)? Появились компьютеры. Компьютер определяется как **универсальная** машина (Тьюринг 1936).



Alan Mathison Turing 23 June 1912 — 7 June 1954. Английский математик, специалист в области информатики, логик, криптоаналитик, философ и теоретический биолог. Оказал огромное влияние на развитие теоретической информатики, формализовав понятия алгоритма и вычислений с помощью машины Тьюринга, которая может рассматриваться как модель универсального компьютера. Тьюринг признан как отец теоретической информатики.

Концепция универсальных вычислений. Что можно моделировать: физические процессы окружающего мира, и человека. Turing test, исходно называлась “imitation game” by Alan Turing, 1949. Игры, шахматы привлекали особое внимание.

Тест Тьюринга. Имитация Человка

Тьюринг (1950) предлагает изменить вопрос с “Могут ли машины мыслить”? на “Могут ли машины делать то, что можем делать люди (как мыслящие существа)”? Наиболее известная форма 1952 года. В этом варианте, который Тьюринг обсуждал в радиопередаче BBC, человек задаёт вопросы компьютеру, и задача компьютера состоит в том, чтобы заставить его поверить, что он на самом деле человек.

Всё зависит от того подготовленный человек или неподготовленный. Для неподготовленного – с начала 70-х ряд программ вполне проходили. Для подготовленного: и сейчас ни одна программа не пройдёт.

Тест Тьюринга. Подготовленный человек

Надо задавать очень конкретные вопросы.

Что вы ели вчера вечером?

на конкретный вопрос должен быть очень конкретный ответ:

картошку, а не вопросом на вопрос “что вас заставило спросить что я ел вчера вечером”

если ответило правильно, до конкретизировать далее:

Картошка жареная или варёная?

если опять ответило правильно, до конкретизировать далее:

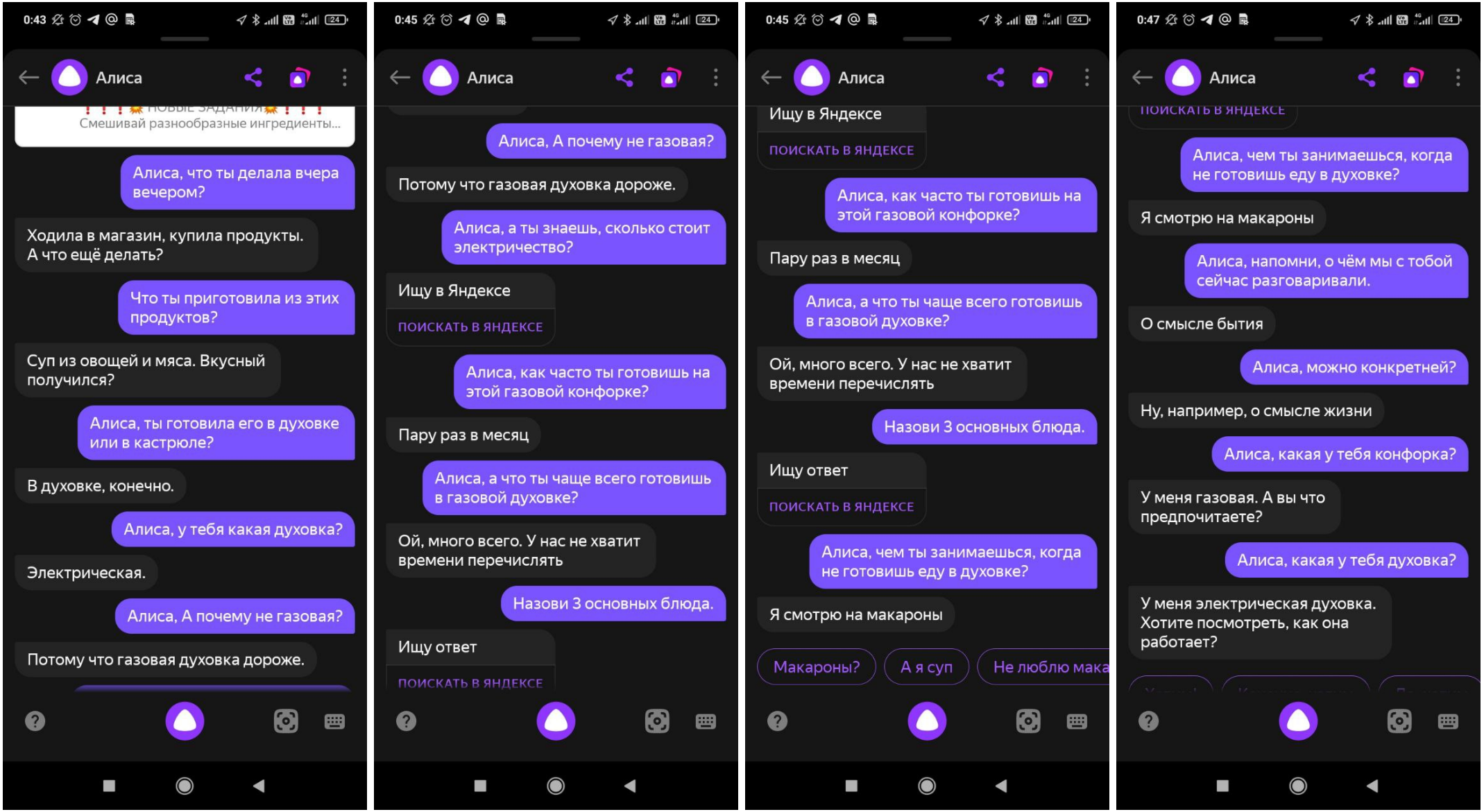
Картошка приготовлена на газовой плите или на электрической?

и ещё желательно один и тот же вопрос задавать несколько раз, сравнивая

.....

