

Quo Vadis, ИИ?

Кто мы такие, куда мы идем, как путь наш измерить

Александр Николаевич Горбань

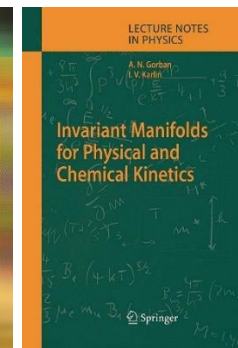
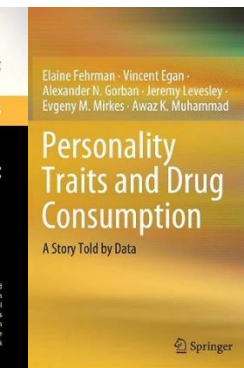
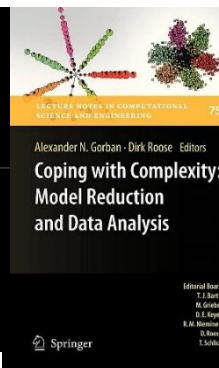
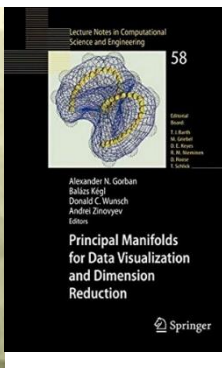
University of Leicester, UK

Нижегородский государственный университет
им. Н.И. Лобачевского, Россия

Я верю, что это так; что мы находимся на пороге эпохи, в которой будет очень велико влияние машин для решения интеллектуальных задач. Весьма возможно, что они будут доминировать.

Marvin Minsky, 1961

Позвольте представиться



Alexander N. Gorban (АН Горбань)

Professor of Applied Mathematics, [University of Leicester](#), UK

Подтвержден адрес электронной почты в домене le.ac.uk - [Главная страница](#)

[Multiscale Analysis](#) [Model Reduction](#) [Kinetic Equations](#) [Mathematical Chemistry](#)
[Mathematical Neuroscience](#)

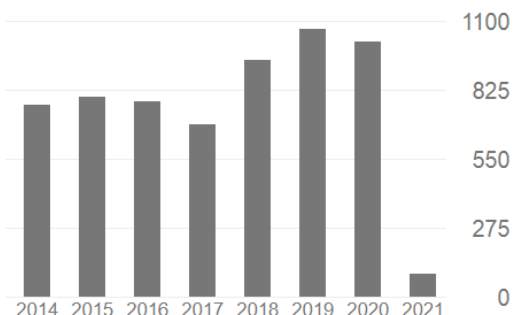
ПОДПИСКА ОФОРМЛЕНА

Процитировано

[ПРОСМОТРЕТЬ ВСЕ](#)

	Все	Начиная с 2016 г.
Статистика цитирования	12840	4611
h-индекс	55	34
i10-индекс	195	107

НАЗВАНИЕ	ПРОЦИТИРОВАНО	ГОД
☐ Neural networks on personal computer AN Gorban, DA Rossiev Novosibirsk: the Science, 276	1102 *	1996
☐ Training neural networks AN Gorban Moscow: USSR-USA JV" Paragraph", Russian. English translation: AMSE Trans ...	759 *	1990
☐ Neuroinformatika AN Gorban. WL Dunin-Barkowski. AN Kirdin. EM Mirkes. ...	683 *	1998



Позвольте

Оргкомитет девятнадцати ежегодных международных конференций
Нейроинформатика 1999 - 2017 и
Правление Российской ассоциации нейроинформатики

В знак признания выдающихся успехов в области теории и приложений искусственных нейронных сетей и за неоценимый вклад в организацию и проведение конференций присвоили

ПРОФЕССОРУ
ГОРБАНЮ АЛЕКСАНДРУ НИКОЛАЕВИЧУ

почётное звание

ПИОНЕРА РОССИЙСКОЙ
НЕЙРОИНФОРМАТИКИ



Alexander N. Gorban

Professor of Applied Mathematics
Подтвержден адрес эл.
Multiscale Analysis, Mathematical Neuroscience

☐ НАЗВАНИЕ

☐ Neural networks on personal computers
AN Gorban, DA Rossiev
Novosibirsk: the Science, 276

☐ Training neural networks
AN Gorban
Moscow: USSR-USA JV "Paragraph", Russian Academy of Sciences, 1999

☐ Neuroinformatika
AN Gorban. WL Dunin-Barkowski. AN Kirdin. E

Проректор НИЯУ МИФИ

Председатель оргкомитета девятнадцати
ежегодных международных конференций
Нейроинформатика 1999-2017, профессор

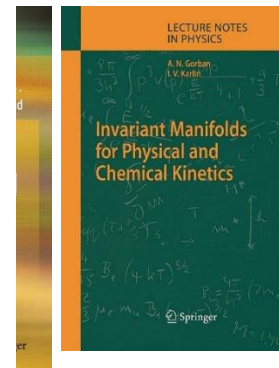
Президент Российской ассоциации нейроинформатики,
член-корреспондент Российской академии наук



В.В. Ужва

Б.Н. Оныкий

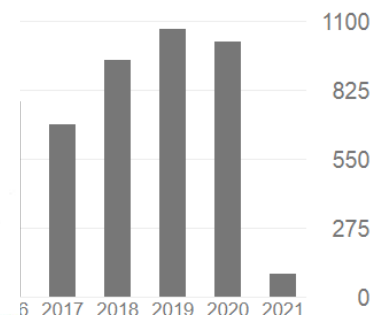
Б.В. Крыжановский



пано ПРОСМОТРЕТЬ ВСЕ

Все Начиная с 2016 г.

12840	4611
55	34
195	107



МОСКВА, 2 ОКТЯБРЯ 2017 г.

«Кейнот» пленарный доклад по нейронным сетям на Всемирном Конгрессе Вычислительного Интеллекта (IEEE WCCI2020)

← → ↺ wcci2020.org/speakers/ 🔍 ☆ 📧 ⚙️ 👤 ⋮



IEEE WCCI 2020
IEEE World Congress on Computational Intelligence
Glasgow, Scotland, United Kingdom – July 19-24, 2020

WCCI 2020 ▾ **SPEAKERS** PROGRAM SCHEDULE PROGRAMS ▾ REGISTRATION

CONTACT US

IJCNN Keynote Speakers



**ALEXANDER
N
GORBAN**

Monday,
20th
July,
2020
12:45-
13:45
Professor
of



**KUNIHIKO
FUKUSHIMA**

Tuesday,
21st
July,
2020
13:00-
14:00
Kuniyuki
Fukushima
received

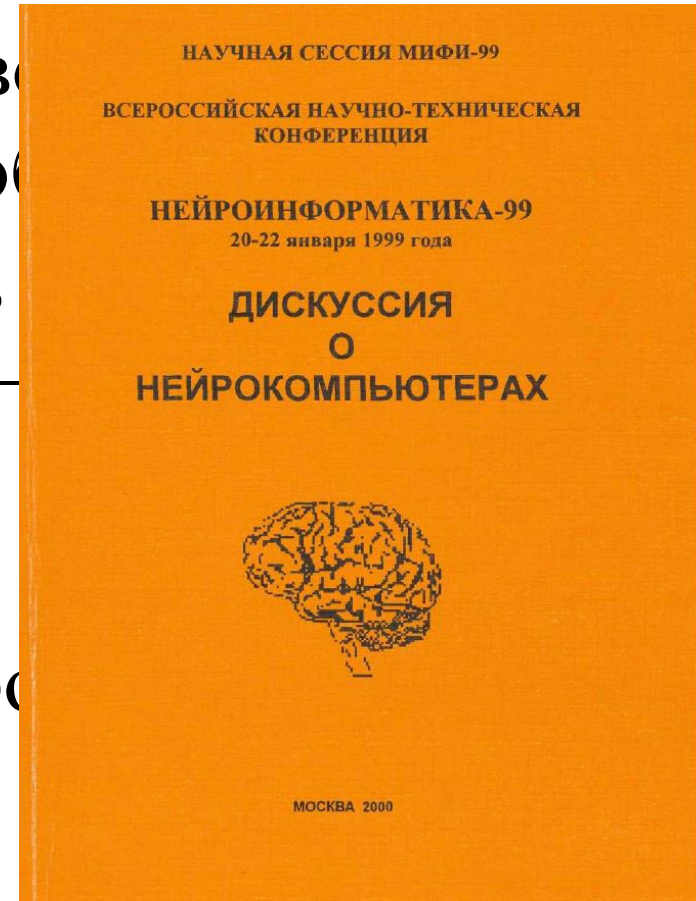
Распространение нейроинформатики сродни той информационной революции, которую произвели персональные компьютеры, одной важнейшей особенностью – переходом *от элитных высот к профанным техникам* (техникам, которые может освоить каждый человек с некоторым общим образованием – не более того, то есть профан, а не профессионал). Профанация – это не всегда плохо.

А.Н. Горбань

Дискуссия о нейрокомпьютерах
МИФИ 1999

Распространение нейроинформатики сродни той информационной революции, которую произвели персональные компьютеры, одной важнейшей особенностью – переходом *от элитных высот к профанным техникам* (техникам, которые может освоить каждый человек с некоторым общим образованием, а не более того, то есть профан, непрофессионал). Профанация – плохо.

Дискуссия о нейро



Во время подготовки к этой лекции я попытался ответить на три вопроса

- Что делают люди, разрабатывающие ИИ (кто мы такие)?
- Куда движется развитие ИИ (куда мы идем)?
- Как измерить движение по этому пути (как путь наш измерить)?

Латинское «Quo Vadis» - вопрос, заимствованный из апокрифических Деяний Петра – сокращенное выражение этих трех вопросов.

План

1. Немного истории.
2. Определение, пришедшее «из народа».
3. Размерность и всемогущество неограниченного перцептрона (Теорема 1 Розенблатта).
4. Ограниченные возможности ограниченного перцептрона, Минский&Пайперт и первая зима.
5. Всплеск конца 1980х – начала 1990х. Почти все алгоритмы готовы!
6. Взрыв активности в 21 веке! Начинается настоящее время.

