

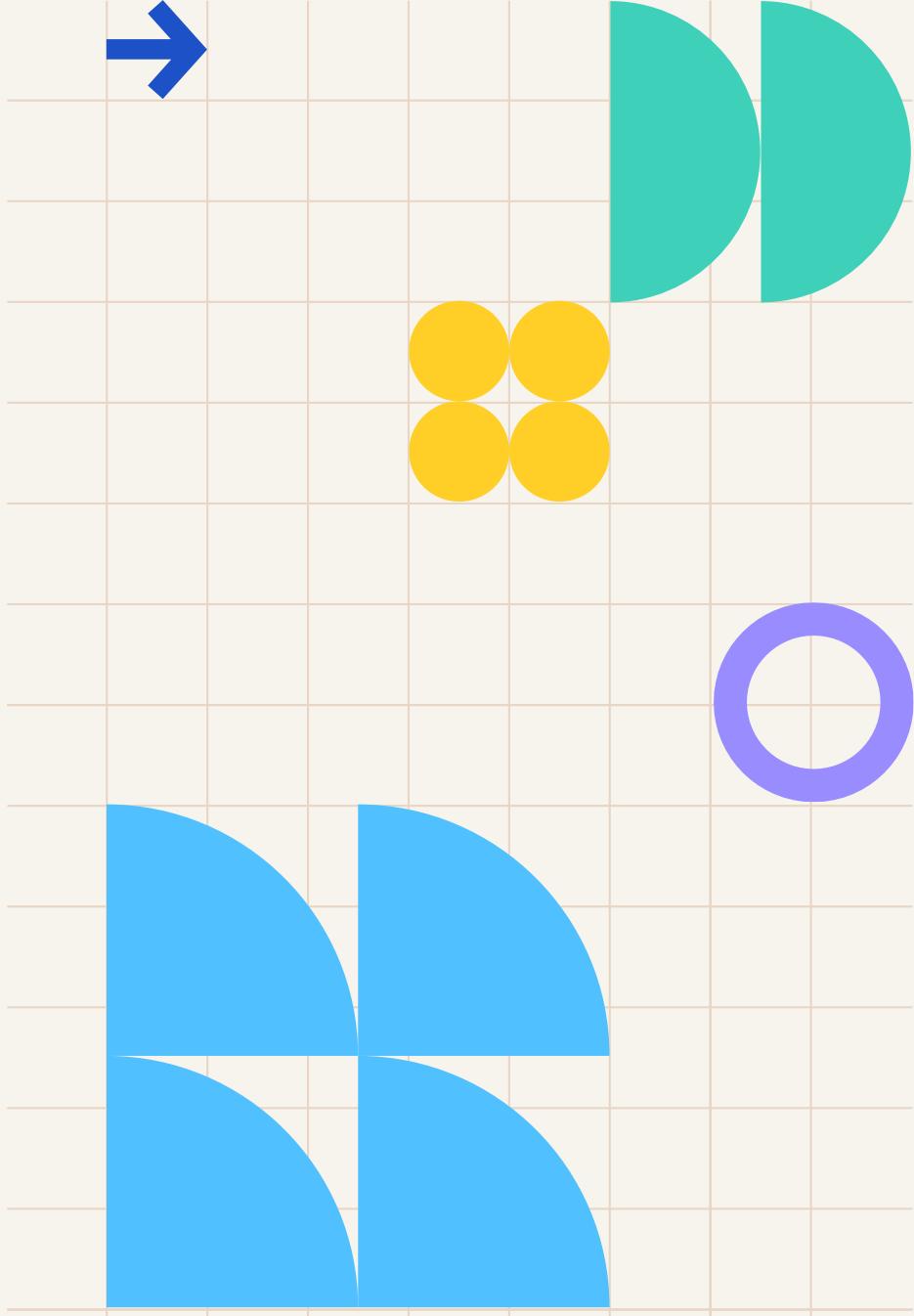
Quō vādis, AI?

лат. Куда ты катишься, ИИ?
(AI - *artificialis intelligentia*)

Михаил Бурцев

к.ф.-м.н.

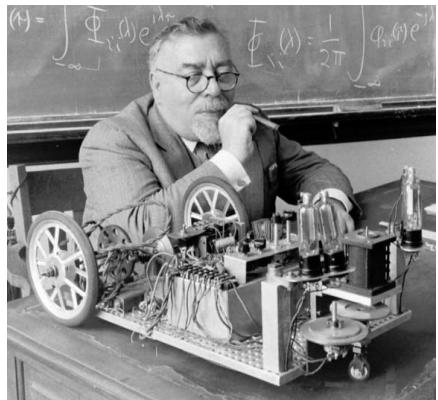
London Institute for Mathematical Sciences



01

Искусственный интеллект:
эволюция идей

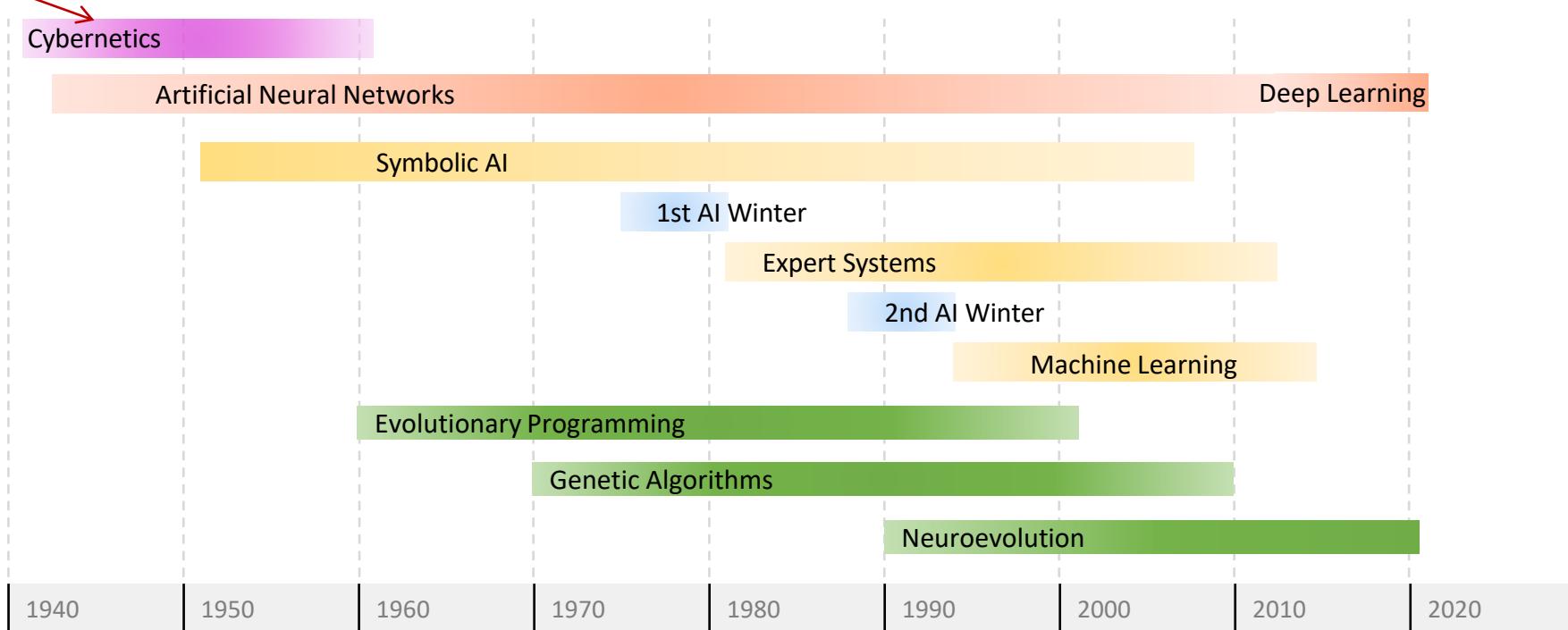
Искусственный интеллект: эволюция идей



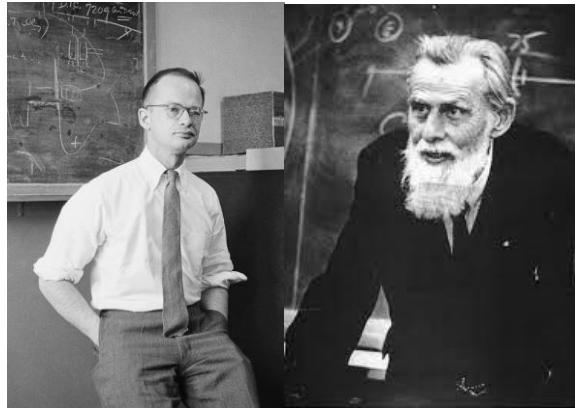
Как машина может адаптироваться к изменяющимся условиям?

Отрицательная обратная связь

- Сложные системы состоят из иерархии саморегулирующихся элементов
- Отдельный элемент имеет механизм управления, приводящий к изменению выхода обратно пропорционально отклонению от целевого значения



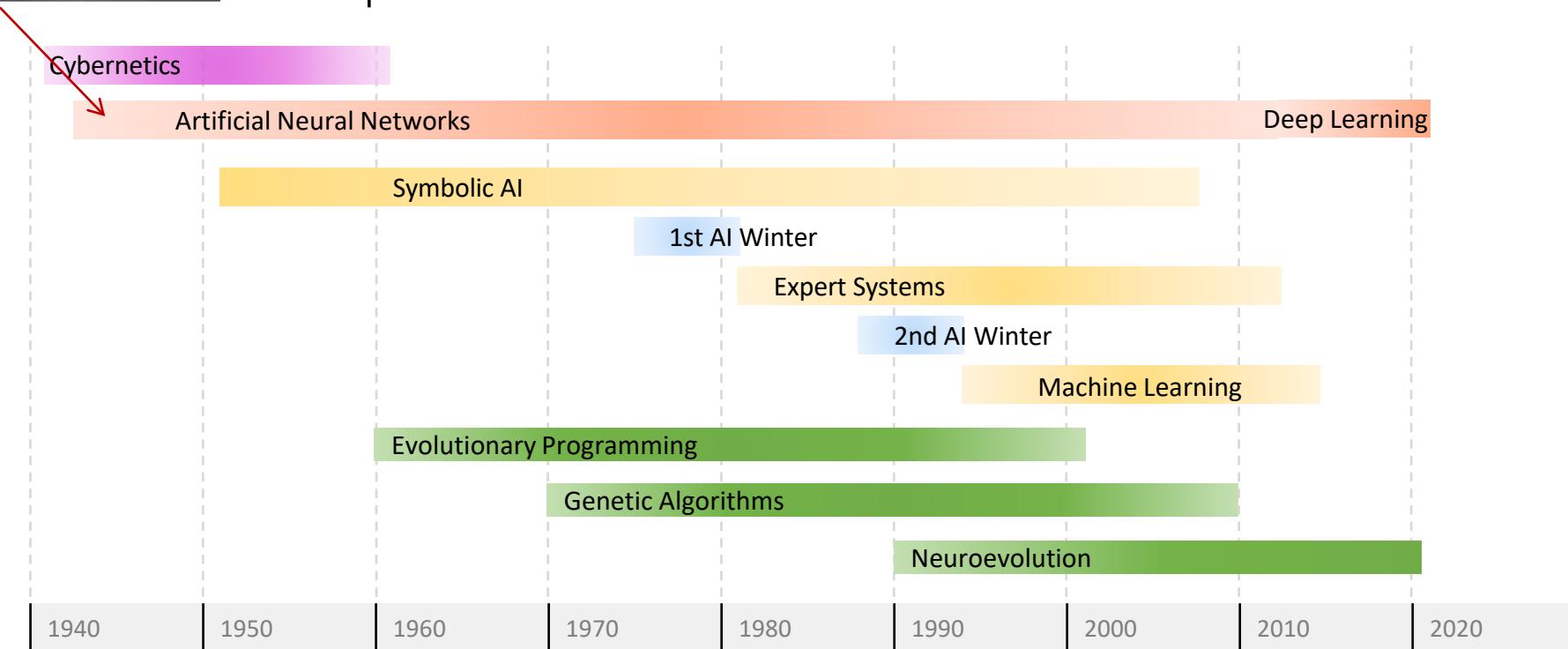
Искусственный интеллект: эволюция идей



Простые схемы обратной связи не подходят для многих реальных задач

Сложные вычисления, как композиция простых

- Отдельный элемент вычисляет простую функцию
- Сложные вычисления реализуются, как последовательная композиция простых



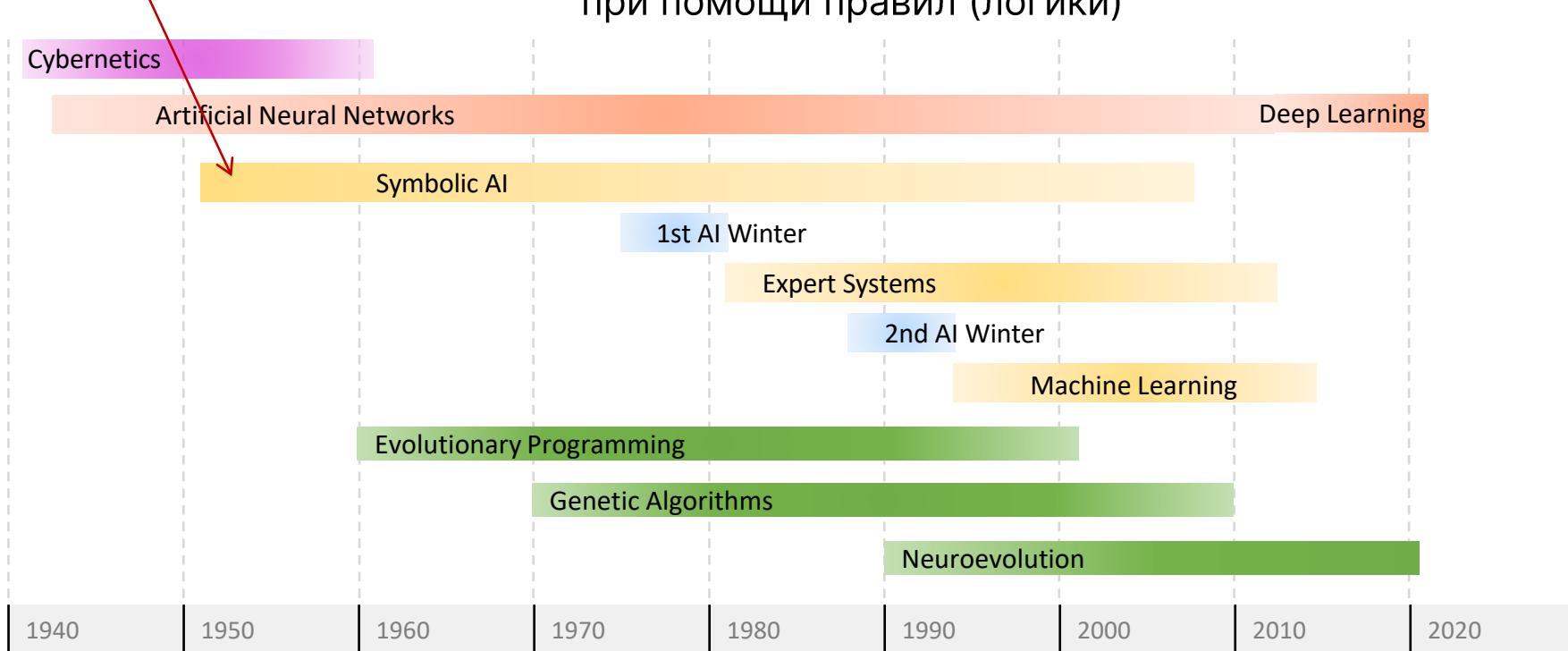
Искусственный интеллект: эволюция идей



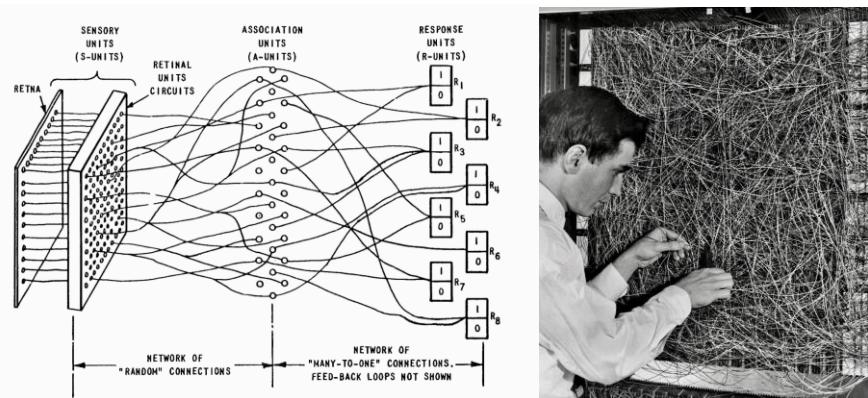
Простые схемы обратной связи не подходят для многих реальных задач

Рассуждения человека при решении задачи могут быть описаны набором правил

- Из начальных утверждений можно вывести решение задачи при помощи правил (логики)