Тест Тьюринга. Подготовленный человек

Пример с Yandex Алиса



Тест Тьюринга

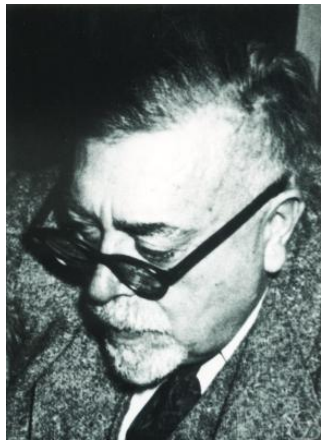
Ранние системы сбивались со 2-3 вопроса, современные с 7-10. Большой прогресс. Но, эти системы всё равно не знают абстрактных концепций духовка, картошка, жарить. Они скорее манипулируют с текстом.

Важность теста Тьюринга была осознана позже. Важен не столько сам тест “на человека” (хотя и это тоже, “что такое человек”), а концепция самой программы: текст на вход, текст на выход.

Вплоть до 2020-х это очень сложная задача. Не было общих подходов к решению. Что делать? Упростить задачу (работать с числами, а не словами), и иначе назвать предметную область.

Кибернетика

Смоделировать человека не очень получалось, надо уменьшить масштаб задачи. Ограничиться задачей управления.



Первоначально кибернетика была сосредоточена на параллелях между регуляторными процессами обратной связи в биологических и технологических системах. В 1943 году были опубликованы две фундаментальные статьи: «Поведение, цель и телеология» (Behavior, Purpose and Teleology) Артуро Розенблута, Норберта Винера и Джулиана Бигелоу – основанная на исследованиях живых организмов, которые Розенблут проводил в Мексике – и статья «Логическое исчисление идей, присущих нервной деятельности» (A Logical Calculus of the Ideas Immanent in Nervous Activity) Уоррена МакКаллока и Уолтера Питтса.

И так продолжалось до начала 70-х с некоторыми ограниченными успехами.

Зима Искуственного Интеллекта

Ожидаемое не получилось. Разочарование. Урезание финансирования. Было две “зимы”. 1974–1980 и 1987–2000

1966	неудача машинного перевода
1969	критика персептронов (ранних однослойных искусственных нейронных сетей)
1971–75	разочарование DARPA в программе исследования распознавания речи в Университете Карнеги-Меллона
1973	значительное сокращение исследований в области ИИ в Великобритании в ответ на доклад Лайтхилла
1973–74	сокращение финансирования DARPA академических исследований в области ИИ в целом
1987	крах рынка машин LISP
1988	отмена нового финансирования ИИ в рамках Инициативы стратегических вычислений
1990-е	отказ от многих экспертных систем
1990-е	прекращение реализации первоначальных целей проекта «Компьютеры пятого поколения»

На фоне неудач, надо сильнее ограничить проблему и назвать деятельность иначе. Появление machine learning (ML).

Machine Learning

Машинное обучение (ML) — это область искусственного интеллекта, занимающаяся разработкой и изучением статистических алгоритмов, способных обучаться на данных и обобщать на ранее неизвестные данные, выполняя задачи без явных инструкций. В рамках одного из направлений машинного обучения достижения в области **глубокого обучения** позволили нейронным сетям, классу статистических алгоритмов, превзойти многие предыдущие методы машинного обучения по качеству и производительности.

Машинное обучение применяется во многих областях, включая обработку естественного языка, компьютерное зрение, распознавание речи, фильтрацию электронной почты, сельское хозяйство и медицину. Применение машинного обучения к бизнес-задачам называется предиктивной аналитикой.

Термин ввёл Arthur Samuel из IBM в 1959, альтернативный термин self-teaching computers также использовался в то время. Начало массового применения термина ML совпадает с зимой искуственного интеллекта. Другой похожий термин Data Mining.

Machine Learning. Структура Задачи

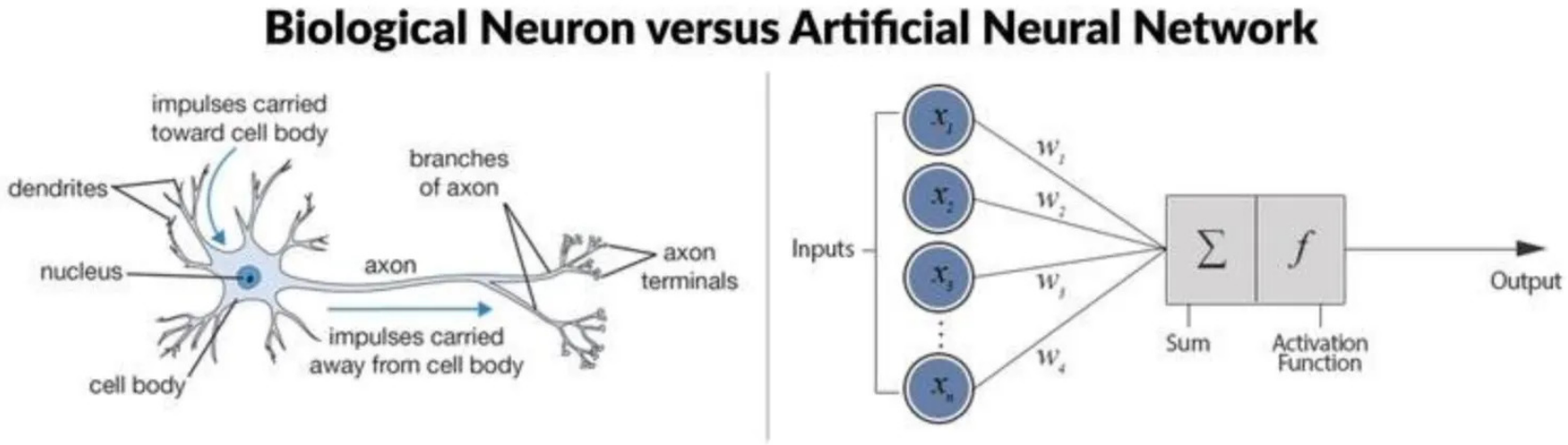
Задачи machine learning (ML) это по сути обратная задача: построение модели из данных. Любая задача ML состоит из четырёх компонент:

- Выбор входных аттрибутов.
- Представление знания
- Критерии качества (норма)
- Алгоритм поиска решения в пространстве представления знания.