План

7. Начало спада и проблема ошибок.
8. Конструкция корректоров.
9. Объяснимость и логическая прозрачность ИИ.
10. Сообщества ИИ агентов.
11. Является ли ИИ интеллектом?
12. Изменение понятия «истина».

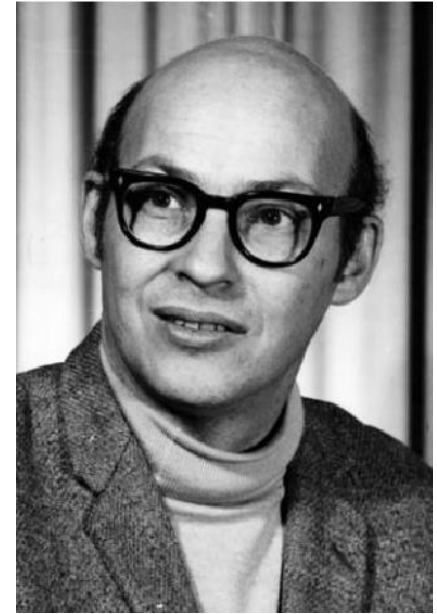
История – ключ к настоящему

Дартмутский семинар – два месяца летом 1956. 10 классиков



Джон Маккарти –
лидер «аккуратных»

Мы попытаемся понять, как
обучить машины
использовать естественные
языки, формировать
абстракции и концепции,
решать задачи, сейчас
подвластные только людям,
и улучшать самих себя.



Марвин Минский –
лидер «неряшливых»

В 1959 Минский и Маккарти основали в MIT лабораторию
информатики и искусственного интеллекта

Две базовых формы ИИ

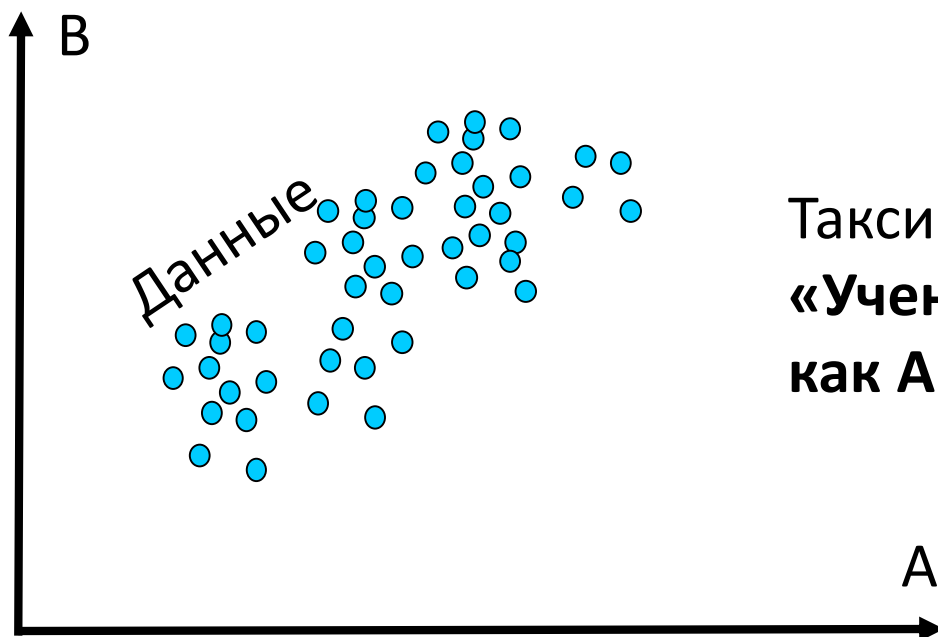
- **ИИ, основанный на знаниях** – это очень 😊 продвинутая форма справочника.



- **ИИ, основанный на данных** – это очень 😊 усовершенствованная форма статистики. По определению 1833 года: «цель статистики заключается в представлении фактов в наиболее сжатой форме» (для учета и прогноза).



Определение «из народа»



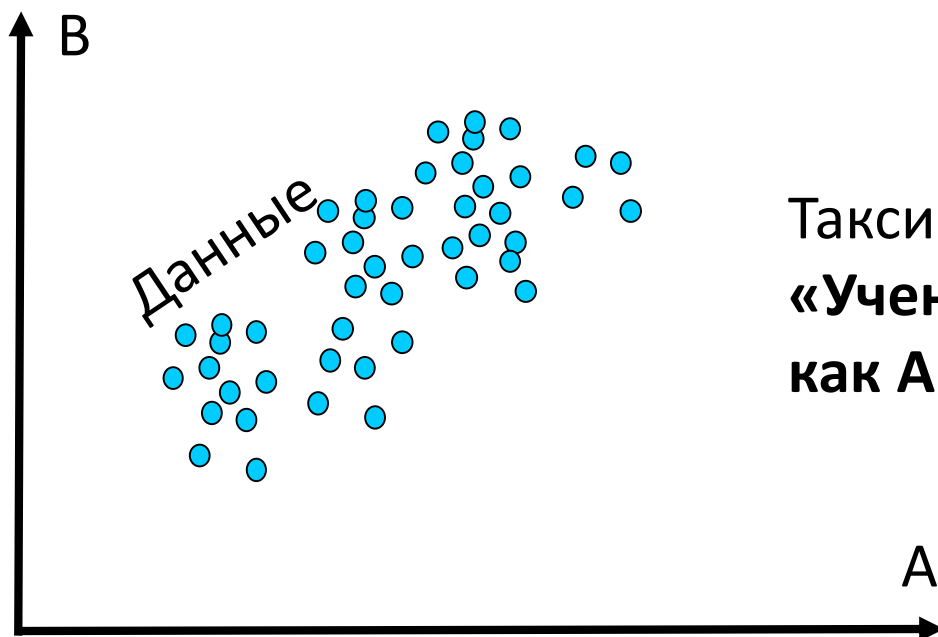
Таксист – стажеру АН:
**«Ученый? Значит изучаешь
как А связано с В?»**

Если А и В многомерны и сложны, то возникают три творческих вопроса:

1. Как оцифровать А?
2. Как оцифровать В?
3. В какой форме искать зависимость В от А?

Остальное – дело техники.

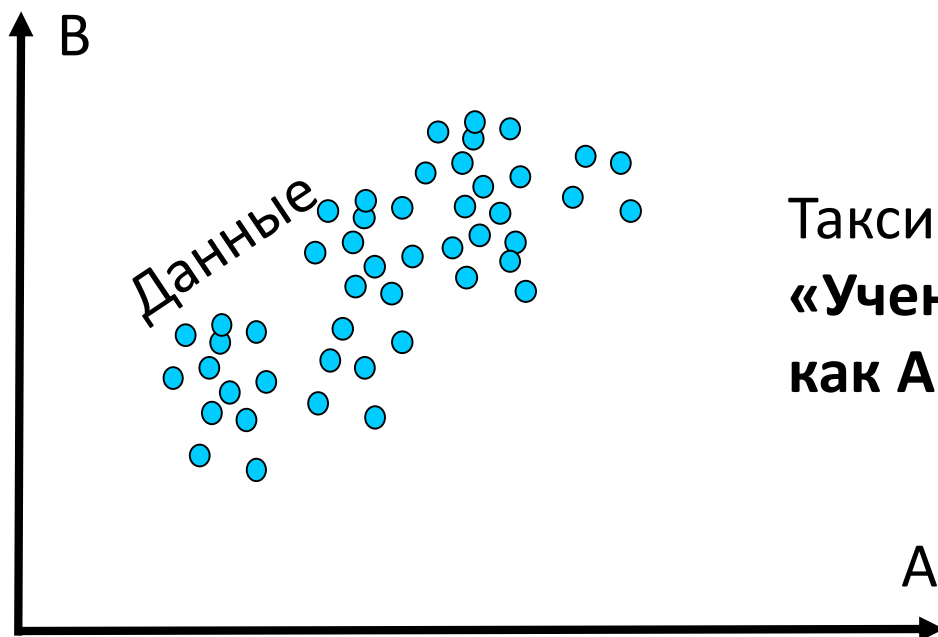
Определение «из народа»



Таксист – стажеру АН:
**«Ученый? Значит изучаешь
как A связано с B ?»**

Представим, что A – предложение на русском языке, а B – его перевод на английский (китайский, венгерский, ...).
Коллекцию текстов найдем в Интернете.
Как оцифровать предложения переменной длины для машинного обучения? В какой форме искать преобразование одного в другой? Целиком? Последовательно (слово за слово)? Еще как-нибудь?

Определение «из народа»



Таксист – стажеру АН:
**«Ученый? Значит изучаешь
как А связано с В?»**

Представим, что А – ситуация на доске игры ГО, а В – выбор хода.

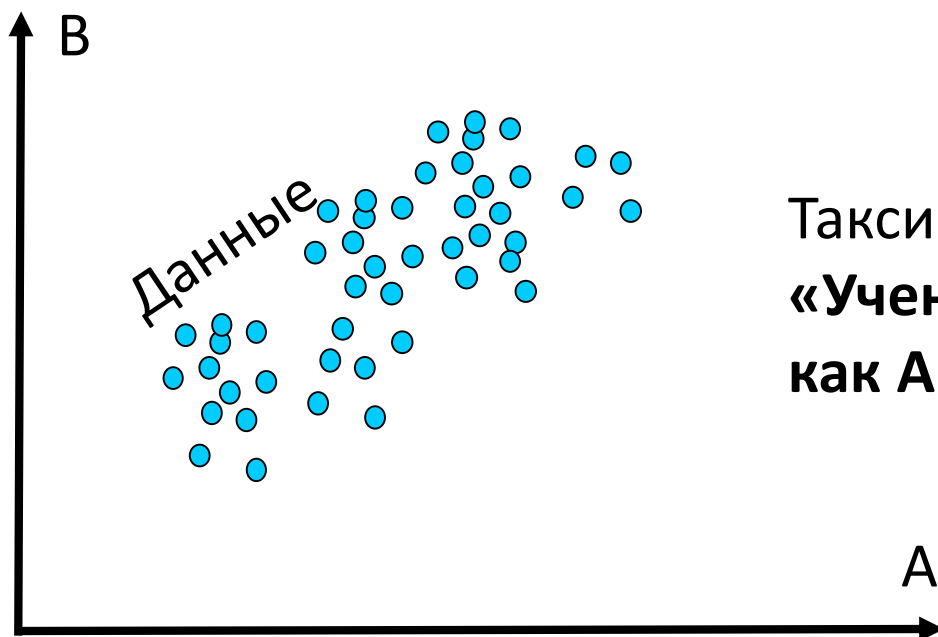
Как оценивать качество хода?