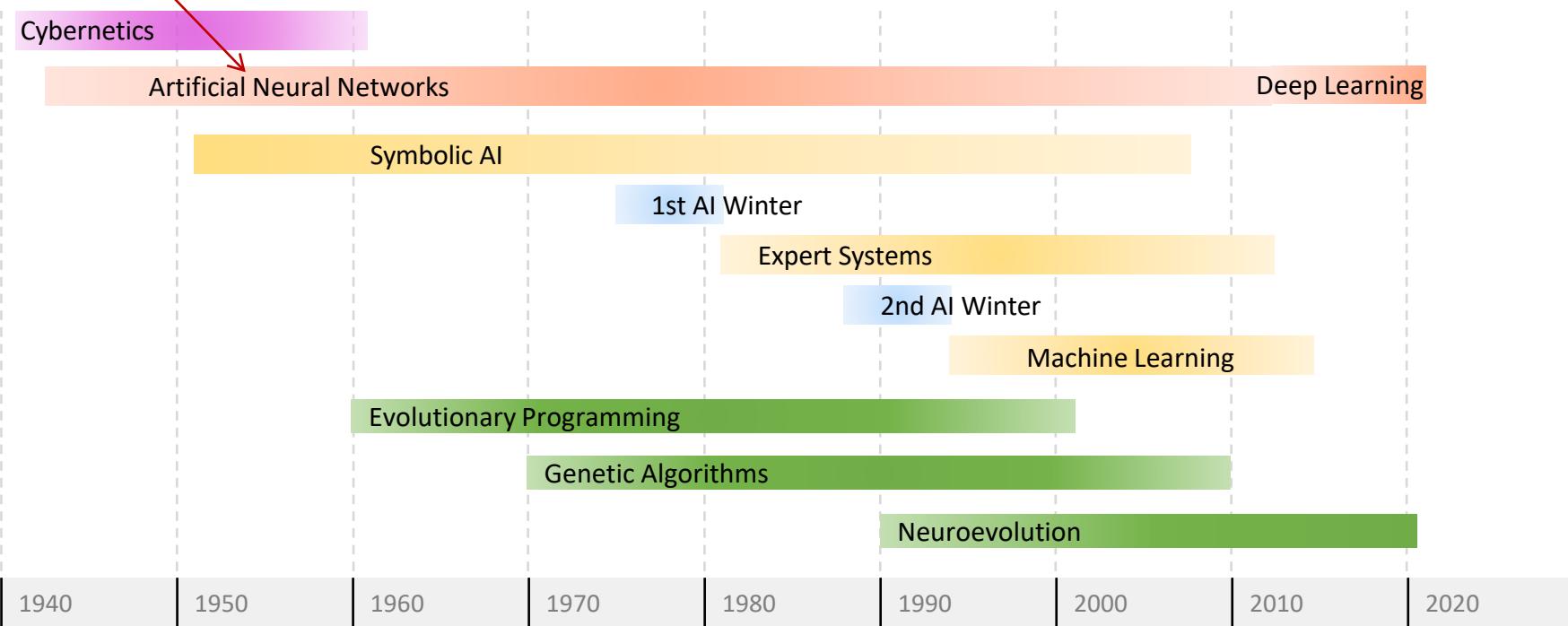
Искусственный интеллект: эволюция идей



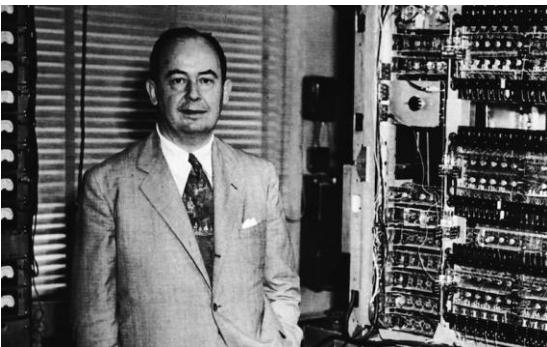
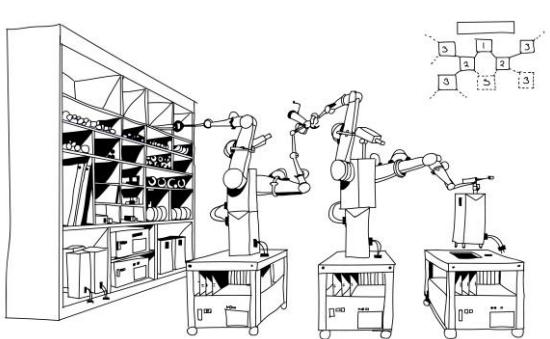
Необходимо не только адаптироваться, но и запоминать решения

Обучение – изменение связей между элементами

- Изменения связей должны повышать вероятность верного ответа



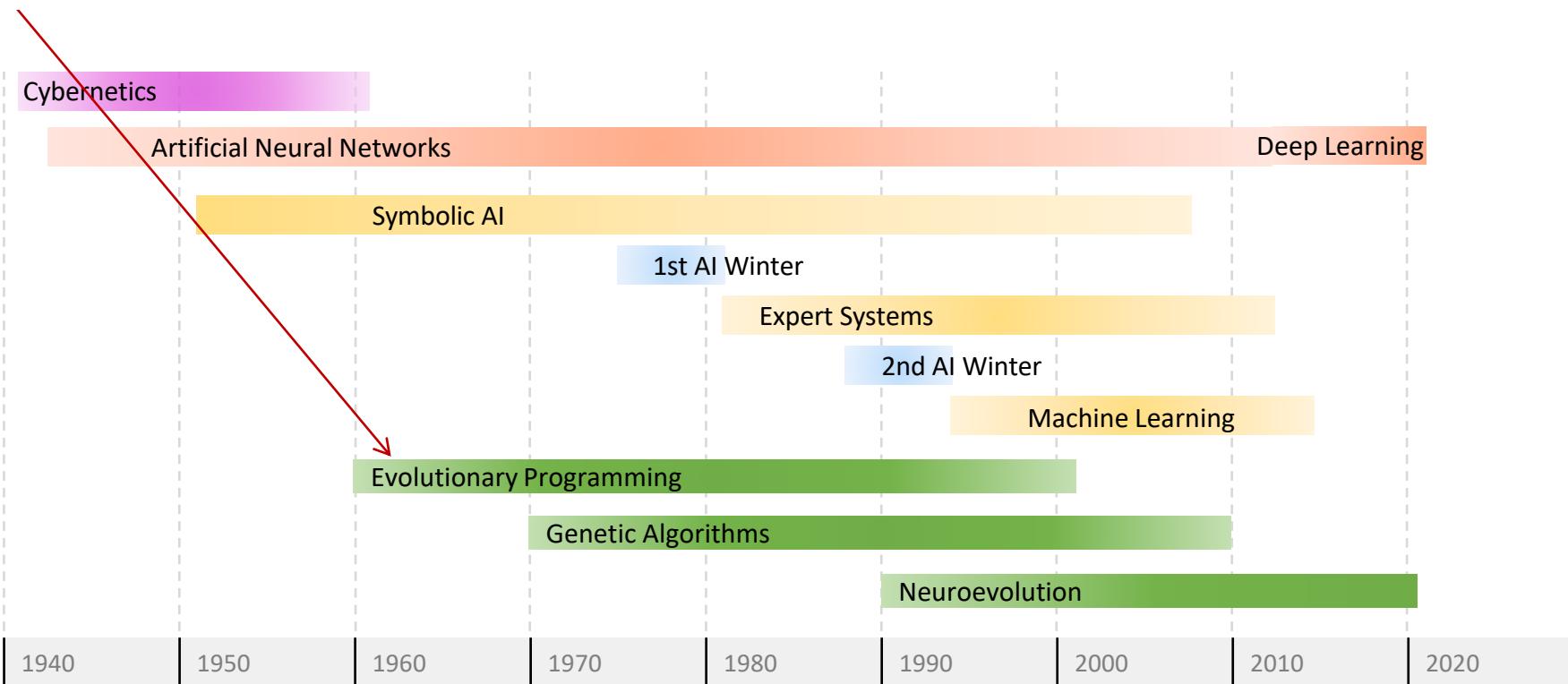
Искусственный интеллект: эволюция идей



Простые схемы обратной связи не подходят для многих реальных задач

Саморепликация машин нужна для эволюции

- Самореплицирующиеся машины могут мутировать и отбираться для лучшей реализации функции



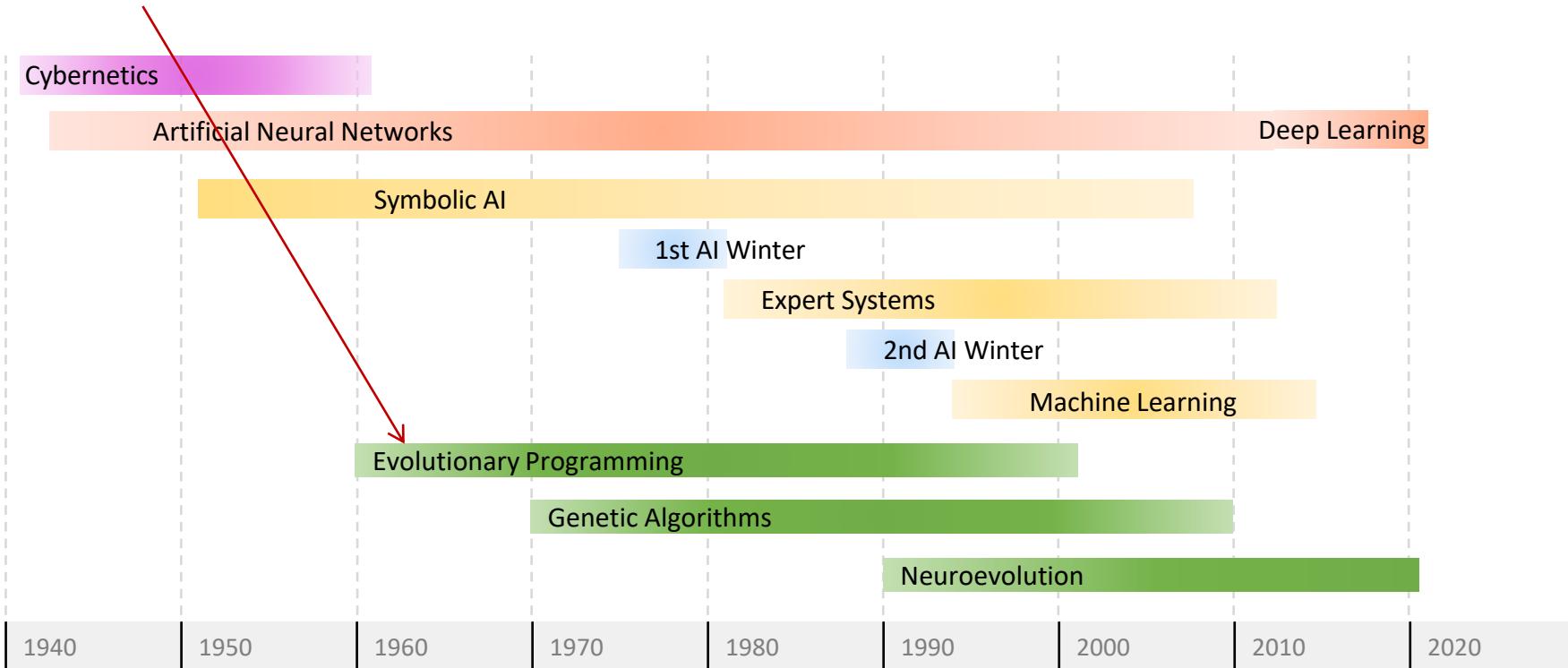
Искусственный интеллект: эволюция идей

1 1 1 0 0 1 0 1 0 0 0 1 1 0
1 1 0 0 0 1 0 1 1 0 0 1 1 0
0.2 0.6 1.2 3.0 0.8 2.4 0.6
0.2 0.7 1.2 3.0 0.8 2.2 0.6

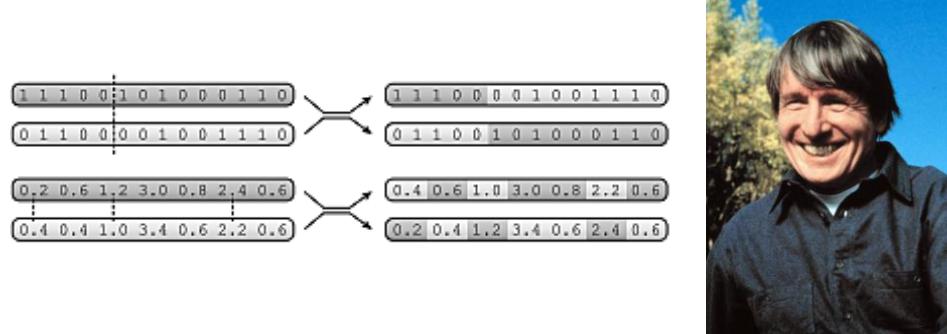


Эволюция происходит за счёт мутаций

- Параметры задачи можно закодировать многомерным вектором вещественных чисел, а затем провести мультикритериальную оптимизацию за счет стохастического поиска



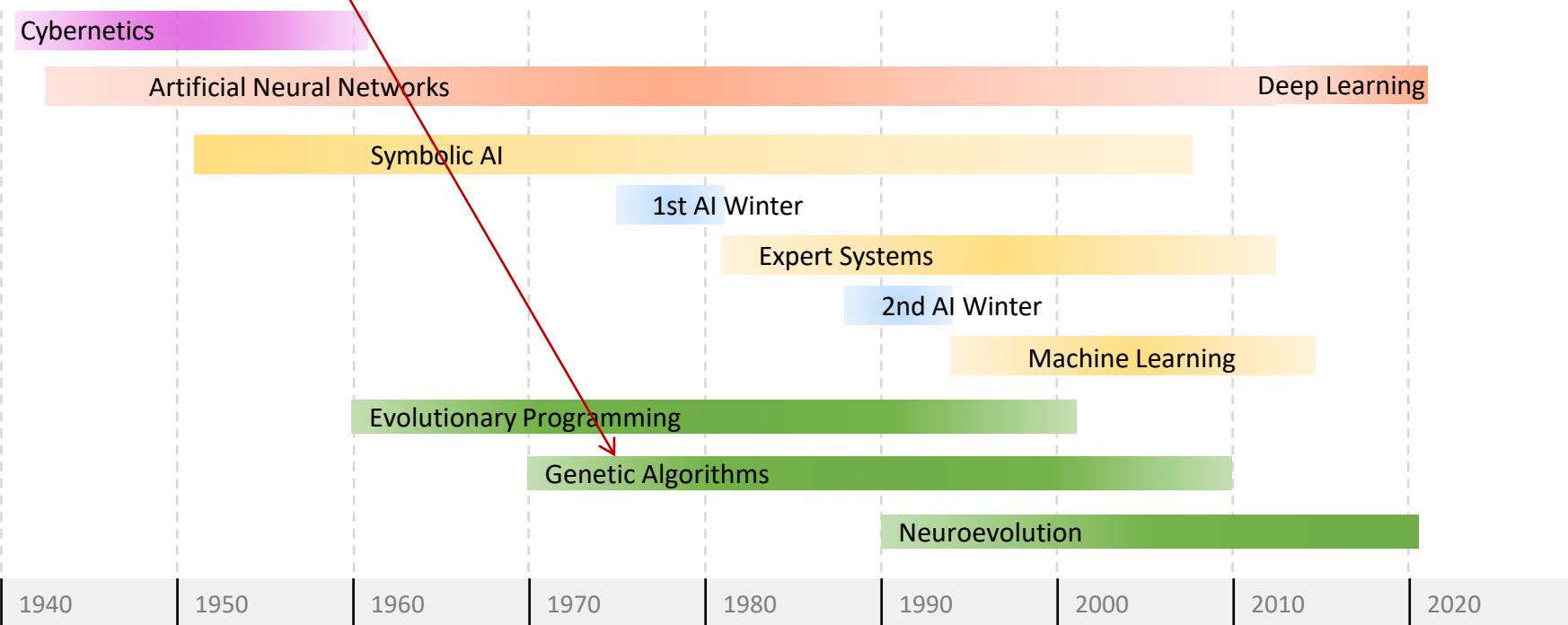
Искусственный интеллект: эволюция идей



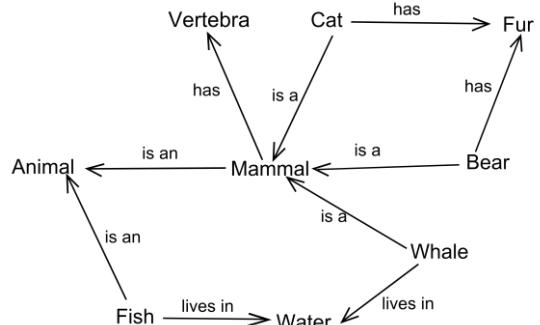
Эволюция за счет мутаций происходит медленно

Рекомбинация обеспечивает ускорение эволюции

- При рекомбинации потомству передаются функциональные «блоки», что открывает возможности для быстрого перебора комбинаций