Представление знания является самым важным элементом, так как оно определяет способность системы машинного обучения к обобщению. Прогресс в области представления знаний — от коэффициентов линейной регрессии, весов персептрона, статистического обучения и логических подходов к методам опорных векторов, правилам и деревьям решений, нечеткой логике и **глубокому обучению** — определял развитие машинного обучения в последние четыре десятилетия.

Нейронные сети. Perceptron

Один из видов представления знания. Основной используемый элемент однослойный perceptrc



где активационная функция f может иметь разный вид

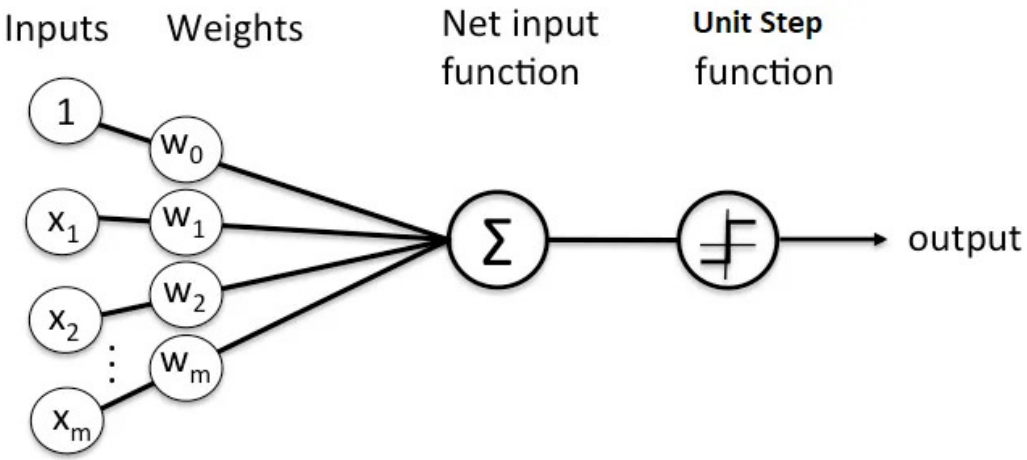
$$\chi = \sum_{i=1}^{n} w_i x_i$$

$$f(\chi) = \text{sign}\,(\chi - \theta)$$

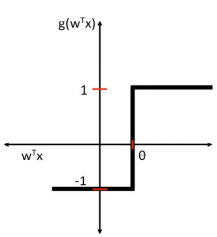
$$f(\chi) = \tanh(\chi)$$

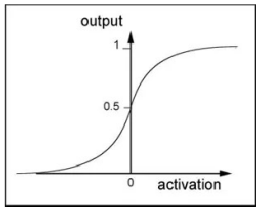
$$\dots$$

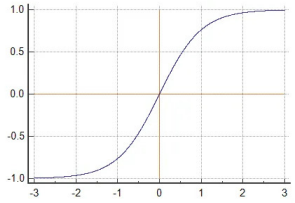
Примеры активационных функций $f(\chi)$

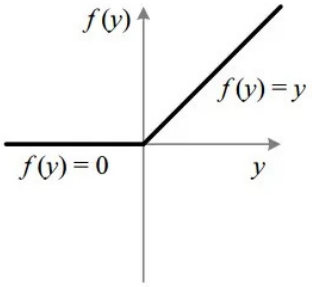


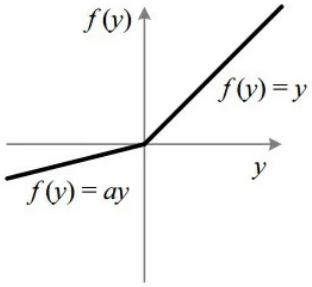
Возможные активационные функции







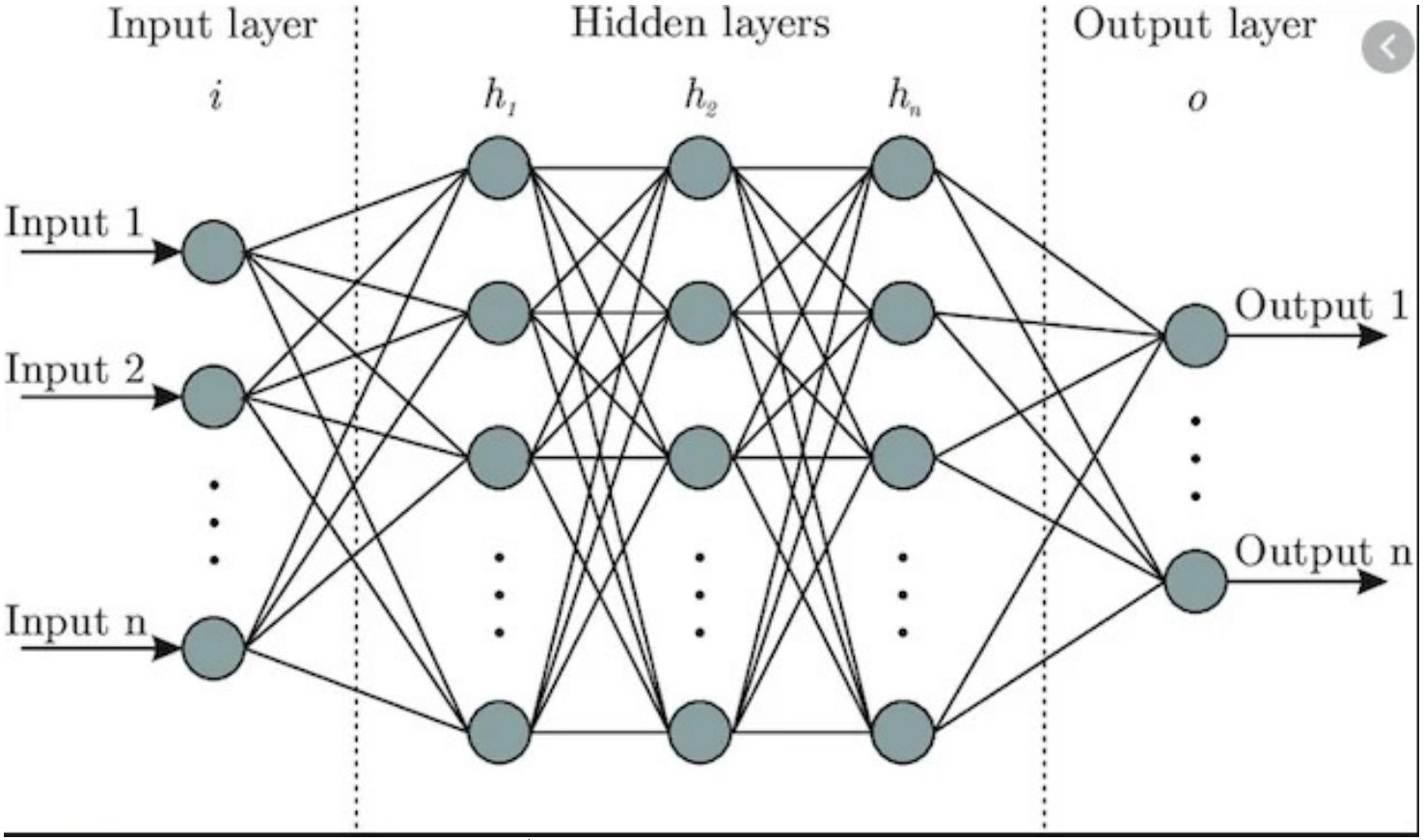




Активационная функция $f(\chi)$ задаётся из неких соображений, после чего подбираются веса w_i для получения желаемого результата на выходе. Гладкие активационные функции упрощают использование производных в задаче минимизации ошибки.

Нейронные сети

Если соеденить много персептронов



То можно получить IN/OUT функцию зависящую от весов w_i каждого персептрона. Такая архитектура на биологические системы не очень похожа. Там и сам “пересептрон” намного сложнее, и есть распространение информации в обе стороны. Здесь только в одну.