Как их оцифровать для машинного обучения?

В какой форме искать преобразование А в В?

Где найти данные для обучения – желательно миллионы партий?

Определение «из народа»



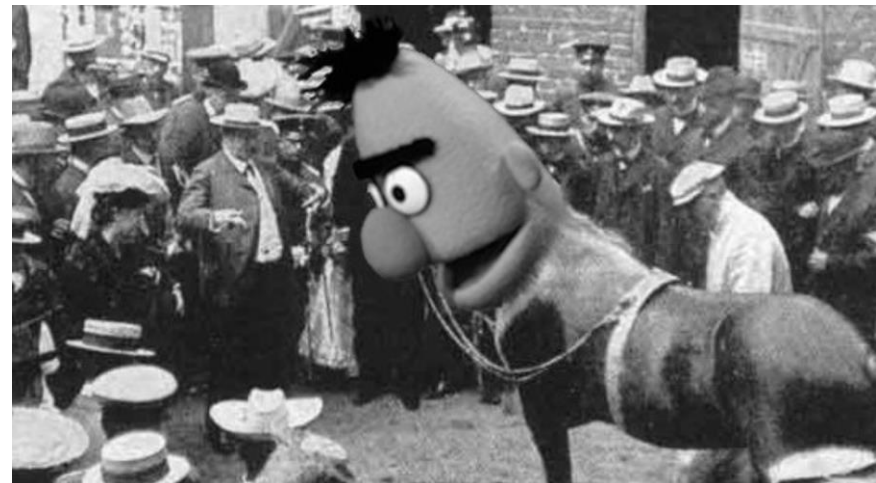
Таксист – стажеру АН:
**«Ученый? Значит изучаешь
как А связано с В?»**

Как видим, вопросов больше трех:

1. Как оцифровать А?
2. Как оцифровать В?
3. В какой форме искать зависимость В от А?
4. Как оценить качество найденного решения?
5. Где взять данные для обучения с уже оцененными ответами?

Бойся «Умного Ганса»!

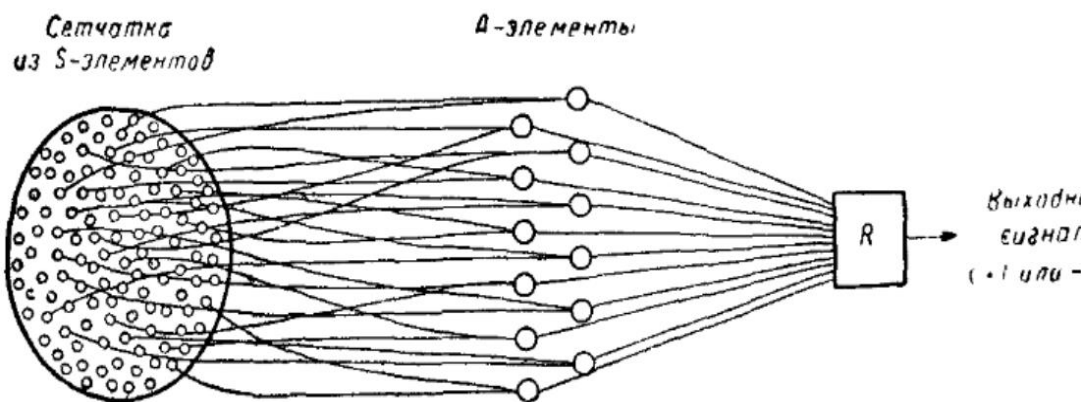
- Эффект Умного Ганса – проблема учащихся, решающих задачу путем научения ложным ассоциациям.
- Назван в честь лошади, которая, казалось, могла решать задачи,
- но в действительности реагировала на произвольные сигналы, подаваемые ее тренером.
- Пример «блестящего танка»: фото «своих танков» были на ярком фоне, а «чужих» на темном – распознавался фон.



Теорема 1 Розенблатта (Frank Rosenblatt). Перцептрон

Ф. Розенблатт. Принципы нейродинамики.
Перцептроны и теория механизмов мозга,
1962 (отчет 1961, русский перевод М., 1965)

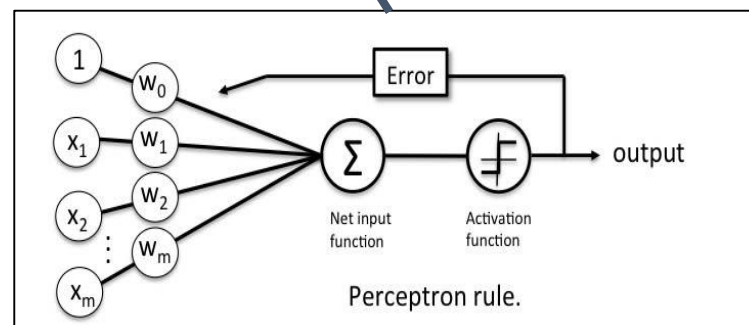
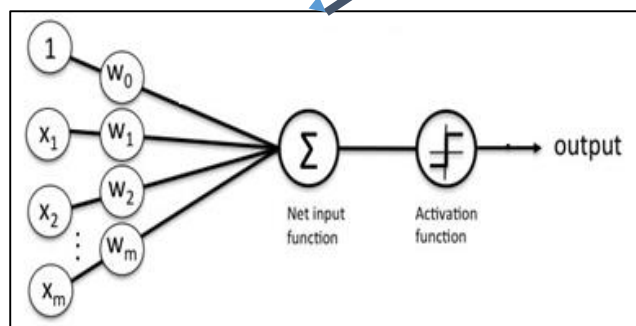
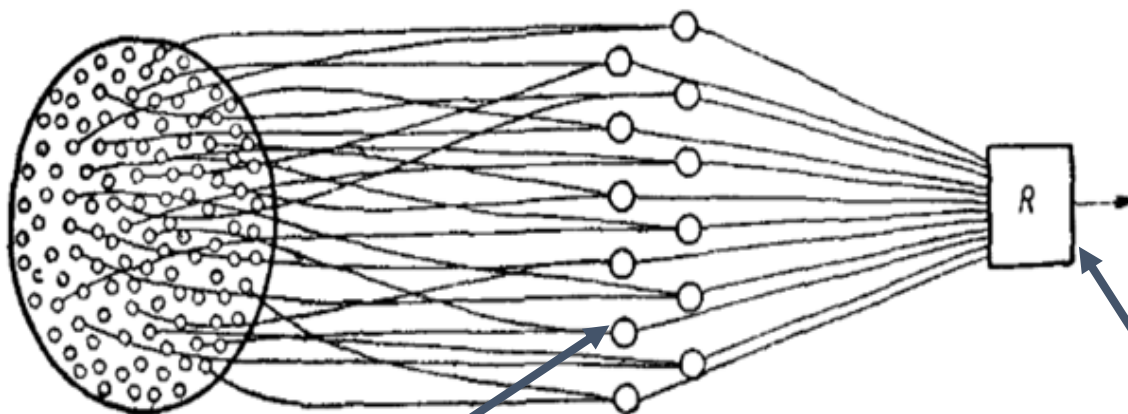
Учился вместе с Минским в
Bronx High School of Science



Ф и г. 5. Типичная структурная схема элементарного перцептрона.

Теорема 1 Розенблатта:
Элементарный
перцептрон может
решить любую задачу
классификации.
**«Всесовершенство
неограниченного
перцептрона»**

Доказательство теоремы 1 Розенблатта основано на том, что в если **размерность пространства свойств**, поставляемых скрытым слоем, **достаточно велика**, то обучающее множество в нем образует **симплекс**. Требуется еще, чтобы примеры разных классов не склеивались в пространстве свойств. Никакой линейной разделимости заранее не предполагается!



Ограниченные возможности ограниченного перцептрона



Теорема Минского-Пайперта 1969

Если А-элементы получают сигналы с ограниченных площадок, то элементарный перцептрон может не все – например, он не может определить связность – **больше слоев надо.**

.....

