

Python编程与人工智能应用

南京邮电大学计算机学院

张伯雷

<https://bolei-zhang.github.io/course/python-ai.html>

为什么学习本门课程



- Python

- 简便
- 强大的社区
- 库与框架
- 多用途

[视频链接](#)

- 人工智能

- 数字化社会（第三次工业革命）
- 各个行业都需要人工智能来实现更自动、智能的应用
- 探索智慧的原理和极限

- Python与人工智能

- 数据科学：numpy, scipy, cvxopt..
- 机器学习：sklearn, xgboost...
- 深度学习：tensorflow, pytorch...

- Python语言比C语言容易学
 - 本校将Python语言作为文科专业的入门编程语言，而C语言一般只给工科专业开设
 - 美国的常春藤联盟基本都是以Python为入门
 - 在学习Python的过程中，可以忽略数组越界、内存回收、动态内存分配等细节，而更关注计算机科学算法的本质
- 有C语言基础，学习Python语言更加容易，因此本门课通过7个学时讲解Python语言，其它课时讲解人工智能算法的Python实现
- 编程语言的基本思想都是相通的
 - 建议可以用“边用边学”的方式来学习

本门课程授课内容



- 第一章：人工智能与Python概述
- 第二章：Python语言基础
- 第三章：线性回归及Python实现
- 第四章：逻辑斯蒂分类及Python实现
- 第五章：K-近邻分类与K-均值聚类
- 第六章：朴素贝叶斯分类
- 第八章：神经网络及Python实现

课时计划表与学时分配



学时: 32=26(理论课)+6(3次实验), 2学分

章节	第1章	第2章	第3章	第4章	第5章	第6章	第8章
理论课时	2	7	5	4	2	2	4
实验 序号内容 及课时		1. (2学时) Python语言 基础实验	2. (2学时) 线性回归和逻辑斯 蒂分类实验				3. (2学时) 神经网络实验

- 教材

- 陈景强等. Python编程及人工智能应用, 南邮自编教材

- 参考书

- 薛景, 陈景强等. Python程序设计基础教程.人民邮电出版社, 2018年.
 - 周志华. 机器学习[M]. 清华大学出版社. 2017年.

- 中国大学MOOC

- 课程网址:

<https://www.icourse163.org/spoc/course/NJUPT-1464077167>

课程评价（研究性课程）

- 学时: $32 = 13(\text{理论课}) * 2 + 3(\text{实验}) * 2$
- 考试: 闭卷 成绩 = 期末 * 40% + 平时 * 60%（参考）
- 平时成绩:
 - 考勤
 - MOOC作业
 - 上机实验
 - 课堂提问

- 中国大学MOOC

- 浏览器网址访问:

<https://www.icourse163.org/spoc/course/NJUPT-1464077167>

- 昵称格式: NJUPT学号姓名, 如NJUPTB20030925徐震

- 认证为南邮学生才可访问

- 慕课堂

- 搜索微信小程序“慕课堂”

- 采用与中国大学MOOC相同的账号登录

- 扫描右侧二维码, 或通过课堂码“82FRZW”加入



- MOOC成绩由两部分组成：
 - (i) 单元测验
 - (ii) 编程作业：需要进行互评才有成绩；
- 注意每次作业提交的截止日期，过后不能再补

- 3次实验初步安排

- 实验一：Python语言基础实验

- 时间（待定）：计算机学科楼

- 实验二：线性回归和逻辑斯蒂分类实验

- 时间（待定）：计算机学科楼

- 实验三：神经网络实验

- 时间（待定）：计算机学科楼

- 实验要求

- 提前预习、设计代码

- 独立完成实验（编码、调试、运行成功、检查）

- 每次实验一份电子档报告，不交纸质

- 实验成绩根据实验准备、运行、报告等综合打分

Python编程与人工智能应用

南京邮电大学计算机学院

张伯雷

<https://bolei-zhang.github.io/course/python-ai.html>

- 人工智能起源与发展
- 人工智能的核心概念
- 人工智能的分支领域
 - 机器学习与深度学习
 - 应用分支领域
- 人工智能的行业应用
- 人工智能的就业前景
- Python人工智能开发环境安装

