进击的AI：从打游戏到做高考题

hululu.zhu@gmail.com

07/02/2023

马上开始...

进击的AI

从打游戏到做高考题

少剑

hululu.zhu@gmail.com

07/02/2023

关于少剑同学

博士实习 + 毕业后在 Google + DeepMind 工作超过 11 年，共 8 个部门

平时喜欢做一些人工智能(AI) 的科普。

- 因为传递和分享知识是一种快乐，也是一种信仰。

声明

这个讲座里所有的内容均基于互联网和学术机构公开的论文、共享代码/模型、博客文章、社交媒体讨论和公开演示。和少剑的本职工作无直接联系。

本幻灯片中的所有观点均只是少剑他个人的观点，与 Google 或 DeepMind 无关

今日安排

- AI 和深度神经网络 DNN
- DNN 的 AI 如何“听说读写”
- 一些有趣的 AI 背后的原理
- 现在的 AI 的局限和挑战
- 讨论和问答

什么是 AI？

"AI" 是 Artificial（人工）Intelligence（智能）的缩写。

- "人工"：由人创造，比如你我
- "智能"：聪明的能力，比如听说读写

AI 有很多实现方法，比如人工预先编辑的行为代码，统计或遗传方法，或者神经网络。但是，2023年的"AI"

- 90%以上的情况下（包括我们的讲座），AI 都特指基于"深度神经网络 DNN"的实现

什么是 AI？

"AI" 是 Artificial（人工）Intelligence（智能）的缩写。

- "人工"：由人创造，比如你我
- "智能"：聪明的能力，比如听说读写

AI 有很多实现方法，比如人工预先编辑的行为代码，统计或遗传方法，或者神经网络。但是，2023年的"AI"

- 90%以上的情况下（包括我们的讲座），AI 都特指基于"深度神经网络 DNN"的实现

那么，什么是深度神经网络DNN？

深度神经网络 DNN

- DNN = Deep（深度）Neural（神经）Network（网络）
- 受到人类大脑神经元结构的启发而设计
  - 多：有很多的神经元，得到输入后进行复杂的计算，然后汇总
  - 深：很多神经元表示为一个“层”，深度网络就是有很多层的神经元的组合
  - 部分激活：模拟脑神经元，不需要所有的神经元都完全参与每一次计算和决策，只有被激活的神经元需要参与下一层的信息传递
    - 想象一下，视觉神经元大概率不需要对声音或者味道做出反应

你到底在说什么？好吧，我们来看一个形象的示例

深度神经网络 DNN 游乐场

- 想象一个例子，我们有 X1 代表身高 和 X2 代表体重
- 我们想用 x1身高 和 x2体重 预测小朋友能不能过体育测试

谷歌 DNN 游乐场

深度神经网络 DNN 的 AI

- 通过给 DNN 输入大量的数据（*比如身高和体重*）和答案（*比如有没有通过体育考试*），DNN 可以逐渐学会正确的人工智能决策。
  - 人工：我们定义了问题，数据，和 DNN 的结构，以及训练的方法
  - 智能：DNN 通过反复训练来提高性能，实现聪明的预测

深度神经网络 DNN 的 AI

- 通过给 DNN 输入大量的数据（*比如身高和体重*）和答案（*比如有没有通过体育考试*），DNN 可以逐渐学会正确的人工智能决策。
  - 人工：我们定义了问题，数据，和 DNN 的结构，以及训练的方法
  - 智能：DNN 通过反复训练来提高性能，实现聪明的预测

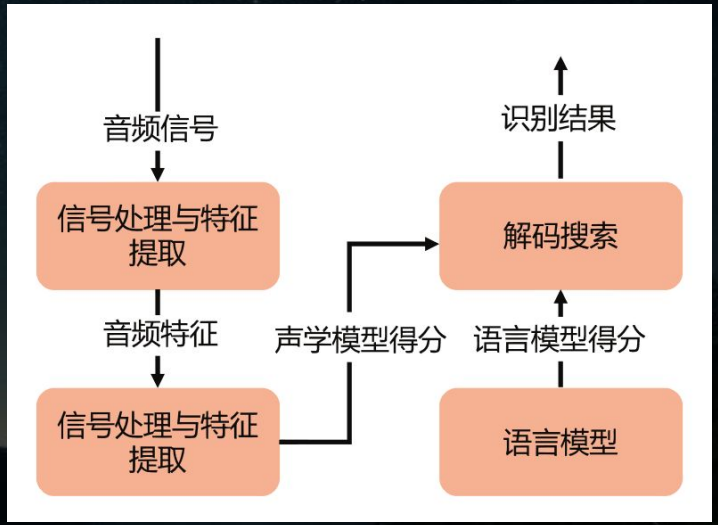
你知道吗？基于DNN 的 AI 可以"听说读写"（还有"看"和"画"）！

我们来认识一下！（注意，从现在开始，"AI" 都指代 "基于DNN的AI"）

AI 怎么“听”？

- 声音通过震动产生，在空气（或其他媒介）传播
- 通过麦克风，声音信号被转换为电流模拟信号
- 然后，通过模数转换器，电流模拟信号被转换为数字信号
- 数字信号就是AI能接收的数据
  - 有时候还需要一些预先的处理，比如对声音的频率、强度、声调的分析
- 同时，我们能告诉AI这些数据（声音）代表那些文字
- 然后，我们就可以“训练”语音识别的AI

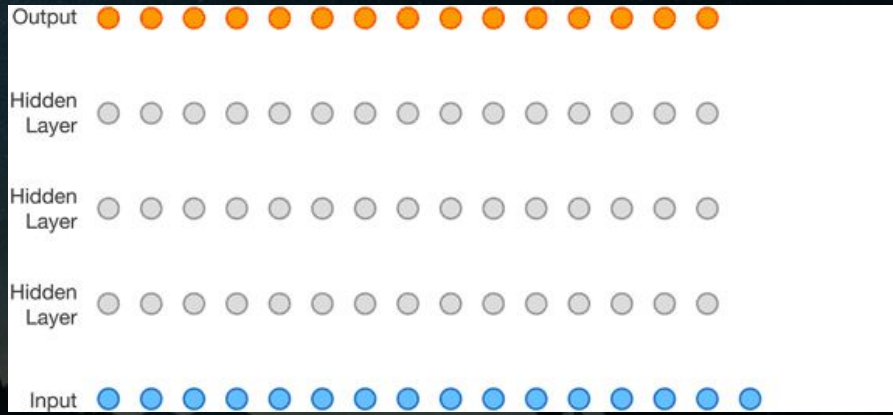
右侧的图是一个“经典”的语音识别的实现，但现在最新的AI几乎可以用一个网络架构（比如Transformer）实现所有的橙色部分。