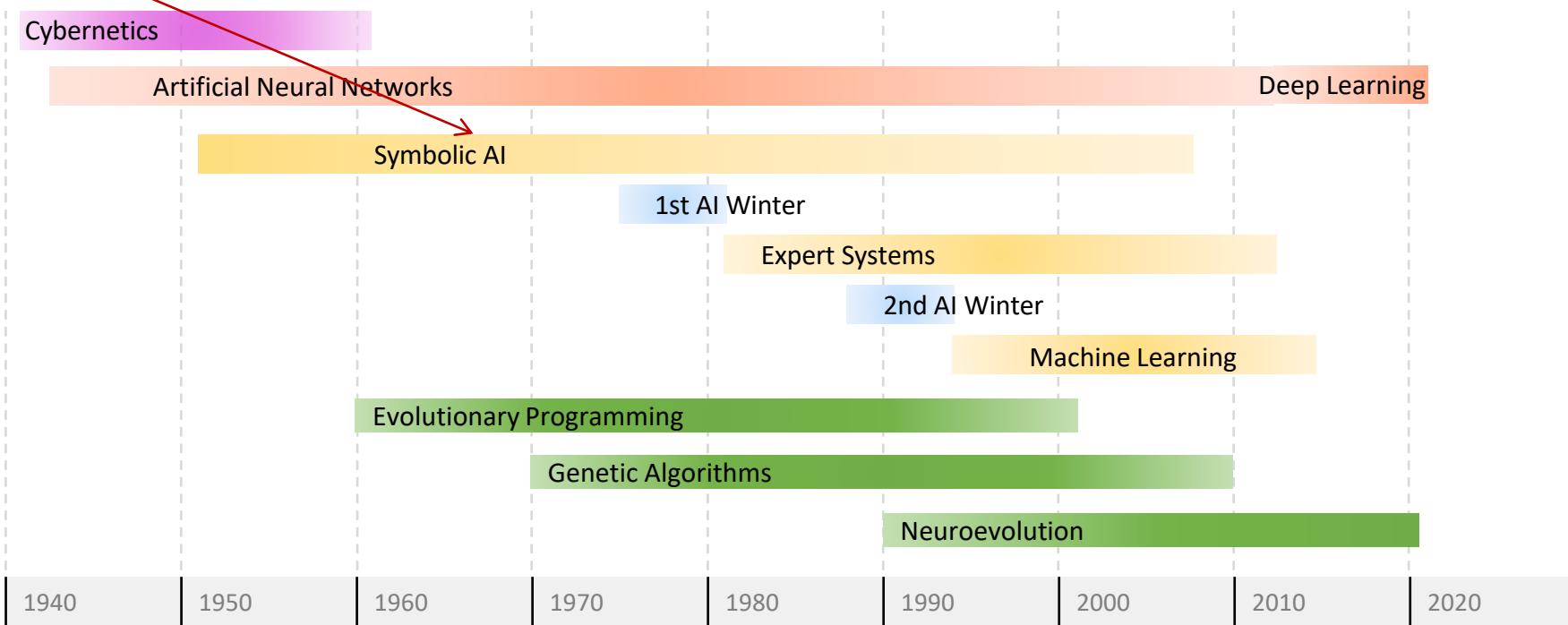
Искусственный интеллект: эволюция идей



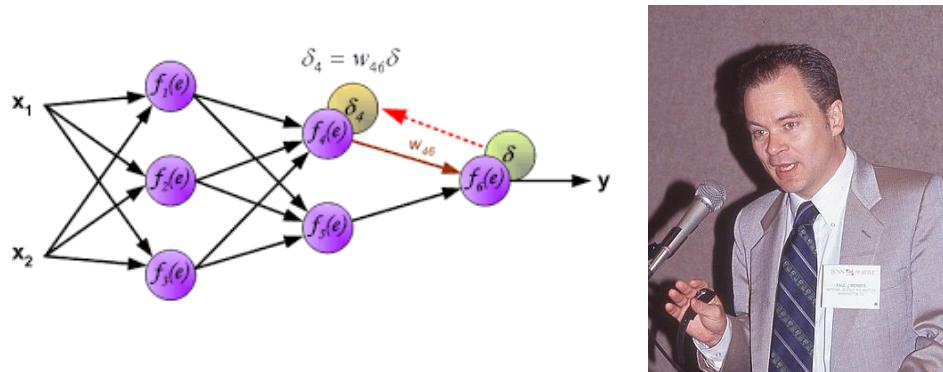
Только рассуждений недостаточно нужны факты об окружающем мире

Семантические сети и графы знаний

- Знания о мире можно представить в виде набора понятий и отношений между ними
- На основе семантической сети и правил вывода можно моделировать рассуждения в некоторой предметной области



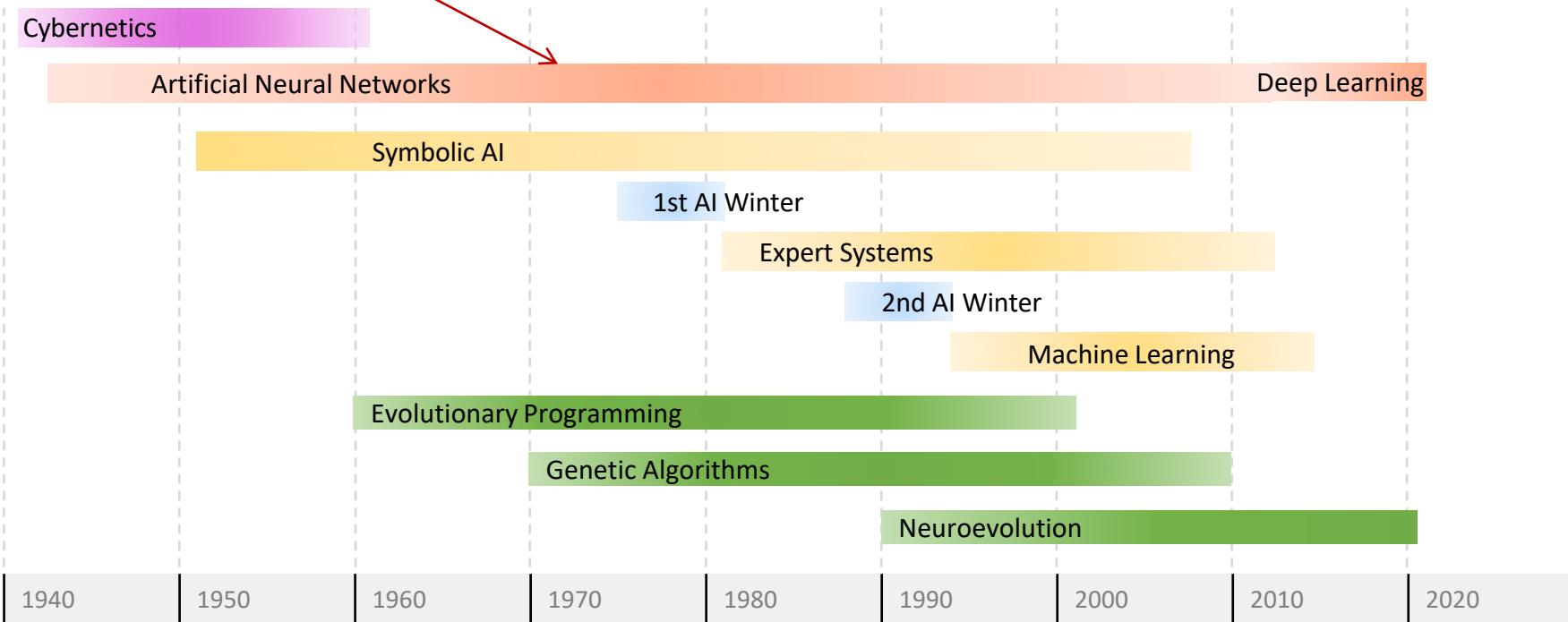
Искусственный интеллект: эволюция идей



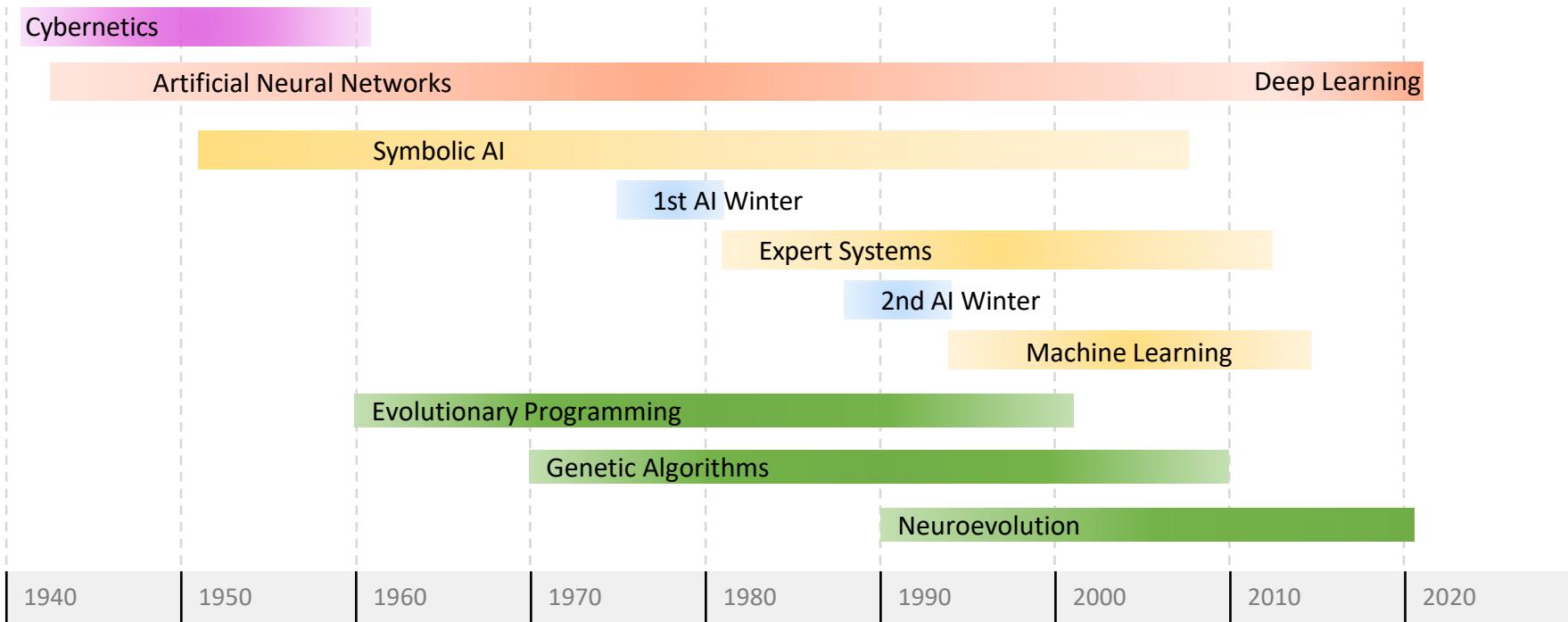
Как подстроить параметры сложной комбинации функций?

Метод обратного распространения ошибки

- Для каждого веса многослойной сети можно посчитать вклад в конечную ошибку, и изменение, позволяющее эту ошибку уменьшить



