

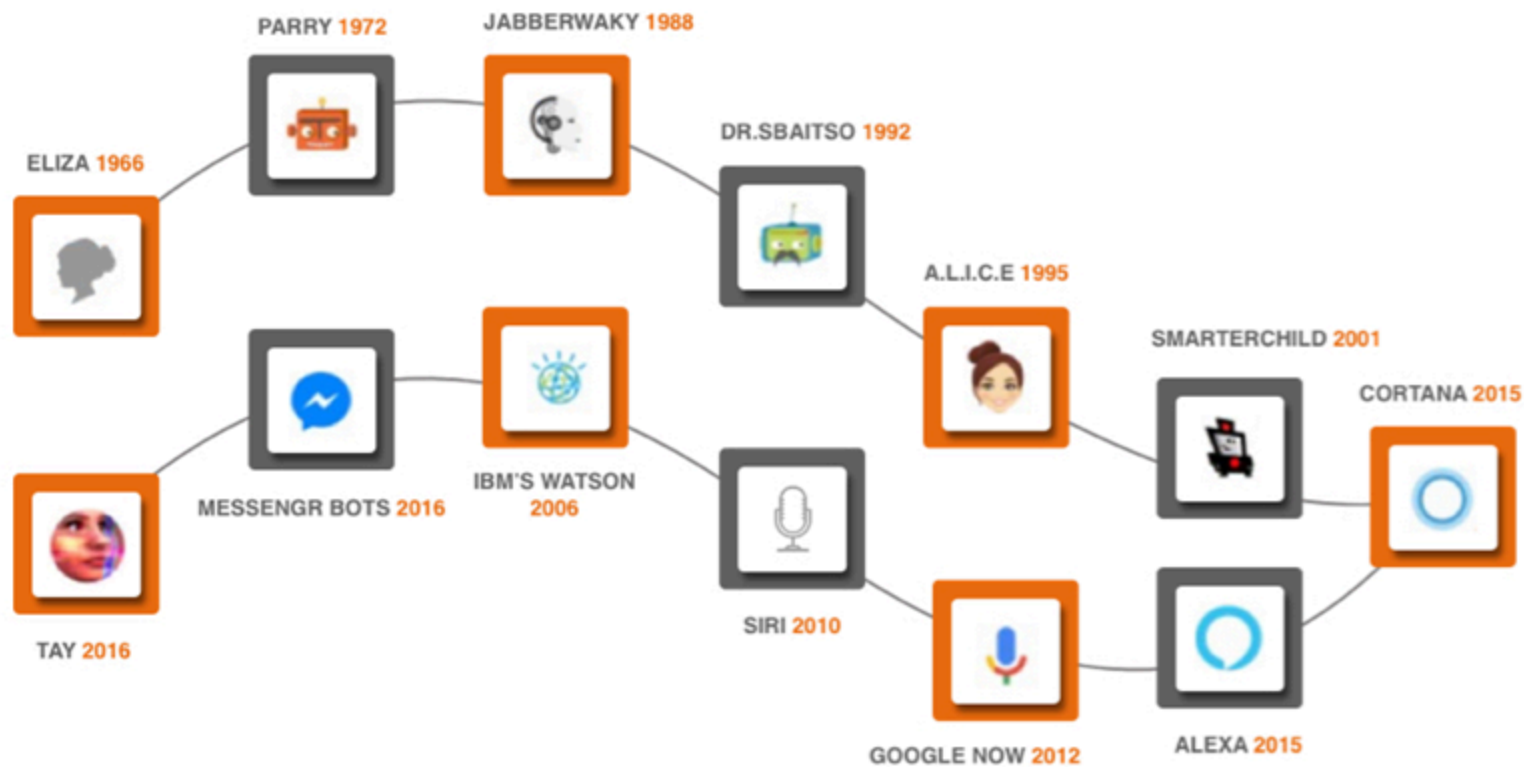
计量世界的奇妙——算法，这也是人工智能？

引子——Chatbot的变迁

- 什么是Chatbot?
- Chatbot（聊天机器人）是经由对话或文字进行交谈的计算机程序，现在我们熟知的苹果的Siri,华为的小艺 以及 Chatgpt 都是一种Chatbot.