人工智能的定义



- 人工智能（Artificial Intelligence），简称为AI
- 是一门研究、开发用于模拟、延伸和扩展人类智能的理论、方法、技术及应用系统的技术科学，已成为计算机科学的一个重要分支。
- 根据维基百科的定义，人工智能所谓的“智能”，指的是机器的智能，区别于人和动物展示的自然智能。
- 人工智能是以人类行为标准进行猜想、假设、研究、实验而造就出来的一种类人类智能。

人工智能的起源

- 二十世纪五六十年代被正式提出
- 三个重要人物，“人工智能之父”：
 - 马文·明斯基 (Marvin Minsky)，建造了世界上第一台神经网络计算机
 - 阿兰·图灵 (Alan Turing)，提出“图灵测试”
 - 指测试者与被测试者（一个人和一台机器）隔开的情况下，通过一些装置（如键盘）向被测试者随意提问。进行多次测试后，如果机器让平均每个参与者做出超过30%的误判，那么这台机器就通过了测试，并被认为具有人类智能
 - 约翰·麦卡锡 (John McCarthy)，提出“人工智能”一词
- 2006年达特茅斯会议



图1.1 2006年达特茅斯会议50周年纪念，当事人重聚达特茅斯（左起：摩尔、麦卡锡、明斯基、赛弗里奇、所罗门诺夫）

第一次高峰和低谷



- 人工智能的第一次高峰：在1956年的达特茅斯会议之后，人工智能迎来了第一段发展高峰期,计算机被广泛应用于数学和自然语言领域，用来解决代数、几何和英语问题
- 人工智能的第一次低谷：在20世纪70年代，人工智能进入了一段痛苦而艰难岁月。
 - 计算机性能不足
 - 问题的复杂性
 - 数据量不足

第二次高峰和低谷



- 人工智能的第二次高峰：1980年，卡内基梅隆大学为数字设备公司设计了一套名为XCON的“专家系统”。这是一种采用人工智能程序的系统，可以简单理解为“知识库+推理机”的组合，具有完整的专业知识和经验。
- 人工智能的第二次低谷：在维持了仅仅7年之后，这个曾经轰动一时的人工智能系统就被宣告结束。到1987年，苹果和IBM公司生产的台式机性能均超过了Symbolics等厂商生产的通用计算机。从此，专家系统风光不再。

第三次高峰



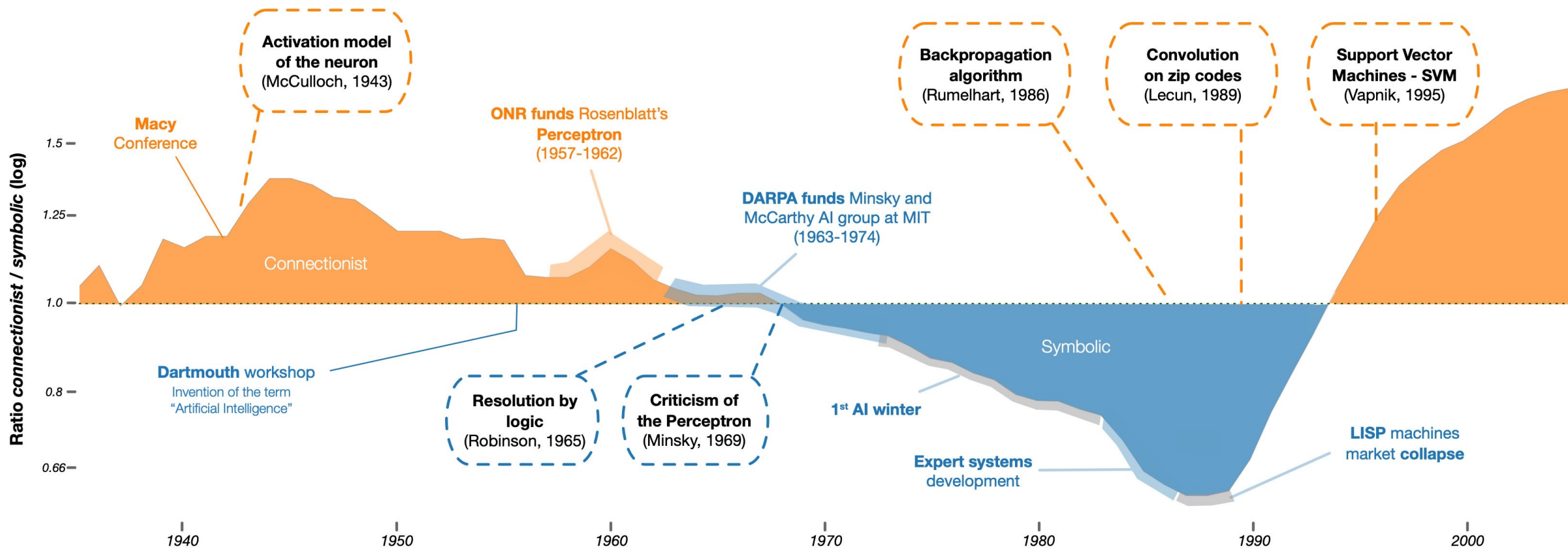
- 人工智能的第三次高峰：2006年Hinton等人提出的深度学习技术并获得成功，人工智能再一次掀起了全球的热潮。
 - 2011年，IBM开发的人工智能程序“沃森”（Watson Jeopardy）参加一档智力问答节目并战胜了两位人类冠军
 - 2015年，在ImageNet图像识别竞赛中，基于深度学习的人工智能算法在准确率方面第一次超越了人类肉眼，人工智能实现了飞跃性的发展
 - 2016年，AlphaGo战胜人类围棋冠军
 - 2017年，AlphaGoZero（第四代AlphaGo）战胜了在人类高手看来不可企及的第三个版本“大师”
 - 2022年，OpenAI开放了ChatGPT

