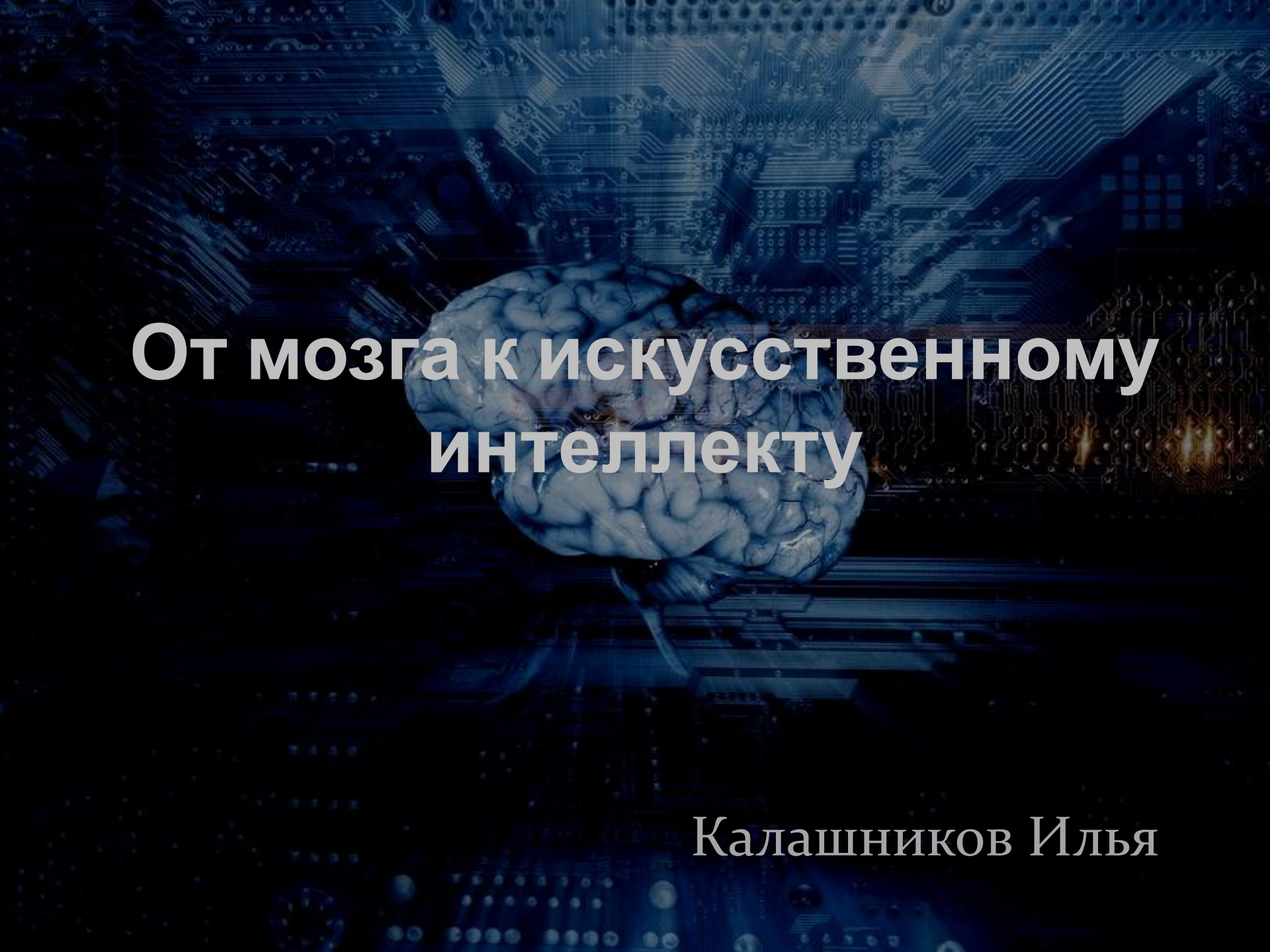


The image features a human brain, rendered in a light blue, almost translucent style, positioned centrally. It is set against a dark blue background that is a complex, glowing circuit board or microchip pattern. The circuitry consists of numerous fine lines, dots, and larger rectangular components, some of which are illuminated with a warm, golden-yellow light, creating a sense of depth and technological sophistication. The overall aesthetic is futuristic and high-tech.

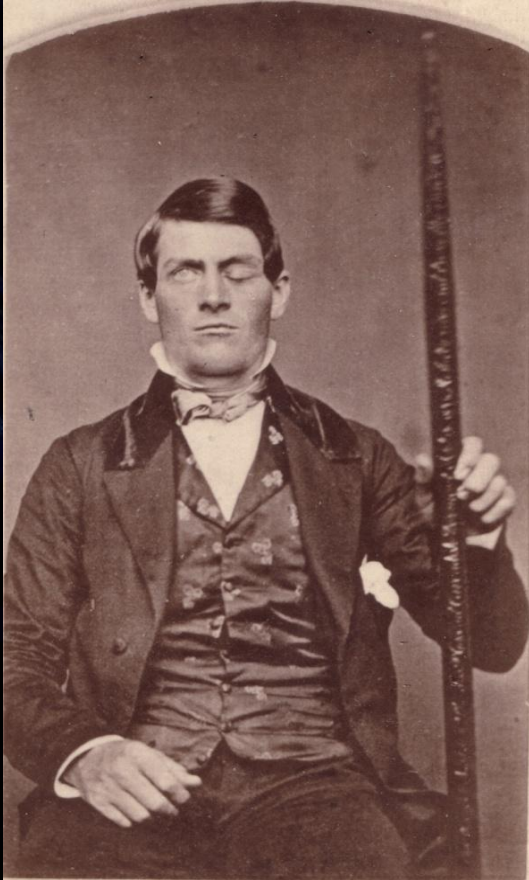
От мозга к искусственному интеллекту

Калашников Илья

Содержание

1. Строение и функционирование мозга
2. Вычислительная техника
3. Скользкие философские вопросы
4. Нейронные сети
5. Модель «Память – предсказание»
6. Перспективы
7. Литература

Загадочная история Финеаса Гейджа



В 1848 году штырь насквозь пронзил голову Ф. Гейджа, повредив кору в области лобной доли левого полушария, нарушив ее связи с левой височной и правой лобной долями и лимбической системой. Это кардинально изменило его поведение, сделав бедолагу раздражительным, капризным и нерешительным человеком с низким интеллектом.

Этот случай вызвал сдвиг научной парадигмы и подтолкнул исследователей к изучению функций различных областей мозга.

