?????

- 生物脑（人脑）的优势_____
- 意识（自我）_____
- 人脑的工作原理_____
- 仿真人脑的基本条件_____
- 人脑的推理模型_____
- 人脑的本能_____

??????????

1. 抽象
    1. 抽象出高效的，合理的，可复用的　　底层抽象
    2. 抽象出高效的、非常抽象的　　高层级概念　　，并应用于看起来毫不相关领域
2. 记忆　**:　**记忆及快速的索引
3. 类比：强大的类比功能
    1. 神经训练的收敛过程可以类比晶体管放大，训练要达到放大区，而不是在截止区和饱和区，不然会过拟合，或者欠拟合
        1. 饱和　　类比　　边际效应
        2. 放大区　　类比　　线性关系
    2. 国足出国踢球，输了球但是个人行李很多：差生文具多
4. 联想：高效的联想思维能力
    1. 抽象概念的联想
        1. 自动构建一个　　逻辑树　　，逻辑组成的树，作为匹配的特征
        2. 每个节点表示逻辑关系，逻辑运算
    2. 具象概念的联想
        1. 猫和猫科动物的类似
5. 意识，自洽
    1. 而人脑第一前提是　　“自洽　”，同时尽量高效组织信息，产生高层抽象
    2. 预测：为了自洽，不断的进行逻辑推理和判断
    3. LLM追求极致压缩，尽量减少冗余信息
6. 学习
    1. 模仿能力：一种快速的、高效的学习方法
        1. 贪心：只要是自洽的，没有矛盾的　　“事物　”就可以被快速整合吸收
    2. 通过时间积累，不断的整理和思考已有的知识，提高效率
    3. 高效获取新技能的能力　　：ARC-AGI-2任务
7. 创造力、创新
    1. 递归抽象　　能力
    2. 高层级的抽象概念，跨领域的互相借鉴
    3. 大范围（深度、广度）的思维推理：不断尝试可能的新概念
    4. 演绎：应用现有的规则和原理进行推演

        归纳：对现象、事物进行归类

        宿因：根据结果推测可能的原因，可能的现象

        抽象：对已知事物和规律进行归纳与宿因，并以此定义出更高层级的新概念或新规则，能用于演绎

??????

?????

1. 意识，可以解释成　　对自我的一种认可，我是我，一般说一个智能具有意识，表示能表现出　　“自我　”的行为。
2. 自我 是一套复杂　　价值观、行为规则　　，是一种关于人文、社会、基本生存等的复杂算法的高层级抽象、表示　　。
3. 由于包含很多非理性的、直觉式的规则和规律，所以难以量化和计算。但是　　自我　/意识就很好的表达了这一类的规则　　/规律

具体来说就是，首先假设一个主体，基本需求是自尊、自信、自强、生存等一系列的目标　　/需求，这些抽象的目标（价值观）才是人类进行活动的价值函数。

特别的用于处理社交，人性规则，等比较人文的需求。

1. 工具类的应用非必须具有意识，但是，　　对于　“玩具　”，“模仿动物　　”，“机器人　　”，“情感类处理　　”非常有用，而且是必须的。
2. 机器人如果不仅仅是产线的工具，需要有一些自我决策的行动，就需要意识。要不然就和汽车产线的机械臂没有区别。
3. 这里的区别不是体现在执行准确任务还是非确定任务，而是能用一些高层级（自我意识）的抽象规则，和当前的最新更新的状态，临场判断和执行一些任务

???

一个无限理智的人，就是一个纯粹的算法，所有的输出都是可预测的，可计算的。

品味，世界观，价值观，说是人类最后的，最擅长的东西。

但是，实际上也是一套复杂的算法而已，肯定能被实现，可以被复制。
只是这中间的训练方法我们还不能完全搞清楚。

但是因为影响最终输出（反应）的因素太多了，而且和大量的历史状态有关，所以这个算法的输入的变量特别多，

人类为了高效的决策，就采用了，经验　　＋模仿　＋贪心的方法来处理输入。

所以看起来很神奇，很特殊，不能理解，但是效果还不错，我们管这个叫做意识。

意识是对人性这个算法的具体实现

而且，完全的理智也可以体现出人性，也可以说是一种算法。

??