语音识别技术概述 louwill12的博客

AI 怎么“说”？

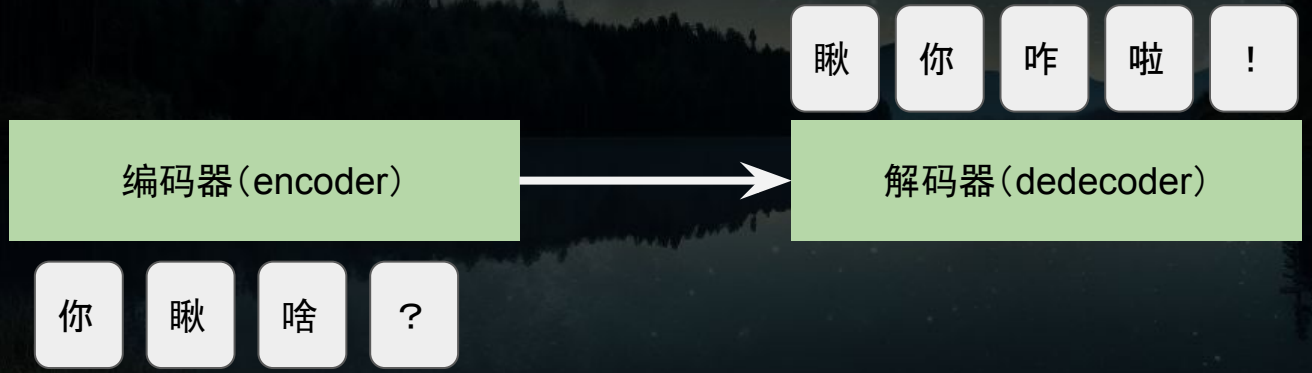
- 输入的数据就是要“说”的文字
- 需要的“答案”就是语音
- 比较简单的是类似“拼音”的方法
  - 把文字转化为音节和声调
  - 拼接为一段语音
  - 问题是比较生硬
- 直接输出声音非常难（昂贵）
  - 效果非常好
  - 更加自然
- Wavenet：在10ms一步的基础上，一步一步“说”



DeepMind WaveNet

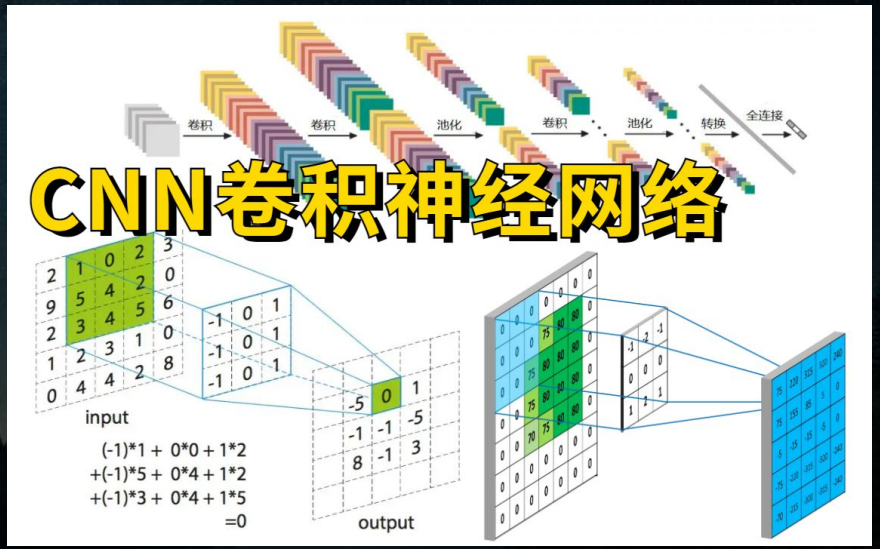
AI 怎么“读”和“写”？

- 我们合并了读和写，因为一般情况下两个任务是相连的
- 读：输入文字 + 写：输出文字
- 我们一般用一个“编码+解码”的架构
  - 编码（encoder）：输入文字串，理解语义
  - 解码（decoder）：在理解语义的基础上，输出文字



AI 怎么“看”？

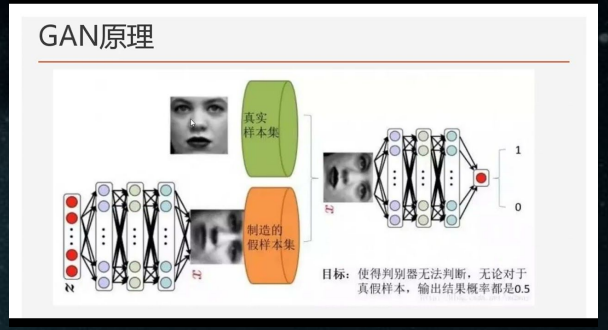
- 可以简化为 AI 怎么理解照片
  - 因为我们接收的外界视觉信息可以理解为连续的照片信息
- 首先，RGB代表红 Red，绿 Green，蓝 Blue
- RGB 以不同的强度混合在一起，就可以形成各种 颜色
- 而我们也可以把图片切成很多很小的块，每一块就是一个RGB混合色，在电脑或手机上，就是一个“像素”
- 组合起来，我们可以用类似1024 X 768 X 3个数字表示一个图片，1024 X 768是像素，3是RGB，传入AI作为数据
- 然后我们可以告诉AI这个图片的意义（比如文字）
- 用特殊的AI架构（比如卷积神经网络CNN或者视觉Tranformer ViT）来训练



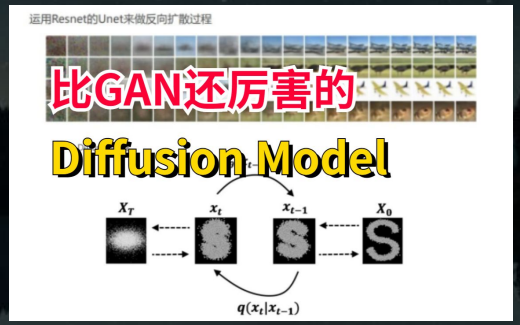
https://www.bilibili.com/video/BV16R4y1w7g4/

AI 怎么“画”？

- 输入的数据可以是文字或者图像或者其他（比如命题画，或者临摹画）
- 输出就是我们需要的图像表示，比如1024 X 768 X 3个数字代表的一个图片，1024 X 768是像素，3是RGB
- “画画”这几年有两个大的思路
  - 一个叫GAN（生成对抗网络）
  - 一个叫Diffusion（扩散模型）