Люди с разделенными полушариями

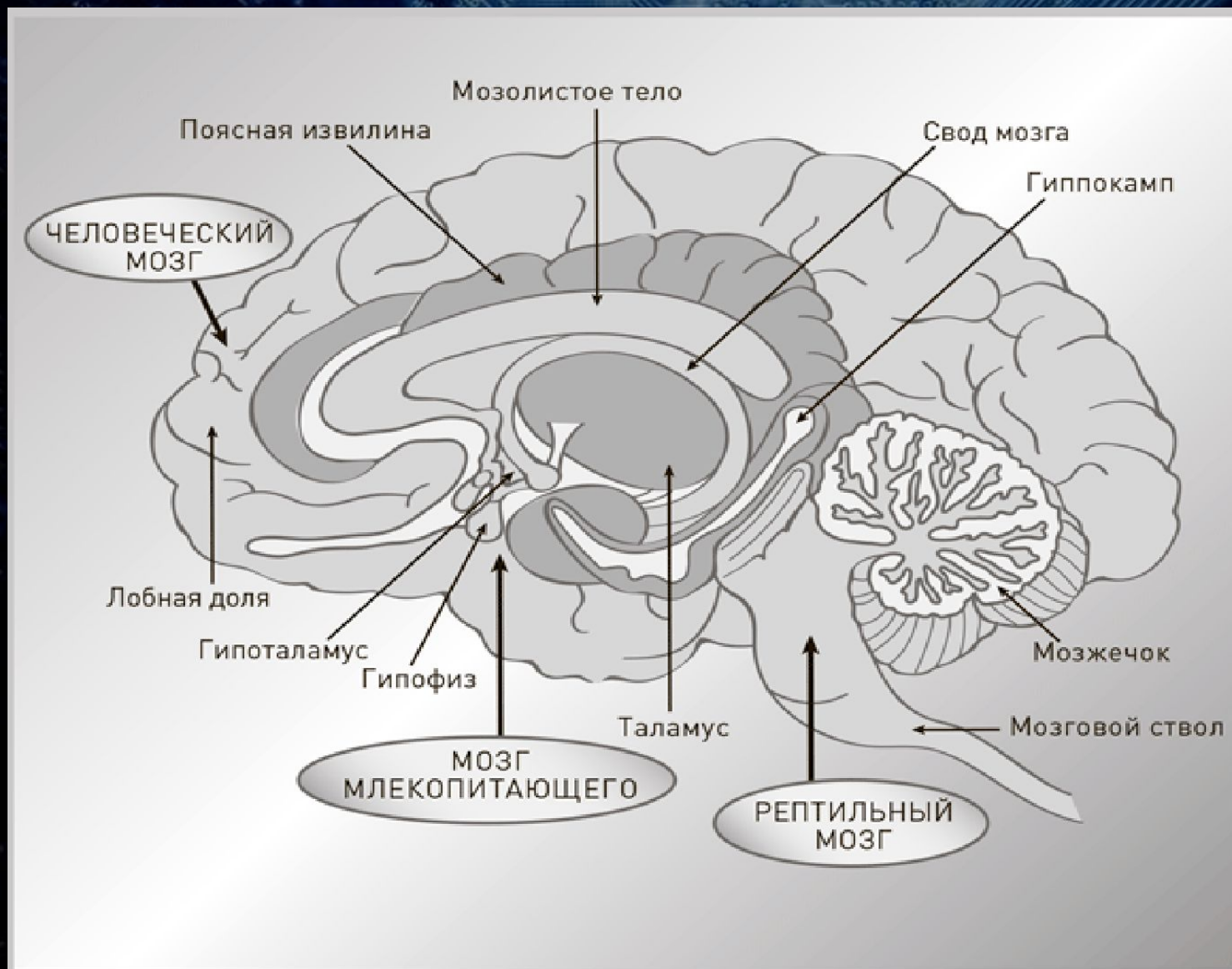


Эксперименты с людьми, у которых рассечено миндалевидное тело или даже удалено одно из полушарий показали, наличие видимых различий в отношении к чему-то между сознательным и бессознательным. При этом когнитивные функции полностью сохраняются, но различные полушария показывают разные результаты тестирований и имеют различные желания и пристрастия. Левое, речевое и аналитическое полушарие оправдывает поведение правого.

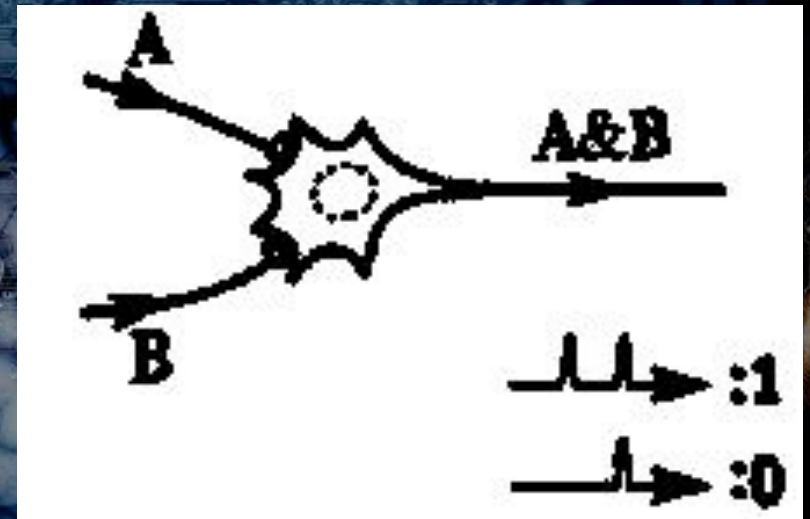
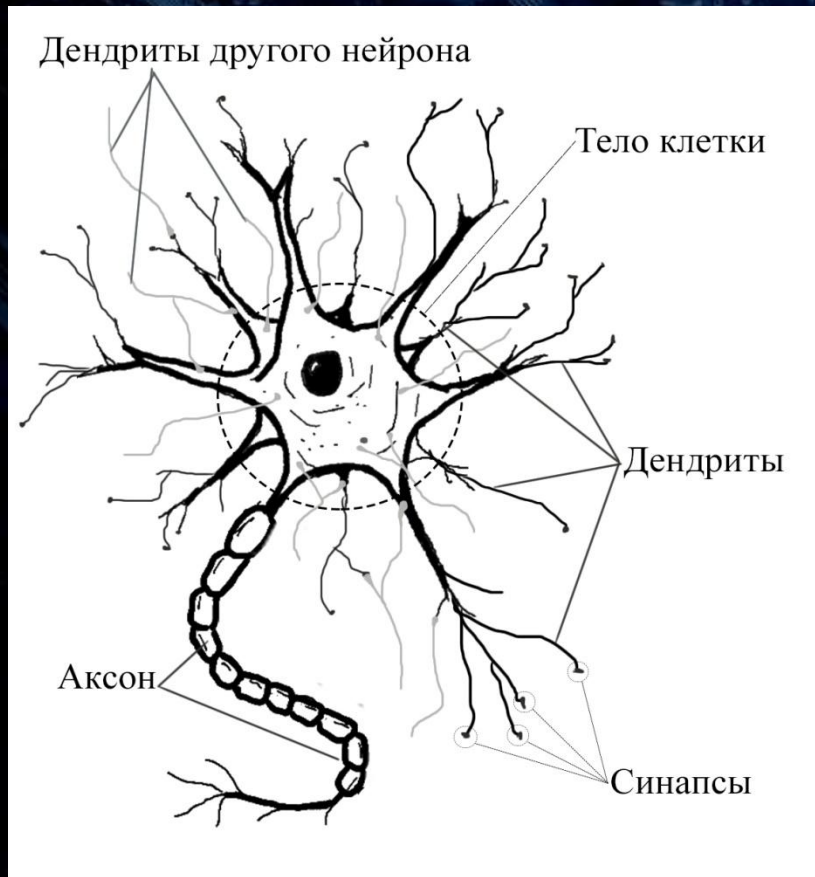
Современные методы исследования деятельности мозга