Chatbot的变迁

The History of Chatbot



Chatbot的起源：ELIZA

- ELIZA是第一个聊天机器人，由 Joseph Weizenbaum 于 1966 年创建，它使用模式匹配和替换方法来模拟对话，最初的目的是帮助心理咨询医生来解决患者的精神问题。
- ELIZA的关键方法涉及认出输入里的提示字词，并且找出相关的、预先设定的回答，在显然有意义的方式下，让谈话继续下去（例如，输入里有“母亲”，回答是“多说一点你的家庭”）。

Chatbot的起源: ELIZA

Welcome to

EEEEEE	LL	IIII	ZZZZZZ	AAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LL	II	ZZZ	AAAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LLLLLL	IIII	ZZZZZZ	AA AA

Eliza is a mock Rogerian psychotherapist.

The original program was described by Joseph Weizenbaum in 1966.

This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?

YOU: Men are all alike.

ELIZA: What is the connection, do you suppose ?

YOU: They're always bugging us about something or other.

ELIZA: Can you think of a specific example ?

YOU: Well, my boyfriend made me come here.

ELIZA: Is it important to you that your boyfriend made you come here ?

YOU: He says I'm depressed much of the time.

ELIZA: I am sorry to hear that you are depressed.

YOU: It's true. I am unhappy.

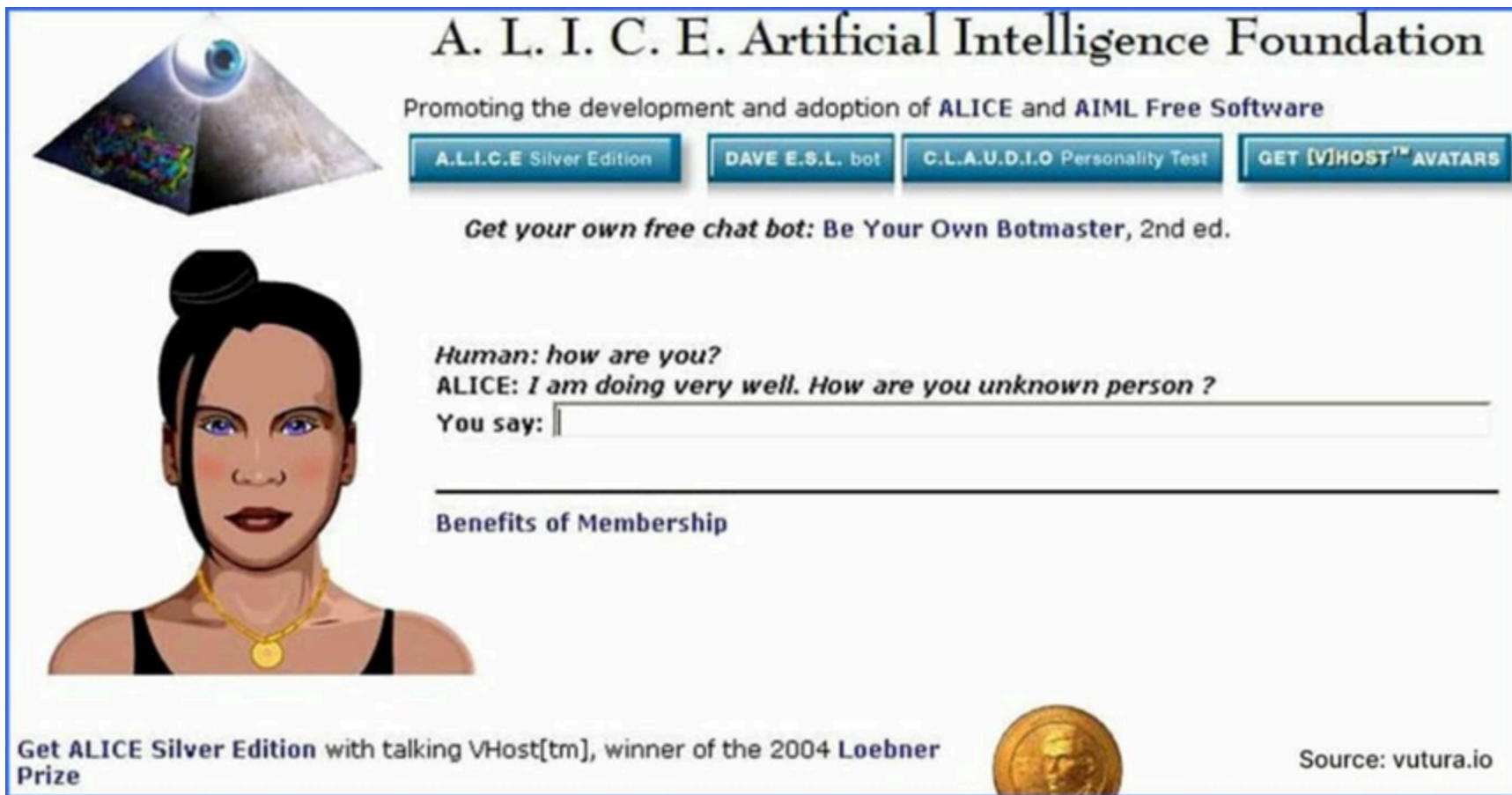
ELIZA: Can you explain what made you unhappy ?