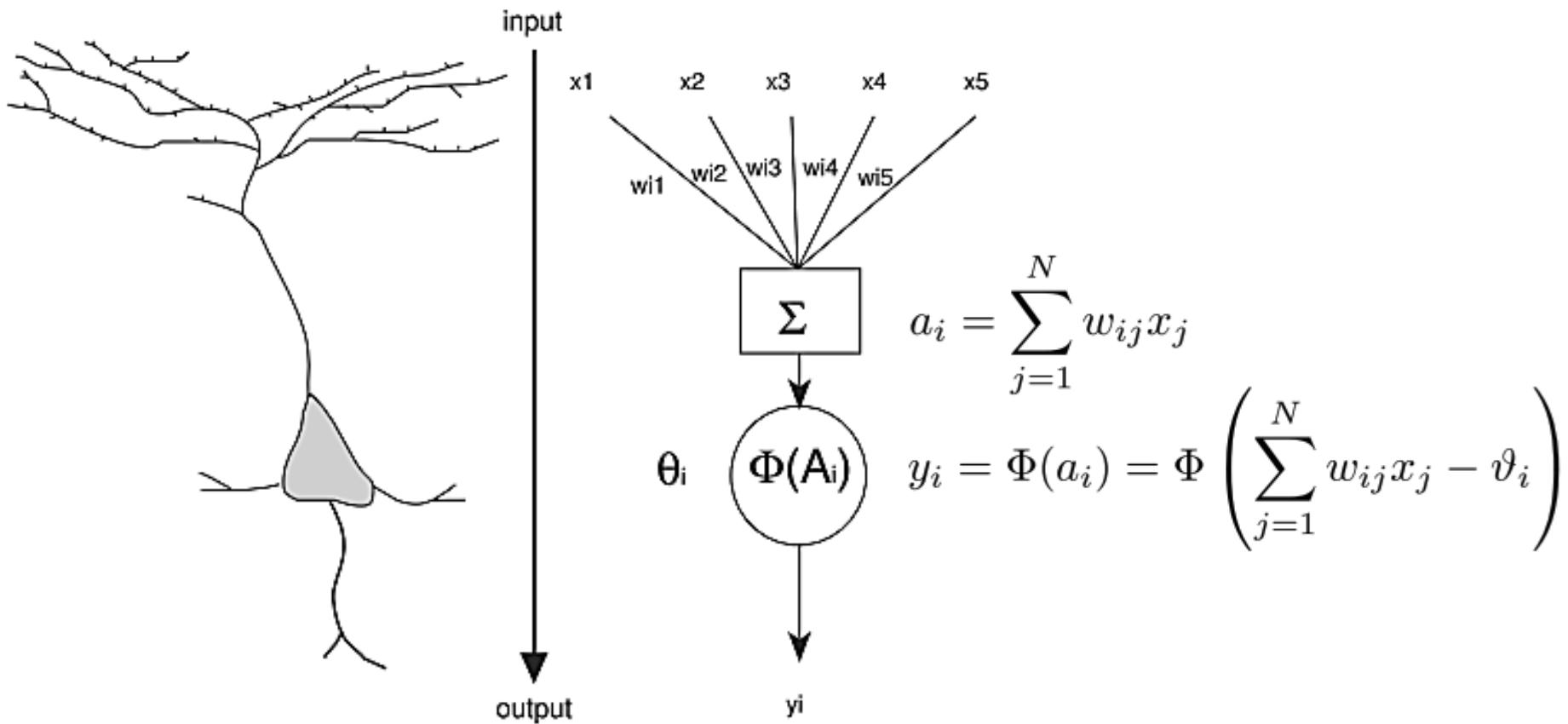
Искусственный интеллект: эволюция идей



02

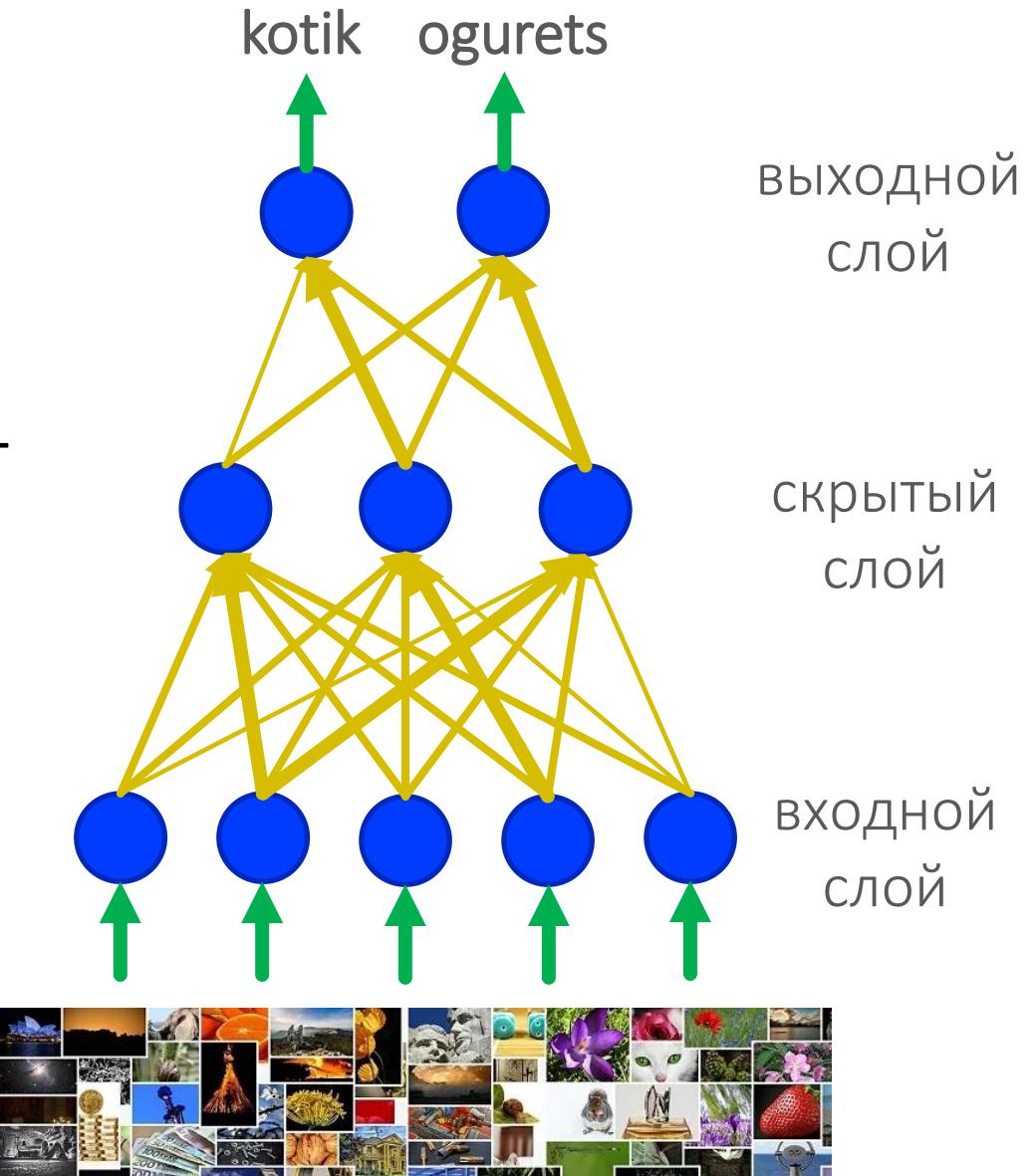
Глубокое обучение сегодня

Искусственный нейрон



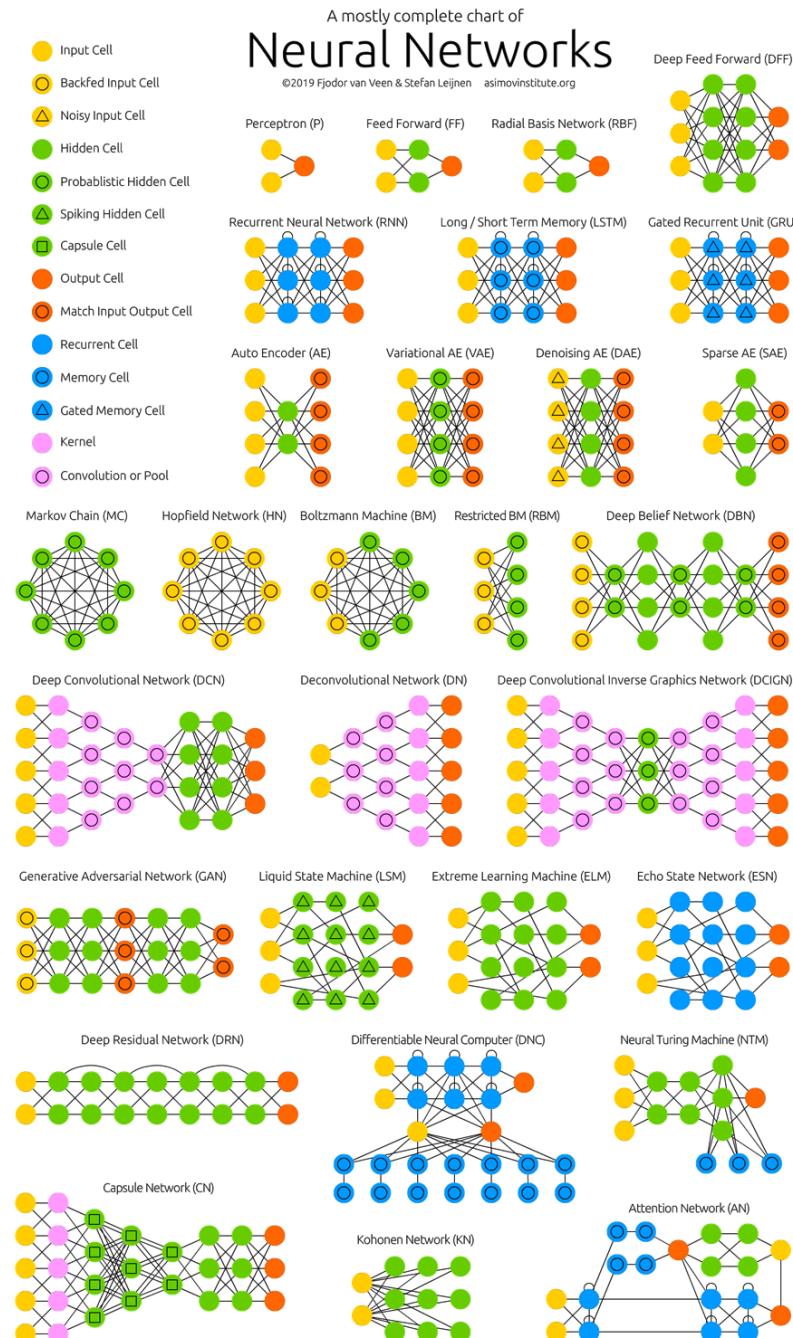
Обучение нейросети

- Нейросеть учится на примерах правильных ответов
 - обучающая выборка
- Обучение нейросети происходит за счет изменения весов
- Для проверки того, насколько хорошо прошло обучение используются примеры, которые не были использованы при обучении
 - тестовая выборка

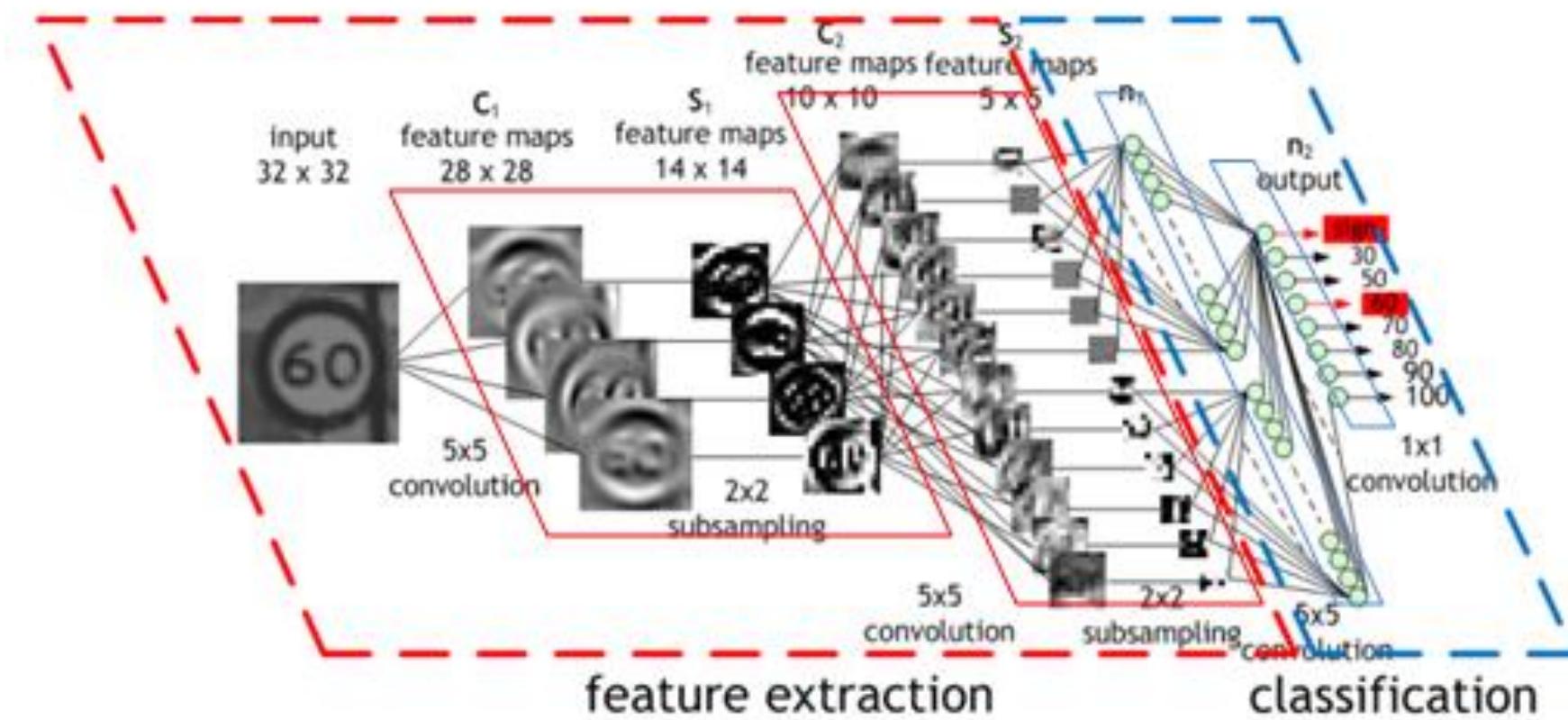


Нейросети удивительно хорошо работают

- Обучение сходится почти всегда
- Чем больше данных используется для обучения, тем лучше
- Чем больше слоев и параметров в нейросети, тем лучше
- Сложность архитектуры слоев и функции отдельных нейронов не мешают обучению
- Решают задачи в широком круге прикладных областей – компьютерном зрении, обработке естественного языка, системах управления