- МРТ и фМРТ – регистрация тока крови
- Электрэнцефалограмма – детектирование электрических сигналов
- Позитронно-эмиссионная томография – продукты распада натрия-22
- Нейростимуляция – введение электродов в мозг
- Оптогенетика – введение в нейрон светочувствительного гена
- Транскраниальное ЭМ сканирование – отключение околочерепных областей
- Магнитоэнцефалография - измерение магнитного поля, порожденное токами мозга

Строение мозга



Нейрон

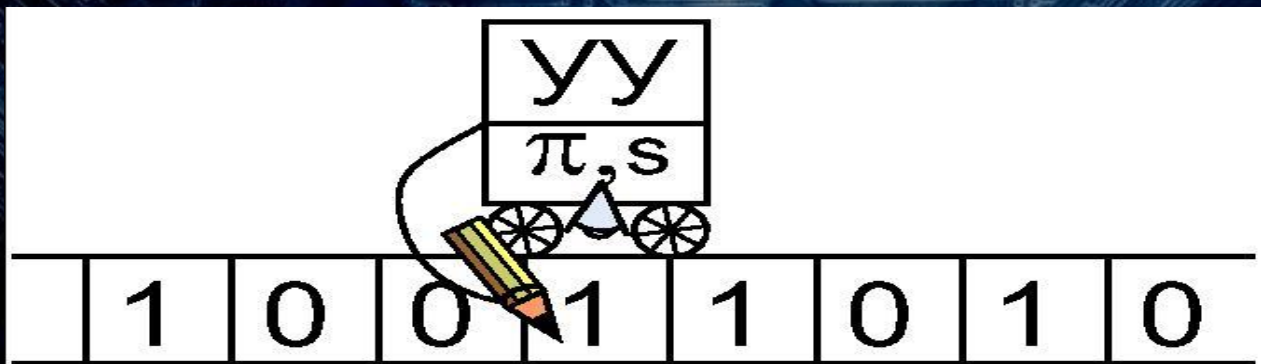


Правило ста шагов: мозг может решить существенную задачу за полсекунды. За это время задействуется цепочка не более чем из 100 нейронов.

A human brain is centered in the image, rendered in a light blue, almost translucent style. It is superimposed on a dark blue background that features a complex, glowing circuit board pattern. The circuitry consists of numerous lines, nodes, and small rectangular components, some of which are illuminated with a warm, golden light. The overall composition suggests a connection between biology and technology.

Слайд про биохимию

Машина Тьюринга



МТ - бесконечная лента, разделённая на ячейки, и управляющее устройство с конечным числом состояний. Управляющее устройство может перемещаться влево и вправо по ленте, читать, стирать и записывать в ячейки символы некоторого конечного алфавита.

Универсальная МТ - МТ, которая может заменить собой любую машину Тьюринга.

Вопросы вычислимости

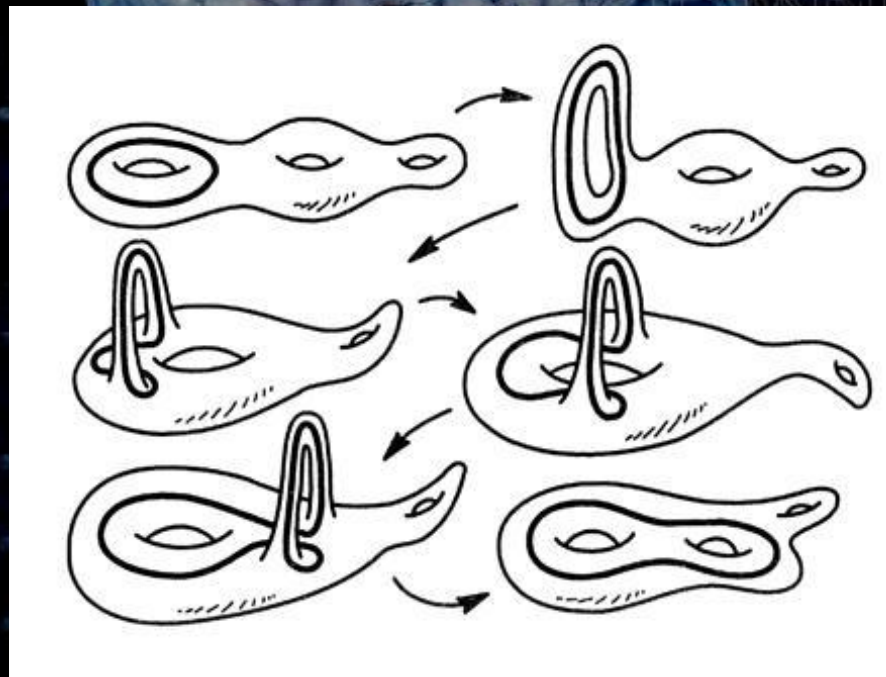
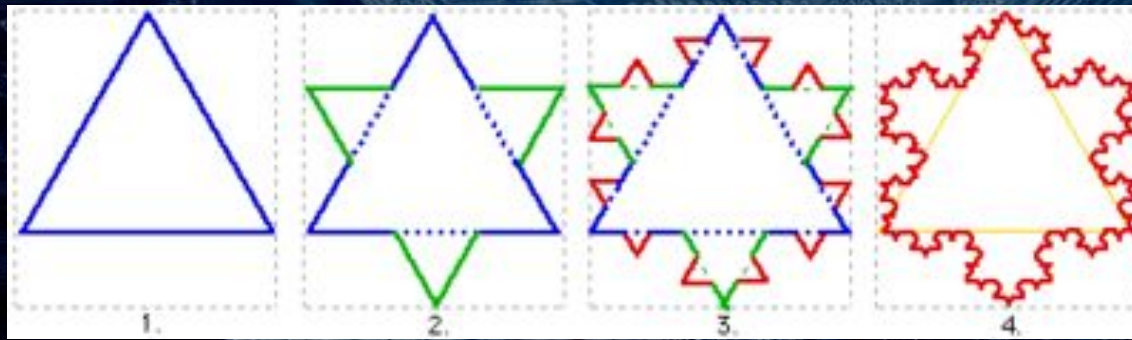
- Тезис Чёрча -всякий алгоритм может быть задан в виде соответствующей машины Тьюринга, класс вычислимых функций совпадает с классом функций, вычислимых на машинах Тьюринга.
- Проблема остановки – дана процедура и входные данные, требуется определить, завершится ли выполнение такой процедуры.
- Сверхтьюринговые вычисления.