1. 表现为能一直统一个完整的， 自洽 的行为表现
1. 保持行为、价值观的一贯性
2. 能体现出个人的风格特色，价值观
1. 有一些内在的价值观设定比较难以受外界的输入而改变
2. 具有不完全受现有观念影响的判断逻辑。

??

1. 当人工智能的研究者终于制造出会使用语言的机器时，却并没有在这些机器中发现意识，会说话的 Chat GPT却没有意识

???????

??

1. 人脑具有反馈回路，反馈回路形成正反馈，不断训练神经元，加强神经元
    1. 视觉神经能识别图像，也能通过提示想象出图像
    2. 语言作为抽象能力很强的表示工具，被充分用于协助思考，但不是智力的全部 @维特根斯坦
2. 所有的神经元时刻待命，适当的输入技能激活输出
    1. 不激活的状态需要主动抑制皮层神经的激活
    2. 闭眼睛的时候，视觉神经会 抑制 视觉皮层的激活
3. 动态 激活
    1. 神经元默认是静息状态，根据输入，逐级激活需要激活的神经元
4. 学习和实践（训练和推理）同步进行， 训推同时 进行
5. 能量不足的时候，大脑不是不能工作，而是有 低能耗模式
    1. 低血糖时候，大脑能通过降低视觉识别的精度来降低能耗
        1. 老鼠识别天敌 “猫 ”就可以，不用识别出来具体是哪种猫

??/?????

1. 贪心、模仿
2. 高层级的抽象带来高效率
3. 不断的对输入进行推理，保证自洽

???????

1. 生物神经元或者说大脑的训练目标，可能是让神经元处在一个 有规律的环境中，不能接受杂乱无章的信号，比较喜欢有规律的受控的处罚和激活
2. 能量效率，节能
3. 只有在快要忘记的时候去努力回忆，记忆才会被深深地刻在大脑里

???

1. 支持非常大的处理空间：每个神经元大约有 1000-10000的突触，海马体神经元更多，约 98% 的突触位于树突
2. 对很多（ N=~ 10000 ）个输入信号处理，神经元通过学习，调整之间的固定连接，构造多个中间层的处理机制，形成高层级的抽象，实现高效动态计算。 大模型通过矩阵乘法将所有输入做映射计算。
3. 这 N个信号，实际存在的实际组合样本很少，大部份组合情况是不存在的，非常稀疏，
4. 如果 N个信号，只会存在一种情况，一个神经元就能表达，通过模仿、贪心学习快速进行抽象，而 AI 需要一个非常大矩阵乘法。

5. AI的大矩阵映射出来的结果和人类的单个抽象 (语义）概念不对应
1. 通过梯度下降求解，自然会自动提高权重的信息密度，每个单点的输出可能是多种概念的某个
共同特征的激活
2. 人类则更注重适应性和丰富性，强调保持灵活性和上下文的完整性，而 LLM 侧重于统计压缩，力求最大程度地减少冗余信息

?????????

??

1. 能够持续学习
2. 建立世界模型并进行规划的能力
3. 能够进行新的高层级的抽象
4. 能够进行 “元 ”概念的学习和泛化

?????

1. 理解自我、个体的概念
2. 维持 自洽 的基本能力
3. 行动的目标都是内源性，外部不会直接给定一个明确的目标

??

1. 创新的原理，需要形成 递归抽象 的思维方式和能力
2. 维护自洽，需要能做到 实时动态 的递归抽象

???????