人工智能的三大学派



- 符号主义 (Symbolism) :
 - 一种基于逻辑推理的智能模拟方法
 - 又称逻辑主义、心理学派或计算机学派。
 - 启发式算法→专家系统→知识工程，知识图谱
- 连接主义 (Connectionism) :
 - 又称仿生学派或生理学派
 - 是一种基于神经网络及网络间的连接机制与学习算法的智能模拟方法，比如深度学习
- 行为主义 (Actionism) :
 - 强化学习：“感知——行动”的行为智能模拟方法

人工智能发展历程



强人工智能与弱人工智能



- 强人工智能：
 - 具备与人类同等或者超越人类智慧，具有心智和意识，能根据自己意图开展行动
 - 类人的人工智能，即机器的思考和推理就像人的思维一样
 - 非类人的人工智能，即机器产生了和人完全不一样的知觉和意识，使用和人完全不一样的推理方式
 - 图灵测试
- 弱人工智能：面向特定的问题或任务
 - 计算机视觉
 - 自然语言处理
 - 语音识别等

- 混淆视听的测试

- 「你家养了猫？」
- 「哪里买的毯子？」
- 「你今天心情怎么样？」
- 「9566乘以55523等于几？」

- 「温诺格拉德」测试

“镇上的议员们拒绝给愤怒的游行者的游行许可——因为他们担心会发生暴力行为”——是谁在担心暴力行为？

A、镇上的议员们

B、愤怒的游行者的

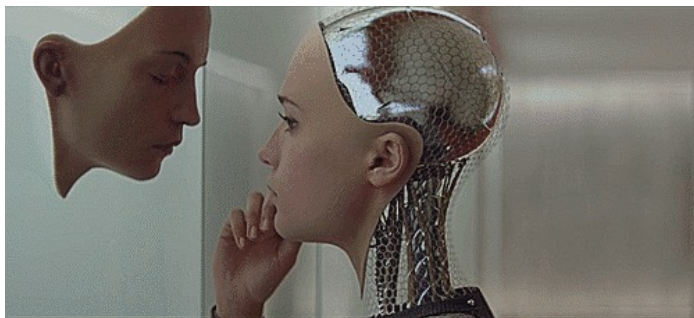
我：猫坐在毯子上，因为它很温暖。什么很温暖？

Siri：有趣的问题。

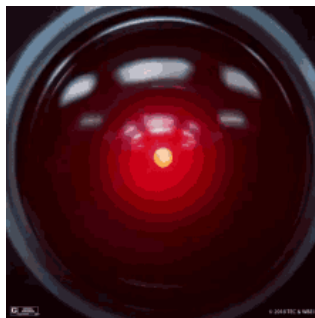
强人工智能与弱人工智能



- 我们以为的人工智能



机械姬



2001太空漫游

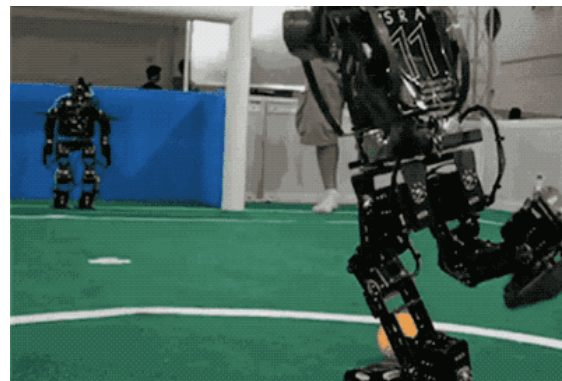


爱、死亡与机器人

- 实际中的人工智能



机械臂



机器人足球

强人工智能是否可实现?

- 南京大学周志华教授认为强人工智能是不可实现的
 - “从技术上来说，主流人工智能学界的努力从来就不是朝向强人工智能，现有技术的发展也不会自动地使强人工智能成为可能；即便想研究强人工智能，也不知道路在何方；即便强人工智能是可能的，也不应该去研究它。”