Вопросы вычислимости

- Тезис Чёрча -всякий алгоритм может быть задан в виде соответствующей машины Тьюринга, класс вычислимых функций совпадает с классом функций, вычислимых на машинах Тьюринга.
- Проблема остановки – дана процедура и входные данные, требуется определить, завершится ли выполнение такой процедуры.
- Сверхтьюринговые вычисления.

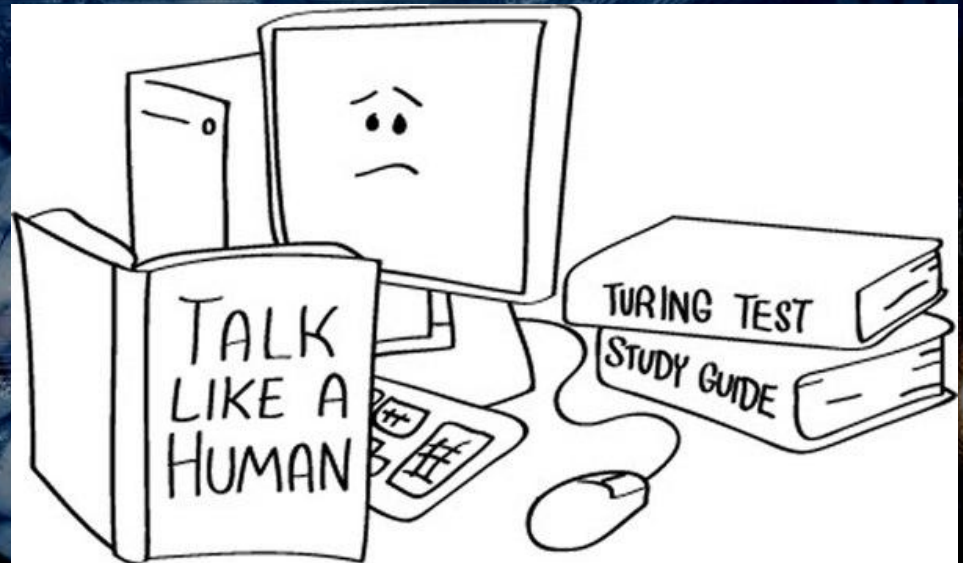


Невычислимые функции



Алгоритмы и понимание

- Тест Тьюринга



- Китайская комната



Сильный и слабый ИИ

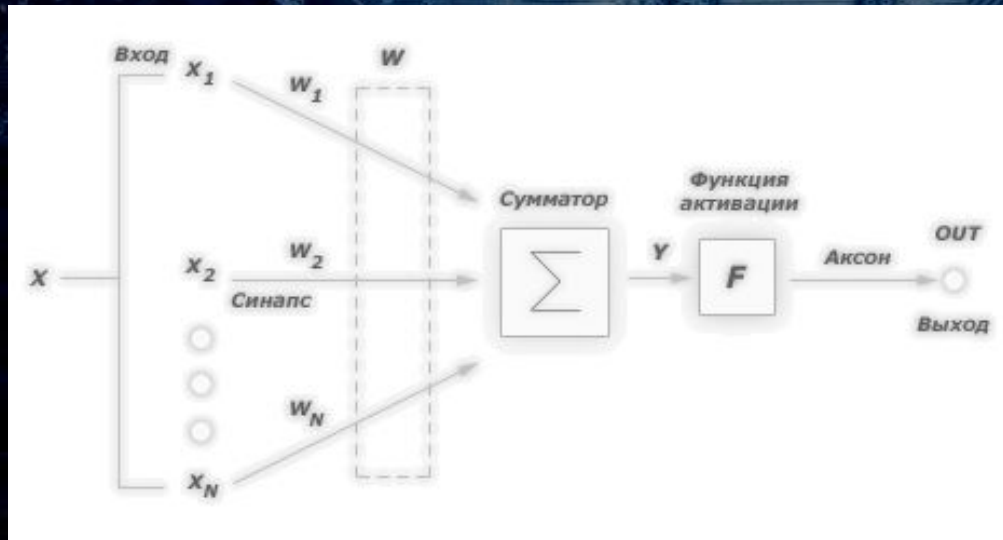
- Гипотеза сильного ИИ: “Соответствующим образом запрограммированный компьютер с нужными входами и выходами и будет разумом, в том смысле, в котором человеческий разум — это разум.”
- Гипотеза слабого ИИ: не будет.
- Функционализм - наличие интеллекта или обладание разумом свойство организации, не имеющее ничего общего с составными элементами.

Что такое интеллект?



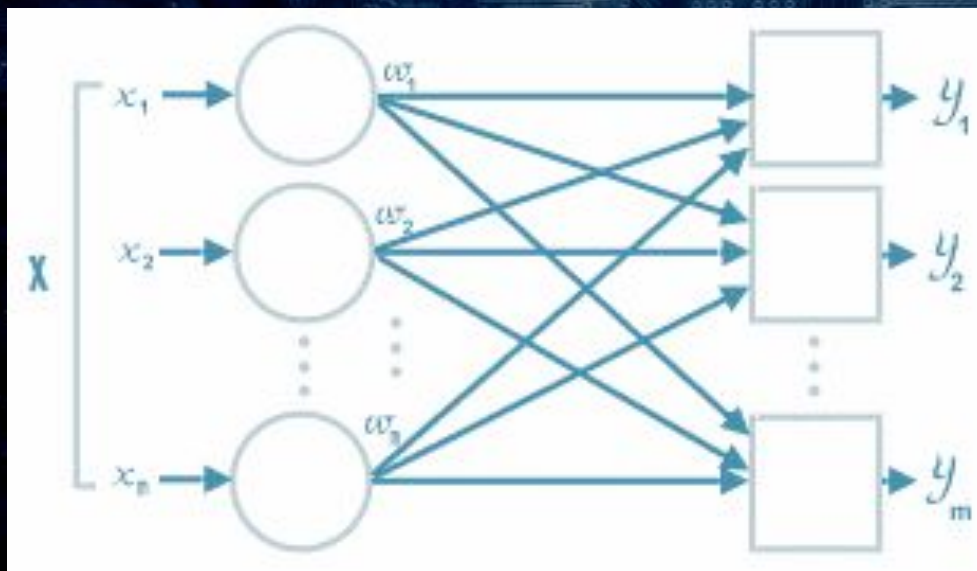
УРОВЕНЬ	ВИД	ПАРАМЕТР	СТРУКТУРА МОЗГА
0	Растения	Температура, освещенность	—
I	Пресмыкающиеся	Пространство	Мозговой ствол
II	Млекопитающие	Социальные отношения	Лимбическая система
III	Люди	Время (особ. будущее)	Префронтальная кора

Искусственный нейрон



- Вычисления выхода нейрона предполагаются мгновенными, не вносящими задержки.
- Не появляются эффекты синхронизации, когда скопления нейронов обрабатывают информацию синхронно, под управлением периодических волн возбуждения торможения.
- Нет механизмов, регулирующих работу сети в целом (например гормональная регуляция в мозге).
- Существует большое разнообразие биологических синапсов