https://www.youtube.com/watch?v=iTGW1V_B8mg



https://www.bilibili.com/video/BV1dY411o7of/

好了，AI 可以"听说读写看画"，然后呢？

- 现在的 AI 以及可以做很多非常酷非常难的事情
  - 玩游戏
  - 斗地主
  - 帮助科学研究
  - 做高考题

- 我们一起来看看 AI 怎么能做这些事情！

能玩超级玛丽的 DQN

曾经，通关超级玛丽是我的梦想

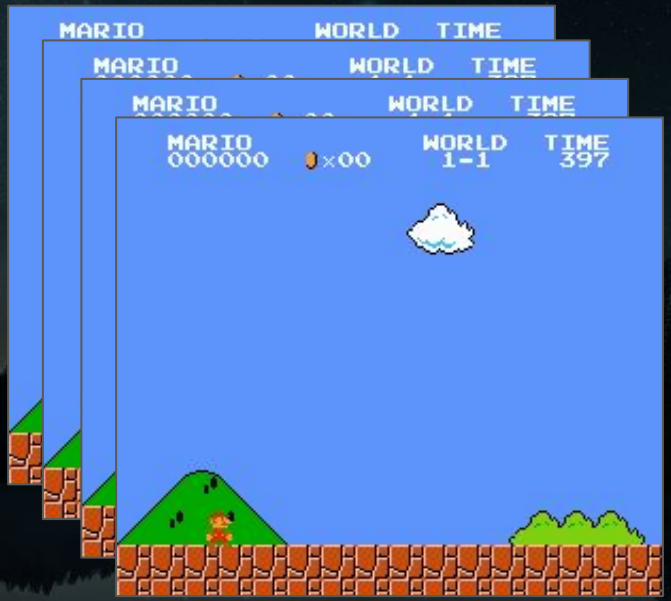
现在，只要2个小时训练，我就可以用我的 AI 帮我来实现！

- AI 怎么了解游戏状态
- AI 如何做出决策
- DQN 到底是什么

来自DeepMind

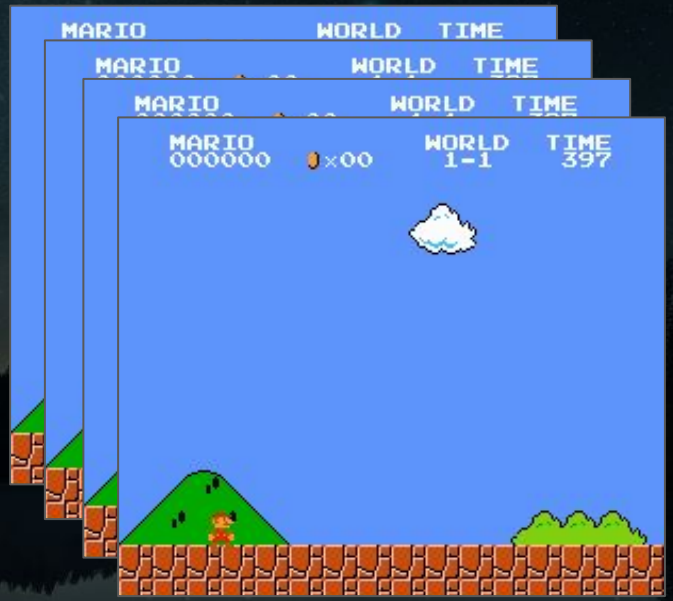
AI 怎么了解游戏状态？

- 游戏有图像和声音，为了简化，我们只需要图像信息
- 但我们需要将连续的图像传入 AI，而不是依靠静态的一张图片，为什么？



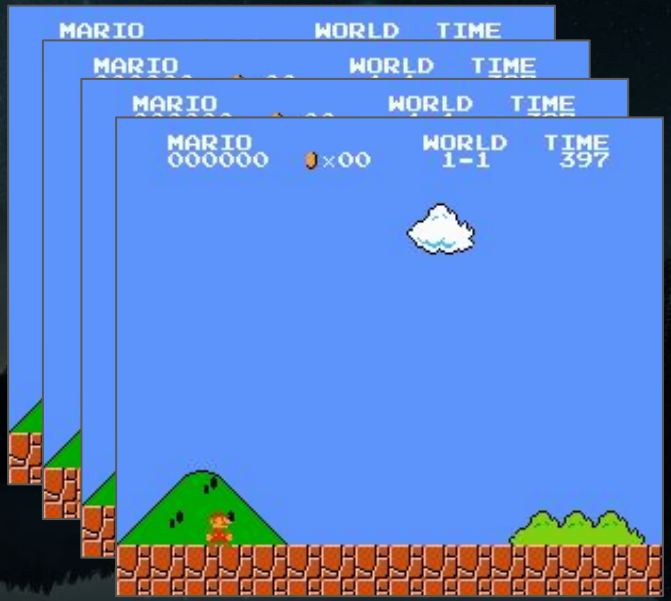
AI 怎么了解游戏状态？

- 游戏有图像和声音，为了简化，我们只需要图像信息
- 但我们需要将连续的图像传入AI，而不是依靠静态的一张图片，为什么？
- 答：因为如果只看一个图，你不知道马里奥和敌人的运动方向！连续的图像能告诉AI运动的轨迹



AI 如何做出决策

- 这里AI需要什么决策？
  - 每一个时间点，AI 要决定往前，还是往后，还是不动，还是跳，还是攻击。。
- 所以AI的数据是连续的游戏图像
- 我们还需要用一个方法告诉AI怎么去找到答案（往前还是其他）
  - 因为这里没有现成的答案！
  - 我们会使用DQN



DQN？

DQN = Deep Q Network（深度Q网络，Q是一个目标延迟回报函数）

DQN是强化学习的一种，强化学习是一个特殊的 AI 学习

- 简单来说，就如同我们在课堂学习知识准备高考
- 我们不会在学每一个科目或者做每一个题目的时候被告知每一步行为的"答案"
- 我们只会在未来考试的时候得到分数，来了解自己的学习
- 这里的Q可以理解为对高考分数的估计，而DQN的目的是为了找到平时的学习最佳策略，来最终实现高考分数Q的提高

DQN 如何帮助超级玛丽通关？

- 这里，Q网络可以理解为预估游戏的得分（或类似，比如不掉血通关）
- 给定游戏图像的输入
- DQN会尝试一个可能的策略（往前，往后，攻击等等），直到游戏结束
- DQN之后会学习（训练）
  - 学习成功的高分经验
  - 远离失败的低分经验
- 在训练很多次以后，DQN就能实现通关！
- 如果你想自己训练，可以参考[我的代码]