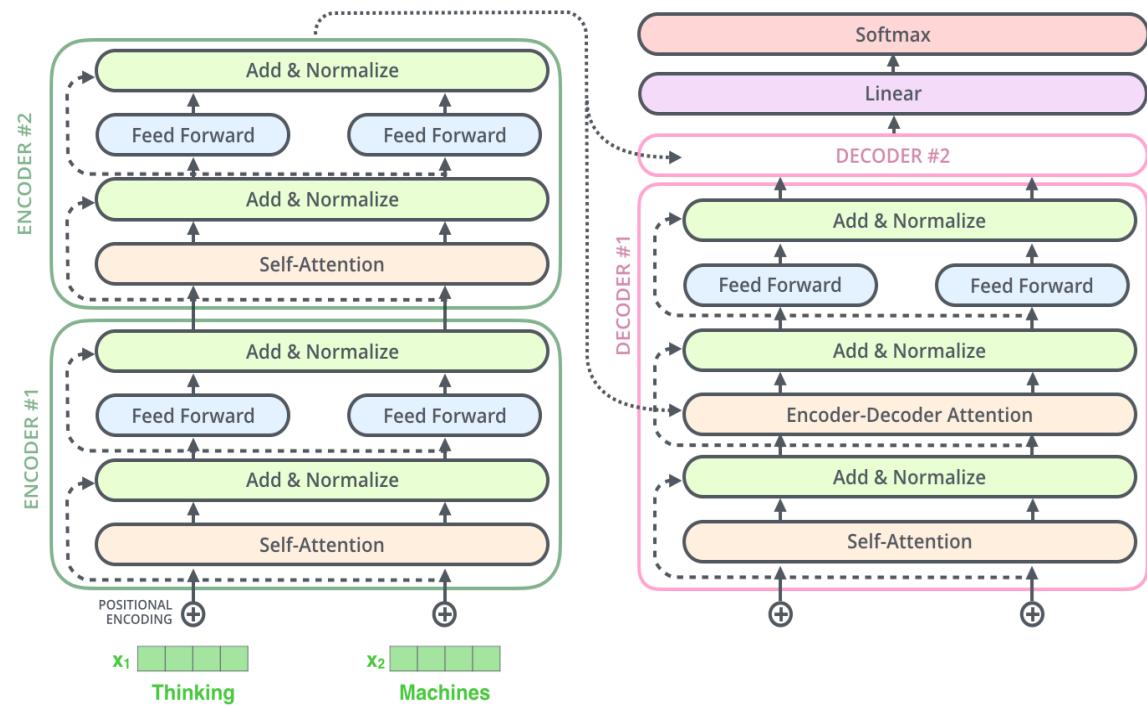


Обучение с учителем: сверточные нейронные сети



Моделирование последовательностей

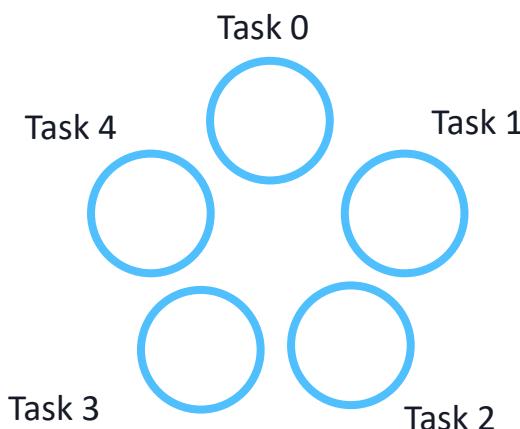
Трансформер



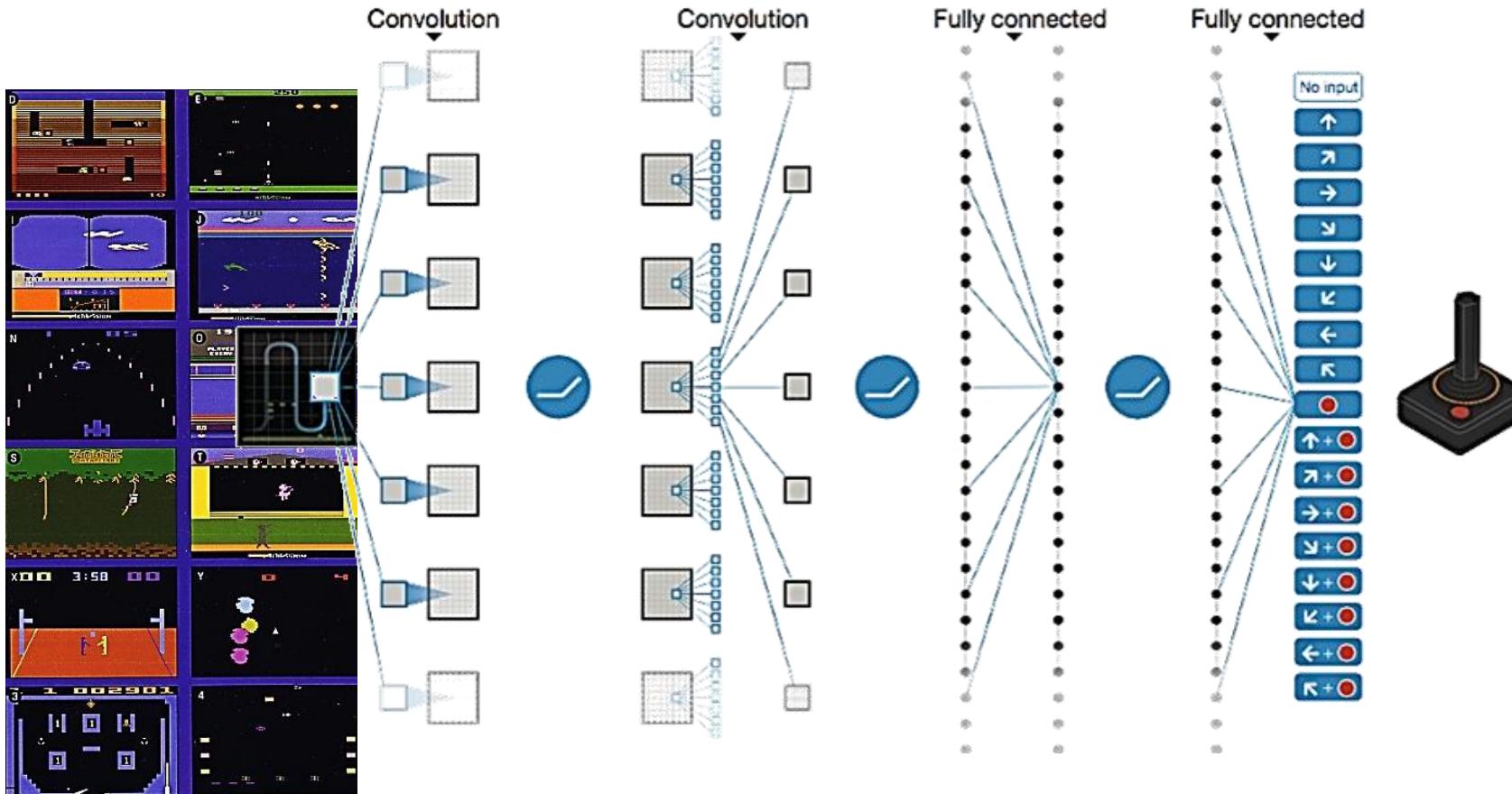
Универсальность методов обучения



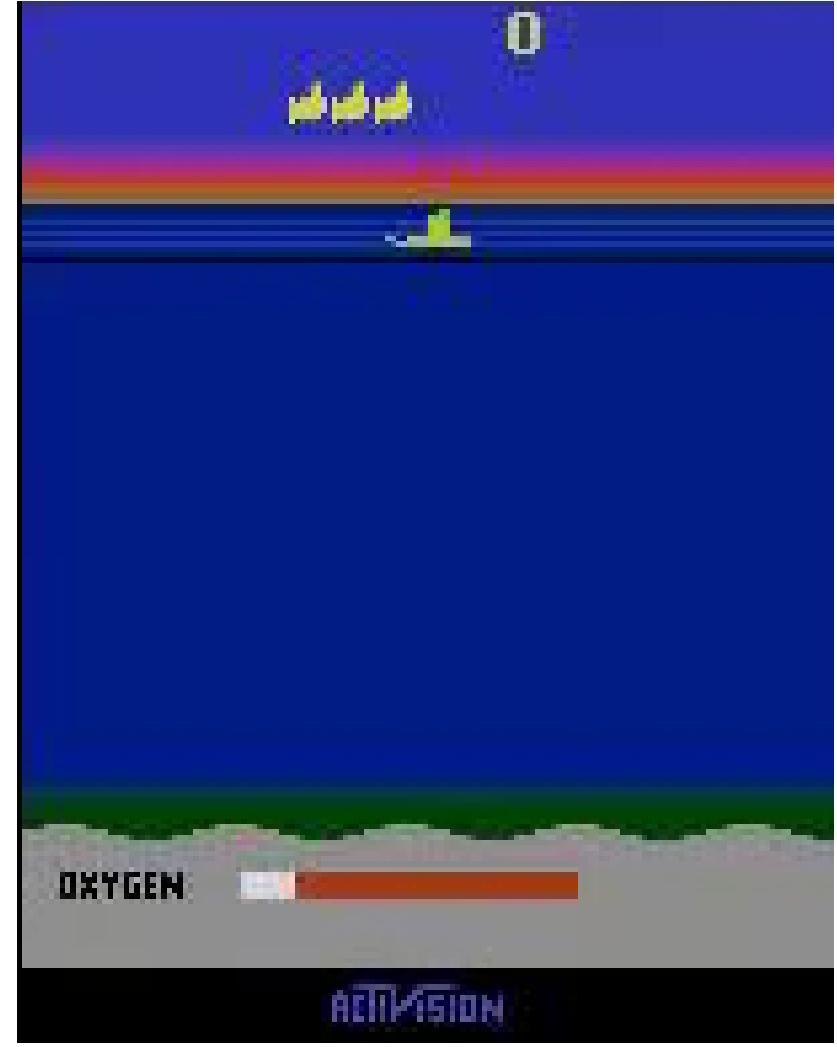
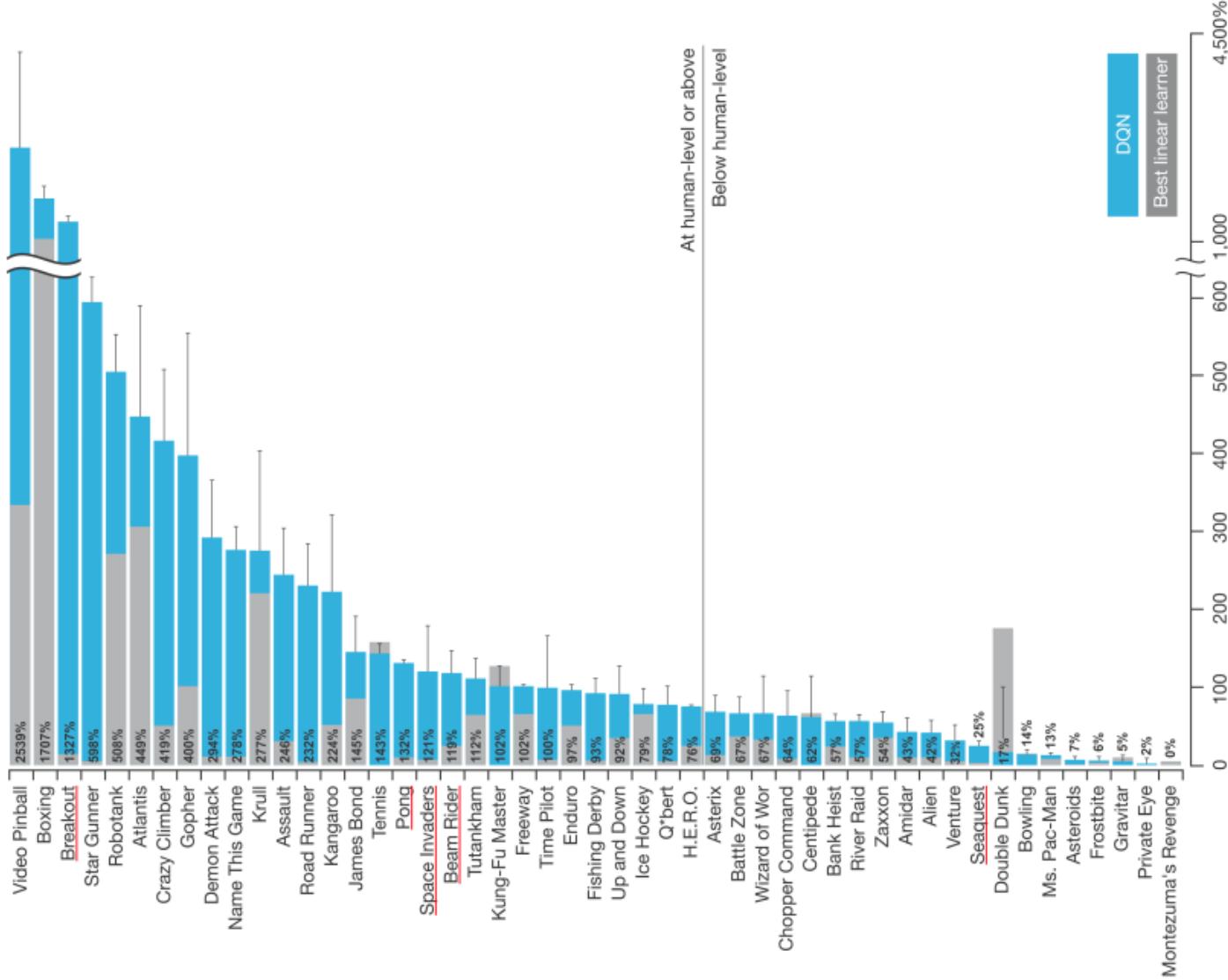
- одна предметная область
 - обучение с учителем
 - требуется много размеченных данных



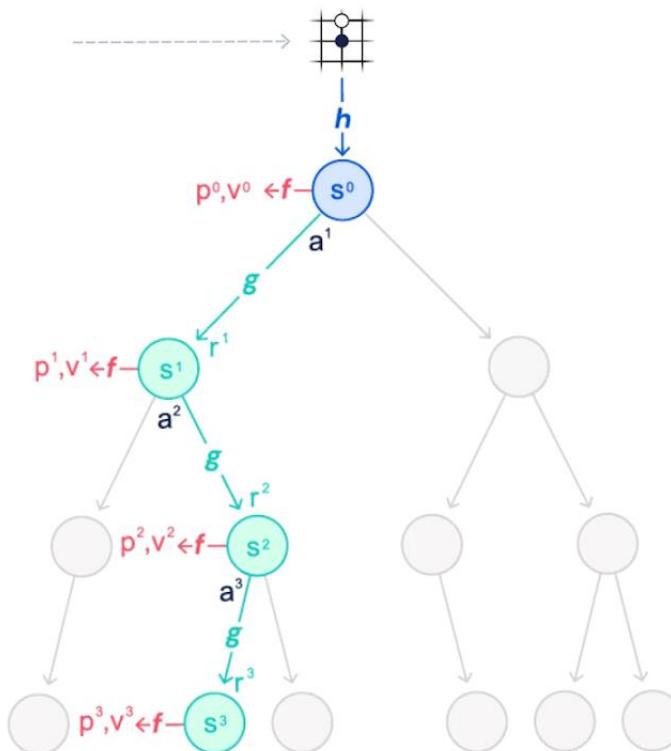
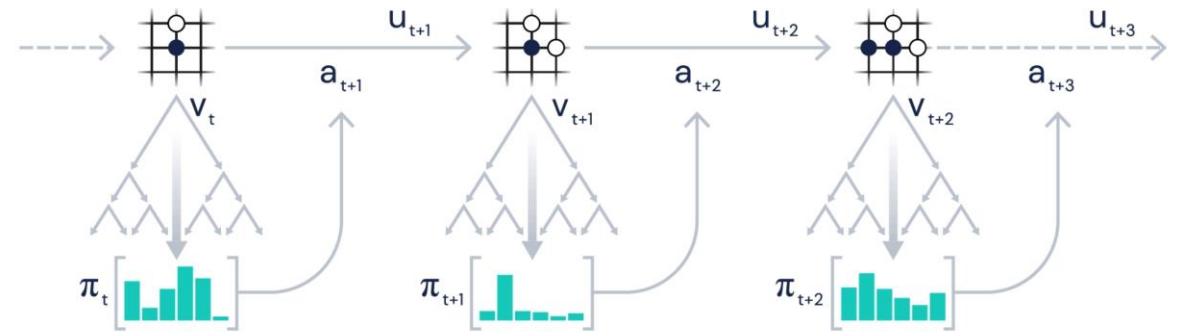
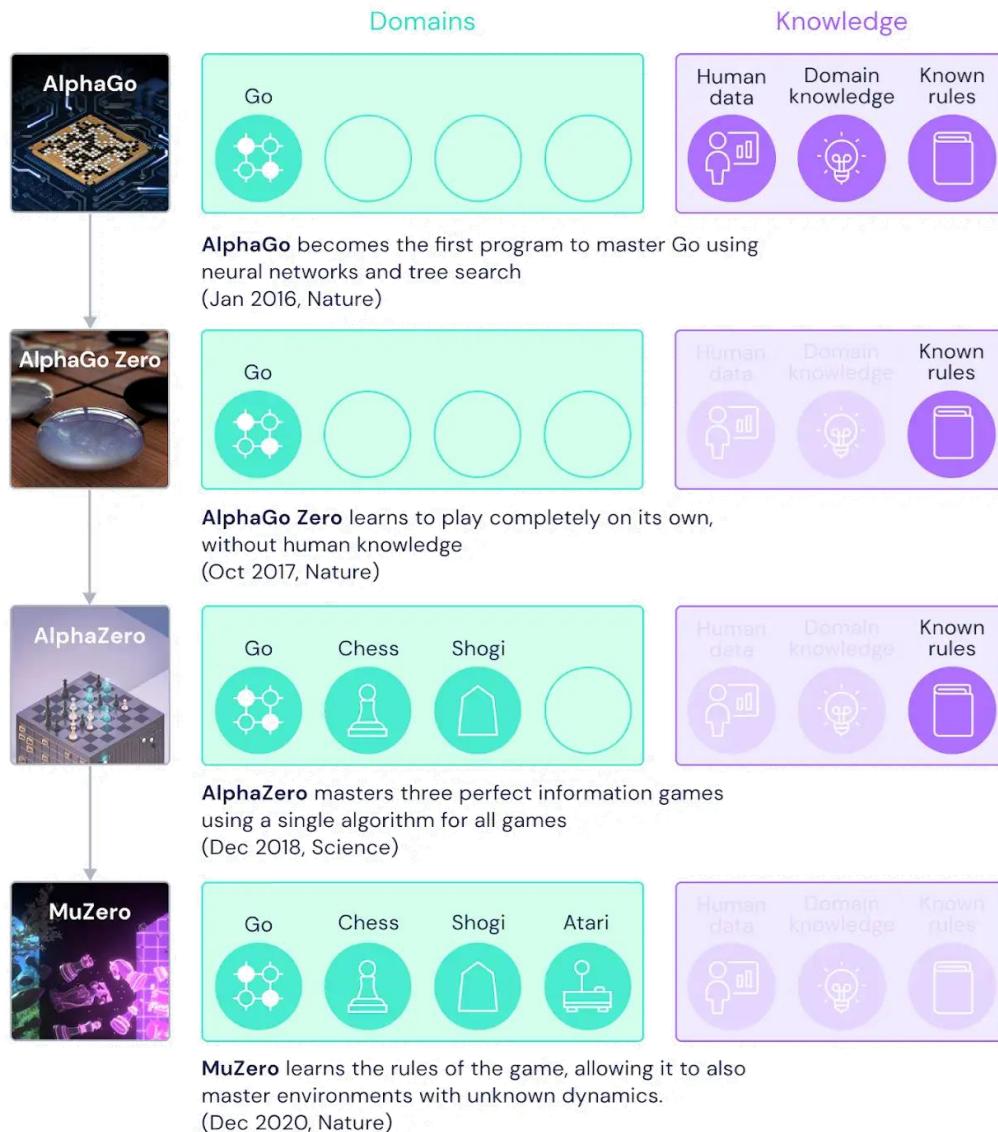
Обучение с подкреплением: универсальный игрок



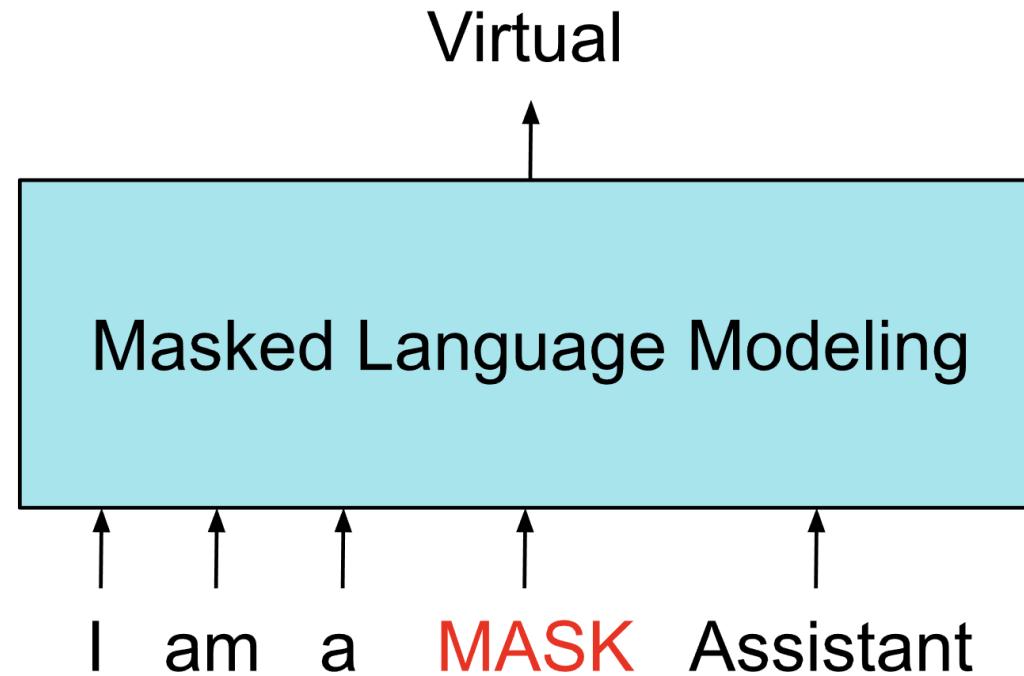
Обучение с подкреплением: универсальный игрок