1. 神经元
    1. 可以被输入激活，激活后可以激活其他的连接的神经元
    2. 记忆区神经元
        1. 大部分时间是静息状态
    3. 逻辑推理神经元
        1. 视觉识别、语音识别，皮肤传感器等，人脸识别等等
        2. 正向，顺序激活，快速，没有递归
    4. 规划神经元
        1. 一团细胞，反复递归的激活，循环
2. 区域
    1. 短路径，高并行
        1. 小脑，视觉识别
    2. 大量的边缘分布
        1. 记忆的节点
    3. 状态区
        1. 一堆的能主动激活其他人的细胞组成的临时记忆区，就是意识的本体
        2. 大量连接了记忆区的节点，一个连接多个，可以根据当前其他的状态进行动态的选择，从而表示一个临时的状态
            1. 比如，当前是在编程，那么当前这个细胞连接的编程相关的记忆会激活当前，其他的记忆的激活会被屏蔽
    4. 长路径，低并行，决策区
        1. 接收状态区的细胞的信号
        2. 清醒地时候，这些细胞就有足够的能量激活其他的细胞，形成一次迭代，和行动决策，激活下一步动作的信息源头
3. 运行的本质
    1. 有一堆循环连接的神经元
        1. 只要没有在休息状态，就会不断根据输入状态激活对应的神经元
        2. 不断循环得保持激活状态，维持当前的意识

?????

从人类的生理机制、认知心理学与现实行为
视角来看，这三层架构完全对应着人类从生物本能到高阶思维的演化与发育逻辑。

核心范式：基因预设结构（先天），经验演化行为（后天）

人类大脑的解剖结构与神经网络拓扑（如海马体、前额叶、杏仁核）是基因硬编码预设的“骨架”；而我们在社会与自然环境中积累的知识、习惯和专业技能，则是后天通过神经可塑性（Neuroplasticity）动态涌现并沉淀下来的“内容”。

1. ????????????

- 生理根基：由脑干、下丘脑与杏仁核等古老脑区控制，不经过大脑皮层的理性思考即可直接触发。
- 人类行为表现：
  - 生存与稳态维持
    - 对剧痛的瞬时缩手（避害）
    - 饥渴带来的焦虑感（促使求生）
    - 极度疲劳时的强制入睡机制
  - 威胁与中断机制
    - 遇到突发巨响或危险时瞬间触发“战或逃（Fight or Flight）”反应
    - 强制打断当前正在进行的任何复杂思考
  - 熵减驱动（好奇心），自我认知的补全，自洽
    - 对未知环境产生天然的警觉与探索欲望
    - 通过消除不确定性来降低潜在的生存风险

????“????”?**?????** (Survival & Energy Efficiency)

生物脑的一切高级功能都服从于这个终极指令。大脑只占人体重量的2%，却消耗了20%的能量。因此，大脑极其“吝啬”计算资源。

- 本能体现：
  习惯化（把重复动作固化到潜意识以节省算力）、避免认知失调，以及在非必要时倾向于使用“直觉”而非“深度逻辑”。

???? (The Reward / Loss Function)

在有逻辑之前，大脑先有情绪。情绪是大脑用来快速评估环境并做出反应的全局权重参数。

情绪是一套非常高效的处理复杂环境和外部情况的规则/算法，用来快速评估行动的好坏。

?????**??**

- 本能体现：
  生存、保护自己，趋利避害。多巴胺驱动探索和目标达成（奖励函数），恐惧和痛苦驱动逃避（损失函数）。当你“思考下一步行为”时，你的大脑实际上是在计算哪一条路径能带来最大的预期收益或最小的生存威胁。
- 奖励系统（多巴胺）
  成功预测、吃到奶、被抱起来时，多巴胺系统会释放愉悦信号，强化刚才的行为和认知。这是所有学习的化学基础。这种“行为 -> 反馈 -> 调整权重”的闭环是本能。

奖励驱动的“强化学习”(Reinforcement Learning)

??????????????