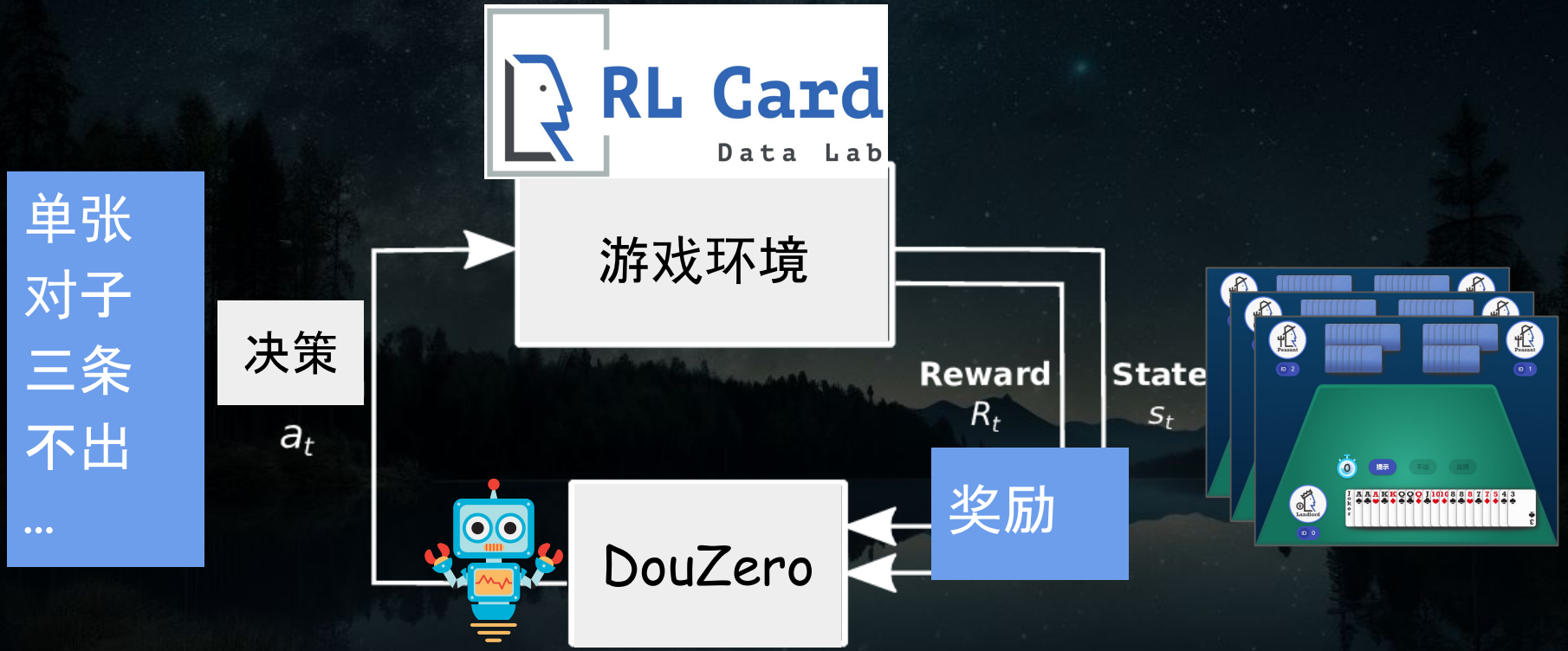
能打斗地主的 DouZero

来自 快手

- 超有影响力，因为全中国超过3亿人玩斗地主！
- 超多数据状态
  - 17 or 20 cards out of 54 (4.7e13 - 3.2e14)
- 超大"动作"（出什么牌）空间
  - 超过 2万多种出牌方式！
- 不完美信息（看不到对手牌）
  - 相比AlphaGo看到所有信息
  - 所以类似AI打星际或者王者荣耀，但是不处理图片
- 同时存在竞争和合作！
  - 地主要1v2
  - 农民要合作2v1

Action Type	Number of Actions
Solo	15
Pair	13
Trio	13
Trio with Solo	182
Trio with Pair	156
Chain of Solo	36
Chain of Pair	52
Chain of Trio	45
Plane with Solo	21,822
Plane with Pair	2,939
Quad with Solo	1,326
Quad with Pair	858
Bomb	13
Rocket	1
Pass	1
Total	27,472

DouZero 斗地主的数据和决策



DouZero的算法

- 每一家(地主，上家农民，下家农民)有自己的DouZero AI
- 用斗地主模拟器开始游戏
- 每一个DouZero根据牌局信息(手牌和台面上的)来选择出牌
- 玩到游戏结束，分数胜负
- 存储牌局信息，开始训练！
  - "赢者为王":对于赢下来的牌局(比如地主AI最后赢了，对于地主是赢，对于农民是输)，简单认为每一步赢盘都是"好"的，要深入，要学习
  - "输了，连呼吸都是错的":对于输掉的牌局，简单认为每一步都是"差"的，要总结，要远离
- 重新开始游戏，再总结，再训练
- 直到DouZero无敌

DouZero 是否真的这么厉害？

百闻，不如一试！

https://www.douzero.org/

能玩王者荣耀的 绝悟 AI

来自 腾讯

- 王者荣耀是一个5v5的实时对战游戏
- 可以类比之前两个游戏的AI
  - 和超级玛丽类似，输入连续的图像，选择动作策略
  - 和斗地主类似，只看到部分信息，存在合作和 竞争
- 更加复杂的问题
  - 王者一般要打20分钟以上，按照每一秒 20帧，就需要20 x 60 x 20 = 24000 步预测后，才知道结果！
  - 而每一步都只有100-200毫秒要做出判断！
  - 每一步游戏画面和动作的**复杂**度高很多

能玩王者荣耀的 绝悟 AI

来自 腾讯

- 王者荣耀是一个5v5的实时对战游戏
- 可以类比之前两个游戏的AI
  - 和超级玛丽类似，输入连续的图像，选择动作策略
  - 和斗地主类似，只看到部分信息，存在合作和 竞争
- 更加复杂的问题
  - 王者一般要打20分钟以上，按照每一秒 20帧，就需要20 x 60 x 20 = 24000 步预测后，才知道结果！
  - 而每一步都只有100-200毫秒要做出判断！
  - 每一步游戏画面和动作的**复杂**度高很多

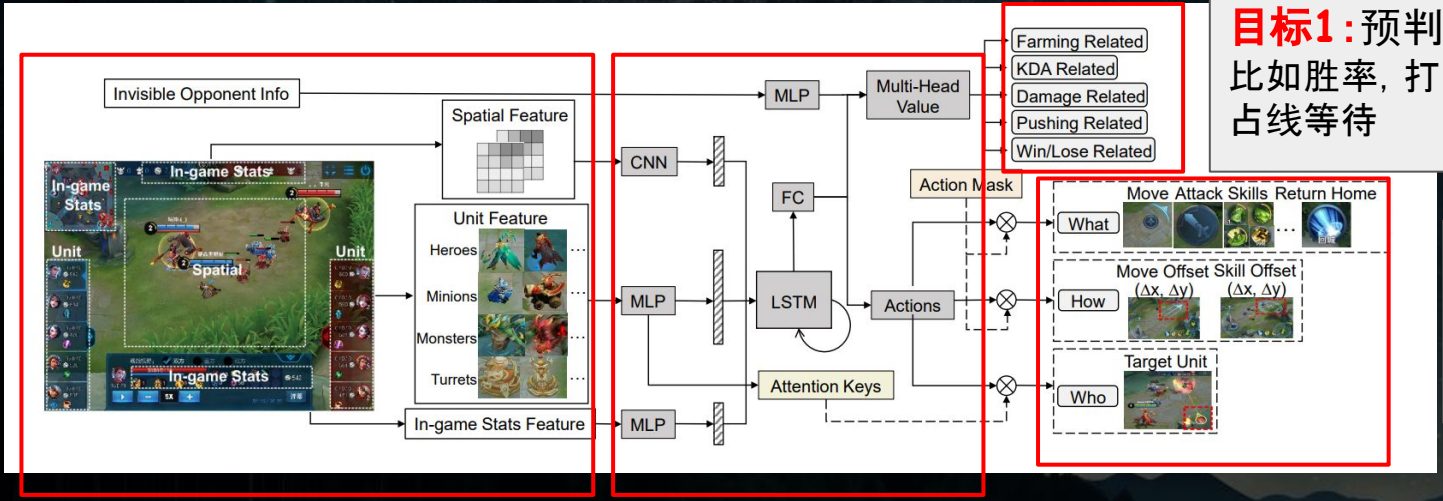
https://zhuanlan.zhihu.com/p/161751312

能玩王者荣耀的 绝悟 AI

- 首先，它是一个类似于AlphaStar（星际争霸）的架构
- 使用了很多最新研究的成果，包括
  - 围棋AlphaGo的MCTS
  - 课程学习（先易后难）
  - 强化学习（PPO算法，比DQN更强大更稳定）