Обучение с подкреплением: универсальный игрок

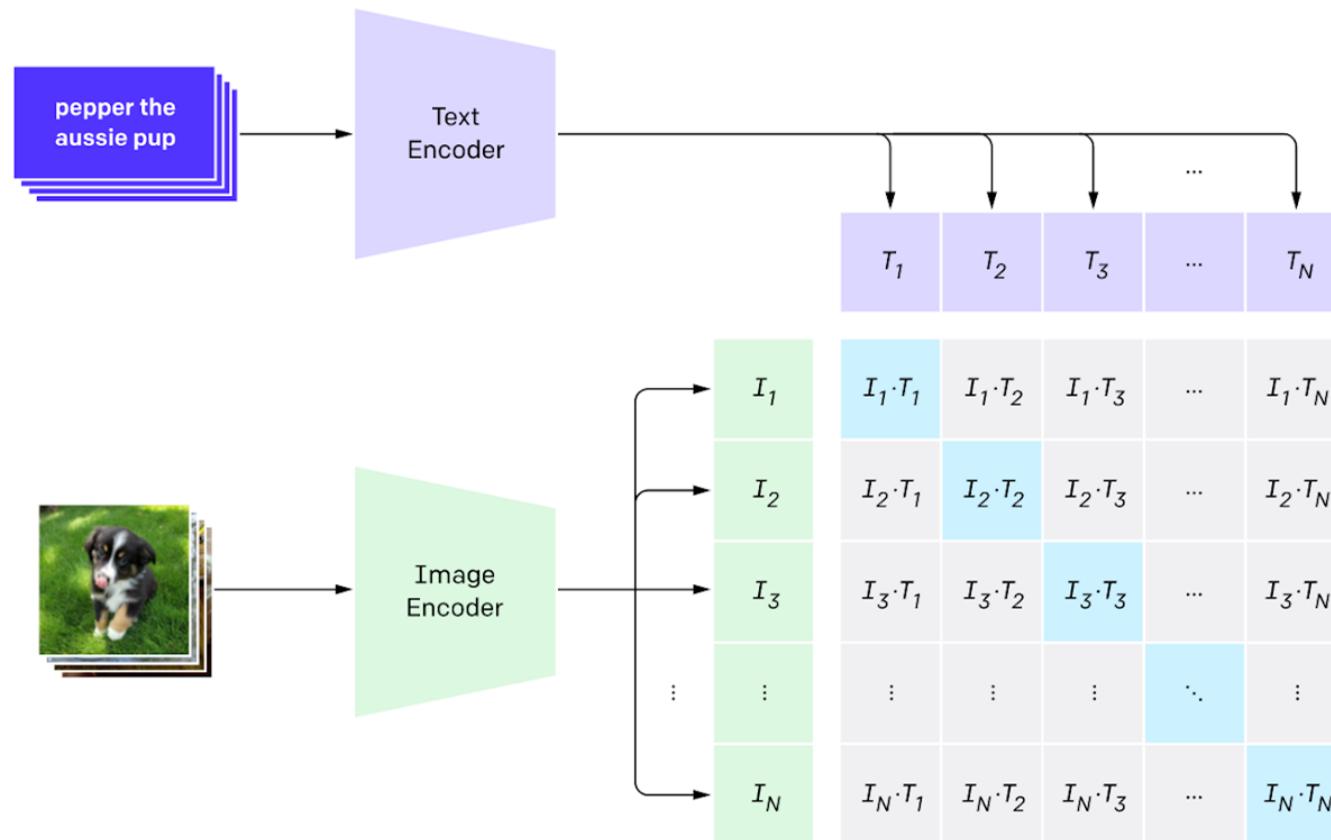


Языковые модели с маскированием



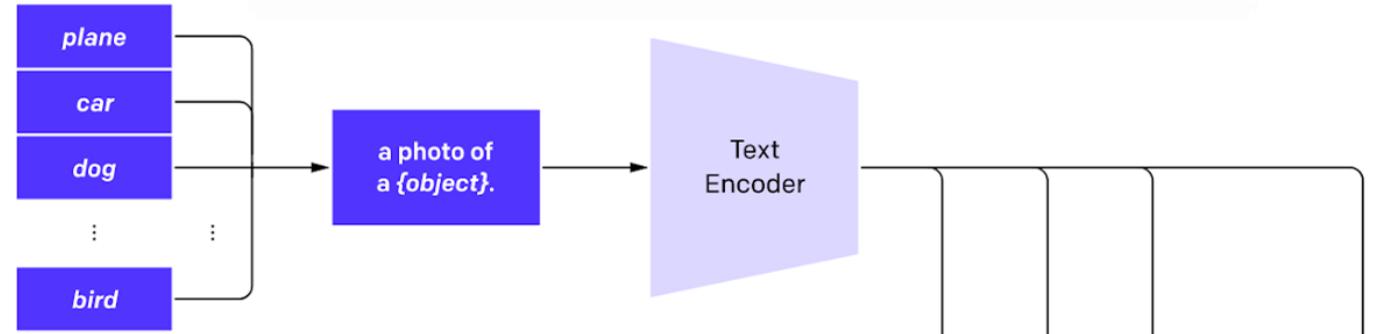
Мультимодальное обучение

1. Contrastive pre-training

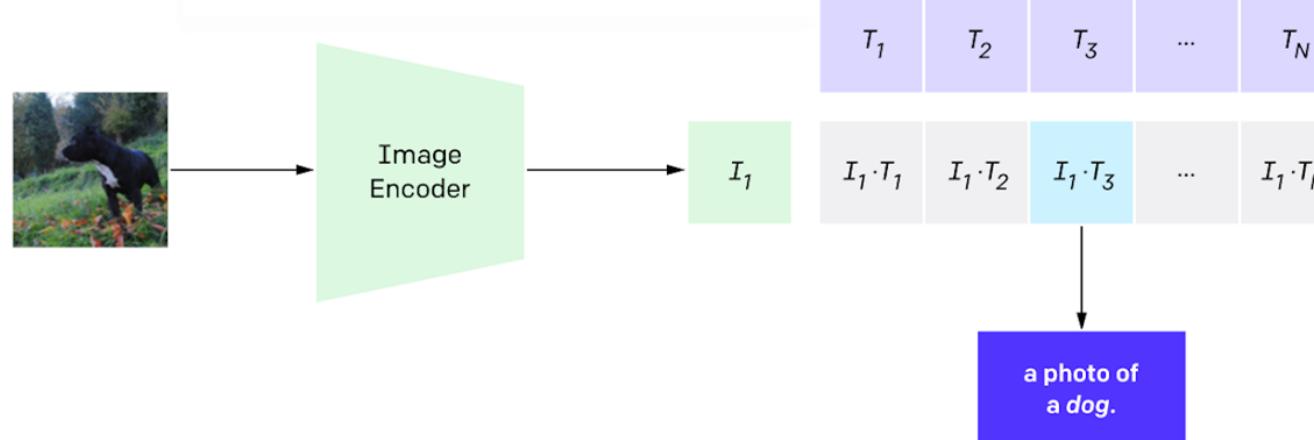


Мультимодальное обучение

2. Create dataset classifier from label text



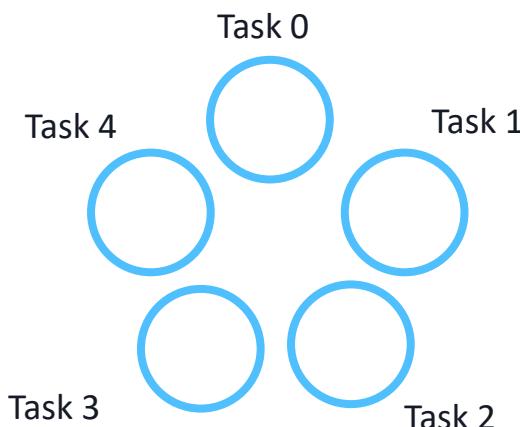
3. Use for zero-shot prediction



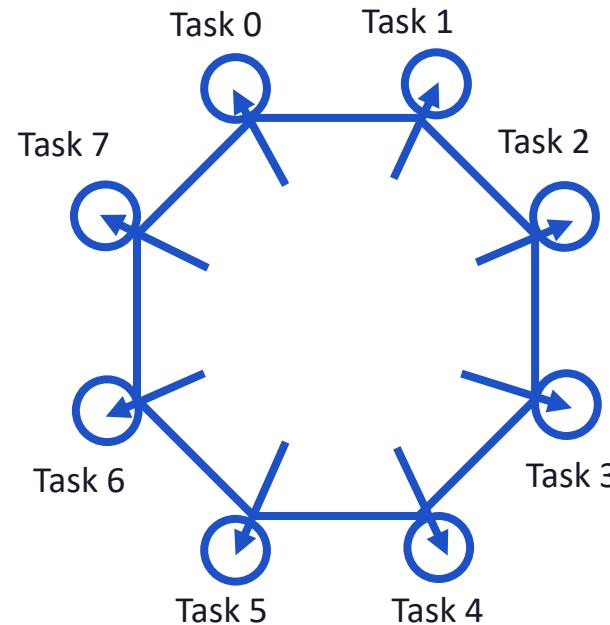
Универсальность методов обучения



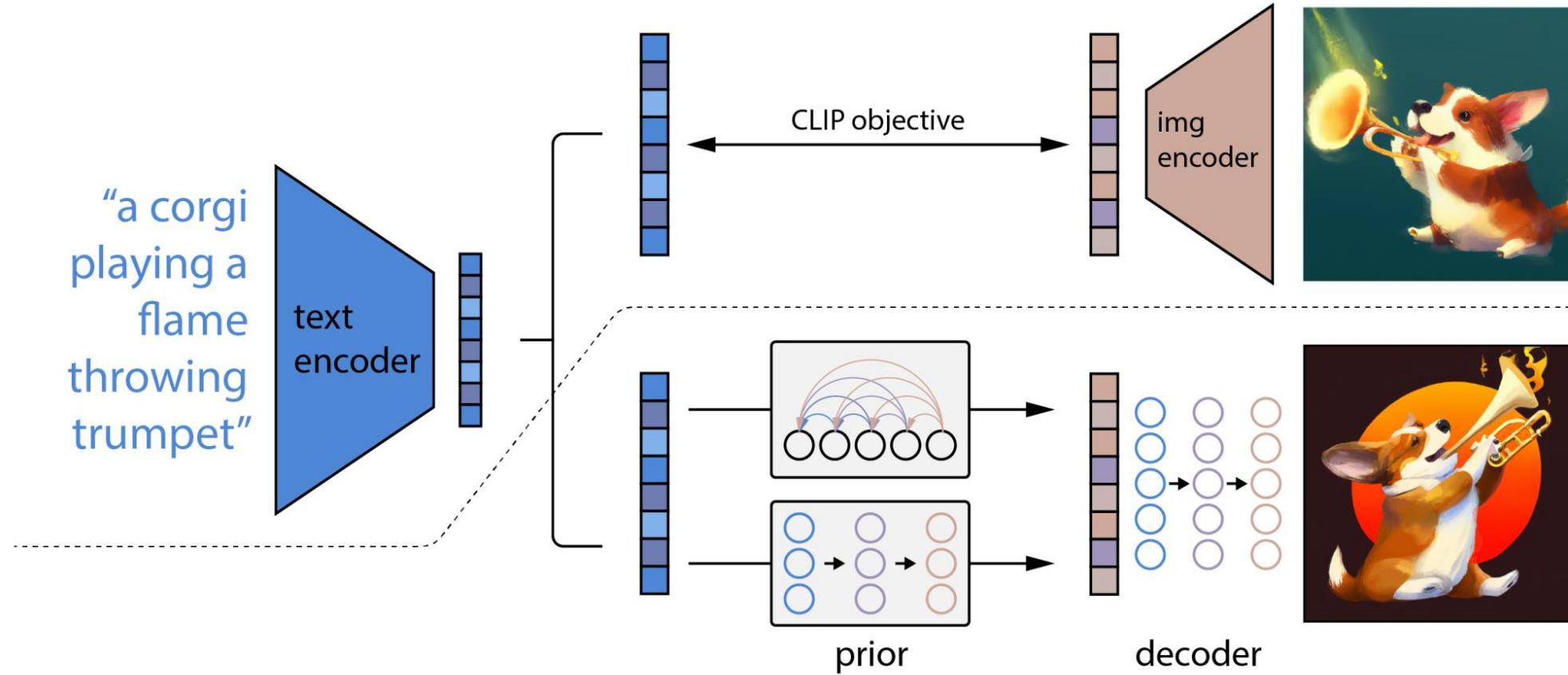
- одна предметная область
 - обучение с учителем
 - требуется много размеченных данных



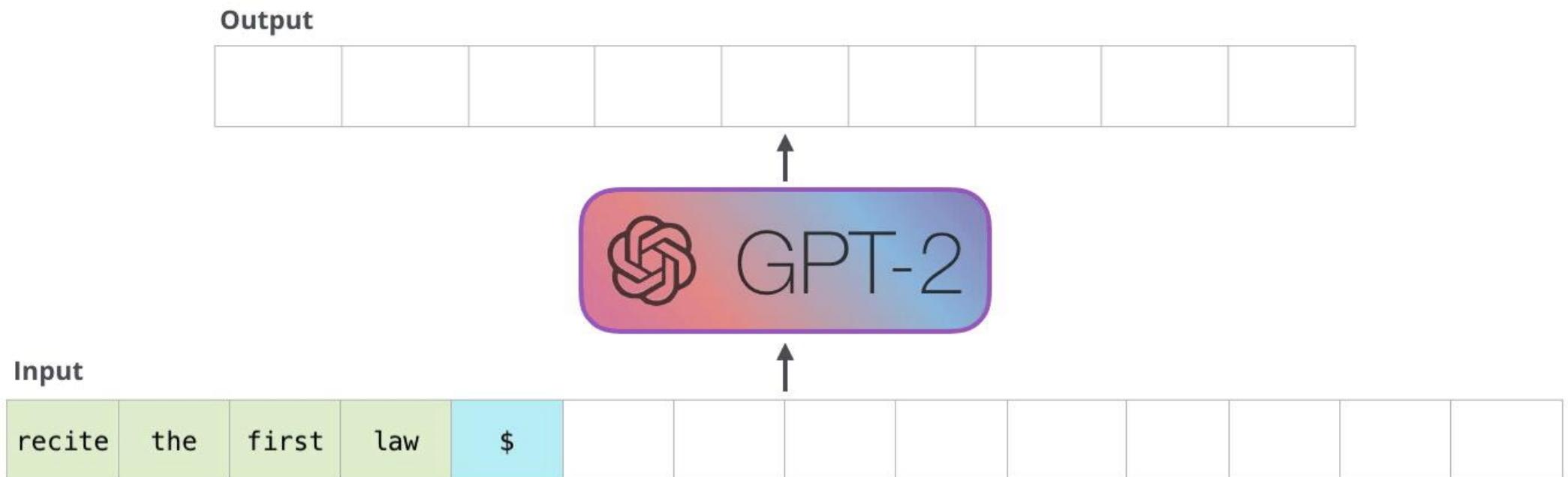
- несколько задач
- предобучение + дообучение
 - очень много данных
- мало размеченных данных



Моделирование (генерация) изображений

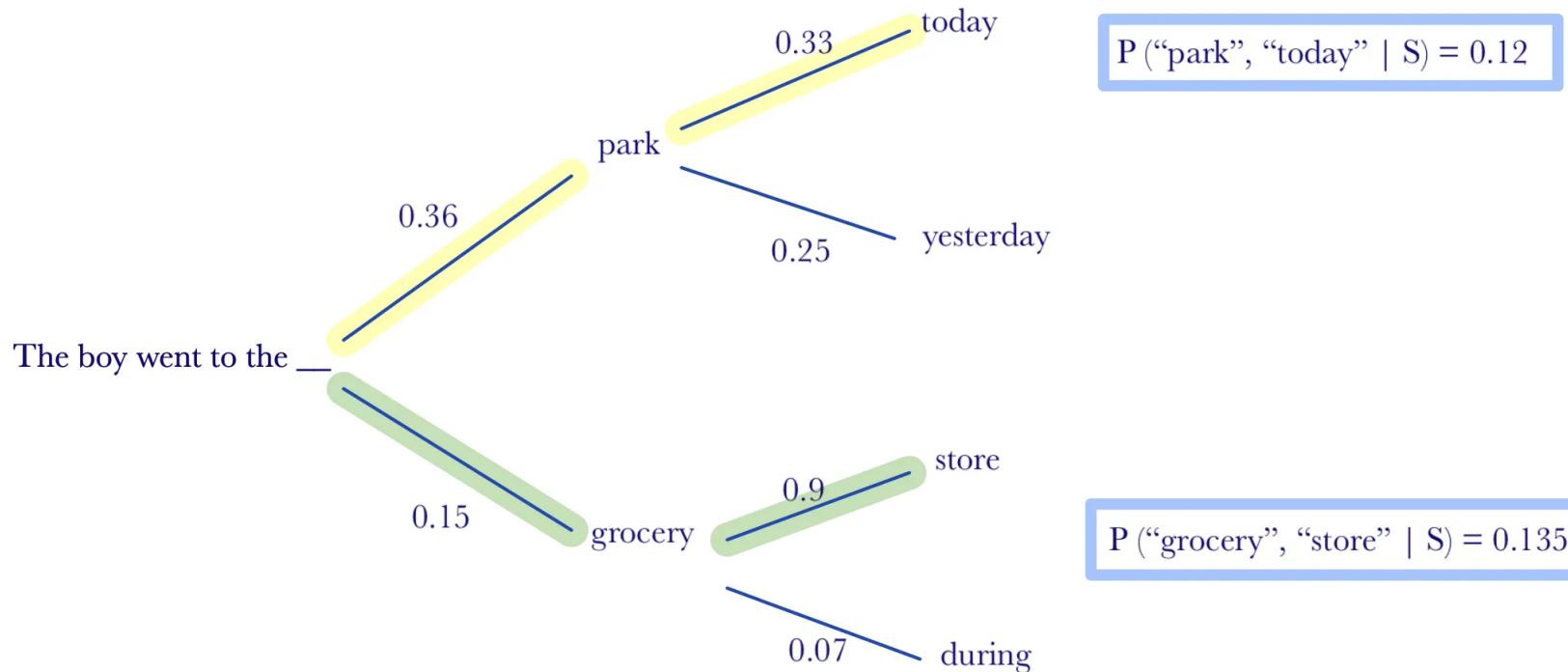


Генеративные языковые модели



Большие языковые модели (Large Language Models)

LLM ≈ “Средний” текст из интернета
Как сделать модель более специфичной?



Обучение с обратной связью от человека (RLHF)

Step 1

**Collect demonstration data
and train a supervised policy.**

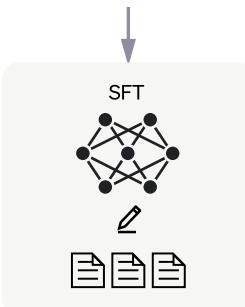
A prompt is
sampled from our
prompt dataset.

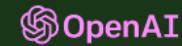


A labeler
demonstrates the
desired output
behavior.



This data is used to
fine-tune GPT-3.5
with supervised
learning.





ChatGPT: Optimizing Language Models for Dialogue

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.

TRY CHATGPT ↗

November 30, 2022
13 minute read

BBC Sign in Home News Sport Weather iPlayer

NEWS

Home | Cost of Living | War in Ukraine | Coronavirus | Climate | UK | World | Business | Politics | Tech