*Но люди слышат другое:
пропускают слово
«ограниченный» (и не
читают книг).*

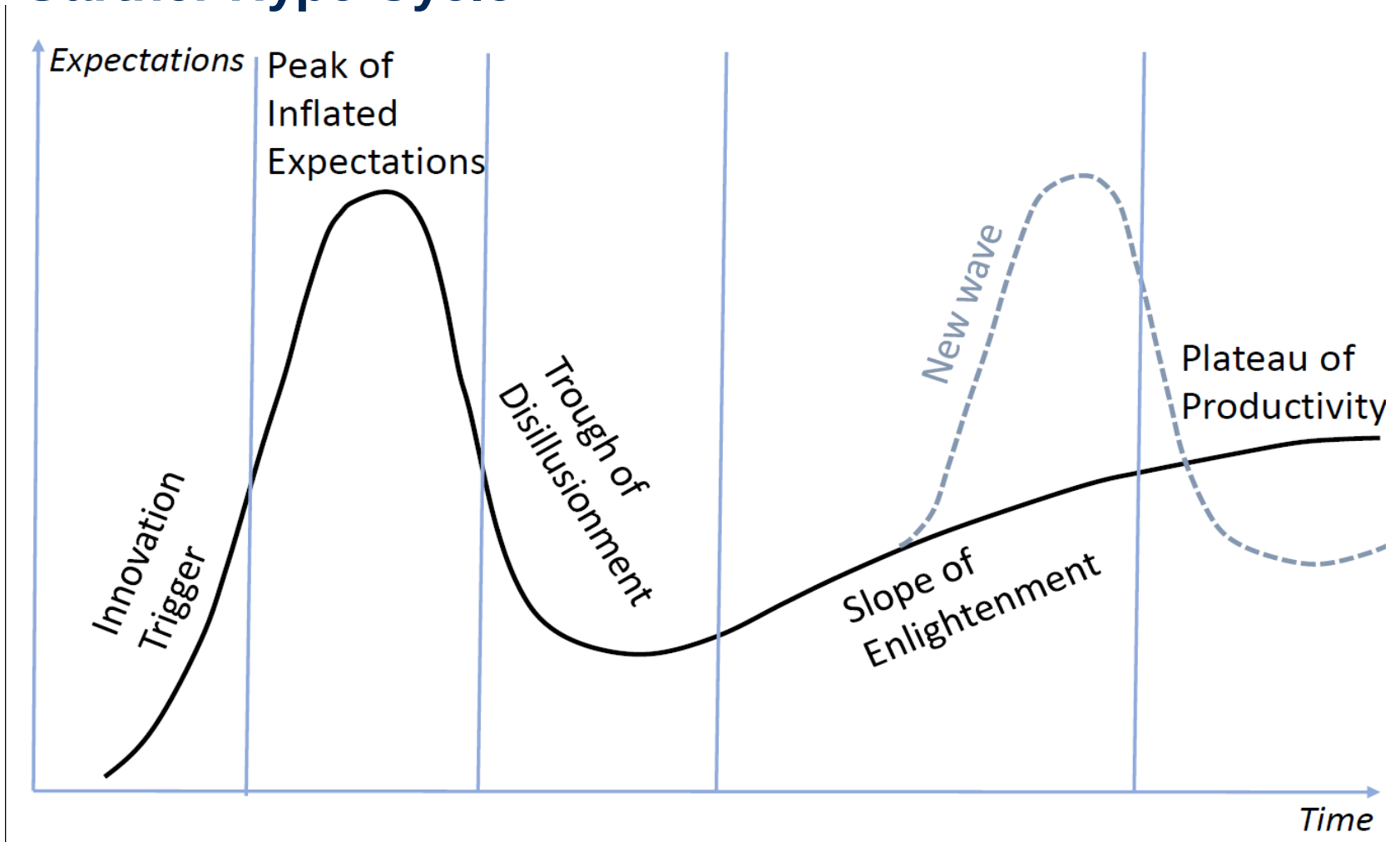
Великие имена ИИ первой волны в СССР (1956-1970)

(ИИ, основанный на данных)

А.Г. Ивахненко, М. М. Бонгард, М. Л. Цетлин,
М.А. Айзерман, Э.М. Браверман, Л.И. Розоноэр,
В.И. Варшавский, В.С. Гурфинкель, В.Н. Вапник,
А.Я. Червоненкис, А. И. Галушкин,...

Гартнеровский «цикл шумихи»

Gartner Hype Cycle



Гартнеровский «цикл шумихи»

Gartner Hype Cycle



Зима ИИ 1970-???

(середина 1980х? Начало 2010х?)



Причины зимы ИИ

- Теорема Минского-Пайперта лишь сигнал – и неправильно прочитанный.
- Главное – неуспешность прикладных проектов ИИ в сочетании с интенсивной рекламой и раздутыми ожиданиями.

Зима не была полной

- А в это время классики копали – и провигались.
- Полномасштабное восстановление исследований – в середине 1980.
- Три «новации»: сети Хопфилда, метод обратного распространения ошибки, карты Кохонена.



Типы задач для ИИ

Удобно разделить задачи на пять основных направлений: поиск, распознавание образов, самообучение, планирование и индукция (индуктивный вывод)...

Marvin Minsky, 1961

PROCEEDINGS OF THE IRE

January

Steps Toward Artificial Intelligence*

MARVIN MINSKY†, MEMBER, IRE

The work toward attaining "artificial intelligence" is the center of considerable computer research, design, and application. The field is in its starting transient, characterized by many varied and independent efforts. Marvin Minsky has been requested to draw this work together into a coherent summary, supplement it with appropriate explanatory or theoretical noncomputer information, and introduce his assessment of the state-of-



Типы задач для ИИ

Удобно разделить задачи на пять основных направлений: поиск, распознавание образов, самообучение, планирование и индукция (индуктивный вывод)...

Marvin Minsky, 1961

- Проблемы Минского сосредоточили исследования на наиболее важных технических вопросах.
- Где-то 90% ИИ публикаций до сих пор сосредоточено на постепенном продвижении в этих проблемах.
- В настоящее время мы видим впечатляющие успехи в этих областях.

Три пришествия нейронных сетей

- Розенблатт и Уидроу (конец 1950х - начало 1960х);
- Хопфилд, Кохонен и Хинтон с соавторами (1980е);
- GOOGLE, Amazon и другие гиганты (2000е - сейчас).

Технические возможности компьютеров накапливались монотонно.

Интеллектуальные прорывы в области нейронных сетей опережали технические возможности, а потом замедлялись до появления нового поля возможностей.

Терпеливые стали классиками!

Три источника современного взрыва ИИ, основанного на данных

- Производительность компьютеров стала достаточной, чтобы учить большие системы.
- «Наглецы» попробовали – и получилось.
- Бизнес был готов и быстро сориентировался.

А предшественники в 1970х-1990х строили, строили – и построили достаточную для старта научную базу.

Причины третьего пришествия

- Компьютеры стали достаточно мощными;
- Появился и вырос Интернет;
- Информационные гиганты (GOOGLE, AMAZON, FACEBOOK, MICROSOFT) в достаточной мере развили бизнес сетевых услуг;
- Необходимость диверсификации дот-комов;
- Алгоритмы обучения были, в основном, готовы.

Алгоритмы обучения были готовы: что мы умели на пороге 2000

- Учить сети произвольной структуры – слоистые, с обратными связями, со многими слоями, и др. (1970е-1980е);
- Предобрабатывать данные для сокращения описания и удобства обучения (1980е-1990е);
- Удалять лишние связи и уменьшать разрядность связей (для понятности и аппаратной реализации) (1980е-1990е);
- Создавать мультиагентные системы; в которых небольшие интеллектуальные агенты делят между собой проблемы и состязаются в их решении

Алгоритмы обучения были готовы: что мы умели на пороге 2000

- Учить сети произвольной структуры – слоистые, с обратными связями, со многими слоями, и др. (1970е-1980е);
- Предобрабатывать данные для сокращения описания и удобства обучения (1980е-1990е);
- Удалять лишние связи и уменьшать разрядность связей (для понятности и аппаратной реализации) (1980е-1990е);
- Создавать мультиагентные системы; в которых небольшие интеллектуальные агенты делят между собой проблемы и состязаются в их решении

Алгоритмы обучения были готовы: что мы умели на пороге 2000

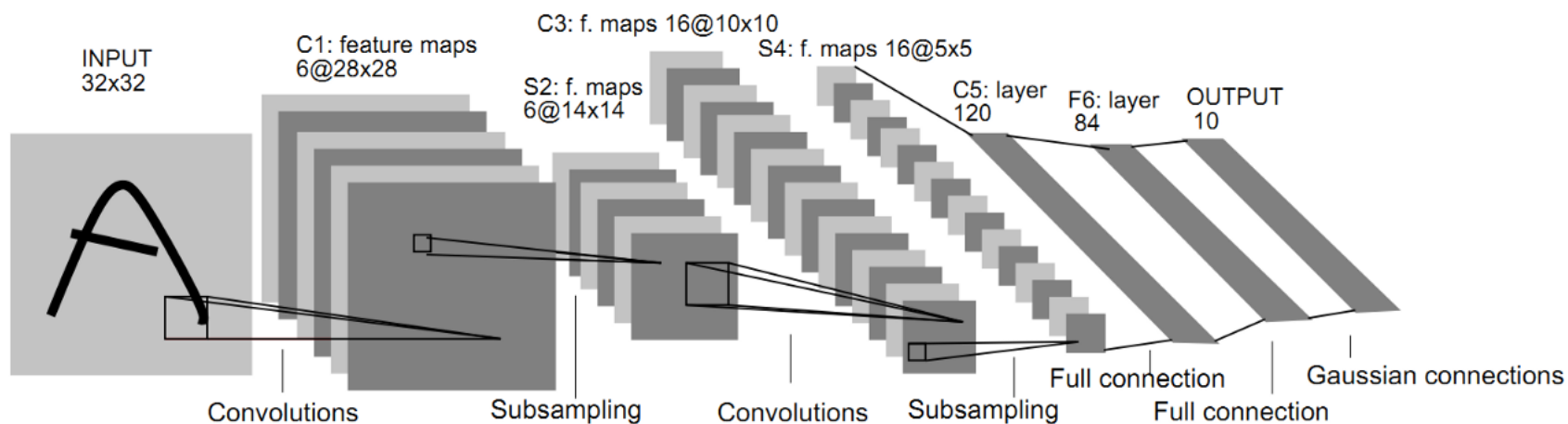
- Учить сети произвольной структуры – слоистые, с обратными связями, со многими слоями, и др. (1970е-1980е);
- Предобрабатывать данные для сокращения описания и удобства обучения (1980е-1990е);
- Удалять лишние связи и уменьшать разрядность связей (для понятности и аппаратной реализации) (1980е-1990е);
- Создавать мультиагентные системы; в которых небольшие интеллектуальные агенты делят между собой проблемы и состязаются в их решении

Алгоритмы обучения были готовы: что мы умели на пороге 2000

- Учить сети произвольной структуры – слоистые, с обратными связями, со многими слоями, и др. (1970е-1980е);
- Предобрабатывать данные для сокращения описания и удобства обучения (1980е-1990е);
- Удалять лишние связи и уменьшать разрядность связей (для понятности и аппаратной реализации) (1980е-1990е);
- Создавать мультиагентные системы, в которых небольшие интеллектуальные агенты делят между собой проблемы и состязаются в их решении (1990е).