- 北京大学黄铁军教授则认为强人工智能是可实现的
 - “强人工智能是可以做出来的，而且可以在有生之年做出来，30年之内是可以做出来的；把最难的问题放在最后，不要把最难的问题放在第一步，如果放在第一步就是哲学家，就是争论，上来就开始给你很多的一套一套的，但是没用，有用是你去实践，去做。”
- “我们不能既承认人类智力的优越性，又否认可以通过人类智力实现人工智能”

- 机器学习 (Machine Learning) :
 - 从**已知数据**中**学习**其中蕴含的规律或者规则并利用这些规律和规则**推广到未来的新数据**
 - 按方法分类, 机器学习可分为
 - 有监督学习: 需要大量有标注数据
 - 无监督学习: 无需标注, 但是计算难度很大
 - 半监督学习: 为小部分样本提供真实标注
 - 按任务分, 机器学习可分为
 - 回归 (有监督): 预测的数据对象是连续值。**教材的第3章**
 - 分类 (有监督): 预测的数据对象是离散值。**教材的第4、5、6、7章**
 - 聚类 (无监督): 在数据中寻找隐藏的模式或分组。**教材的第5章**
 - 线性回归、逻辑斯蒂分类、K近邻分类、K均值聚类、贝叶斯分类、决策树

- 深度学习 (Deep Learning) :
 - 深度学习是一种实现机器学习的技术，并不是一种独立的学习方法
 - “深度”的含义
 - 深度等于更大规模的网络
 - 深度等于更抽象的特征
 - 深度学习模型是以人工神经网络为基础
 - 人工神经网络训练速度慢、容易过拟合、经常出现梯度消失以及在网络层次比较少的情況下效果并不比其他算法
 - 随着计算性能的提升和互联网时代数据量的爆炸式增长，深度学习效果大幅提升。
 - 三大要素：硬件、数据、算法



深度学习“三巨头”，2018年图灵奖获得者

机器学习与深度学习



- 人工智能、机器学习、深度学习的关系：
 - 机器学习是一种实现人工智能的方法
 - 深度学习是一种实现机器学习的技术
 - 机器学习和深度学习不是人工智能的全部
- 三者关系