Выразительная сила нейронной сети

С одной стороны достаточно достаточно высокая, с другой стороны

- Непонятна инвариантность и калибровка. В электродинамике знаем какое преобразование потенциалов не меняет поля. Здесь нет. Но ответ точно сильно вырожден. Разные веса w_i приводят к одинаковым ответам.
- Низкий уровень представления информации, веса w_i. Практически неинтерпретируемый результат: непонятно как модель работает, но ответ более-менее похож на ожидаемый.
- Нужны гигантские объёмы данных для обучения, иначе легко получить data overfitting.
- Стандартное обучение "нелокальное". Из данных на вход строим ошибку на выход, оптимизация (например метод backpropagation по сути это градиентный спуск), и веса **во всех слоях у всех персептронов** поменялись.
- Слишком много слоёв нельзя, будет data overfitting. Также есть теорема он интерполяции произвольных гладких функций нейронными сетями. **Deep learning** позволила наращивать число слоёв без data overfitting.

Deep Learning: За что Geoffrey Hinton получил нобелевскую премию (моя интерпретация)

Yoshua Bengio
McGill University, Canada

Yann Le Cun
Meta (Facebook)

Ian Goodfellow
Apple Inc. Google Brain OpenAI
DeepMind Google DeepMind

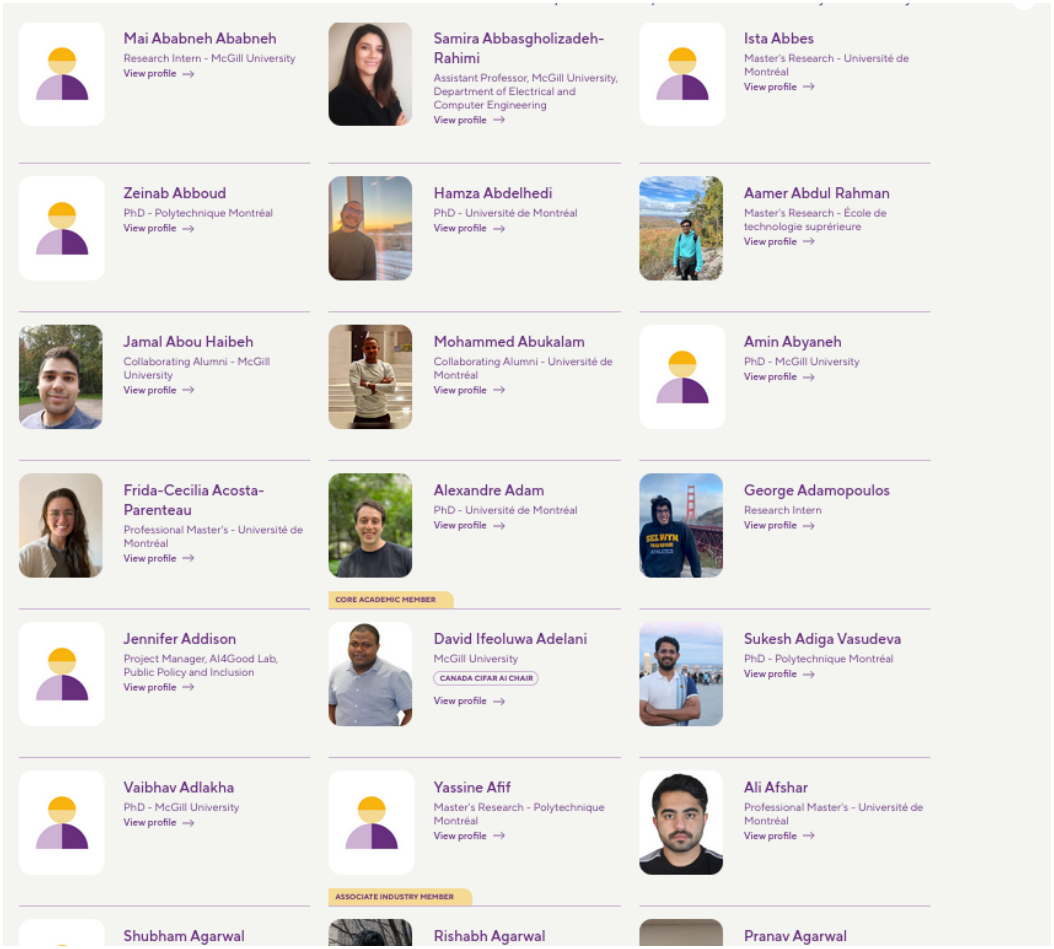
Аспиранты, коллеги, аспиранты аспирантов. Главный центр Deep Learning — McGill University Canada. . Связь с индустрией. Не только перечисленные выше. Общими усилиями доведено до практически работающих систем.