YOU:

Chatbot的发展：ALICE

- 从 20 世纪 60 年代快进到 90 年代，第一个人们熟知的能在线交流的聊天机器人——ALICE出现了
- 但就像 ELIZA 一样，ALICE 也是一个根据规则建构起来的计算机程序，接收输入并产出输出。事实上，ALICE 在以下三个方面优于 ELIZA：
 1. 它以一种被称为人工智能标记语言（AIML）的编程语言编写，这种语言类似于 XML，它允许 ALICE 在更为抽象的层面上做出回应；
 2. 它拥有成千上万种可能的回应；
 3. 它会存储之前与用户的对话，并将对话存储在数据库中。
- 虽然 ALICE 和 ELIZA 相比更为先进，但它的输出依然是由人写成的，算法只是来挑选什么样的输出更适合某一输入。从根本上说，人们写出了回答，也写出了用来挑选什么回答合适的算法，这都是为了模仿一场真实的对话。

Chatbot的发展：ALICE



- 该程序模拟通过 Internet 与真人聊天。Alice被定义为一位年轻的人类女性，她会告诉用户她的年龄、爱好和其他有趣的事实，并回答用户的对话。

Chatbot的现在: Chatgpt

- 2022年11月30日, OpenAI发布了一个名为ChatGPT的自然语言生成式模型, 它以对话方式进行交互。
- 不同于前两种模型, ChatGPT是一种基于自然语言处理技术的对话生成模型, 它通过学习大量的语言数据和预测下一个单词的能力, 可以自动地生成人类可读的自然语言文本, 从而实现自然、流畅的对话生成。
- 简单来说, Chatgpt是一种**生成式人工智能**, 回复的有可能是从未在语料库中出现的、由聊天机器人自己“创造”出来的句子

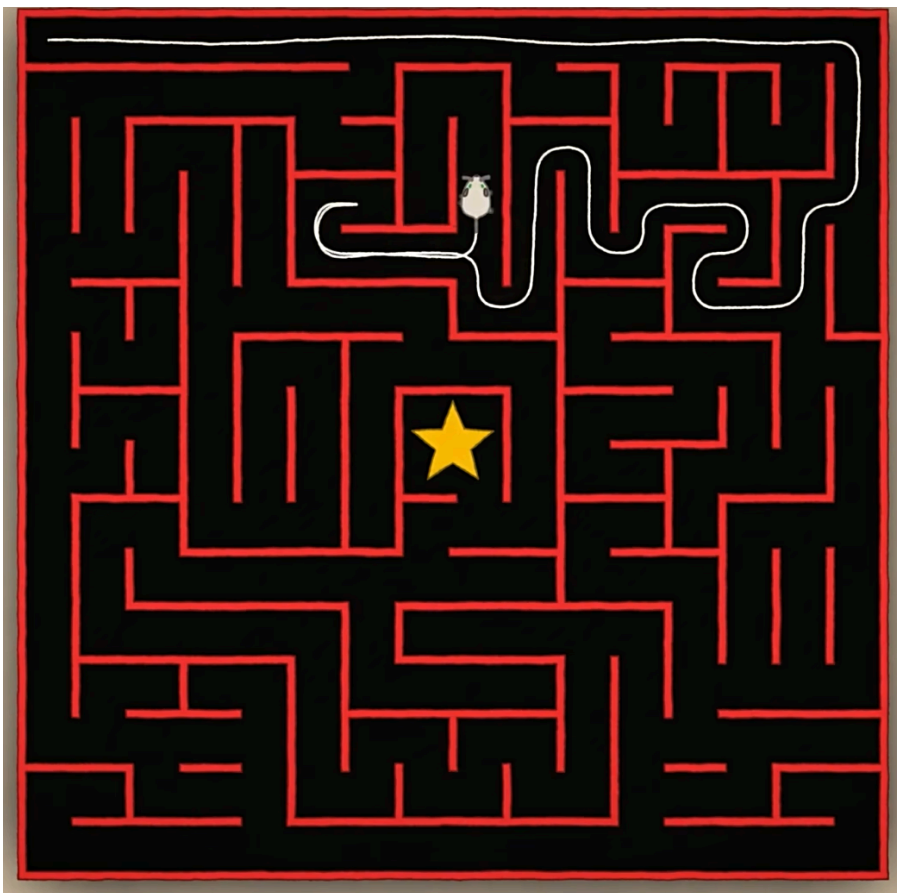
什么是AI?