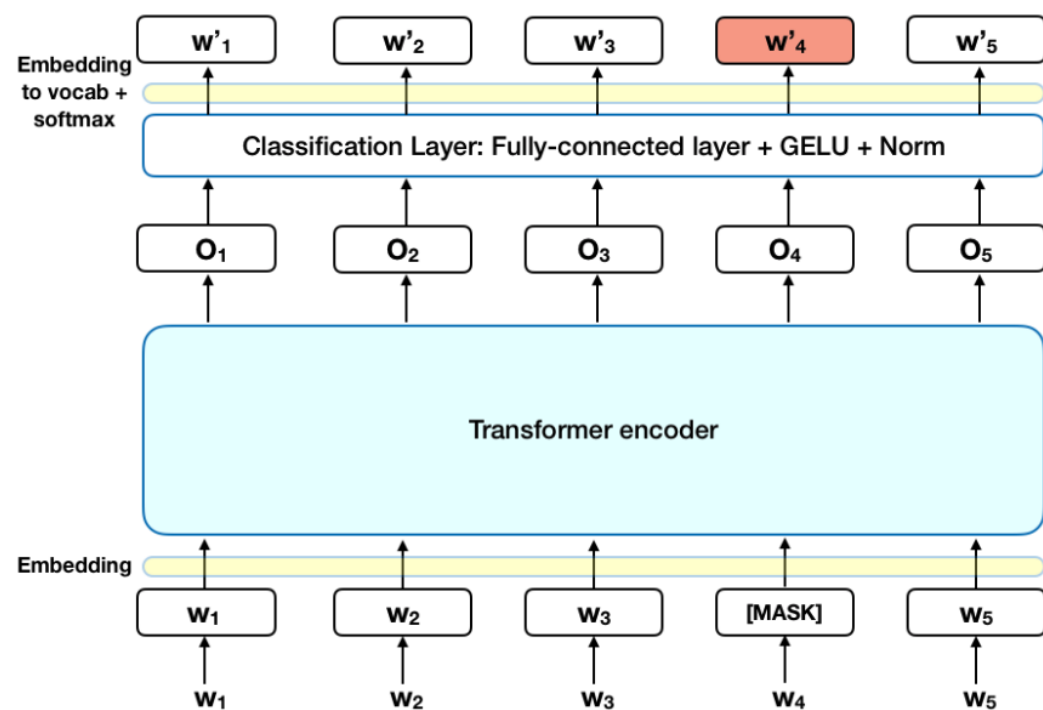
人工智能的应用分支领域



- 人工智能还从各个方面模拟人的智能。
- 人最基本的智能包括语言、视觉、听觉等，让机器分别拥有这些智能，便产生了人工智能技术的应用分支领域
 - 自然语言处理
 - 计算机视觉
 - 语音识别
 - 推荐系统
 -

- 自然语言处理

- Natural Language Processing, 简称NLP
- 利用计算机对自然语言进行智能化处理。
- 研究目标是让机器理解并生成人类语言, 用自然语言与人类平等流畅地沟通交流, 最终拥有“智能”
- 基础自然语言处理技术主要围绕语言不同层级展开
 - 包括音位、形态、词汇、句法、语义、语用、篇章
- 自然语言处理任务
 - 机器翻译、人机对话、自动问答、文档摘要等
- 目前, NLP的主流方法是基于深度学习的方法