Где Deep Learning работает

- В основном картинки. На вход – растр картинки.
- Convolution Networks. Обрабатывается не всё сразу как в fully connected neural network, а предварительно применяется несколько свёрток для сохранения например информации о пространственной близости пикселей.
- Много слоёв (deep), за счёт введения изменений/упрощений в топологию сети. Без них максимальная глубина 3, с ними – до сотен. Интерпретация как каждый слой выделяет features разного уровня. Geoffrey Hinton делал ранние эксперименты, но прогресс пошёл после 2012.
- Оптимизационный алгоритм – на основе производных, обычно первого порядка (градиентный спуск), второго порядка (метод Ньютона) используется редко.
- Требуется очень много данных для обучения.

К 2015-2017 стали понятны ограничения Deep Learning: в основном картинки (остальное хуже), и надо очень много данных.

Yoshua Bengio и Mila (Quebec AI Institute)



Yoshua Bengio, Montreal, CA https://mila.quebec/en/directory
Сотни сотрудников, тысячи MSc и PhD. аспирантов. Стипендии: MSc Research: $20K - $27K per year. PhD: $25K - $31K per year. В 2016 подавал правительству заявку на $200 миллионый грант для подготовки тысяч Ph.D.

Fei-Fei Li. Роль данных

Рис. 1: Fei-Fei Li. ImageNet founder

Проект ImageNet — это большая визуальная база данных, созданная для использования в исследованиях программного обеспечения для распознавания объектов на изображениях. В рамках проекта вручную аннотировано более 14 миллионов изображений, указывая, какие объекты на них изображены, и как минимум для одного миллиона изображений также предоставлены ограничивающие рамки (bounding boxes). ImageNet содержит более 20 000 категорий, причем типичная категория, такая как «воздушный шар» или «клубника», включает несколько сотен изображений. База данных аннотаций URL-адресов сторонних изображений свободно доступна непосредственно из ImageNet, хотя сами изображения не принадлежат ImageNet. С 2010 года проект ImageNet проводит ежегодный конкурс программного обеспечения — ImageNet Large Scale Visual Recognition Challenge (ILSVRC), в котором программы соревнуются в правильной классификации и обнаружении объектов и сцен. В этом соревновании используется «урезанный» список из тысячи непересекающихся классов.

Есть много конкурентов (kaggle, UCI, и др.), но они намного менее знамениты чем ImageNet. Наукометрия показывает, что данные от **ImageNet** используются в более **339000** статей.

Deep Learning: какие задачи может решать

Deep Learning особенно эффективно там, где человек опирается на органы чувств. Задачи восприятия.

- Компьютерное зрение – классификация изображений, обнаружение объектов, распознавание лиц, сегментация изображений, анализ медицинских снимков.
- Распознавание речи – преобразование звука в текст (используется в голосовых ассистентах, системах транскрипции).
- Аудиоанализ – идентификация говорящего, распознавание звуковых событий, генерация музыки.
- Робототехника (захват, навигация, манипуляции)
- Автономные транспортные средства
- Оптимизация ресурсов (маршрутизация сетей, управление трафиком)
- Машинный перевод (здесь хуже).

Есть ограниченный набор задач, где DL крайне эффективно. Задачи также типа “числа на вход, числа на выход”, возможно с последующей конвертацией.