- 好的，讲到现在，你可能仍然对AI没有一个很清晰的概念
- 但事实上，上述所有的Chatbot都是一种AI，而且只是AI应用的一个方向
- **一个经典的AI定义是：“智能主体可以理解数据及从中学习，并利用知识实现特定目标和任务的能力。(A system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation)”**

AI是什么？——从算法说起

算法是什么？

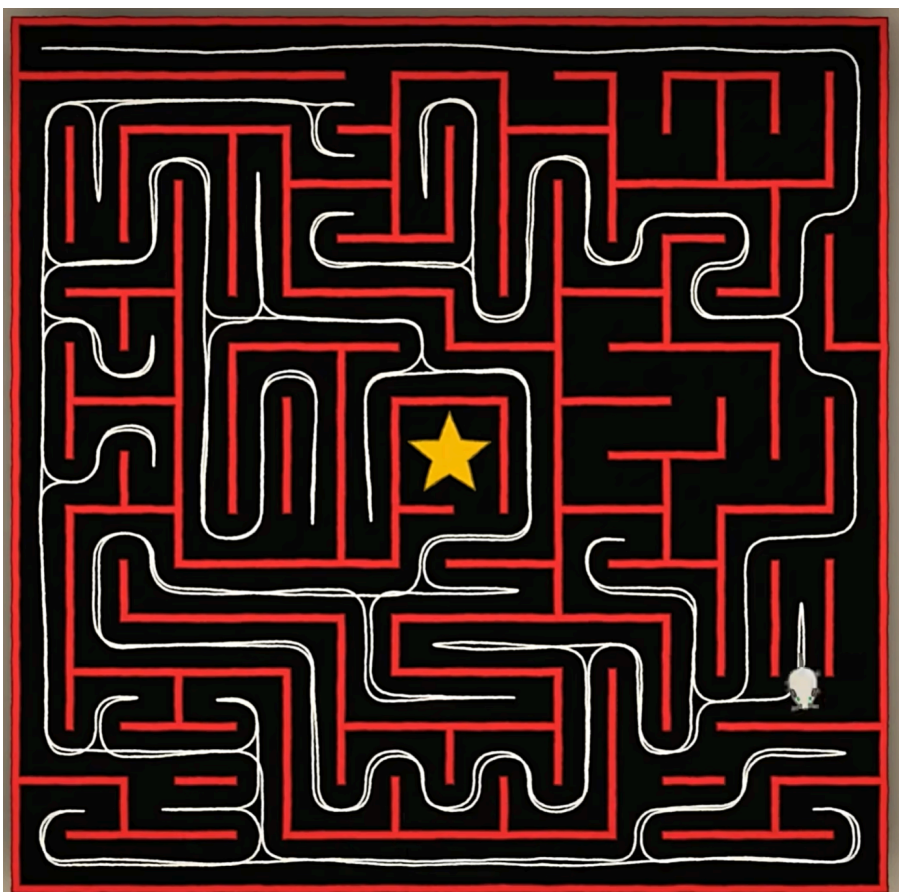
Example: 寻路



- 假设迷宫左上角有一只小鼠，如何寻找一条最短/最快路径能够到达迷宫的中心？

Example: 寻路

- 一个很容易想到的方法就是穷举法，通过遍历整个迷宫的所有路径，从而寻找出最短/最快的路径。
- 经典的遍历算法:深度优先遍历(DFS)和广度优先遍历(BFS)