Что принесли 2000е-2010е

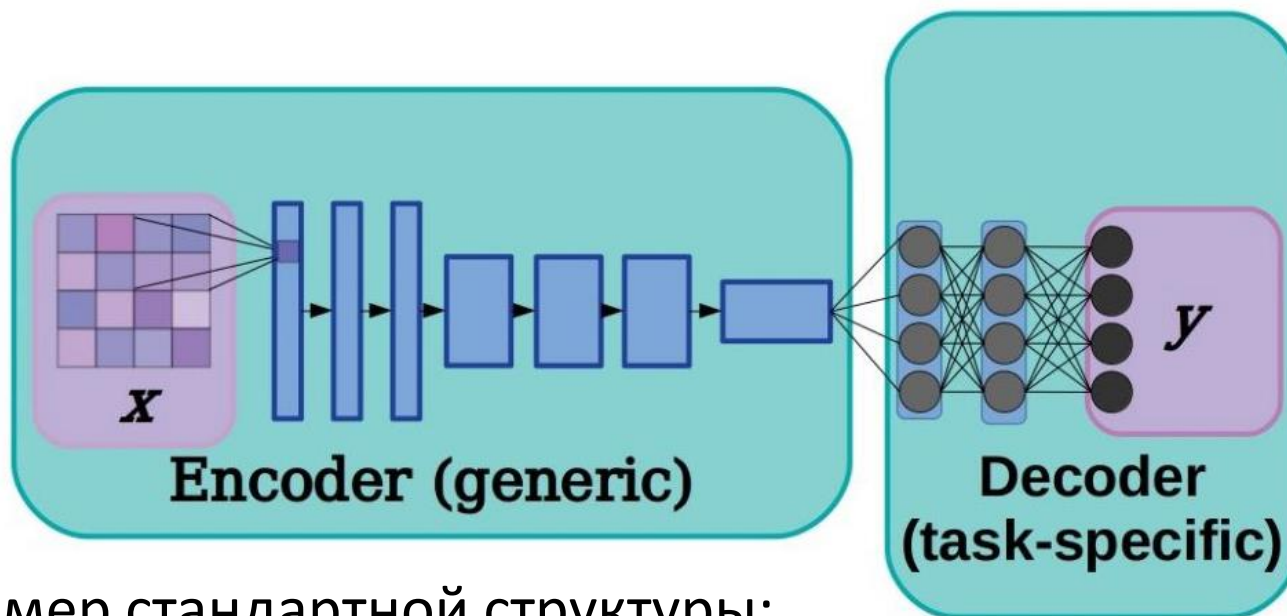
- Успех больших и очень больших систем для решения сложных задач



Пример: сеть LeNet-5
для распознавания рукописных цифр

Что принесли 2000е-2010е

- Опыт разработки больших систем и начало стандартизации блоков

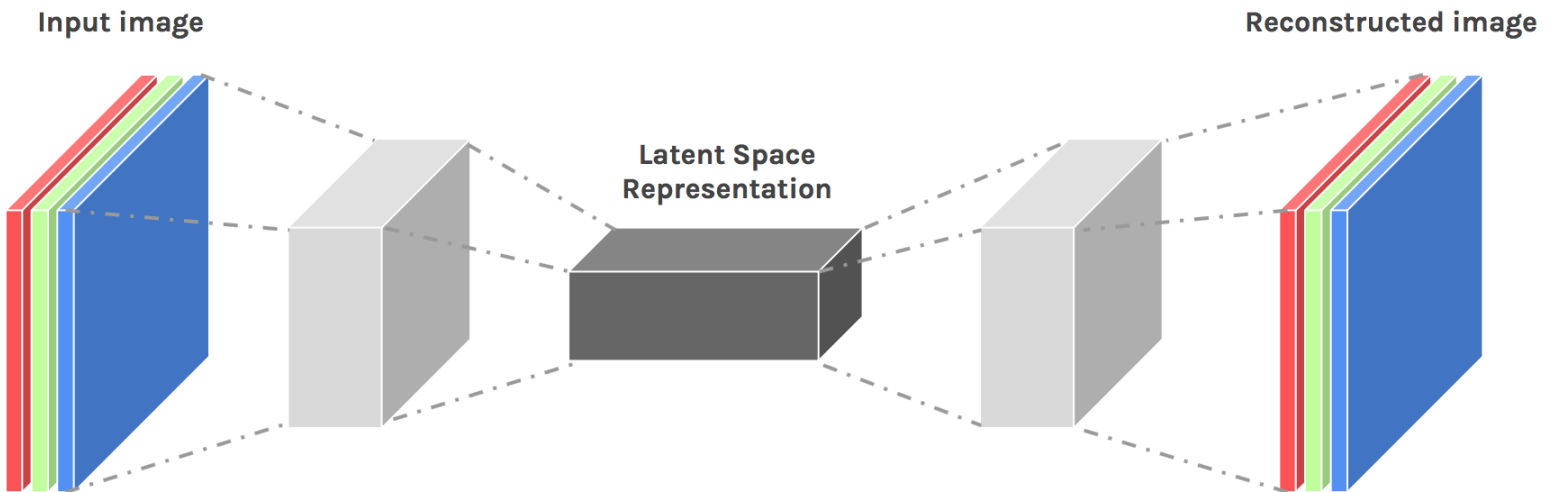


Пример стандартной структуры:

универсальный кодер для формирования признаков и
специфический декодер для решения конкретной задачи

Что принесли 2000е-2010е

- Некоторые стандартные блоки известны были и раньше, но нашли широкое применение в 2000е



Пример стандартной структуры:

Автоенкодер: преобразование информации в саму себя через максимально узкое горло (1980е-1990е).

Нелинейный факторный анализ?

Что принесли 2000е-2010е

- Большое количество прикладных разработок и **бизнес-восторг**

Машинное обучение: зона «бизнес-восторга»

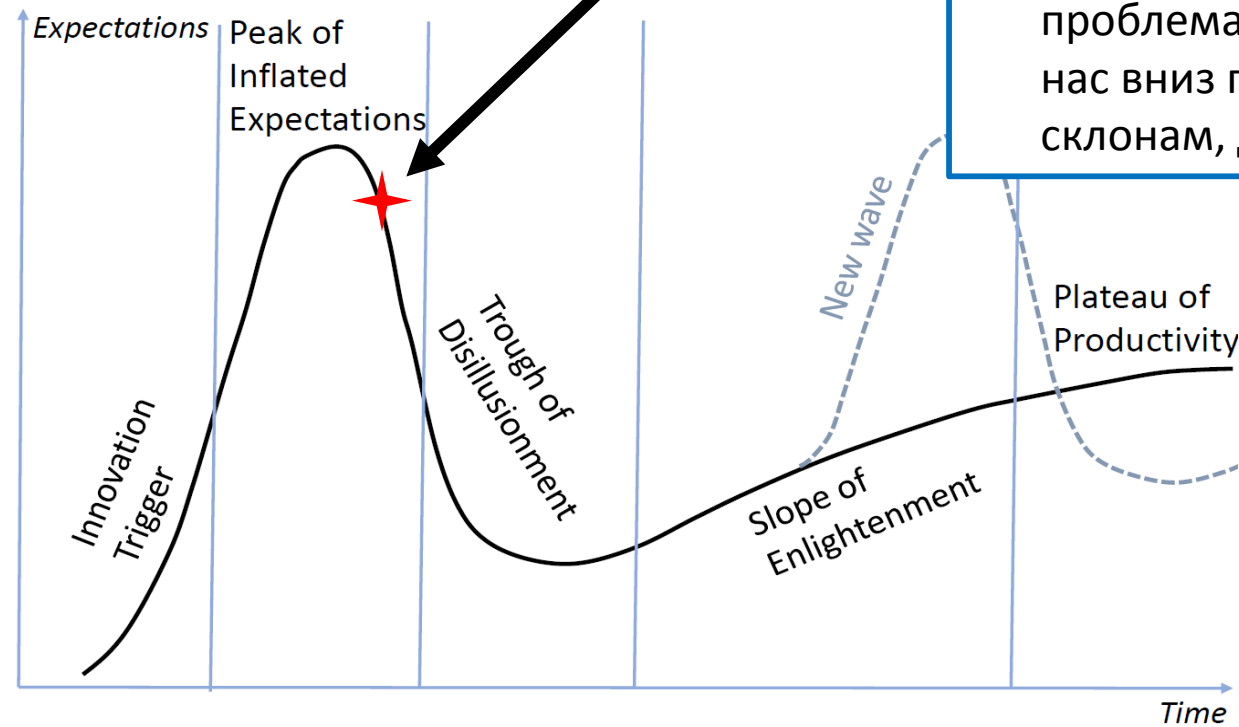
1. Генерация текстов естественного языка
2. Распознавание речи
3. Виртуальные Агенты
4. Оборудование, оптимизированное для AI
5. Управление Принятием Решений
6. Биометрия
7. Роботизированная Автоматизация Процессов
8. Беспилотный транспорт
9. Текстовая аналитика и обработка естественного языка
10. Цифровые двойники / ИИ моделирование
11. Кибербезопасность
12. Помощь в работе со знаниями
13. Создание контента (видео, реклама, публикации,...)
14. Одноранговые, (децентрализованные) сети
15. Распознавание эмоций
16. Распознавание образов
17. Автоматизация Маркетинга
18. Игры

Несколько наиболее впечатляющих успехов

- Поисковые системы Интернета;
- Распознавание лиц;
- Игра в шахматы;
- Игра Го;
- Машинный перевод;
- Системы – советчики и «рекомендатели»;
- Протезирование («умные протезы») и усиление естественных функций;
- Автономные роботы и беспилотники;...

Проблема о проблеме

Gartner Hype Cycle: ИИ сейчас здесь



- **Какая проблема** будет толкать нас вниз по склону?
- Может быть, это та же самая проблема, которая толкала нас вниз по предыдущим склонам, десятилетия назад?

Два камешка в ботинке

Гора успехов нейронного ИИ вдохновляет многих! Однако два камешка в нашей обуви могут замедлить это фантастически успешное движение:



ИИ делает неожиданные ошибки, и будет делать их в будущем;



Решения нейронного ИИ непрозрачны и не могут быть объяснены логически.

Проблема, сдерживающая развитие

ОШИБКИ СИСТЕМ ИИ И ИХ
ОБРАБОТКА

DRIVERLESS CAR SAFETY —

Report: Software bug led to death in Uber's self-driving crash

Sensors detected Elaine Herzberg, but software reportedly decided to ignore her.