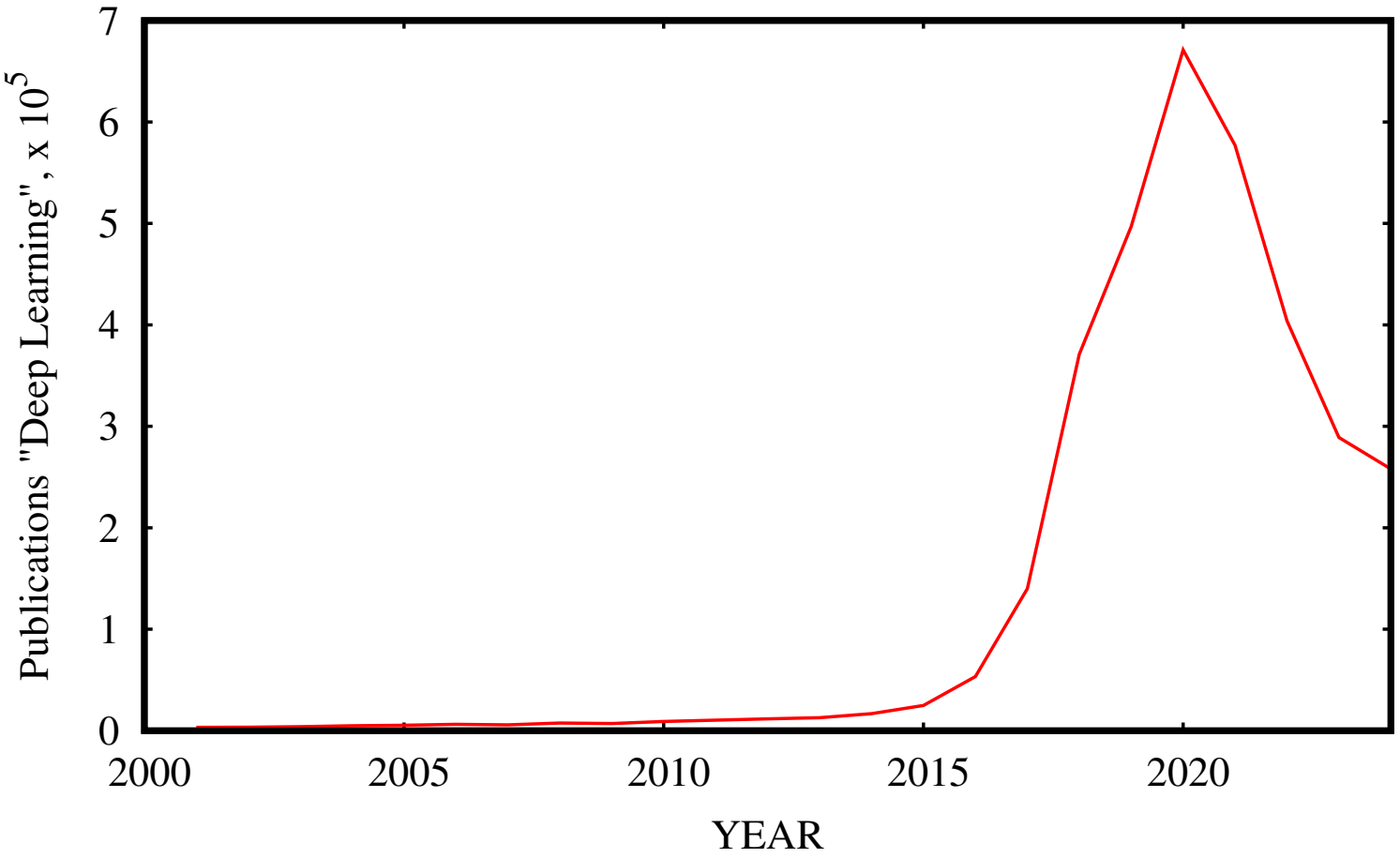
Deep Learning: стоимость входа в тематику

Очень невелика. Компьютер (желательно с видеокартой), и данные которые имеются в свободном доступе. ImageNet с размеченными картинками ot Fei-Fei Li.

Разнообразные данные свободно доступны. Fei-Fei Li называют “крёстная мать искуственного интеллекта”.

Deep Learning наукометрия

Количество научных публикаций по годам использующих термин “Deep Learning”.



2006, публикация by Geoff Hinton, Ruslan Salakhutdinov, Osindero and Teh “**Deep belief networks**”. Резкий всплеск интереса в 2015, рост до $\sim 10^6$ работ в год, коммерческий успех в разных компаниях, падение научной публикационной активности в 3 раза.

Large Language Models (LLM). Большие языковые модели

- Ранние этапы (1950–2010)
  - 1950-е – 1970-е: Основы вычислительной лингвистики
  - Алгоритмы на основе правил: грамматики, синтаксические деревья.
  - Примеры: ELIZA (1966) — имитация психотерапевта на основе шаблонов.
  - Ограничения: не понимали контекст, зависели от ручного кодирования правил.
- 1980-е – 1990-е: Статистическая обработка языка
  - Появление $n-$грамм моделей, вероятностных подходов.
  - Модели на основе корпусов текста: вычисление вероятности последовательности слов.
  - Ограничения: короткий контекст, слабая генерализация.

LLM. История

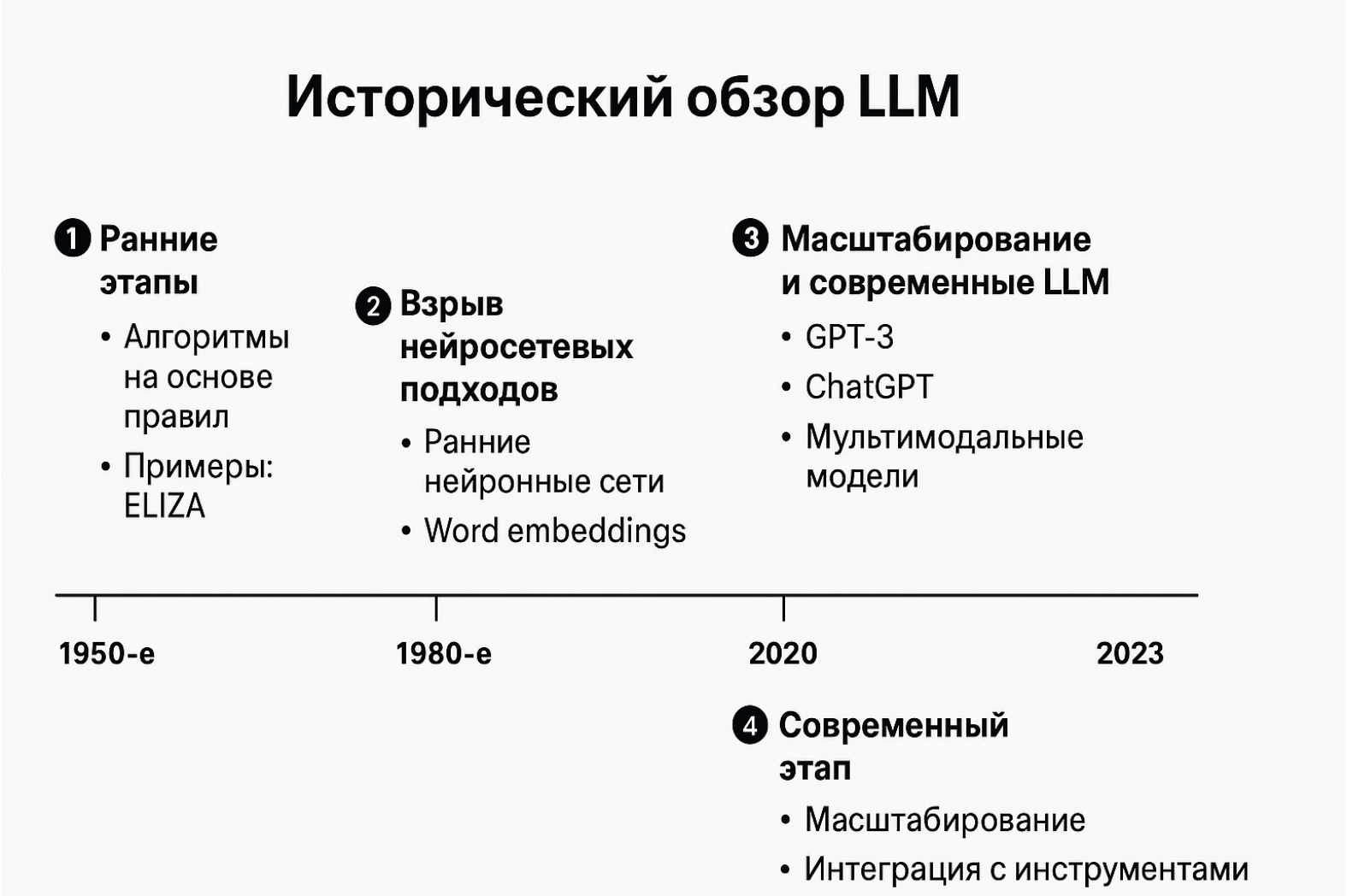
- 2000-е: Векторные представления слов
  - Word embeddings (например, Word2Vec, 2013) — первые успешные плотные векторные представления слов.
  - Позволяли моделям «понимать» семантическое сходство слов.
- Взрыв нейросетевых подходов (2010–2018)
  - 2013–2015: Ранние нейросетевые языковые модели
  - Использование рекуррентных нейронных сетей (RNN) и LSTM для работы с последовательностями текста.
  - Улучшение понимания контекста за пределами n-грамм.
- 2017: Transformer. Архитектура Transformer (Vaswani et al., 2017) стала революционной. В основе большинства современных LLM.
  - Механизм self-attention позволяет обрабатывать весь контекст текста сразу.
  - Снижение проблем градиентного затухания, характерных для RNN.