遍历太低效了，能不能优化一下？

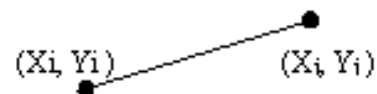
- 高中我们普遍用的是欧式距离

$$d = \sqrt{(x - x_0)^2 + (y - y_0)^2}$$

- 曼哈顿距离和欧氏距离的意义相近，也是为了描述两个点之间的距离，不同的是曼哈顿距离只需要做加减法，这使得计算机在大量的计算过程中代价更低，而且会消除在开平方过程中取近似值而带来的误差

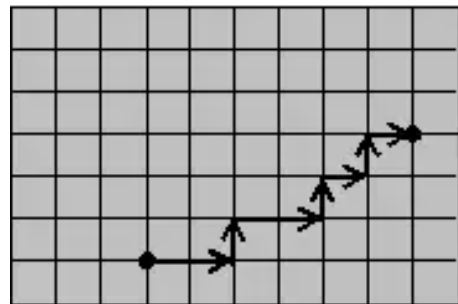
$$d = |x - x_0| + |y - y_0|$$

- 事实上，在走迷宫的过程中小鼠只能往上下左右四个方向走，因此曼哈顿距离是衡量迷宫中两点距离的更好的方式
- 由曼哈顿距离衍生出的**A*算法**能以更高效的方式找出一条相对短的路径



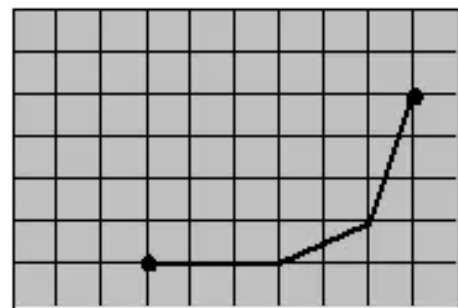
欧式距离

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$



曼哈顿距离

$$d = |X_2 - X_1| + |Y_2 - Y_1|$$



非欧式距离

$$d = [(X_2 - X_1)^{0.6} + (Y_2 - Y_1)^{0.6}]^{1/0.6}$$

算法是什么？

- 算法是为了解决特定问题而规定的一系列操作
- 算法与函数类似，都能接受输入产生输出

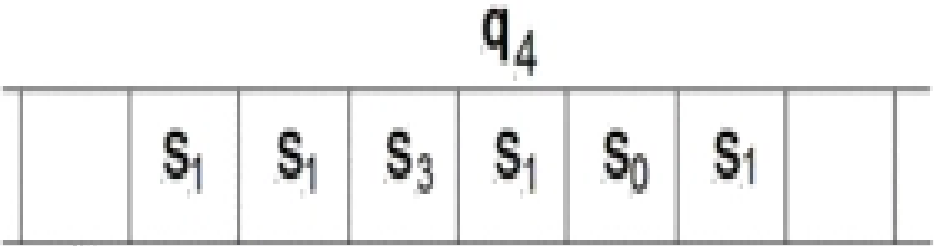
- 事实上，算法有复杂有简单，有些快有些慢
- 优秀的算法能够在较短的时间内找到问题的解，或者在相同的时间内处理更多的数据。
- 算法研究者们更关注算法的**渐进时间复杂度**（Big O表示法），即随着问题规模的增长，算法执行时间或所需资源的增长趋势。这是因为在实际应用中，问题规模往往很大，而渐进时间复杂度能够反映算法在大规模问题上的性能表现。
- 算法的时间复杂度只是评判算法优劣的一个相对指标，在实际应用中，我们还需要考虑其他因素，如算法的具体实现细节、计算机硬件的性能、数据的实际分布等
- 因此，在实际应用中，我们通常会通过实验来测试算法在特定条件下的性能表现，并选择一个在大多数情况下都能表现良好的算法