大脑是一台"预测机器"，它存在的根本目的是不断预测下一刻会发生什么，并让预测与现实的误差最小化。

这是生物脑最核心的动力。大脑会自动对下一秒发生的事做预测，如果预测失败，就会产生强烈的电信号。

也就是说，大脑并不是被动地接收世界，而是主动地建构一个关于世界的模型，然后：

1. 用感官输入去校验模型；
2. 当预测出错时，要么更新模型（学习），要么采取行动改变世界（行为），让世界符合预测。

记忆、注意、情绪、决策——全部可以被解释为这台预测机器的不同子模块。

???????????? (Predictive Coding & Planning)

现代认知科学认为，大脑本质上是一个预测机器（**Prediction Machine**）。预测是为了更好得避免可能的危险，更好得生存。

- 本能体现：大脑不会被动等待外界输入，而是每时每刻都在利用长期记忆，在内部生成对外界环境的预期。
- 模拟与剪枝：在面临复杂决策时，大脑会在意识中构建未来的**计划树（Plan Tree）**或决策树。它在沙盘中推演不同行为可能导致的结果，并通过奖惩机制对不可行的分支进行快速“剪枝”，最终选择出最优路径。这是一种极高阶的计算本能。

“????”????? (Prediction Error & Curiosity)

- 功能：
  当婴儿看到一个球违背引力漂浮在空中（违背了大脑初步建立的物理模型），他们的注视时间会明显变长。这种对“意外”的关注就是好奇心的生物学基础。
- 成长价值：这种本能强制大脑去关注那些“还理解不了”的信息，从而不断修正内建的模型，实现认知的迭代。
- 婴儿天生就会把目光投向新出现的、变化的、对比强烈的东西。这不是教出来的，是神经系统的默认设置。它确保婴儿持续采集新信息。
  以便更好得了解这个世界，做出更好的预测，更好得适应环境

- 预测与惊讶机制　　即使是新生儿，大脑也在不停地　　"猜下一秒会发生什么　　"。当现实和预测不符时，会产生　　"惊讶信号　　"，这个信号就是学习发生的地方　　——它告诉大脑："这里有新东西，记下来，更新模型。"

2. ??????????????

- 生理根基　　：感觉皮层、运动皮层与海马体在发育早期建立的　　最小粒度动作与感知单元（**Primitives**）。
- 人类行为表现　　：
  - 工作记忆与焦点调度
    - 将注意力短暂维持在眼前的　　3–5 个事物上（如听一句话、盯着一个目标）
  - 模式识别与语义抽取
    - 自动信号处理，听觉，视觉等
    - 在杂乱的环境中瞬间识别　　“那是人脸　　”，“那是喊我的声音　　”
  - 基础工具与肢体控制
    - 抓握、行走、发声，以及对基本物理规则（如重力、物体碰撞）的直觉感知
  - 信息存取与映射
    - 将听到的声音映射为语言概念
    - 将看到的文字转译为抽象含义

???????**????** (Hierarchical Memory System)

没有记忆，就没有时间概念，也就无法进行逻辑推演。大脑的记忆本能并不是为了　　“记录过去　　”，而是为了预测未来。

- 短期　/工作记忆:　　相当于大脑的内存（　　Context Window）, 容量极小，负责处理当前正在发生的信息和维持暂时的逻辑推理。
- 长期记忆:
  相当于庞大的向量数据库。大脑的本能是自动将高频出现的、带有强烈情绪标签的短期记忆进行　　“权重固化　　”，转化为长期记忆，以便在未来遇到类似情景时瞬间调用，无需重新计算。

?????“????”?? (Statistical Learning)

大脑本质上是一个极其强大的　　概率计算引擎　　。

- 功能:
  婴儿天生会不断计算环境中各种刺激出现的频率。例如，在听成年人说话时，他们能通过统计音节之间出现的概率，在完全不懂词义的情况下，准确地切分出哪里是一个词的开始和结束。
- 成长价值:　　这是语言习得和逻辑归纳的起点。

??????"????"