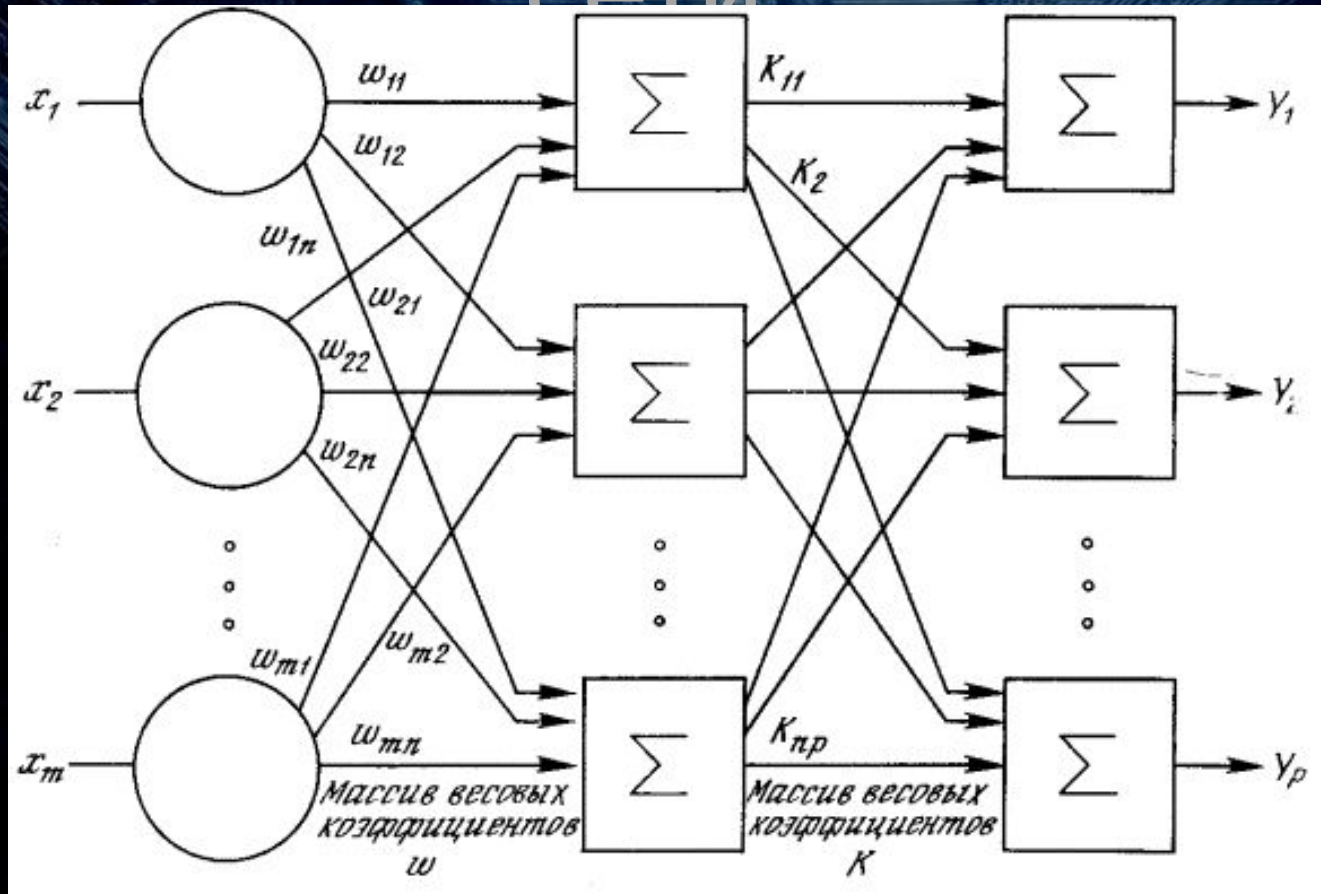
Однослойная нейронная сеть



Можно считать веса элементами матрицы W , имеющей m строк и n столбцов, где m - число входов, а n - число нейронов.

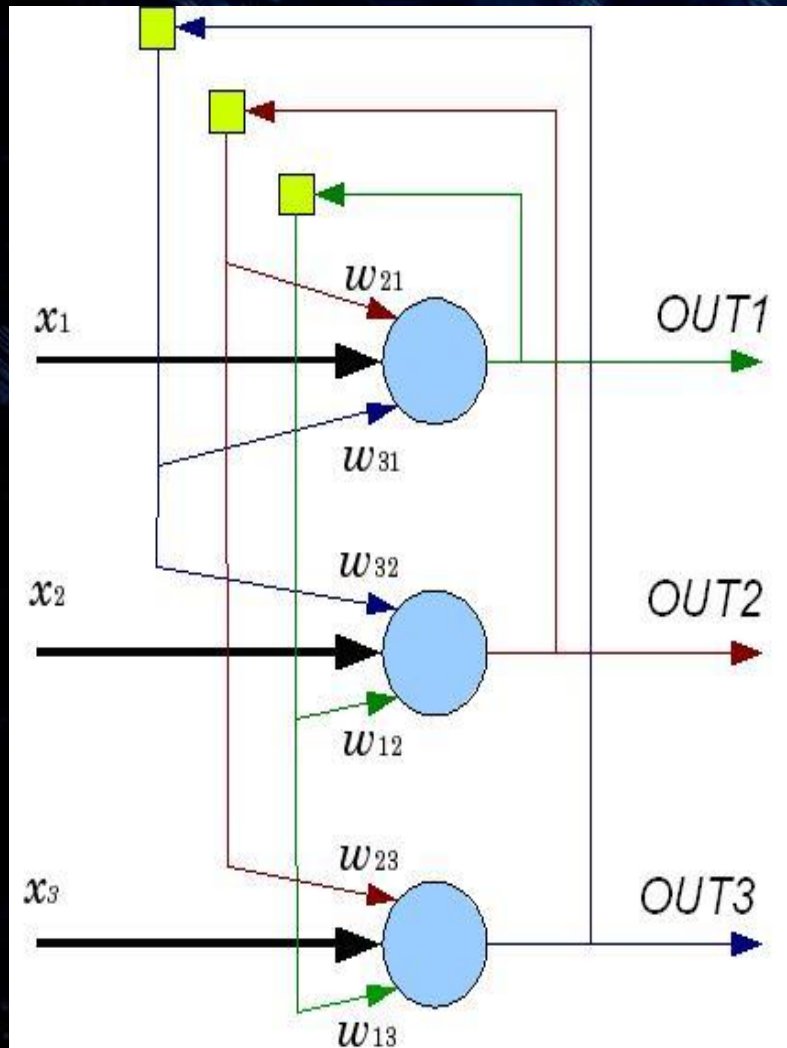
Вычисление выходных сигналов каждого нейрона – вычисление вектора $N = XW$, где X – входные сигналы.

Многослойные нейронные сети



$(XW)K = X(WK)$, значит многослойные нейронные сети сводятся к однослойным.