说了这么多，算法与AI到底什么关系？

- 早期的人工智能大多是通过固定指令来执行特定的问题，并不具备真正的学习和思考能力
- **AI在很大程度上是算法设计的结果**

传统AI的典型：图灵机

- 图灵机由图灵于1936年提出的一种抽象的计算模型，即将人们使用纸笔进行数学运算的过程进行抽象，由一个虚拟的机器替代人类进行数学运算
- 图灵把这样的过程看作下列两种简单的动作：
 1. 在纸上写上或擦除某个符号；
 2. 读写头从纸的一个位置移动到另一个位置。
- 而在每个阶段，人要决定下一步的动作，依赖于 (1) 此人当前所关注的纸上某个位置的符号和(2) 此人当前思维的状态。



读取	写入	移动
空白	无	无
0	1	向右移
1	0	向右移

AI在很大程度上是算法设计的结果

- 早期的AI其实就是人机，在人工设计的算法下能够简单的执行一些命令
- 那它就不是我们现在常说的AI吗？

传统AI与机器学习

传统AI

- 它需要一些输入和一些代码形式的逻辑，并为您提供输出



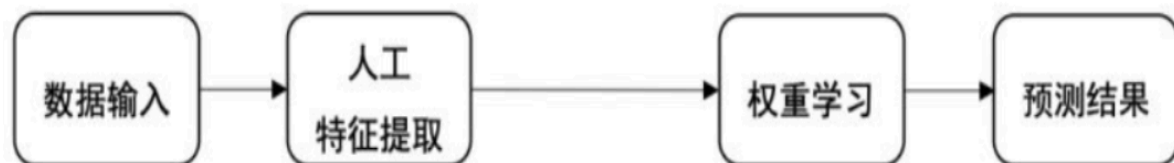
- 传统算法基于算法中描述的步骤产生输出。给出算法输入，它根据您自己硬编码的规则和参数生成输出。

传统AI与机器学习

机器学习

- 机器学习是一门关于数据学习的科学技术，它能帮助机器从现有的复杂数据中学习规律，以预测未来的行为结果和趋势。
- 例如：当我们在淘宝购物时，机器学习算法会根据我们的购买历史来推荐可能会喜欢的其他产品，以提升购买概率

经典的机器学习过程



机器学习的简单原理

- 假如我们正在教小朋友识字（一、二、三）。我们首先会拿出3张卡片，然后便让小朋友看卡片，一边说“一条横线的是一、两条横线的是二、三条横线的是三”。



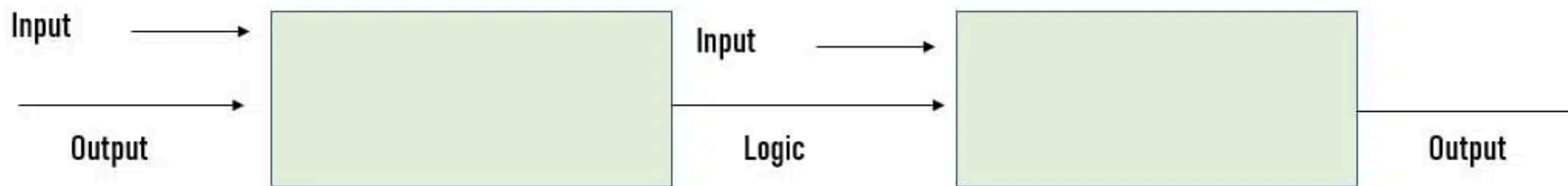
- 不断重复上面的过程，小朋友的大脑就在不停的学习：
- 当重复的次数足够多时，小朋友就学会了一个新技能——认识汉字：一、二、三

机器学习的简单原理

- 我们用上面人类的学习过程来类比机器学习。机器学习跟上面提到的人类学习过程很相似。
- 认字的卡片在机器学习中叫——训练集(training set)
- “一条横线，两条横线”这种区分不同汉字的属性叫——特征(feature)
- 学会了识字后总结出来的规律叫——模型(model)
- 通过训练集，不断识别特征，不断建模，最后形成有效的模型，这个过程就叫“机器学习”

机器学习

- 需要输入和输出，并会根据输入和输出生成一些逻辑，然后可以使用这个新逻辑来处理新输入以提供输出。



- ML算法基于通过提供给它的输入进行学习来预测输出。通过输入进行的学习称为训练过程。
- 给出要学习的算法数据，并调整参数来解释数据。然后可以使用这些参数集来解释/预测新数据