我们来详细看一下它的基本架构和训练目标

基于强化学习自我博弈的训练目标



目标1：预判形势：
比如胜率，打钱，KDA，占线等待

目标2：预测
Who：什么目标
How：如何移动
What：什么动作

所有相关游戏信息的输入
- 但没有全局信息输入，故没有作弊

复杂的中间计算

能玩王者荣耀的 绝悟 AI

有哪些特点？我们来看一下这个评测文章

https://www.zhihu.com/question/391039689

进击的AI：从打游戏到做高考题

hululu.zhu@gmail.com

07/02/2023

休息一下，马上回来...

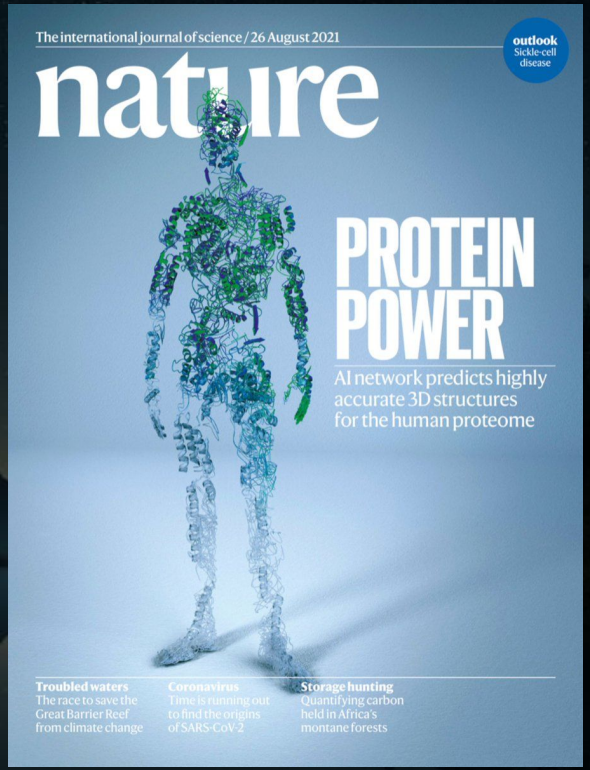
- 预测蛋白质结构的 AlphaFold
- 图片输出的 Stable Diffusion
- 做高考题的 ChatGPT

能预测蛋白质结构的 AlphaFold

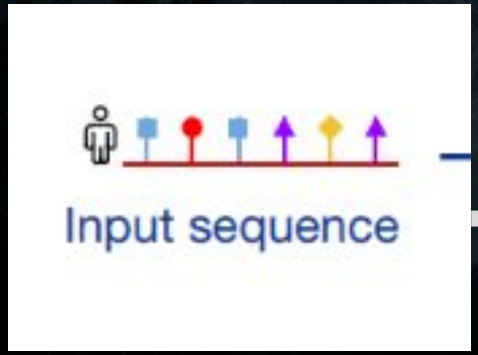
来自 DeepMind

- 蛋白质是我们身体里非常重要组成，它们帮助我们的身体运作正常。
- 研究发现蛋白质的结构（而非氨基酸的组成）极大影响蛋白质的功能，所以如果能预测蛋白质结构，将会帮助我们人类
- 我们几十年前就可以很测量蛋白质的组成序列，但我们很难去知道它的结构
    - 就好比，给你很多乐高积木，但不告诉你它能搭成的物体，你能不能用全部的乐高积木来搭一个稳定的结构？