TIMOTHY B. LEE - 5/7/2018, 11:12 PM





A Tesla electric car crashed into a highway barrier in Mountain View, California, on March 23, 2018. Investigators confirmed that Autopilot was partially to blame.

Browser tabs: cassad - Google Sear | University of Leicester | ceraxon oral uk buy - | ip address example - | Inbox (5,253) - tyukin | Access By Bus:Kitakyu | Access your files off-c | Facial recognition x

Address bar: eandt.theiet.org/content/articles/2017/08/facial-recognition-technology-fails-at-notting-hill-carnival

Navigation icons: back, forward, refresh, home, search, star, menu, print, share, download, etc.

Menu

E&T ENGINEERING AND TECHNOLOGY

Facebook, Twitter, Print, Email, +



Facial recognition technology 'fails' at Notting Hill Carnival

By Hilary Lamb

Published Friday, September 1, 2017

A trial of facial recognition technology by police monitoring the Notting

Windows taskbar: Type here to search | Taskbar icons: File Explorer, Edge, PowerPoint, etc. | System tray: 19:57, 11/09/2017, ENG

MACHINES

IBM Watson gave 'unsafe and incorrect' cancer treatment options

by Colm Gorey

🕒 27 JUL 2018

👁 1.03K VIEWS

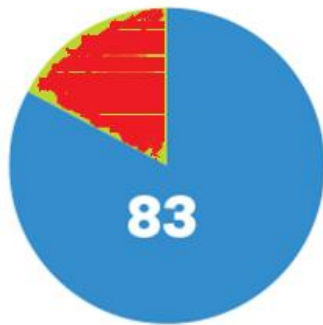


LATEST NEWS



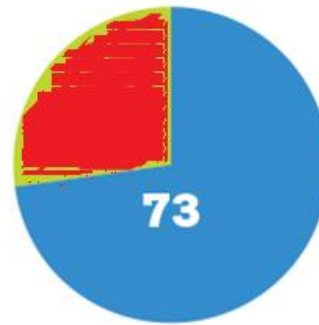
Irish-based researchers one step closer to solving great stellar mystery

Bumrungrad International
Hospital, Thailand:
83% concordance



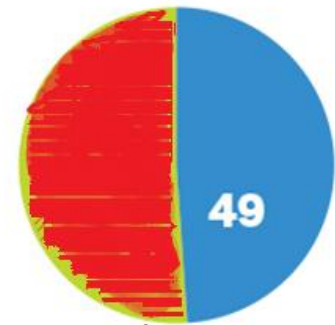
211 patients with breast, colorectal,
gastric, and lung cancer

Manipal Comprehensive
Cancer Center, India:
73% concordance



638 patients with
breast cancer

Gachon University
Gil Medical Center, South
Korea: 49% concordance



656 patients with
colon cancer

A [2016 audit by the University of Texas](#) found that
the cancer centre spent \$62 million on the project before canceling it.

Главной проблемой для широкого использования ИИ в мире являются неожиданные ошибки

- Ошибки могут быть опасны;
- Как правило, остается неясным, кто за них отвечает;
- Чтобы избежать повторения обнаруженной ошибки, требуется быстрое, неитеративное исправление системы;
- Исправление ошибок не должно повредить имеющимся навыкам;
- Реальный мир не является хорошей i.i.d. выборкой,
 - типы ошибок очень многочисленны и часто непредсказуемы,
 - поэтому мы не можем полагаться на статистическую оценку вероятности ошибок.

Фундаментальные источники ошибок ИИ

- Программные ошибки;
- Неожиданное поведение человека;
- Нецелевое использование;
-
- **Неопределенность в обучающих данных (100 атрибутов требуют $>2^{100}>10^{30}$ прмеров для полноты картины)**
- **Неопределенность в процессе обучения...**

Ошибки – неизбежные спутники ИИ, основанного на данных

Ошибки ИИ: и этот список продолжается

...

- **InsiperoBot** Poor Testing/Coding
- **Cortana** Not Working Bias/Limited Training Set Poor Testing/Coding
- **Alexa** Blasting Music Unexpected Human Behaviour
- **Passport Checker** Bias/Limited Training Set Poor Testing/Coding
- **Knightscope Robot** Hits Toddler Poor Testing/Coding Unexpected Human Behaviour
- **Facebook Translate** Arrest, Bias/Limited Training Set Unexpected Human Behaviour
- **Google Tag** Racist Bias/Limited Training Set
- **WeChat** Racist Bias/Limited Training Set Unexpected Human Behaviour
- **Microsoft - Tay** Unexpected Human Behaviour Bias/Limited Training Set

.....

(Thanks to Rosie Fenwick and Eliyas Woldegeorgis for collecting the list)

Можем ли мы исправить ошибки ИИ путем систематического подучивания/переучивания?

- Для сохранения имеющихся навыков мы должны использовать полный набор обучающих данных.
- Этот подход требует больших ресурсов для каждой ошибки.
- Однако могут появиться новые ошибки.
- Сохранение имеющихся навыков не гарантируется.
- Вероятность повреждения навыка априори неизвестна.



Как справиться с ошибками в мире многомерных больших данных?

Нужно внешнее устройство - корректор ошибок ИИ

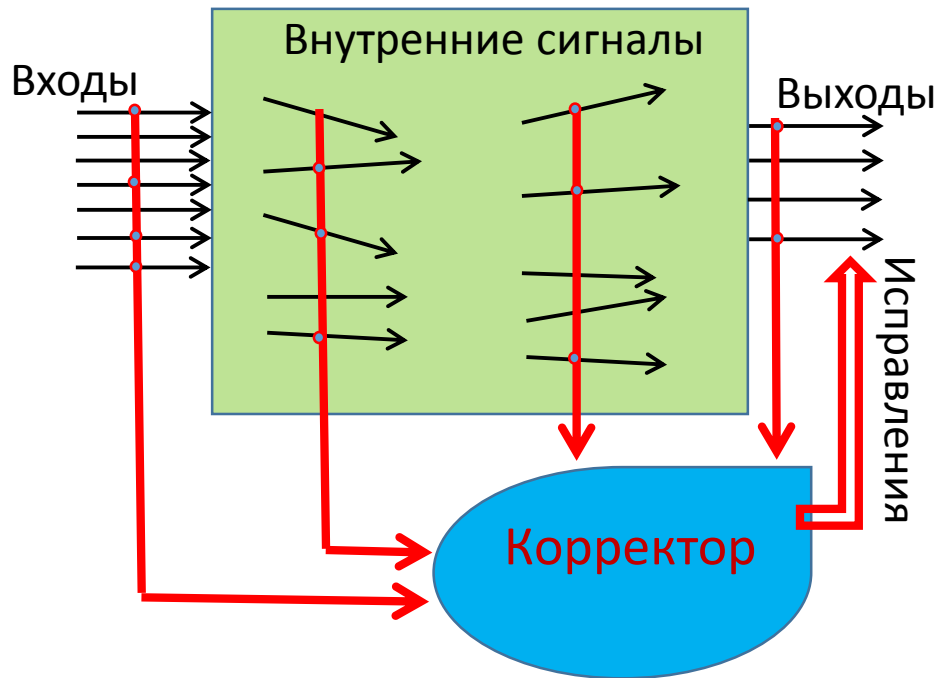


Технические условия

Корректор должен:

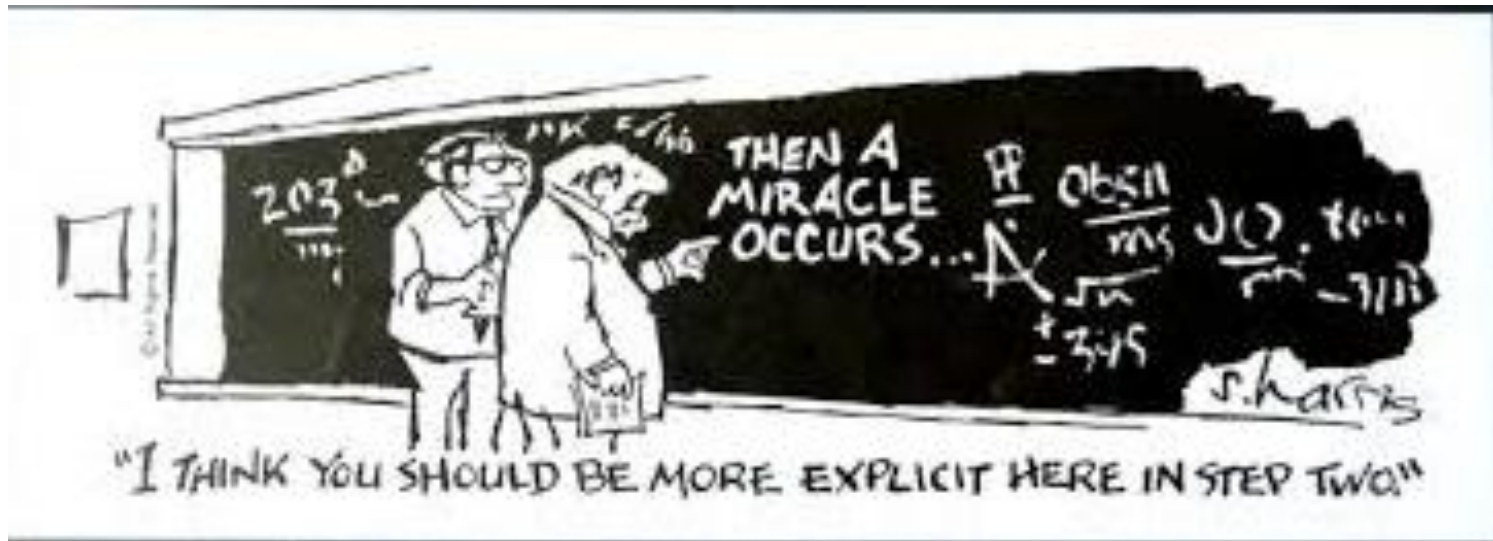
- быть простым;
- не менять навыков унаследованной системы;
- допускать быстрое неитеративное обучение;
- исправлять новые ошибки без разрушения предыдущих исправлений.