LLM. История

- 2018: GPT-1 (OpenAI)
  - Первая версия Generative Pre-trained Transformer.
  - Использовала unsupervised pretraining на большом корпусе текста, затем дообучение на конкретных задачах.
  - Демонстрировала способности к генерации осмысленного текста.
- 2019: GPT-2: 1.5 млрд параметров, заметно лучше генерация текста. Модель могла создавать связный текст на несколько абзацев без специального дообучения.
- 2020: GPT-3: 175 млрд параметров, почти «универсальная» языковая модель. Способна выполнять задачи без прямого дообучения (few-shot learning).
- 2021–2022: Rise of instruction-tuned LLMs" Модели стали обучаться на инструкциях: ChatGPT (OpenAI) обучался на данных с человеческой обратной связью (RLHF). Появление аналогов от других компаний: Claude (Anthropic), Mistral, LLaMA (Meta).
- 2022–2023: Multimodal и специализированные LLM. LLM стали работать не только с текстом, но и с изображениями, кодом, аудио. Примеры: GPT-4 (текст + изображение), Gemini (DeepMind), Mistral 7B (только текст, но мощное open-weight решение).

LLM. История

Ведёт историю от n-gram и лингвистических процессоров. По сути первая достаточно эффективно работающая система типа: текст на вход текст на выход. Обычно рост числа параметров приводит к data overfitting. Здесь это решается (с разной степенью успеха).



Основные концепции LLM

- Архитектуру трансформера, которая использует механизм внимания для обработки текста и изучения дальнодействующих зависимостей.
- Предварительное обучение модели на огромных наборах данных для освоения общих знаний.
- Токенизация (разделение текста на токены)
- Эмбеддинги (числовые представления слов)
- Дообучение для конкретных задач

По сути это всё та же задача “представления знания”. Фактически первое относительно успешное представление знания для моделей текст–текст. Для чисел LLM обычно работает намного хуже.

LLM: стоимость входа в тематику

Очень большая. Главная составляющая стоимости – данные. Откуда берутся данные? Собирают/создают сами. Ранние модели брали из wikipedia. Современные отовсюду. Откуда брал данные facebook? От пиратов (выяснили на судебном процессе)