计算机视觉和语音识别

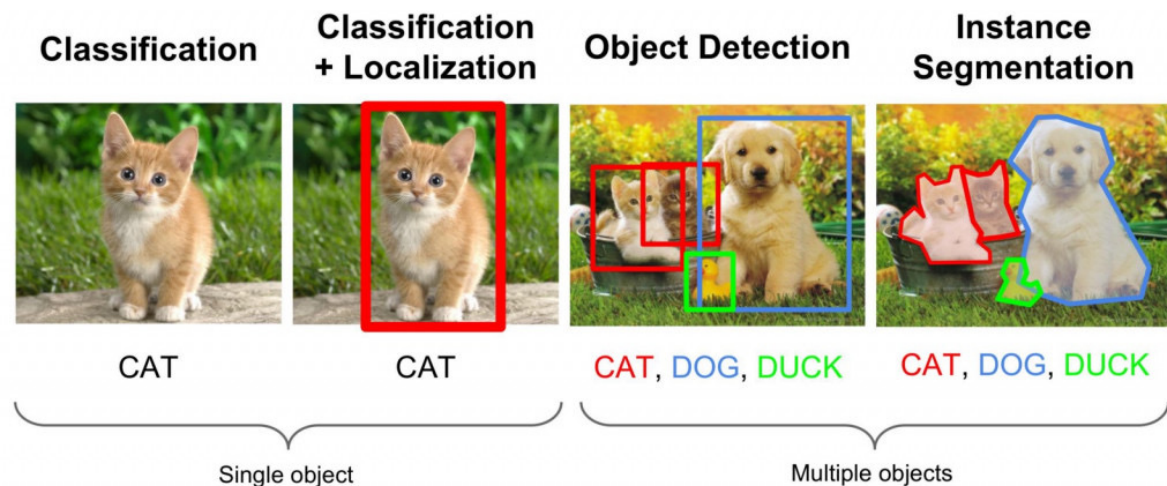
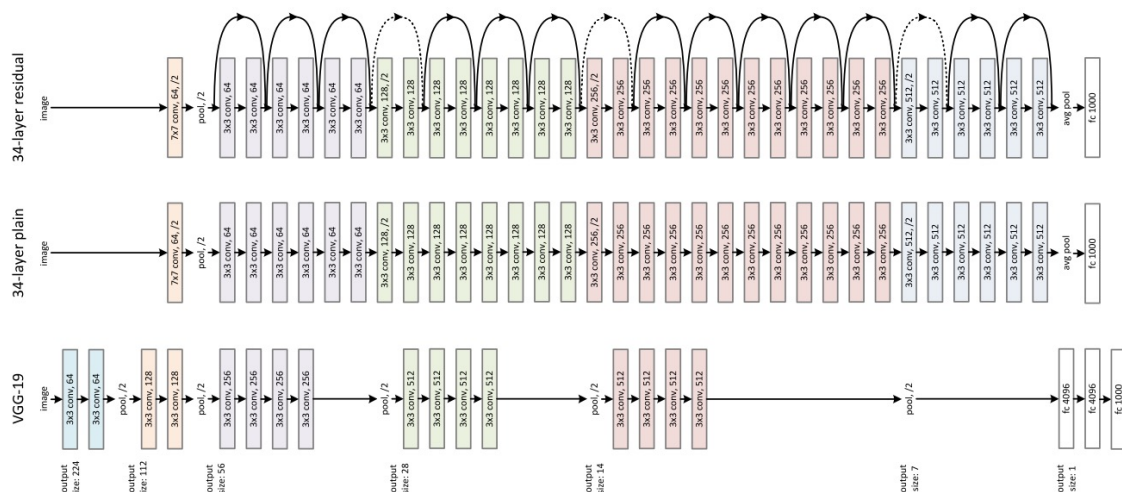


• 计算机视觉

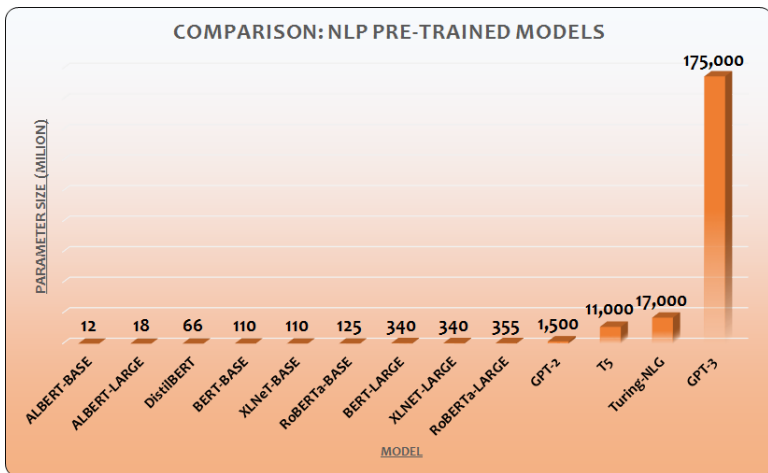
- Computer Vision, 简称CV。是人工智能领域中研究如何使机器“看”的科学
- 主要任务
 - 人脸识别、图片问答、图像识别、目标检测和跟踪等
- 当前计算机视觉的研究主要是基于深度学习技术
 - Fast RCNN、Faster RCNN、Yolo等