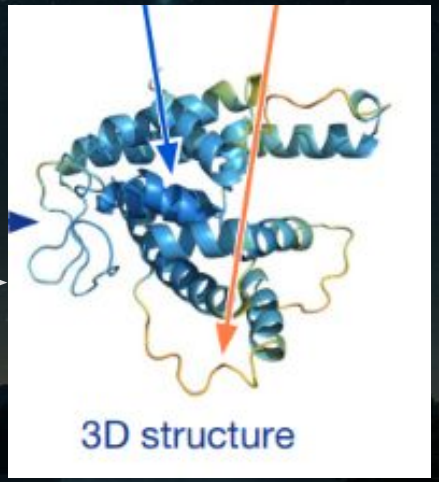
所以，这是一个困扰人类50年的问题，AlphaFold解决了一大块！



AlphaFold 要解决的具体问题？



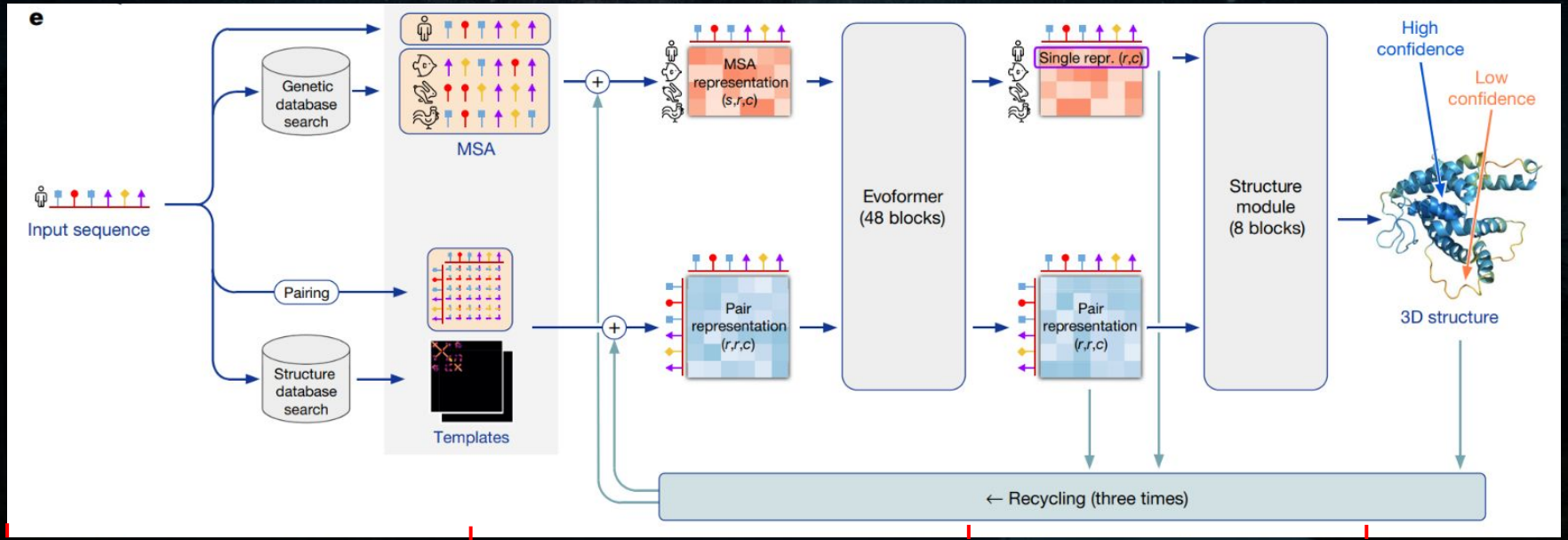
你行你上啊！



数据输入：
组成蛋白质的氨基酸序列

预测的"答案"：
蛋白质的3D结构组成

AlphaFold 的结构和方案



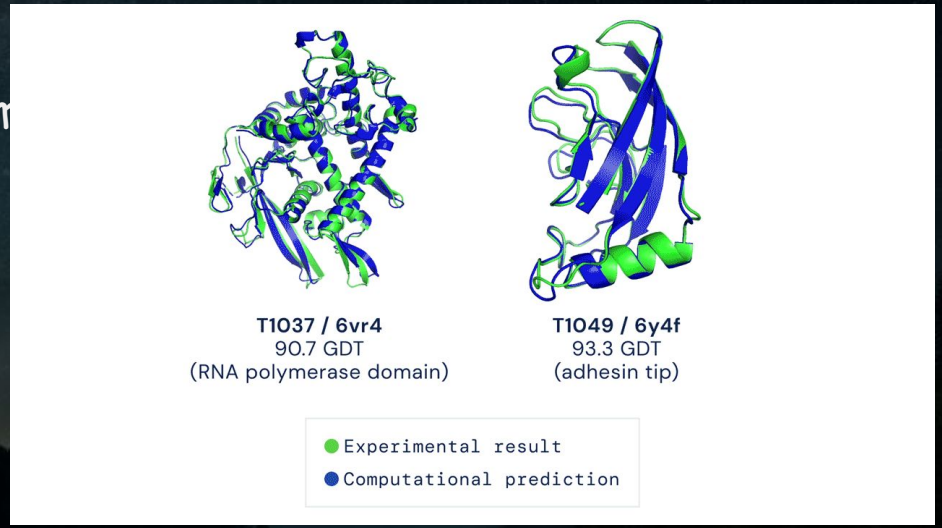
各种方法提取氨基酸碱基序列的特征

特征的融合和编码（encoder）

预测结构的解码（decoder）

AlphaFold 的成功

- 前人几十年数据和经验的积累
- 近几年DNN的突破，特别是Transform
  构的帮助
- 各种模型训练技巧的叠加
- 最多的计算资源的投入
- 多学科专家的合作
- 执着和信念和对社会的责任感
  - 地球超2亿蛋白质结构全预测，AlphaFold引爆「蛋白质全宇宙」-36氪

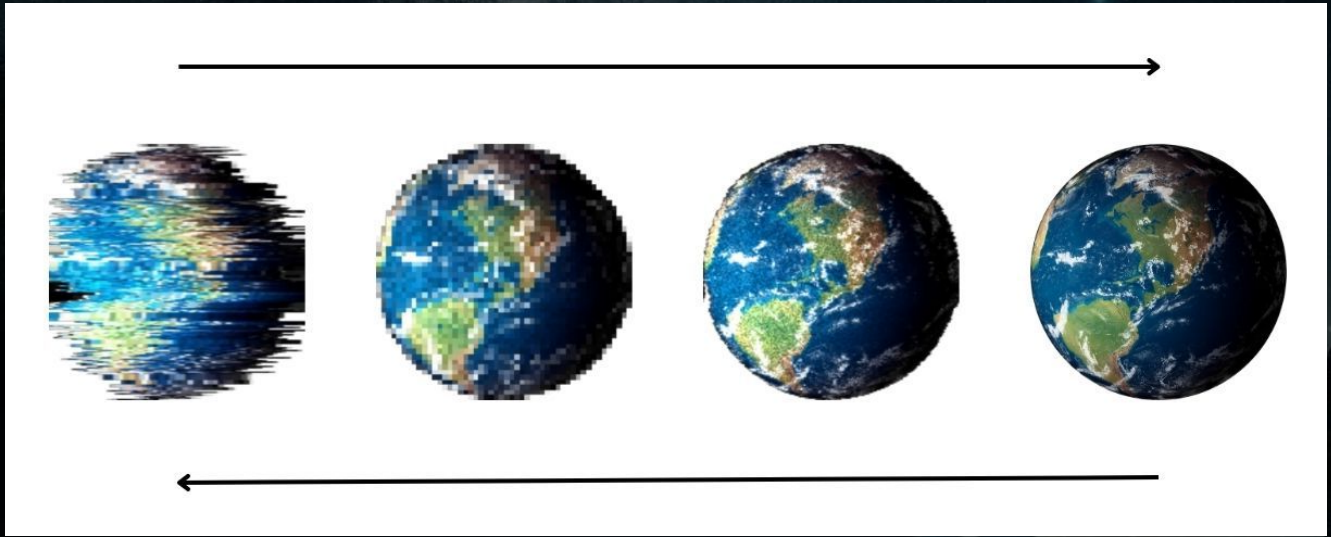


文字输出图片的 Stable Diffusion 来自Stability AI

- 这个一个多"模态"的问题
  - 输入文字(需要去"读"或者说"理解"), 所以需要一个文字 编码器Encoder
    - *我们在这里简化一下文字部分, 类似的chatgpt有更多相关内容*
  - 输出图像(需要去"画"), 所以需要一个 图像生成解码器
- Diffusion (扩散)是什么?
  - Diffusion Model (扩散模型)的简称
  - 为什么叫扩散?因为大致工作原理是先画 轮廓, 然后每一步"扩散"去描述更多细节(我们会给出例子)
- 那为什么叫stable diffusion?
  - 因为stable(稳定)是一个比较合适的形容这个工作的词