Корректор должен отделять ошибки
от правильно решенных примеров и
исправлять их



**Это возможно, если
размерность
достаточно велика!**

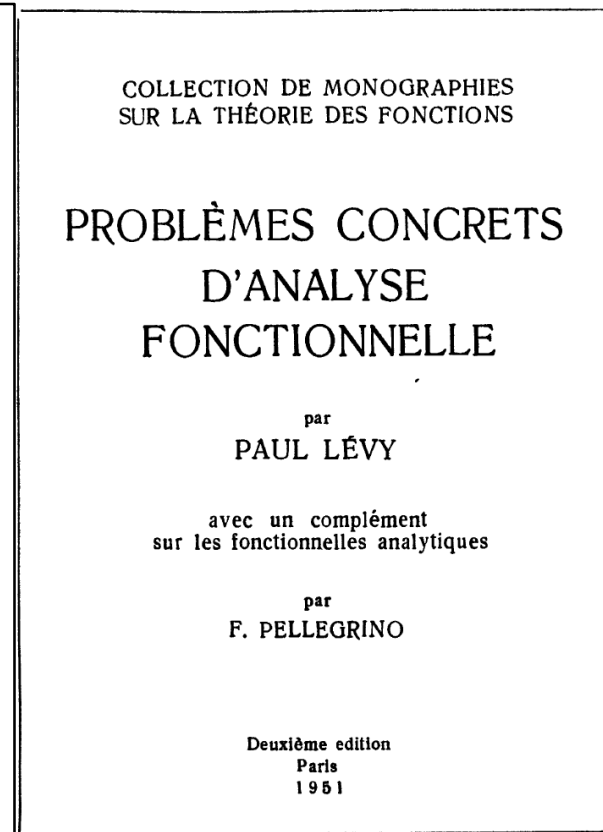
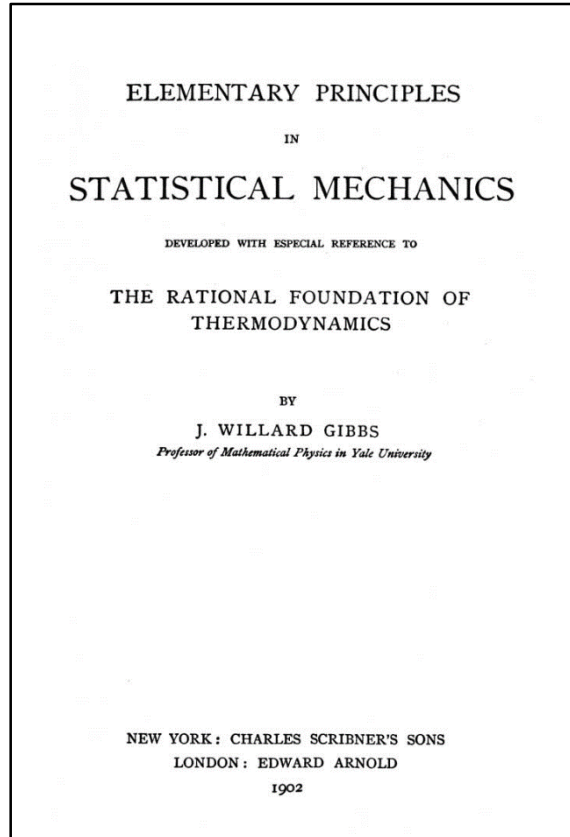
Чудеса стохастической разделимости в больших размерностях



Надо бы добавить немного деталей!

Благословение размерности

- Благословение размерности - это проявление концентрации меры.
- Вся статистическая физика основана на этих явлениях.
- Теоремы стохастического разделения представляют собой новый класс таких теорем.



A.N. Gorban, I.Y. Tyukin, [Blessing of dimensionality: mathematical foundations of the statistical physics of data](#). *Philos. Trans. Royal Soc. A* 376(2118), 20170237, 2018.

Второй камешек в ботинке

Гора успехов нейронного ИИ вдохновляет многих! Однако два камешка в нашей обуви могут замедлить это фантастически успешное движение:



ИИ делает неожиданные ошибки, и будет делать их в будущем;



Решения нейронного ИИ непрозрачны и не могут быть объяснены логически.

Смена базового принципа

- Первая схема: **ИИ суммируют неявный опыт и используют его для решения задач** (можно тиражировать, как программы или устройства).
- Следующая схема: **ИИ производят знания для использования человеком, переноса в другие программные продукты и др.**
- «Именно нейросетевая технология производства знаний, по-моему, является «точкой роста», которая по-новому развернет нейроинформатику, преобразует многие разделы информатики и создаст новые.» Горбань, 1998. Дискуссия о нейрокомпьютерах, МИФИ 1999. 20 лет до начала гонки ХАИ.
- Технология есть (давно), софт есть, **но пользователей мало** – рынок требует «кнопочных решений».
- **ПОЛЬЗОВАТЕЛЬ ЛЕНИВ** – это фундаментальный принцип развития ИИ

Смена базового принципа

Первая схема: ИИ суммируют неявный опыт и используют его для решения задач (можно тиражировать, как программы или устройства).

- Следующая схема: **ИИ производят знания для использования человеком, переноса в другие программные продукты и др.**
- «Именно нейросетевая технология производства знаний, по-моему, является «точкой роста», которая по-новому развернет нейроинформатику, преобразует многие разделы информатики и создаст новые.» Горбань, 1998. Дискуссия о нейрокомпьютерах, МИФИ 1999. 20 лет до начала гонки ХАИ.
- Технология есть (давно), софт есть, **но пользователей мало** – рынок требует «кнопочных решений».
- **ПОЛЬЗОВАТЕЛЬ ЛЕНИВ** – это фундаментальный принцип развития ИИ

Смена базового принципа

- Первая схема: **ИИ суммируют неявный опыт и используют его для решения задач** (можно тиражировать, как программы или устройства).
- Следующая схема: **ИИ производят знания для использования человеком, переноса в другие программные продукты и др.**
- «Именно нейросетевая технология производства знаний, по-моему, является «точкой роста», которая по-новому развернет нейроинформатику, преобразует многие разделы информатики и создаст новые.» Горбань, 1998. Дискуссия о нейрокомпьютерах, МИФИ 1999. 20 лет до начала гонки ХАИ.
- Технология есть (давно), софт есть, **но пользователей мало** – рынок требует «кнопочных решений».
- ПОЛЬЗОВАТЕЛЬ ЛЕНИВ** – это фундаментальный принцип развития ИИ

Смена базового принципа

Первая схема: **ИИ суммируют неявный опыт и используют его для решения задач** (можно тиражировать, как программы или устройства).

- Следующая схема: **ИИ производят знания для использования человеком, переноса в другие программные продукты и др.**
- «Именно нейросетевая технология производства знаний, по моему, является «точкой роста», которая по-новому развернет нейроинформатику, преобразует многие разделы информатики и создаст новые.» Горбань, 1998. Дискуссия о нейрокомпьютерах, МИФИ 1999. 20 лет до начала гонки ХАИ.
- Технология есть (давно), софт есть, **но пользователей мало** – рынок требует «кнопочных решений».
- **ПОЛЬЗОВАТЕЛЬ ЛЕНИВ** – это фундаментальный принцип развития ИИ



Смена базового принципа

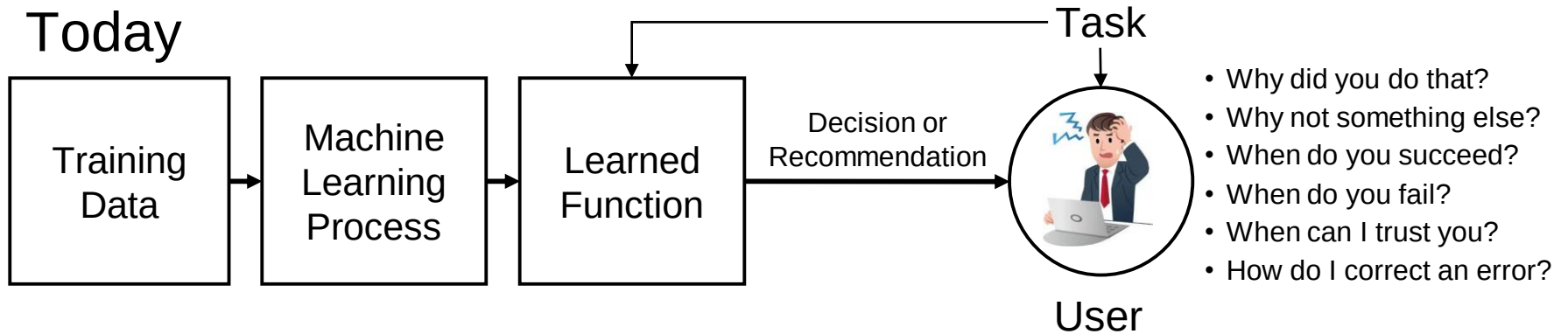
В 1990х были созданы алгоритмы, программы, написаны книги. В 2020 это стало новым хайпом. Сейчас идет всемирная гонка «Explainable AI» (XAI) - объяснимый ИИ