Нейронные сети с обратными СВЯЗЯМИ



Отклик нейронных сетей с обратными связями является динамическим, т. е. после приложения нового входа вычисляется выход и, передаваясь по сети обратной связи, модифицирует вход. Затем выход повторно вычисляется, и процесс повторяется снова и снова. Если при каждой итерации изменения выходных сигналов не уменьшаются, то сеть неустойчива.

Настроенная соответствующим образом сеть с обратной связью может служить моделью ассоциативной памяти человека. Одно входное воздействие вызывает цепочку сменяющихся состояний выходов, т. е. по заданной части нужной информации из памяти извлекается вся информация.

Обучение нейронных сетей

- Обучение с учителем - для каждого входного вектора существует целевой вектор. Весовая матрица итеративно подгоняется для достижения сходства между выходным и целевым векторами.
- Обучение без учителя - обучающий алгоритм подстраивает веса сети так, чтобы получались согласованные выходные векторы, т. е. чтобы предъявление близких входных векторов давало близкие выходы.

Всё есть сигнал

- Маунткастл, “Организирующий принцип мозговой функции”. Кора головного мозга однородна, все зоны выглядят одинаково, значит выполняют одну и ту же операцию.
- Все, чем оперирует мозг, – это пространственные и временные сигналы.

Всё есть сигнал



Отличия памяти неокортекса от компьютерной

- Неокортекс запоминает последовательности элементов, а не отдельные элементы окружающего мира.
- Неокортекс вспоминает последовательности автоассоциативно
- Неокортекс запоминает последовательности в инвариантной форме.
- Неокортекс сохраняет последовательности иерархически.



Новые рамки понимания интеллекта?

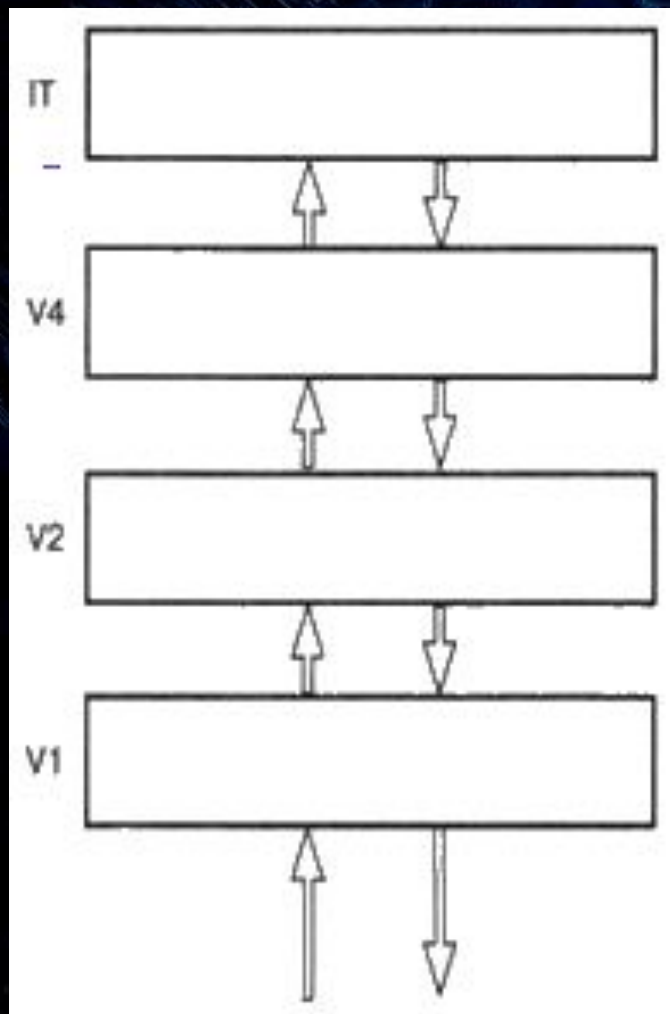
- Поведение не всегда свидетельствует о понимании и интеллекте.
- Мысленный эксперимент “Опытом измененной двери.”
- Прогнозирование - первичная функция неокортекса и основа интеллекта.
- Поведение лишь промежуточный продукт процесса прогнозирования.
- Модифицированная китайская комната.

Модель “Память-предсказание”

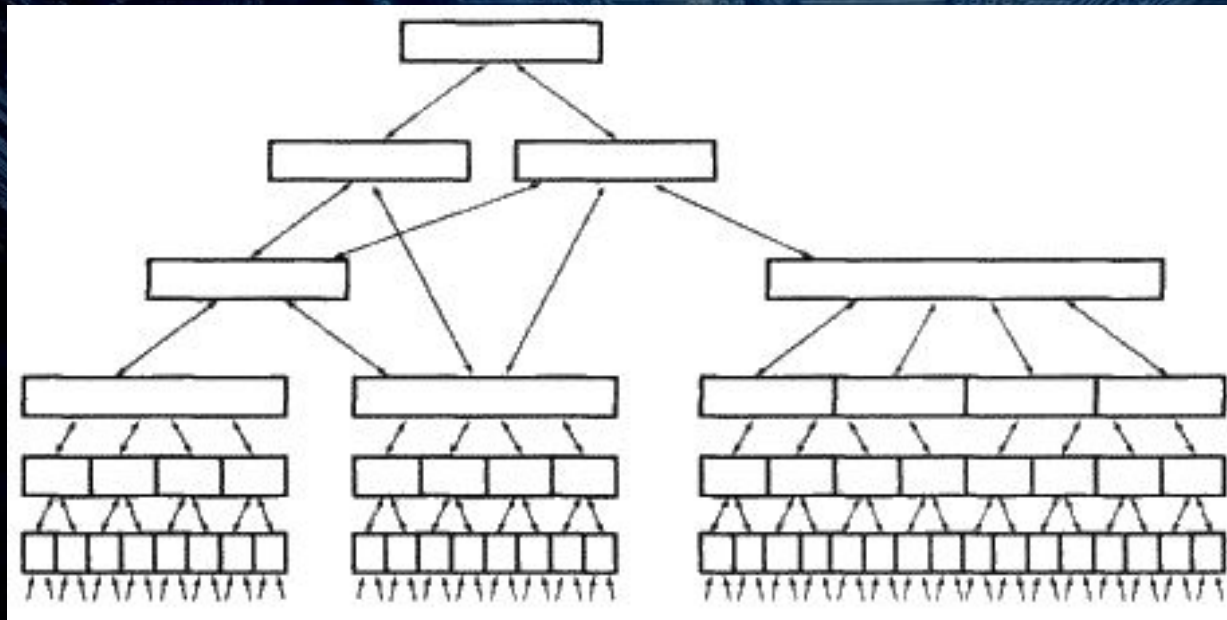
Поступающее на V_1 сильно

искажено из-за саккад и неравномерного распределения рецепторов.

- Каждый нейрон зоны V_1 реагирует на специфический вид сигнала.
- Нейроны каждой последующей зоны возбуждаются благодаря более общим сигналам; IT – нейроны Дженнифер Энистон – создание инвариантных репрезентаций.
- Обратные связи нужны для прогнозирования.