Technology

ChatGPT: New AI chatbot has everyone talking to it

© 7 December 2022

FORTUNE

RANKINGS ▾ MAGAZINE NEWSLETTERS PODCASTS MORE ▾

SEARCH SIGN IN Subscribe Now

NEWSLETTERS · CEO DAILY

The advent of OpenAI's ChatGPT may be the most important news event of 2022

BY ALAN MURRAY AND DAVID MEYER

December 12, 2022 at 2:08 PM GMT+3

Updated January 5, 2023 at 12:35 AM GMT+3

MIT Technology Review

ARTIFICIAL INTELLIGENCE

The ChatGPT-fueled battle for search is bigger than Microsoft or Google

Bloomberg

Europe Edition ▾

Newsletter

ChatGPT Could Be AI's iPhone Moment



World ▾ Business ▾ Legal ▾ Markets ▾ Breakingviews Technology ▾



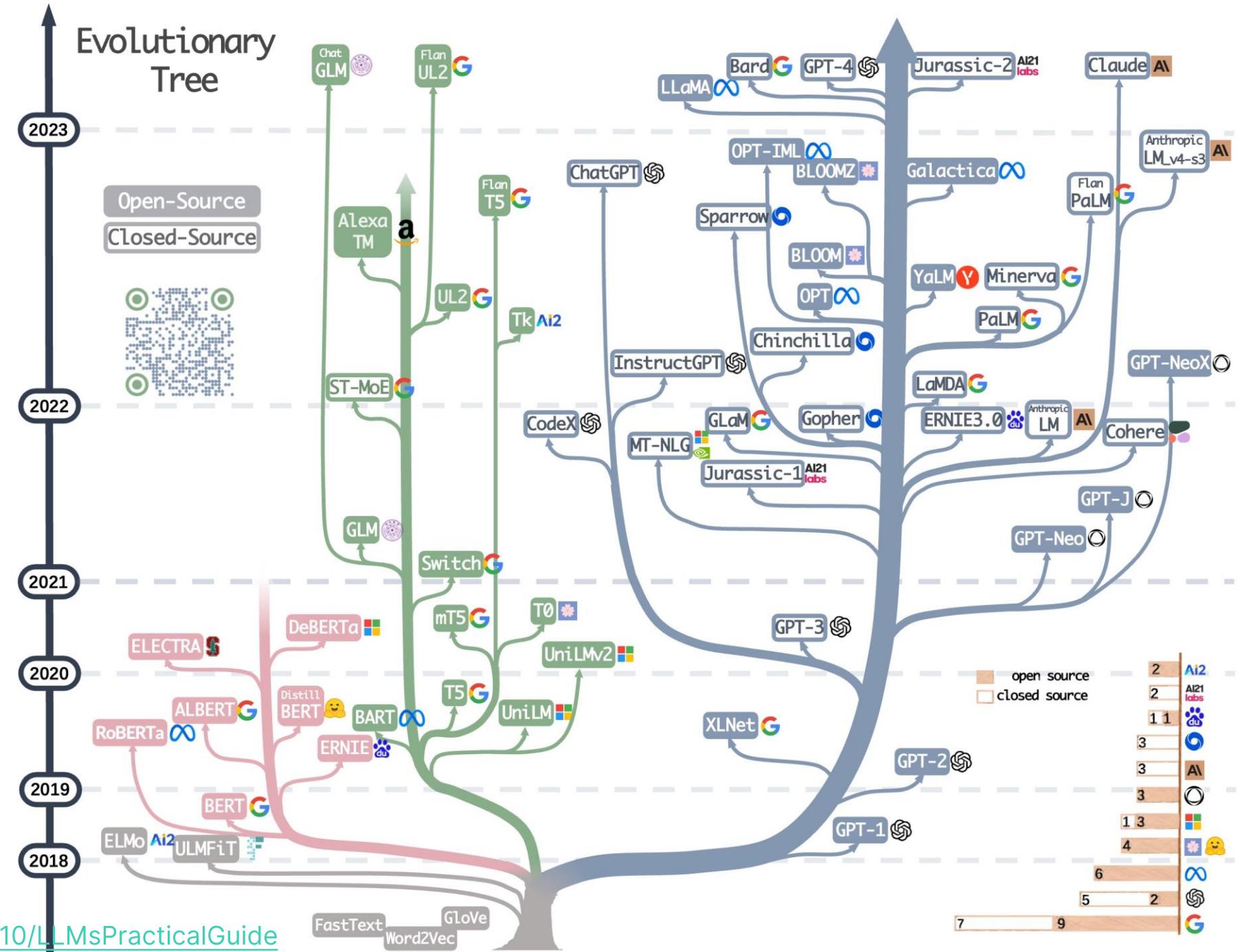
Disrupted

5 minute read · December 15, 2022 6:09 PM GMT+3 · Last Updated 2 months ago



Exclusive: ChatGPT owner OpenAI projects \$1 billion in revenue by 2024

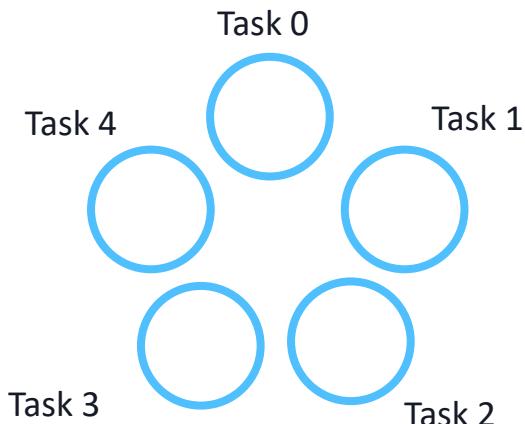
Эволюция языковых моделей



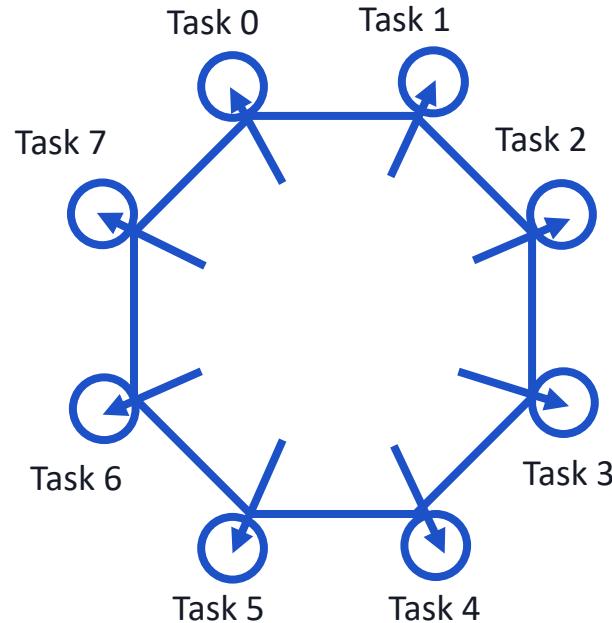
Универсальность методов обучения



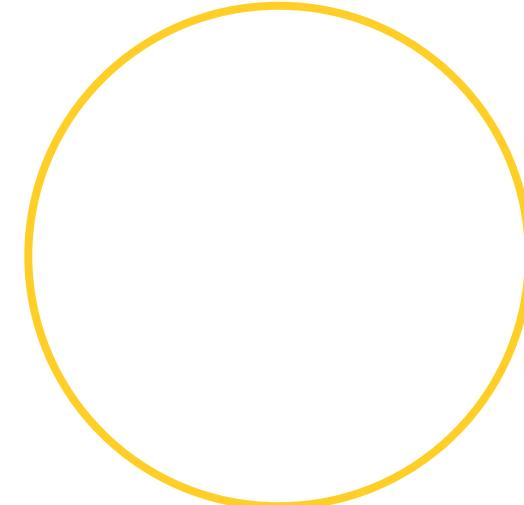
- одна предметная область
 - обучение с учителем
 - требуется много размеченных данных



- несколько задач
- предобучение + дообучение
 - очень много данных
- мало размеченных данных



- промт (zero-shot обучение)
 - предобучение
- очень очень много данных
 - разметка не нужна

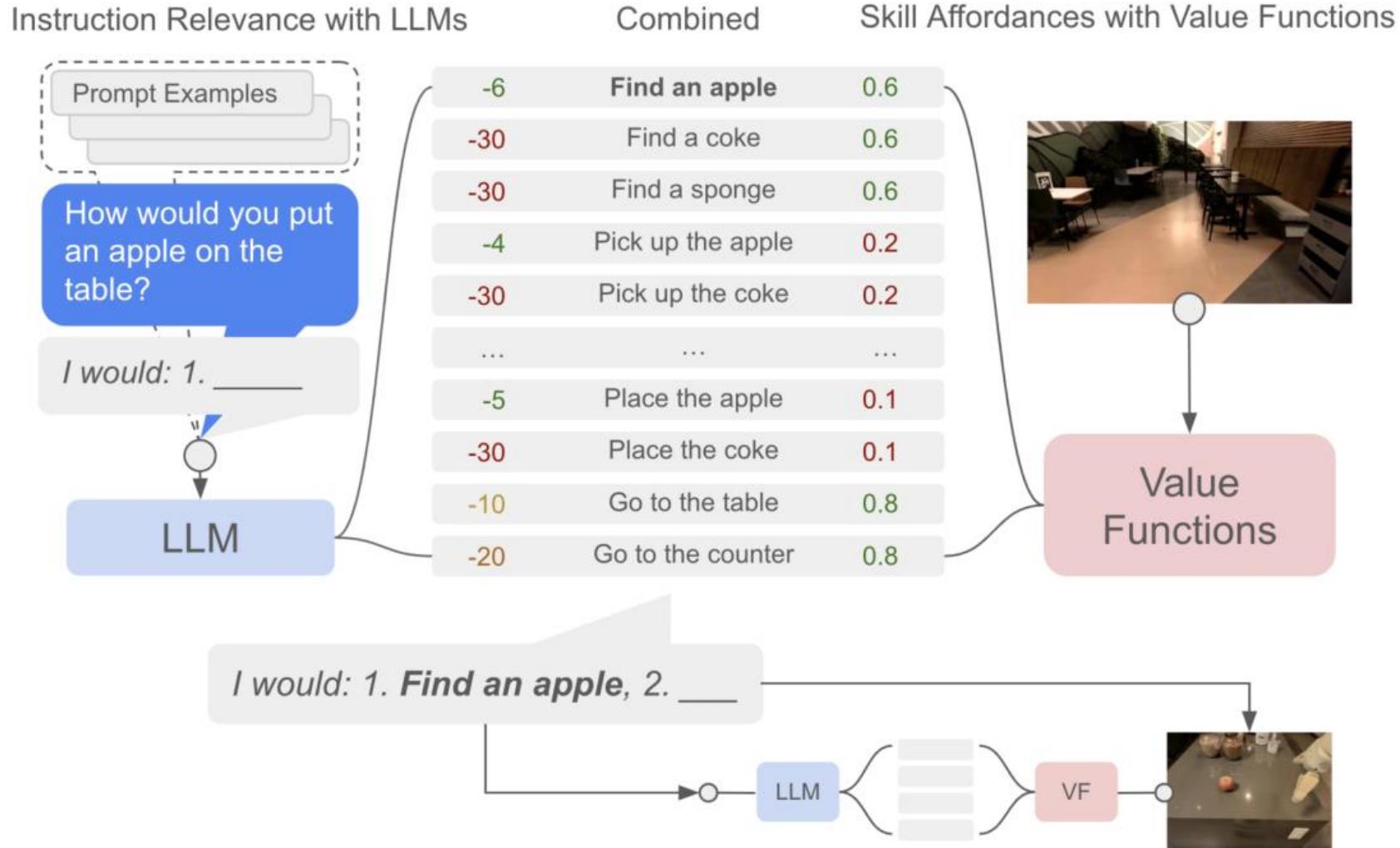


Завтрашний день ИИ

Погружение языковых моделей в реальный мир

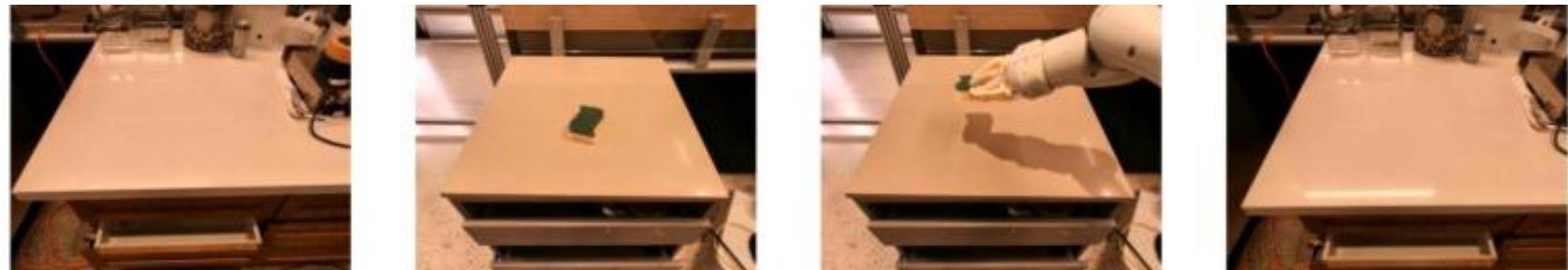


Погружение языковых моделей в реальный мир



Погружение языковых моделей в реальный мир

Human: I spilled my coke, can you bring me something to clean it up?

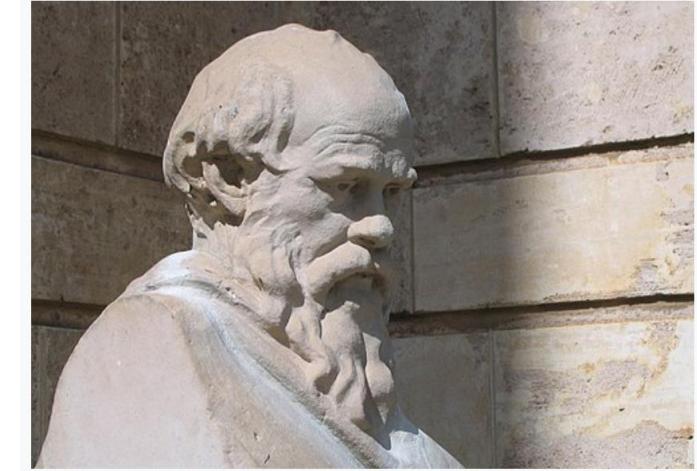


Robot: I would
1. Find a sponge
2. Pick up the sponge
3. Bring it to you
4. Done



УРОКИ АНТИЧНОСТИ

Мéтод Сокráта — [метод](#), названный в честь древнегреческого философа [Сократа](#), основывающийся на проведении [диалога](#) между двумя индивидуумами, для которых [истина и знания](#) не даны в готовом виде, а представляют собой [проблему](#) и предполагают поиск. Этот метод часто подразумевает дискуссию, в которой [собеседник](#), отвечая на заданные [вопросы](#), высказывает [суждения](#), обнаруживая свои знания или, напротив, своё неведение.



Сократические модели

Video Search

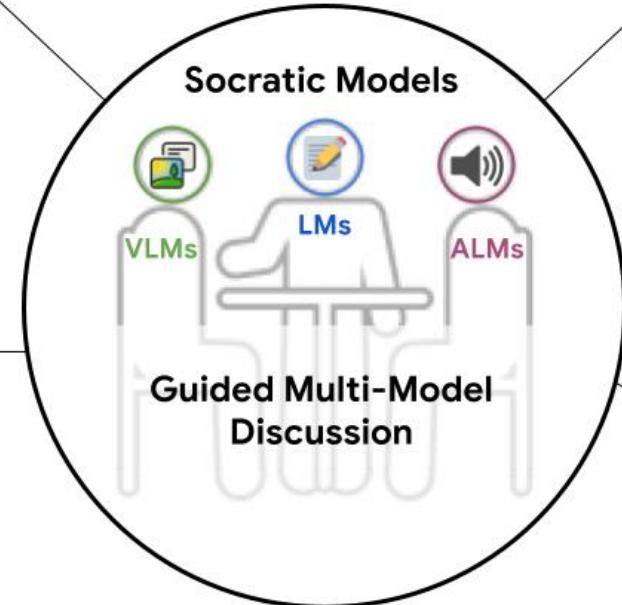
Q: Where did I leave my mug?



Image Captioning



Summary: I am watching netflix in a living room.



Free-form Video Q&A: Visual & Contextual Reasoning

Q: When did I wash my hands? Q: Why was I at the front porch? Q: Why was I chopping wood?



A: I last washed my hands at 3:38 PM.

A: I went to the front porch today to receive a package.

A: Because I needed to get a fire going in the fireplace.

Forecasting: Predicting Future Activities



1:46 PM: I am eating a sandwich in a kitchen.
2:18 PM: I am checking time and working on a laptop in a clean room.
2:49 PM: I am buying produce from a grocery store or market.

3:21 PM: I am driving a car.
4:03 PM: I am in a park and see a playground.
4:35 PM: I am in a home and see a television.

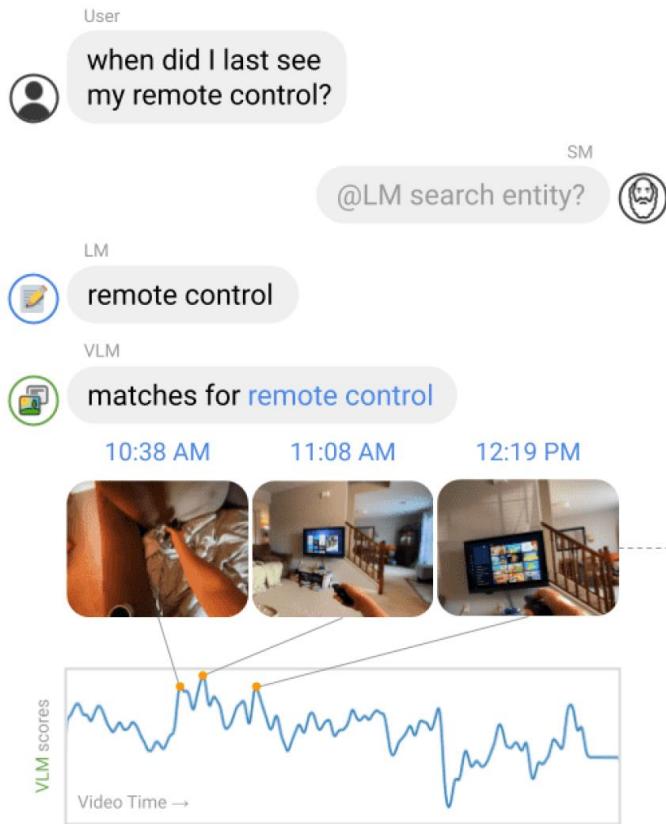
- No training
- Multimodality
- Commonsense
- Reasoning