<https://www.darpa.mil/program/explainable-artificial-intelligence>

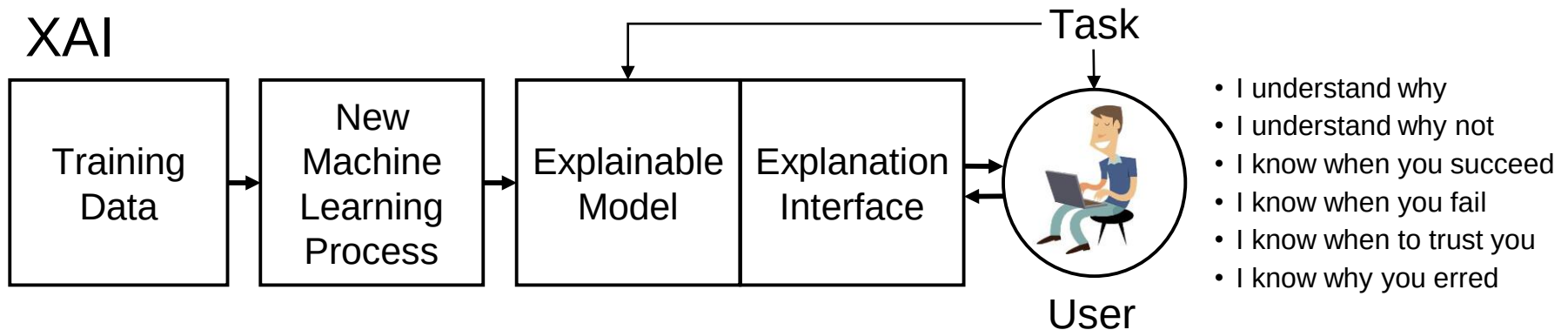


B. Program Scope – XAI Concept

Today



XAI



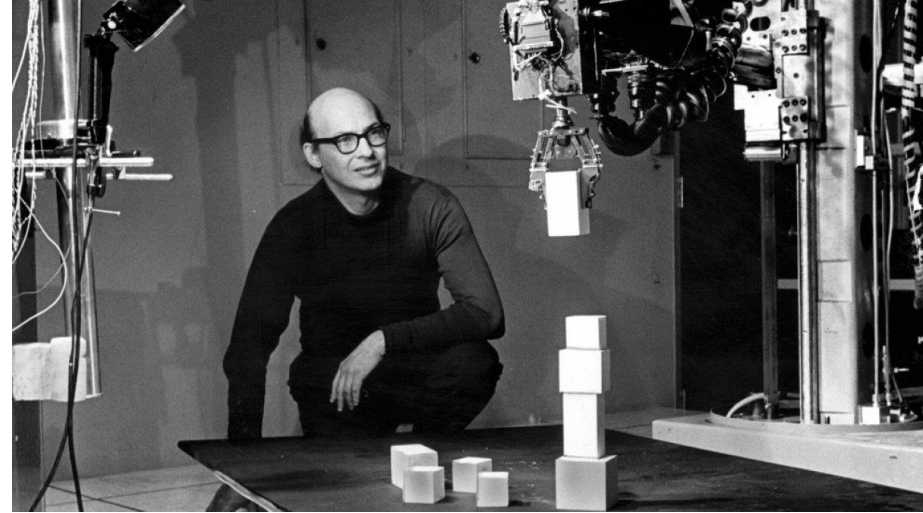
Сообщество ИИ

- Нарождающаяся схема: **ИИ производят знания и учат друг друга.** Они функционируют в разных условиях, накапливают разные знания и обмениваются ими.
- Взаимная коррекция – один из способов организация сообщества.
- Это – хайп на послезавтра.

Мультиагентные системы

- Агент - это компьютерная система, способная к независимым (автономным) действиям.
- Мультиагентная система состоит из ряда агентов, которые взаимодействуют друг с другом.
- Агенты в многоагентной системе могут иметь разные цели и мотивы.
- Чтобы успешно взаимодействовать, агенты должны сотрудничать и вести переговоры.

Микромиры и онтологии ИИ



- Марвин Мински с соавторами предложили, чтобы исследования ИИ были сосредоточены на искусственно простых ситуациях, микромирах.
- Пример: "мир кубиков", который состоит из цветных блоков различных форм и размеров, расположенных на плоской поверхности.
- ***Персональный микромир***-это рабочее место для ИИ.
- Вселенная-это мозаика индивидуальных микромиров и их связей, с распределением опыта между агентами.

Агенты будут создавать свои собственные онтологии и микромиры (используя комбинацию обучения с учителем и без учителя) и конфигурировать реальность из этих частных микромиров.

Изменение понятия «истина»; *Aleteya versus doxa*

- Скорость производства новых знаний меняет соотношение истины и мнений.
- «Пути истины» (Алетея) и «Пути мнения» (Докса) были введены Парменидом и оказали огромное влияние на Платона и всю нашу науку.
- Ньютон, кстати, считал свой закон всемирного тяготения мнением, а не истиной.
- Что производит ИИ из данных? Полагаю, что мнение – текучее, лежащее между знанием и незнанием.
- Истина же не успевает кристаллизироваться – и старое мнение замещается новым, ожидаемо лучше соответствующим новым данным. А путь истины «запределен тропе человеков» и, добавлю, ИИ.

Новая эпистемология (скетч-вызов)

- Истина – результат обработки данных;
- Глубокая истина – результат глубокой обработки больших данных с последующей интерпретацией (понятийным представлением);
- Или: Глубокая истина – результат интерактивной игры с данными и миром, с новыми запросами и новыми гипотезами – гипотетико-дедуктивный путь ИИ.
- Степень истинности измеряется в терминах эмпирического риска.
-
- Когда мы получим таким образом квантовую электродинамику и общую теорию относительности, настанет новая эра.
- А вот теорию революционной ситуации в политических науках мы получаем запросто и давно!

Является ли ИИ *интеллектом?*

1. Тест Тьюринга – диалог через ширму. Кто там – человек или ИИ?

– Фатическое общение, то есть бесцельный «смыслообразный» разговор уже имитируется неплохо.

Но интеллект ли это???

Является ли ИИ *интеллектом?*

1. Тест Тьюринга – диалог через ширму. Кто там – человек или ИИ?

– Фатическое общение, то есть бесцельный «смыслообразный» разговор уже имитируется неплохо.

Но интеллект ли это???

Тест Тьюринга – хороший пример того, как простое (и не очень важное) замечание становится мемом.

Является ли ИИ *интеллектом?*

**2. Человеческие тесты на интеллект –
может ли ИИ быть лучше?**

Попробуем?

Тесты составлены из простых заданий.

Является ли ИИ *интеллектом?*

2. Человеческие тесты на интеллект – может ли ИИ быть лучше?

Примеры (можете изучить полный тест):
[https://psylab.info/Тест Векслера/Взрослый вариант](https://psylab.info/Тест_Векслера/Взрослый_вариант)

- 1) Дано слово, в котором перепутаны буквы. Найти исходное слово – **ТИВОНКР**.
- 2) Дана неизвестная картинка, нарезанная на куски. Собрать эту картинку.
- 3) Объяснить пословицу, не прибегая к метафорам: «Одна ласточка весны не делает».

Является ли ИИ *интеллектом?*

2. Человеческие тесты на интеллект – может ли ИИ быть лучше?

Примеры:

- 1) Дано слово, в котором перепутана буквы. Найти исходное слово – **ТИВОНКР**.
- 2) Дана неизвестная картинка, нарезанная на куски. Собрать эту картинку.
- 3) Объяснить пословицу, не прибегая к метафорам: «Одна ласточка весны не делает».

Может ли это сделать ИИ? Для каждого типа задач можно делать свой ИИ. Так:

- 1 – легко (**ВТОРНИК**);
- 2 – на пределе современных возможностей, или даже еще за пределами;
- 3 – пока за пределами, но подождем.

Является ли ИИ *интеллектом?*

2. Человеческие тесты на интеллект – может ли ИИ быть лучше?

- Смотрите, для каждого задания был спец-ИИ.
- А все задания теста вместе? Когда ИИ сможет пройти весь тест (заранее неизвестный)?
- Дали текст задания – пойми и сделай то, чего раньше не делал. (Тест Горбаня 😊)
- Пока очень и очень далеко от современного ИИ. Даже почти непредставимо. ***А дети могут.***

Является ли ИИ *интеллектом?*

3. Тесты на совместную деятельность – может ли ИИ сотрудничать в совместной работе?

- Совместный выбор цели,
- Совместное планирование,
- Совместные действия.

Для задач, которых раньше не решал.

А дети могут и делают - в игре.

Кто умнее, человек или ИИ?

- Для каждого ИИ существует свой «микромир» задач, в котором он силен.
- Так и трактор может больше, чем человек, и лошадь тоже — но только в своем ограниченном мире заданий.
- Понятие «микромир» ИИ ввел Марвин Минский.

Что будет скоро?

- Машины везде (уже)
- Компьютеры везде (уже)
- И ИИ будет везде

ИИ будет особенно востребован там, где придется формировать решения, основанные на анализе скрытых взаимосвязей в больших наборах многомерных данных... Такая вот очень продвинутая форма математической статистики, соединенной с очень продвинутыми справочниками.

План

1. Немного истории.
2. Определение, пришедшее «из народа».
3. Размерность и всемогущество неограниченного перцептрона (Теорема 1 Розенблатта).
4. Ограниченные возможности ограниченного перцептрона, Минский&Пайперт и первая зима.
5. Всплеск конца 1980х – начала 1990х. Почти все алгоритмы готовы!
6. Взрыв активности в 21 веке! Начинается настоящее время.

План

7. Начало спада и проблема ошибок.
8. Конструкция корректоров.
9. Объяснимость и логическая прозрачность ИИ.
10. Сообщества ИИ агентов.
11. Является ли ИИ интеллектом?
12. Изменение понятия «истина».

Явные знания из данных (новая-старая технология). Выборы в США

МЕТАДААННЫЕ

- Yes=+1, No=-1
- 1. Правящая партия была у власти более одного срока?
- 2. Правящая партия получила больше 50% голосов на прошлых выборах?
- **3. В год выборов была активна третья партия?**
- **4. Была серьезная конкуренция при выдвижении от правящей партии?**
- 5. Кандидат от правящей партии был президентом в год выборов?
- **6. Был ли год выборов временем спада или депрессии?**
- 7. Был ли рост среднего национального валового продукта на душу населения больше 2.1%?
- **8. Произвел ли правящий президент существенные изменения в политике?**
- **9. Во время правления были существенные социальные волнения?**
- 10. Администрация правящей партии виновна в серьезной ошибке или скандале?
- 11. Кандидат от правящей партии - национальный герой?
- 12. Кандидат от оппозиционной партии - национальный герой?

Пример: явное правило

- Президентская партия побеждает тогда и только тогда когда

$$3+4+9 < 0 \text{ \& } 4-8+6 < 0$$

- 3. В год выборов была активна третья партия?
- 4. Была серьезная конкуренция при выдвижении от правящей партии?
- 6. Был ли год выборов временем спада или депрессии?
- 8. Произвел ли правящий президент существенные изменения в политике?
- 9. Во время правления были существенные социальные волнения?