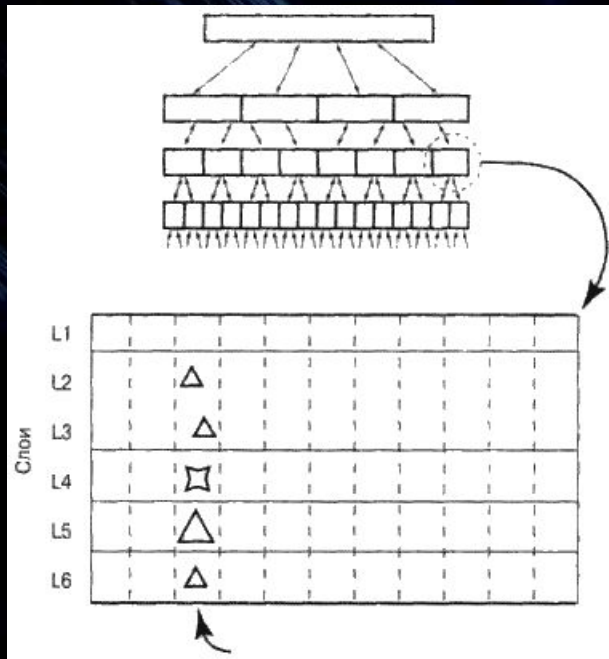
Модель “Память-предсказание”



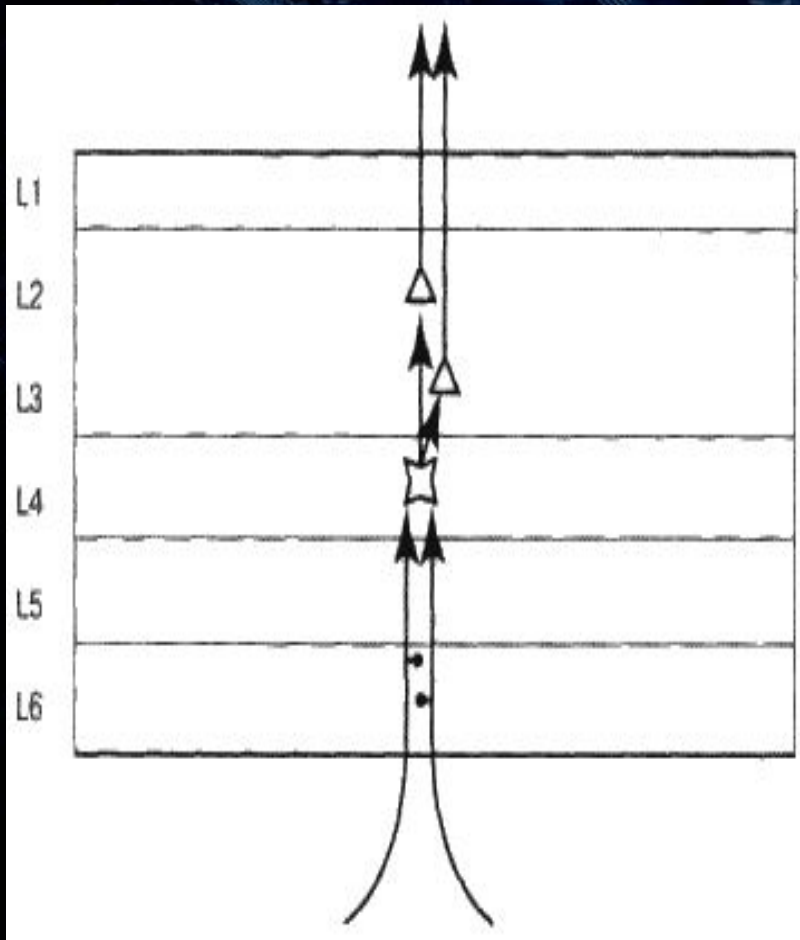
Зоны V разделены на подзоны, сообщающиеся друг с другом посредством вышестоящих зон. Зона IT унифицирована. Каждая подзона формирует инвариантные репрезентации своего уровня.

Структура зон неокортекса

- Вертикально расположенные клетки одной колонки активизируются под воздействием одних и тех же стимулов.
- Наличие колонок нейронов в зонах неокортекса объясняется особенностями формирования коры головного мозга в эмбриогенезе.
- Более 90% синапсов клеток из одной колонки – это связи с клетками, расположенными вне пределов самой колонки.
- Колонка – базовая единица прогнозирования.

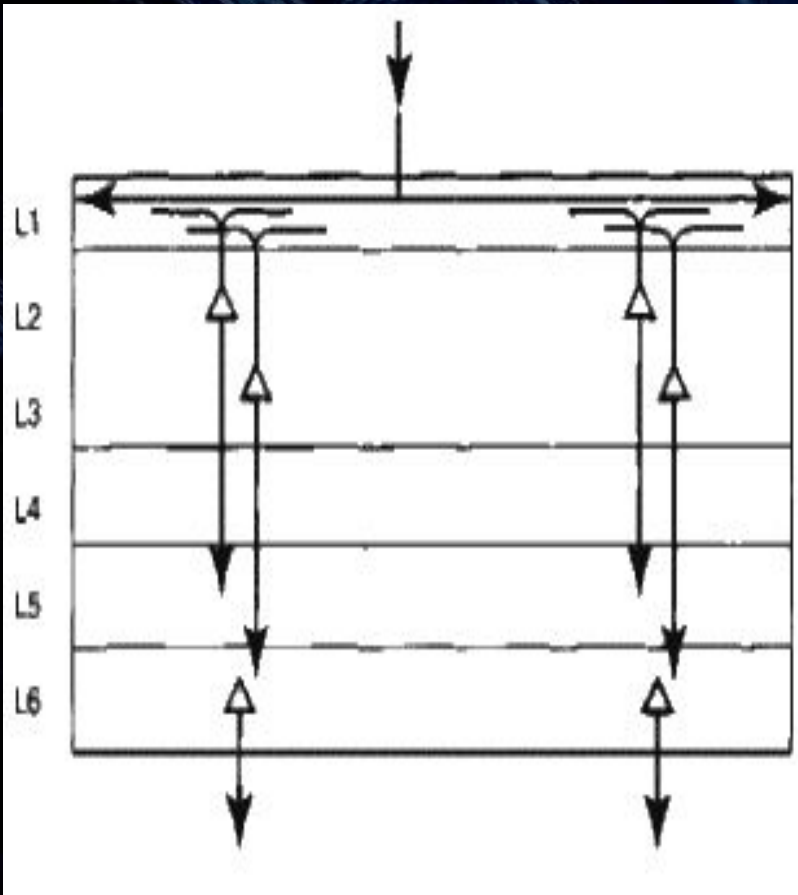


Восходящий поток сигналов



Входные сигналы из нижних зон попадают в слой 4, который является основным входящим слоем. По дороге они «отмечаются» в слое 6 благодаря аксонам. Клетки слоя 4 передают сигналы вверх – к клеткам слоев 2 и 3, находящимся в пределах той же колонки. Далее клетки слоев 2 и 3 передают информацию по аксонам еще выше – к входному слою следующей в иерархии зоны. Информация передвигается от зоны к зоне вверх по иерархии в пределах одной колонки нейронов.