Self-supervised
generative \ metric pretraining

Language modeling

Сократические модели

Socratic Question to Video Search



Socratic Egocentric Image Summaries

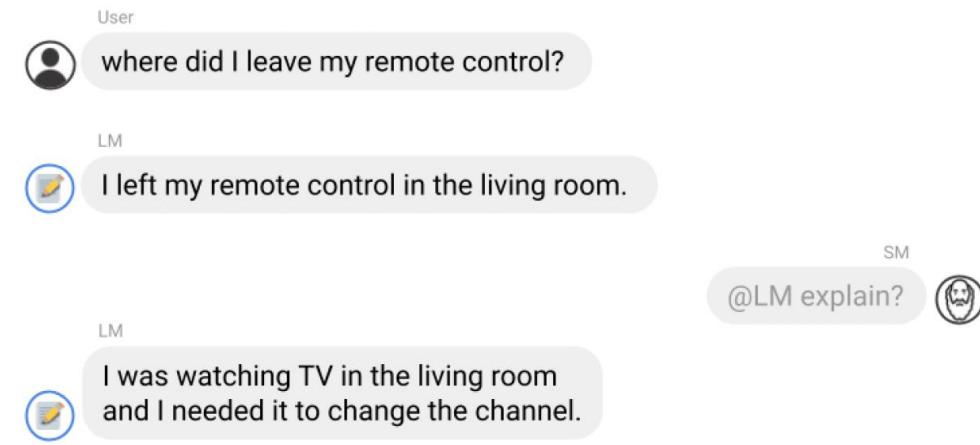


Socratic Video Q&A with Linguistic World-State History

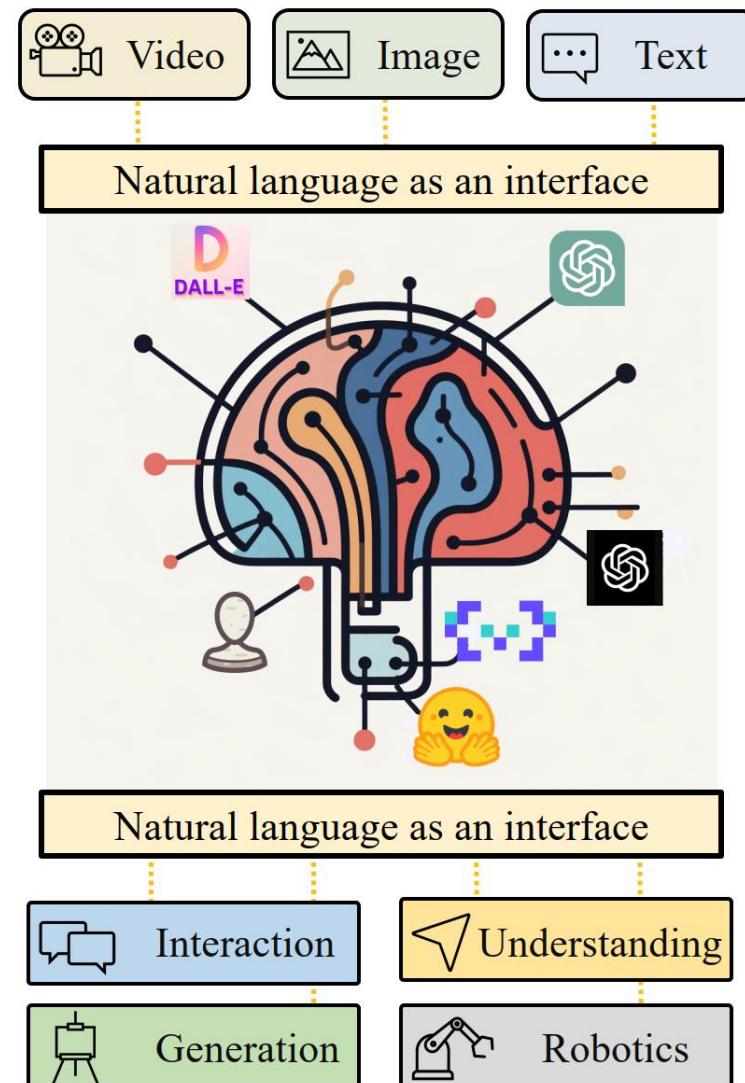
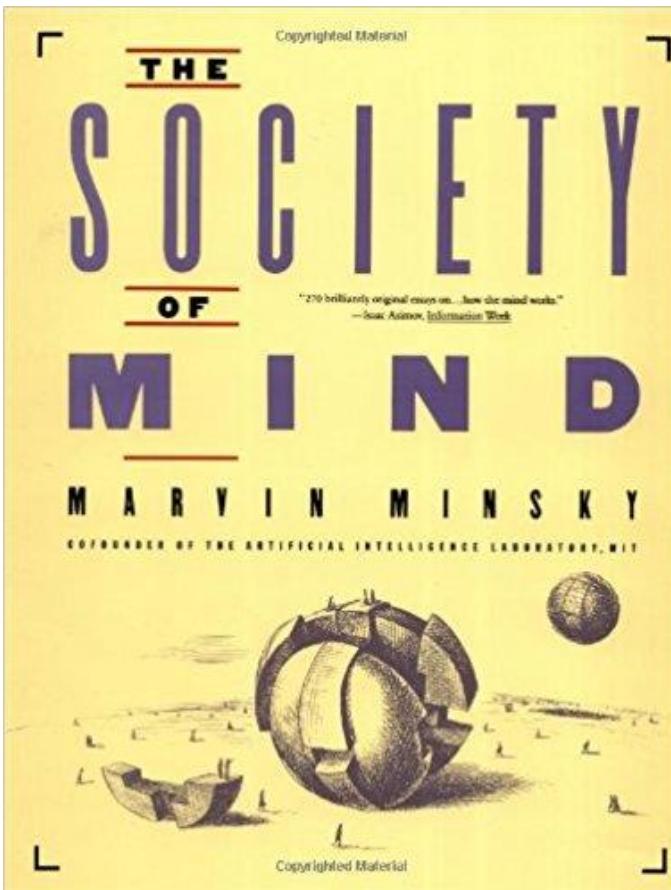
Generated Linguistic World-State History from Video

10:38 AM: Places: living room. Objects: sofa bed, loveseat, coffee table.
Activities: watching TV. I was watching TV in a living room.
11:08 AM: Places: living room. Objects: remote control, television, netflix.
Activities: watching netflix. I was watching netflix on the television.
12:19 PM: Places: television room. Objects: television, remote control,
netflix. Activities: watching netflix. I was watching netflix on a television.

...

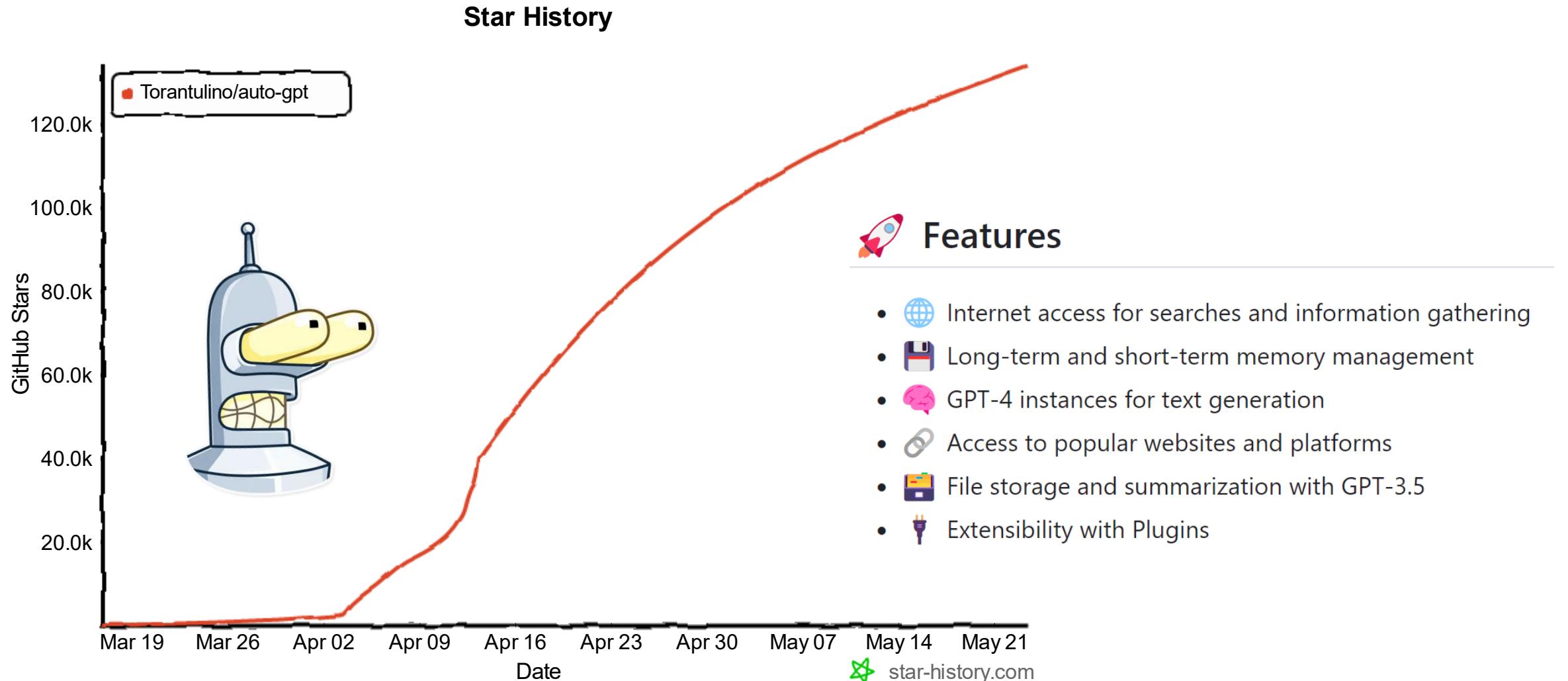


Сообщество ИИ

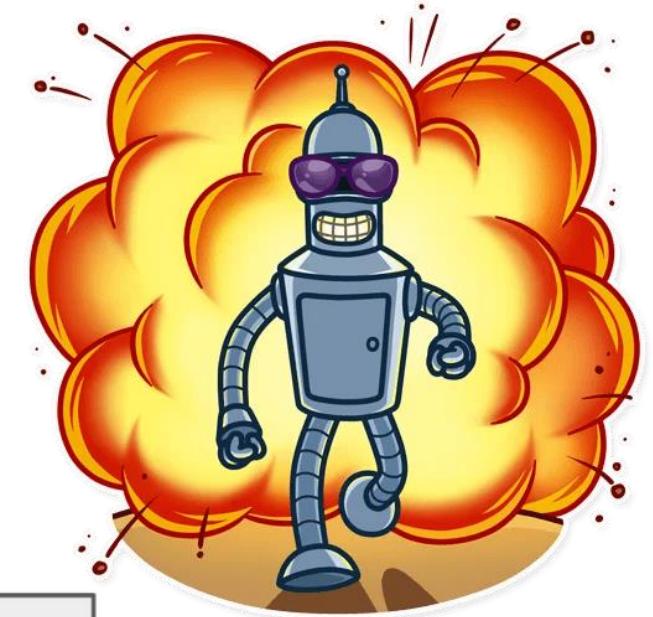
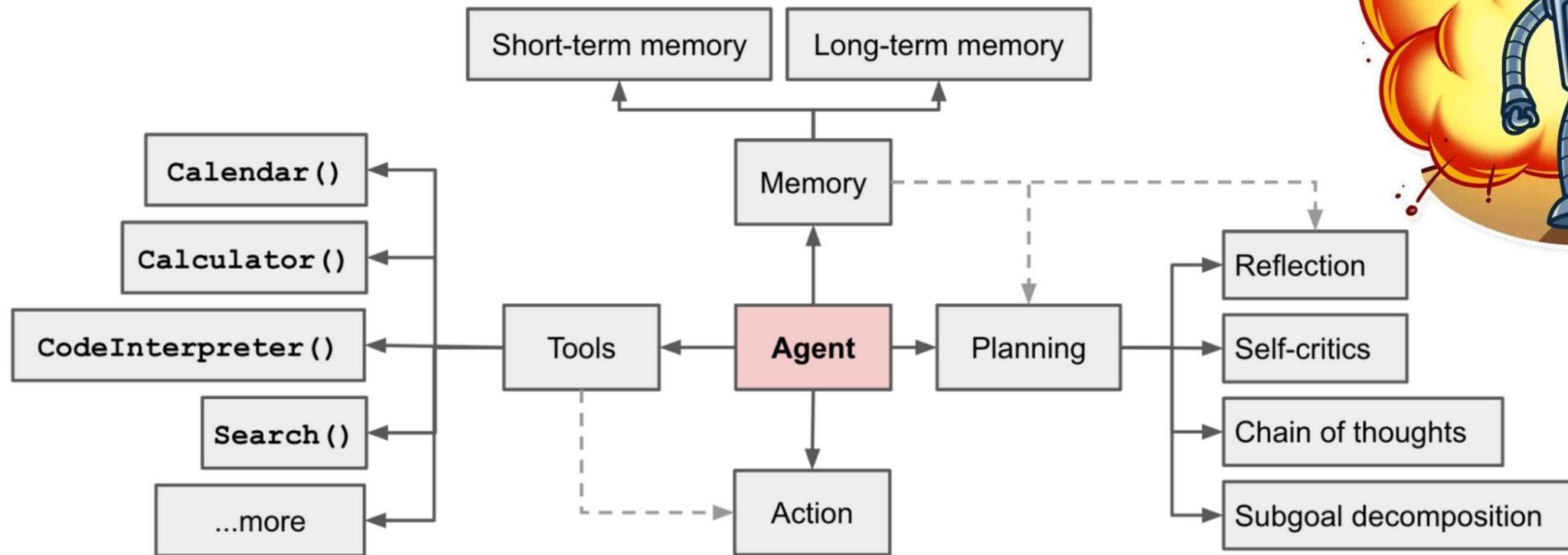


Natural Language-Based Societies of Mind (NLSOM)

ИИ создает самого себя: Auto-GPT



Автономные ИИ агенты



Ненадежность больших языковых моделей

ChatGPT3 on fractions :

Me: what is larger, two fourths or one half?

ChatGPT :

One half is larger than two fourths.

Two fourths can be written as $2/4$, which is equal to $1/2$.
Therefore, one half is equal to two fourths.

Me: didn't you just contradict yourself?

ChatGPT: I apologize, I made a mistake in my previous answer.
Two fourths ($2/4$) is smaller than one half ($1/2$).

Me : Are you sure?

ChatGPT: Yes, I am sure. Two fourths ($2/4$) is equal to $1/2$. Since
 $1/2$ is larger than $2/4$, it follows that one half is larger than two
fourths.



Вызовы для автономных агентов

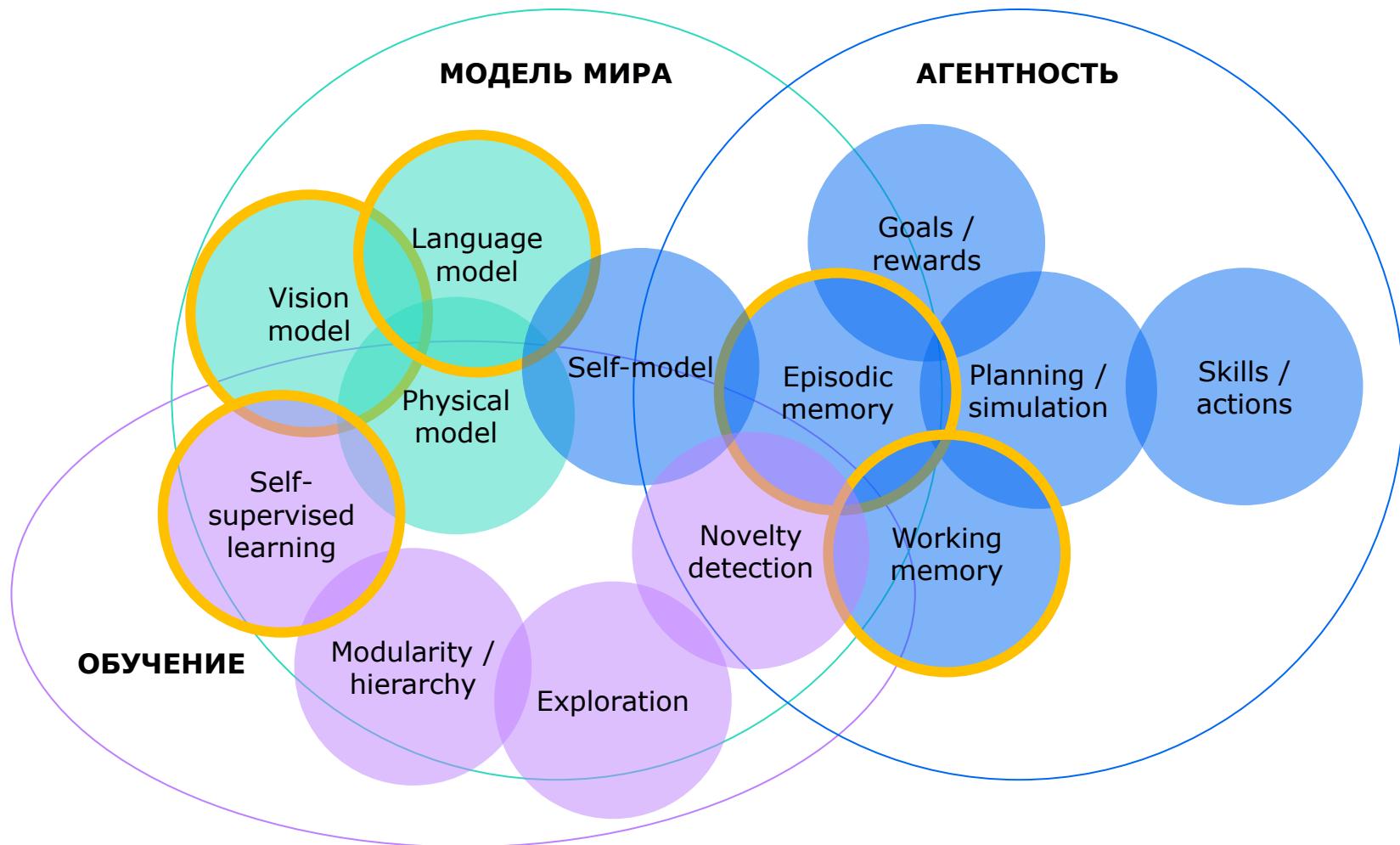
- Длина входного окна ограничена – вся необходимая информация для планирования и промежуточная информация в процессе выполнения задачи не может быть доступна модели для принятия решений
- Декомпозиция задачи и долговременное планирование страдает от недостаточности контекстуальной информации – модель произвольно разрешает неоднозначные ситуации
- Накопление ошибок при работе – на каждом шаге присутствуют ошибки, которые быстро накапливаются при выполнении многоэтапных планов

Будущее ИИ

ИИ сегодня

- Большие данные + предобучение основа современных моделей ИИ
- Специализированные решения получаются дообучением доменных моделей
- Больше параметров модели + больше данных = лучше результат
- Генеративный искусственный интеллект позволяет создавать новые изображения и тексты неотличимые от естественных
- Все больше и больше моделей имеют интерфейс с естественным языком
- Языковые модели позволяют симулировать рассуждение и управлять взаимодействиями между другими моделями для решения задачи
- Начинают исследоваться сообщества ИИ агентов и автономный ИИ
- Системы ИИ становятся все более универсальными

Чего не хватает для создания универсального ИИ



Scientia potentia est!

Knowledge is Power!

Human I x AI = Power²