我们看一下Diffusion和它的逆过程

Diffusion：从模糊到清晰

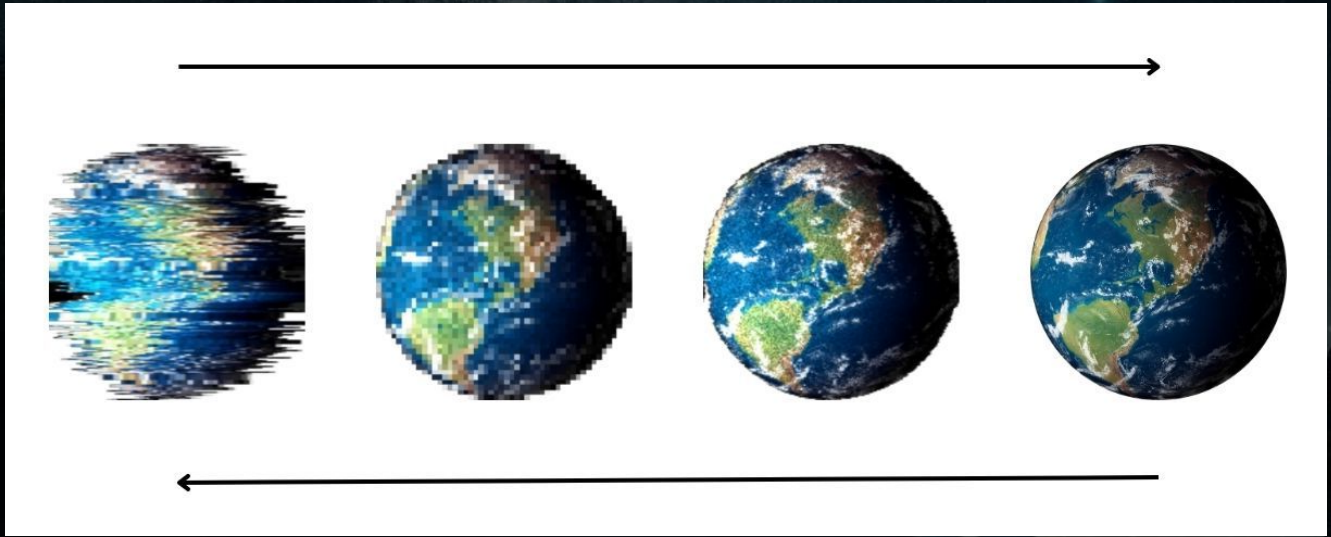


逆 Diffusion：从清晰到模糊

图片：https://renaissancerachel.com/understanding-diffusion-models/

我们看一下Diffusion和它的逆过程

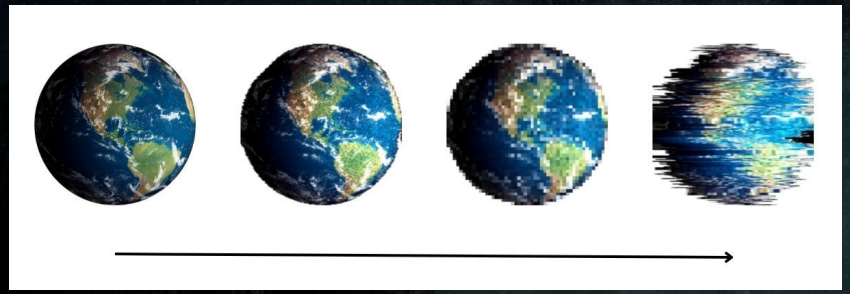
Diffusion：从模糊到清晰



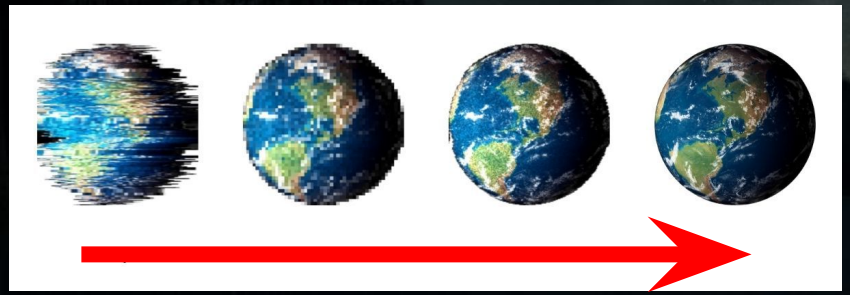
逆 Diffusion：从清晰到模糊

图片：https://renaissancerachel.com/understanding-diffusion-models/

Diffusion如何训练？



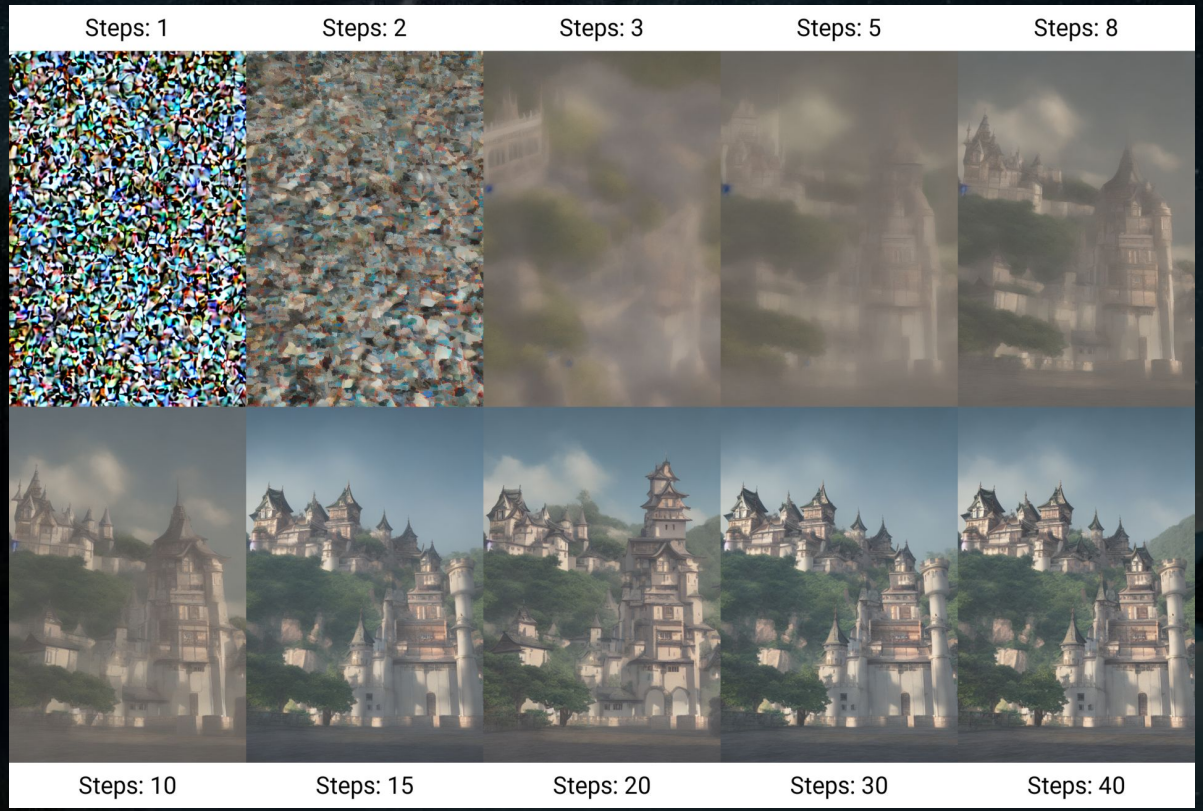
- 对每一个网络图片，我们可以很轻易的加入噪声让图片模糊



- 然后我们可以把上面图片的顺序反过来，每一次用模糊的图片，让模型预测更清晰的版本！

图片：https://renaissancerachel.com/understanding-diffusion-models/

看一下Stable Diffusion的例子



能答高考题的 ChatGPT 来自 OpenAI

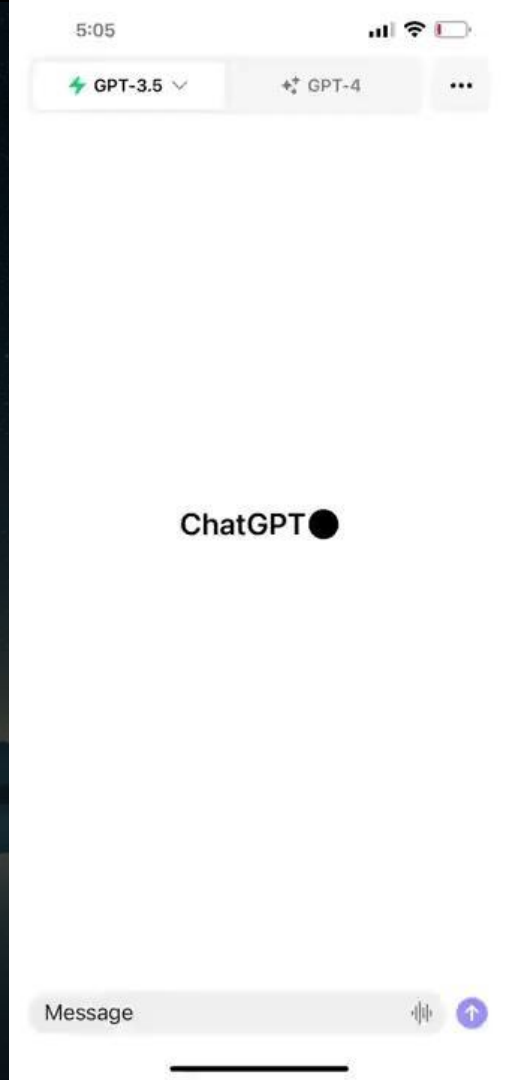
2023年全国卷语文作文：

吹灭别人的灯，并不会让自己更加光明；阻挡别人的路，也不会让自己行得更远。

“一花独放不是春，百花齐放春满园。”如果世界上只有一种花朵，就算这种花朵再美，那也是单调的。

以上两则材料出自习近平总书记的讲话，以生动形象的语言说出了普遍的道理。请据此写一篇文章，体现你的认识与思考。

要求：选准角度，确定立意，明确文体，自拟标题；不要套作，不得抄袭；不得泄露个人信息；不少于800字。



ChatGPT 模型是怎么训练出来的？

一共有4个步骤

预训练
Pretrain

通过大量的"续写下文"来让语言模型掌握各种"知识"的基础

- 语文：今天天气很好，小明说_____
- 数学：请分解 x^2 - 6x + 9，_____
- 编程：import numpy as np，_____
- 英文：Long long day, there is _____
- 经济：根据现在的行情，我们认为黄金_____
- 物理：动量定理是_____
- 政治：论为什么我们要批判胡锡进？_____
- 。。。

ChatGPT 模型是怎么训练出来的？

一共有4个步骤

监督微调
SFT
(supervised finetune)

掌握知识的模型还不知道如何根据人类(比如高考阅卷老师)喜好来交流(chat)，所以需要微调它的风格(style)，SFT(监督微调是第一步)

- 可以理解为我们收集"好学生"和"好老师"的对话
- 我们让模型学习这样的"好"的对话，从而记住这样的风格
- 这样，我们可以通过更好的交流风格或者方式来表述自己学到的知识

但是，这里知识去模仿学习一些对话，我们知道，单纯的模仿不能超越这些好同学或者好老师。如果我们想超越怎么办？