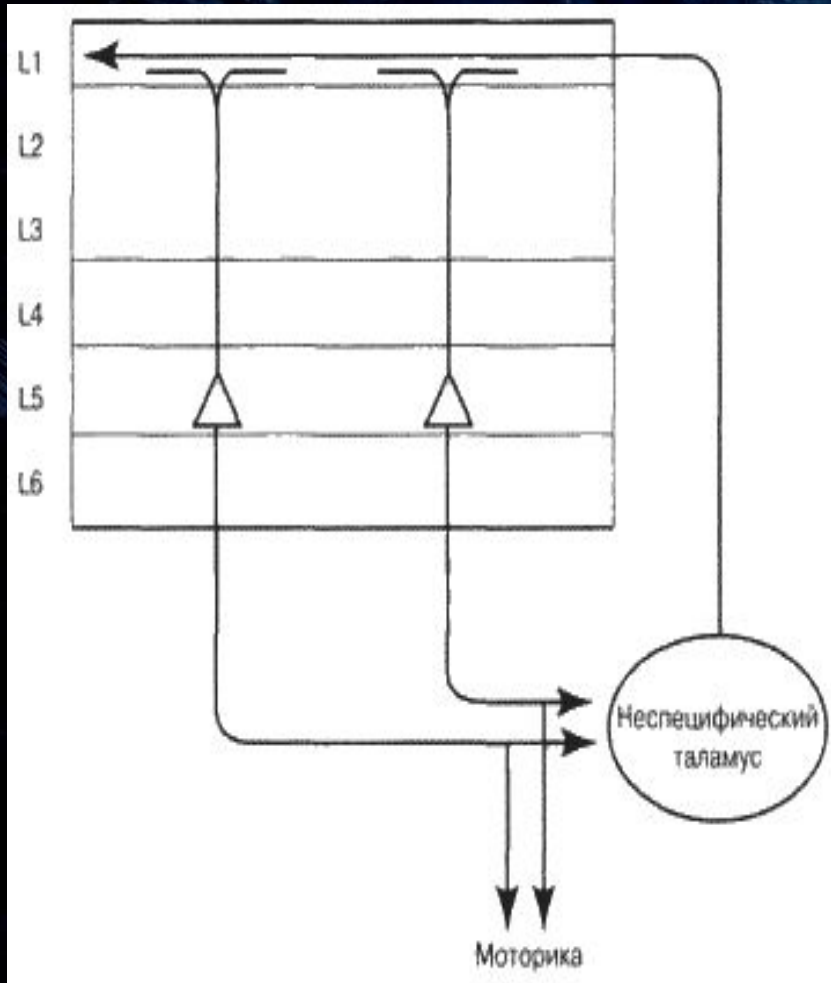
Нисходящий поток сигналов



В слое 1 относительно мало нейронов, но там много дендритов других слоев, а также из других колонок. Поэтому распространяясь из слоя 1 некоторые сигналы могут игнорировать иерархию.

Сигналы распространяются из слоя 1 не только вниз, но и в другие колонки – отражение инвариантной репрезентации.

Непрямой способ связи зон коры



Слой 5 содержит особые крупные нейроны с раздвоенными аксонами. Одно из ответвлений идет к таламусу. Из таламуса сигнал отправляется в ту же функциональную зону. Такой круговорот представляет собой обратную связь с задержкой, которая позволяет автоассоциативной памяти усваивать последовательности сигналов.

Альтернативный путь вверх по иерархии



- “Обыденные” низкоуровневые сигналы не доходят до высших зон иерархии.
- Что бы “обыденный” низкоуровневый сигнал привлек внимание он должен перестать быть “обыденным.”
- Либо найти другой путь – через таламус.
- Привлечь внимание к чему-либо можно сознательным усилием.

Тормозные нейроны, прогнозирование и воображение

- Для создания инвариантных репрезентаций необходимо делать выбор между колонками, отвечающими за идентификацию схожих сигналов. Этим занимаются тормозные нейроны, которые не дают послать сигнал сразу нескольким колонкам.
- Прогнозирование осуществляется благодаря тому, что инвариантные репрезентации сохраняются последовательно.
- В случае неверного прогноза вышестоящей в иерархии зоной тормозные сигналы блокируют дальнейший спуск сигналов.
- Клетки слоя6 могут посылать выходные сигналы назад, к клеткам слоя4 той же колонки. Т.е.

Как происходит процесс обучения?

- Правило Хебба - когда два нейрона активизируются одновременно, синаптические связи между ними усиливаются.
- Суть обучения – формирование классификаций и создание последовательностей.
- Клетки слоя 4 учатся формировать новые классификации, что изменяет сигнал, проецируемый назад, в слой 1.
- Группировка сигналов, следующих друг за другом. Группировка группировок сигналов и т. д.

Гиппокамп– вершина иерархии

- Кора головного мозга взяла на себя многие функции более древних структур.
- Потеря гиппокампа приводит к потере долговременной памяти.
- Абсолютно новый сигнал поднимается все выше по иерархии вплоть до гиппокампа .
- Постепенно гиппокамп передает запоминания в неокортекс.

Человекоподобные машины?

- Человеческий разум – не только функционирование коры мозга.
- Воссоздание других структур мозга – сложнейшее и невыгодное предприятие.
- Что бы общаться подобно человеку нужно быть человеком.
- Затраты на создание и эксплуатацию.

Трудности при создании разумной машины

- Емкость памяти. Неокортекс содержит 32 миллионов синапсов, что примерно равно 8 трлн. байтов двоичной памяти.
- Возможности моделирования описанных процессов на современных ЭВМ сильно ограничены.
- Выход – создание чипов разумной памяти. К тому же система “Память – предсказание” толерантна к дефектам.
- Проблема связности. В мозге каждый нейрон связан с тысячами других нейронов. Реализовать подобное на чипах пока что невозможно.

Преимущества искусственных разумных систем

- Быстродействие нейронов измеряется в миллисекундах, а кремний функционирует со скоростью наносекунд.
- Размер человеческого мозга ограничивается биологическими факторами.
- Добавление новых ощущений и сенсорной иерархии позволит устройству создавать более качественные модели внешнего мира
- Реплицируемость.

Перспективы создания разумных систем

- Интерфейс естественной речи.
- Зрительное восприятие.
- Разумные механические устройства.
- Сенсорные системы:
 - Сигналы могут иметь любую природу
 - Метеорология.
 - Городская инфраструктура.
 - Работа в абстрактных мирах.
- Объединение мыслящих систем в иерархию.

Литература

- С. Блейкли, Д. Хокинс - Об интеллекте.
- Ф. Уоссермен - Нейрокомпьютерная техника: Теория и практика.
- А. Н. Горбань и др. - Нейроинформатика
- Р. Пенроуз – Тени разума.
- Р. Пенроуз – Новый ум короля.
- <http://www.neuroscience.ru/>



SOON.