人类婴儿绝不是被动接收数据的白板，而是一个　　高度优化的、带着预设目标和底层算法的超级学习系统。时间、空间、因果、数量构成了他们理解物理世界的坐标系；而面孔偏好、意图探测和朴素道德，构成了他们理解社会世界的坐标系。而天生自带的短　　/

长期记忆比对与统计学习机制，则是驱动这个系统不断自我更新、最终长成复杂智能的核心引擎。

- 物体恒存的雏形 ：知道物体是连续的、不会凭空消失或穿墙。
- 因果直觉 ：一个球撞到另一个球，第二个应该动。
- 数量感 ：能粗略区分 "多 "和 "少 "。
- 空间感 ：对距离和方向有基本感知。

这些 "先验 "就像一个粗糙的世界模型框架，后续所有学习都是在往这个框架里填细节。

3. ??????????????

- 生理根基 ：前额叶皮层驱动的高阶思维，决定了人类面对全新复杂问题时 “如何思考 ”以及 “如何学习 ”。
- 人类行为表现 ：
  - 第一性原理与解构
    - 面对全新难题（如修车、诊断疑难杂症），能剥离情绪与表象，将其拆解至最基础的物理或 逻辑因果链条 。
  - 试错 -反馈 -修正闭环
    - 学骑车、打网球或做实验时的 “尝试动作 >发现摔倒 /报错 >归因并调整姿势 >再次尝试 ”的演进过程。
  - 技能内化（肌肉记忆化）
    - 用高耗能的 “慢思考 ”（系统 2）反复验证某套解决路径后，将其打包压缩为低耗能的 “肌肉记忆 /直觉快思考 ”（系统 1，如熟练驾车、老手盲打）。
  - 边界感知与自知之明（元认知监控）
    - 清晰意识到 “我的经验在这个领域失效了 ”，从而主动停止盲目猜测，转为搜集资料或向专家请教。

??????**????** (Logic & Abstraction)

现实世界的信息量是无限的，而大脑的算力是有限的。

- 本能体现: 寻找规律，提升效率。大脑本能地将复杂的自然语言、视觉图像等非结构化数据，提取成最简化的符号、规律或 “语义代数”。这种逻辑抽象能力，就是为了将海量信息压缩成最少的规则，从而用一套通用的逻辑去应对千变万化的世界。

?????????(Social Alignment)

人类是极度社会化的生物，婴儿大脑中有专门处理 “人 ”的硬件。

- 功能:
  婴儿出生几天就能本能地识别出人脸形状，并具备模仿本能（你对他吐舌头，他也会尝试模仿）。
- 成长价值: 通过模仿，婴儿跳过了漫长的 “试错 ”阶段，直接从已有的文明成果中（如语言、工具使用、社交礼仪）快速下载 “软件 ”。

??

1. 模仿能力　　：出生几天的婴儿就能模仿大人吐舌头、张嘴　　——这是学习的雏形。
2. 模仿是最低成本的学习，直接模仿标准答案，是风险最低的适应社会的行为

?????

1. 体内稳态（　　**Homeostasis**）　——维持体温、血糖、水分、氧气等在安全范围内。
2. 趋利避害　　——
   靠近奖励（食物、温暖、同伴），远离威胁（疼痛、捕食者）。由边缘系统、杏仁核、伏隔核等负责
   。
3. 学习与记忆　　——把 "什么行为导致了什么结果　　"存下来，　　以便下次预测得更准　　。
4. 社会性认知　　——人是群居动物，理解他人意图、建立关系本身就是　　生存策略　　。
5. 规划与想象　　——前额叶让我们能模拟　　"如果我这样做会怎样　　"，这是　　"思考下一步行为　　"的神经基础。

?????

1. 节能：降低能量消耗，提高生存率
   1. 情绪机制：基本的、高效的指导行动规则
   2. 人性化算法：社会活动能力、提高物种能力
   3. 更快得预测：避免浪费能量
2. 预测危险：提高生存率
   1. 好奇心驱动学习：　　主动学习未知的可能有用的知识
   2. 模仿学习：快速得学习正确答案
   3. 总结规律的学习：学习更高层次的规律，高效得判断和行动
   4. 反思失败：避免反复得错误
   5. 记忆：基础设施
3. 行动
   1. 主动行动
   2. 主动干预环境