ChatGPT 模型是怎么训练出来的？

一共有4个步骤

监督微调
SFT
(supervised finetune)

掌握知识的模型还不知道如何根据人类(比如高考阅卷老师)喜好来交流(chat)，所以需要微调它的风格(style)，SFT(监督微调是第一步)

- 可以理解为我们收集"好学生"和"好老师"的对话
- 我们让模型学习这样的"好"的对话，从而记住这样的风格
- 这样，我们可以通过更好的交流风格或者方式来表述自己学到的知识

但是，这里知识去模仿学习一些对话，我们知道，单纯的模仿不能超越这些好同学或者好老师。如果我们想超越怎么办？

是的，我们可以用强化学习自我博弈来提高！

ChatGPT 模型是怎么训练出来的？

一共有4个步骤

奖励模型
RM (Reward Model)

在使用强化学习之前，我们需要一个类似于高考阅卷老师的自动评分系统！

- 因为只有通过这样的评分机制，我们才能知道自己做的好不好
- 然后我们才能提高！

为什么叫奖励模型？这只是一个管用的属于。你可以理解 为"评判模型"或者"打分模型"。

我们可以另外训练一个奖励（或者说评分）模型

- 输入：在单轮或多轮问答中对话得历史，比如学生和老师的问答文字
- 模型预测：用一个分数（比如0-1或者0-100）来判断答得好不好，比如作文写的怎么样，数学证明有没有正确

ChatGPT 模型是怎么训练出来的？

一共有4个步骤

基于人类反馈的强化学习

RLHF (Reinforcement learning from Human Feedback)

有了一个模型来自 动化"高考评分"或者类似的问答场景，我们就把

- 用"chatgpt自我提高来做高考 题"，复制到了
- 类似于我们"玩游戏要拿高分的场景"

简单说来

- 我们找一些常用的对话场景（比如高考题）
- 我们开启随机性，让模型回答多次，所以每次答案有不同
- 我们使用"评分模型"给所有问题打分
- 我们告诉ChatGpt，如果你要提高，就尽量多学 习你拿高分的案例，多远离你拿低分的案例！

ChatGPT 的训练步骤放在一起，高考冲冲冲

一共有4个步骤

预训练 *Pretrain*	监督微调 *SFT* *(supervised finetune)*	奖励模型 *RM (Reward Model)*	基于人类反馈的强化学习 *RLHF (Reinforcement learning from Human Feedback)*
我要学习 基础知识	我要学习 基本答题风格	我要揣摩 评分规则	我要自我训练和评分 自我提高！
菜鸟	*入门*	*感悟*	*提高再提高*

浅谈 AI 的局限和挑战

- 最聪明的AI也缺乏基本生活常识
- 普遍缺乏情感和价值观
- 训练更多智能和昂贵代价的取舍
- AI 对于白领和蓝领工作的巨大冲击
- AI 教育的缺乏和落后(更新太快)
- AI 的垄断
- AI 的监管

很快的回顾

- AI 和深度神经网络 DNN
- DNN 的 AI 如何“听说读写看画”
- 一些有趣的 AI 背后的原理
- 浅谈AI 的局限和挑战

讨论时间，欢迎问题，批评，想法，建议

进击的AI：从打游戏到做高考题

少剑
hululu.zhu@gmail.com
07/02/2023