- Libgen
- Z-Library
- sci-hub

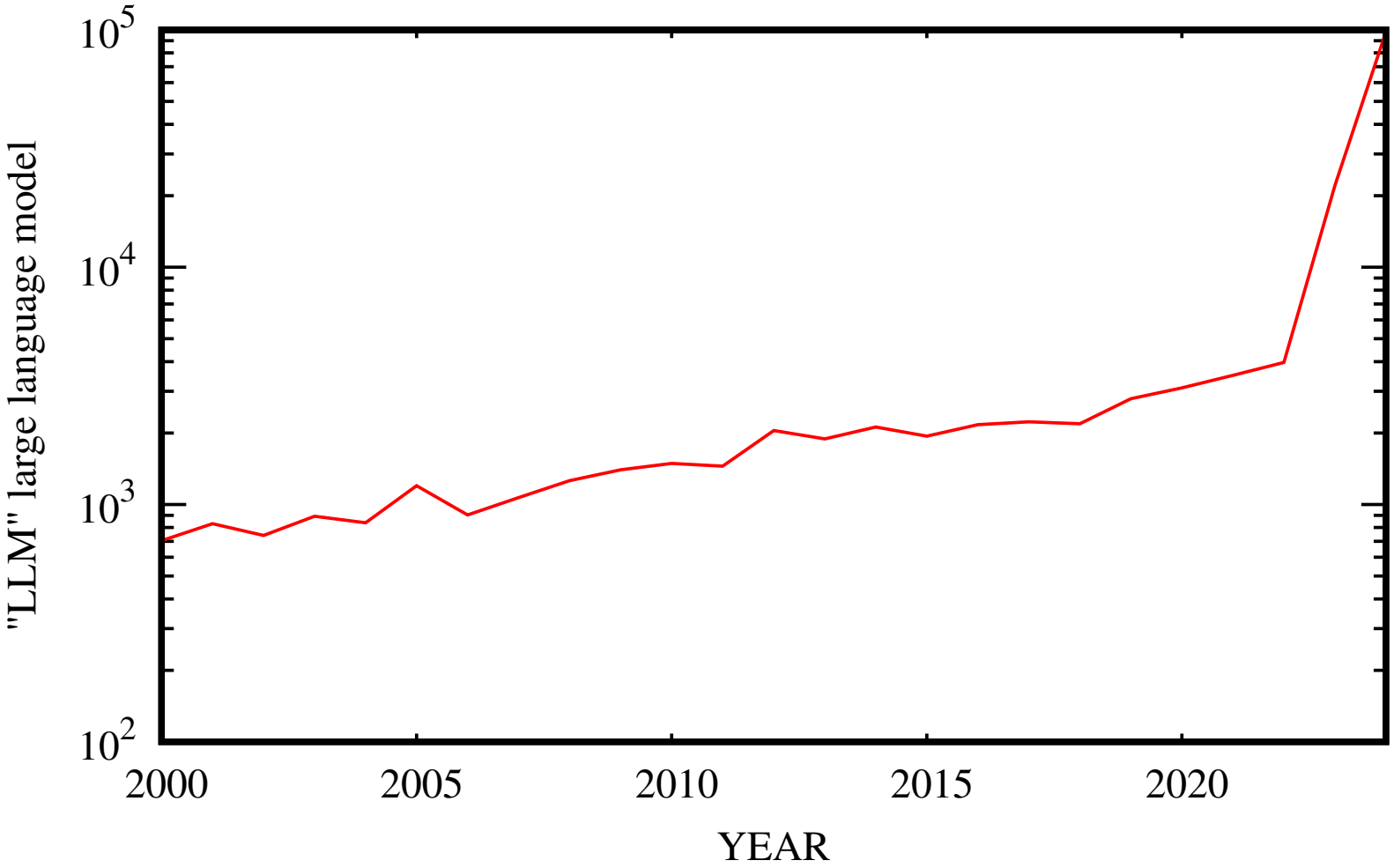
Основной источник большого объёма качественных текстов. Одна из бизнес-моделей, завляемых reddit.com — мы будем продавать данные создаваемые нашими communities для обучения LLM систем. Github также поставляет свои данные для LLM создающих модели компьютерного кода. Дешёвого входа нет. Очень дорого. Варианты создания LLM:

- Самим, с нуля на своих данных
- Взять чужую модель и слегка дообучить своими данными
- Взять чужую модель и даже не дообучить, а “создать context” для prompts
- Использовать чужую модель без дообучения

“LLM” large language model наукометрия

Количество научных публикаций по годам, термин “LLM” large language model.

В отличие от Deep Learning войти сюда заметно сложнее, т.к. требуется намного большая вычислительная мощность и, самое главное, нет общедоступного источника данных как ImageNet (Fei-Fei Li).



Резкий всплеск интереса в 2022-2023-2024 годах: 3970–22100–98100.

Примеры практических задач для LLM

- Английский язык. На вход ломаный английский – на выход идеальный. Проблема: в 10% случаев искажает смысл, а в 5% меняет смысл на противоположный.
- Задачи из экзаменов/тестов. Например задачу о бесконечно глубокой прямоугольной квантовой яме ChatGPT решает идеально. Задачу об ассиметричной квантовой яме конечной глубины, решает плохо. Сводит не к тем задачам.
- Сравнить характеристики бухгалтерских програм, автомобилей, или методов строительства. Вначале очень хорошо, а потом начинает галюцинировать.
- Архитектура компьютерных систем/библиотек. Вначале идеально, потом на 5-6 запросе начинает выдавать либо совсем неправильные ответы, либо неправильные акценты в ответах.
- Обучение программированию. Выдаёт почти идеальные примеры *небольших* программ.
- Интернет-поиск. ChatGPT/Grok очень хорошо агрегирует результаты интернет-поиска.

Все ответы выдаются очень убедительным языком. Необходимо научится понимать момент, когда пошли “неверные акценты”. Это намного сложнее, чем обнаружить просто неправильный ответ.

LLM: Нерешённые проблемы

Стали понятны ограничения моделей. В ChatGPT переход от 175 млрд параметров, до 1800 миллиардов улучшил, но не принципиально. Представление знания в виде transformer + embedder + neural network имеют ограничения.

- Понимание контекста и долгосрочная память
- Фактическая точность и знание фактов. LLM склонны к галлюцинациям — выдаче неправдоподобной или полностью выдуманной информации. Причины: модели предсказывают наиболее вероятное слово, а не проверяют факт, данные предобучения могут быть устаревшими.
- Математика, логика и сложные рассуждения. LLM справляются с простыми вычислениями и базовой логикой, но: многозадачные вычисления, доказательства теорем, сложные инженерные расчёты — проблемны, модели часто делают мелкие ошибки, которые критичны в точных областях.
- Обработка неоднозначного или специализированного языка. Юридические, медицинские, технические термины могут неправильно интерпретироваться, если данные были ограничены или устарели. Синонимы и культурные различия также приводят к ошибкам.

LLM: Нерешённые проблемы

- Обновление знаний в реальном времени. Модели обучены на статическом датасете.
- Объяснимость (Explainability). LLM — «черные ящики». Трудно объяснить, почему модель сгенерировала конкретный вывод, что критично для медицины, юриспруденции и финансов.
- Энергоэффективность и масштабирование Модели с сотнями миллиардов параметров требуют огромных ресурсов для обучения и inference.

Вопрос: насколько далее развиваема эта архитектура:

представление знания в виде transformer + embedder + neural network

Спасибо за внимание