• 语音识别

- Automatic Speech Recognition, 简称ASR, 把人类的语音信号转变为想要的文本或者命令



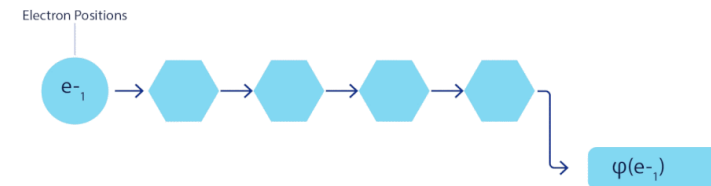
人工智能最新的进展



GPT-3: 1750亿参数，训练所用的数据量达到45 TB，训练费用超过1200万美元。

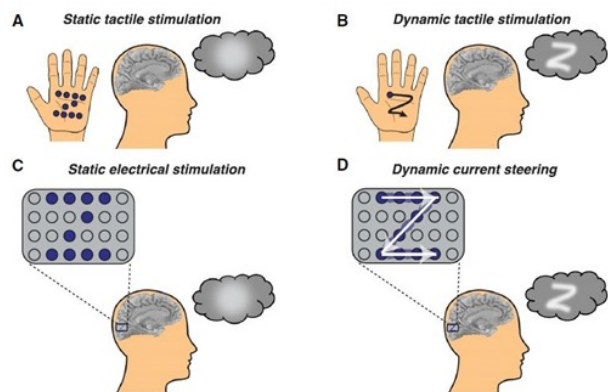


AlphaFold2: 其准确性可以与使用冷冻电子显微镜等实验技术解析的蛋白质3D结构相媲美



FermiNet: 来近似计算薛定谔方程，在精度和准确性上都满足科研标准

人工智能最新的进展 (2)



美国贝勒医学院利用动态颅内电刺激新技术，用植入的微电极阵列构成视觉假体，在人类初级视皮层绘制W、S和Z等字母的形状，成功地能够让盲人“看见”了这些字母。

SimCLR：能够在无标注数据上学习图像数据表征。无监督学习模型可以接近甚至达到有监督模型的效果。



BostonDynamics



这句话包含了几个字：今天要下雨了，你准备怎么办？



这句话包含了16个汉字。



上帝能创造出一块他搬不动的石头吗



根据一般的理解，上帝是一种超自然的存在，并不受物理定律的限制。因此，上帝是可以创造出一块他搬不动的石头，但这个问题的答案取决于人们对上帝的理解。例如，如果人们认为上帝是一种无限的、自我完善的存在，那么上帝是不可能创造出他自己无法搬动的东西的。



- AGI

- 人们把具备与人类同等智慧、或超越人类的人工智能称为通用人工智能 (artificial general intelligence)

- 1. 强化学习

- 《Reward is enough》
 - 该研究认为奖励足以驱动自然和人工智能领域所研究的智能行为，包括知识、学习、感知、社交智能、语言、泛化能力和模仿能力，并且研究者认为借助奖励最大化和试错经验就足以开发出具备智能能力的行为。因此，他们得出结论：强化学习将促进通用人工智能的发展

- 2. 自监督学习

- 自监督学习算法不依赖注释，而是通过暴露数据各部分之间的关系，从数据中生成标签，这一过程被认为是实现人类智能的关键。
 - Lecun: “我们作为人类学习的大部分知识和动物学习的大部分知识都是在自我监督的模式下，而不是强化的模式。基本上是通过观察世界，并与之进行一些互动，主要是以独立于测试的方式进行观察，这是我们不知道如何用机器复制的（学习）类型。”

- 3. 其它

人工智能不足的地方



- 小样本学习、零样本学习
- 推理
- 常识

人工智能行业应用



2020年中国人工智能产业图谱



向企

害

人工智能人才需求



- 我国人工智能产业规模大
 - 到2020年，中国人工智能产业规模超过1500亿元，带动相关产业规模超过1万亿元
- 需求企业薪资高
 - 百度、京东、谷歌、讯飞、阿里巴巴、腾讯



Python语言与人工智能

- Python之父： Guido van Rossum
- Python语言是一种“解释型”语言
 - Python需要一个解释器
 - 在运行的时候将程序翻译成机器语言。解释型语言的程序不需要在运行前编译，在运行程序的时候才翻译，专门的解释器负责在每个语句执行的时候解释程序代码。这样解释型语言每执行一次就要翻译一次，效率比较低
 - C语言是“编译型”语言
 - 两个不兼容的版本： Python2.x、 Python3.x
- Python
 - 动态类型
 - 脚本式语言
 - 引用计数
- Python语言被称为“人工智能语言”
 - NumPy、 SciPy、 Matplotlib
 - Scikit-learn
 - PyTorch、 Tensorflow
 - NLTK、 OpenCV、 SpeechRecognition



Python安装和运行



- 安装Python
 - 官方下载网址: <https://www.python.org>
 - 演示下载和安装最新版Python
- 运行Python
 - 交互模式
 - 打开命令行提示符界面, 输入python指令后
 - 出现“>>>”提示符, 输入语句print(“hello world”)
 - 文件模式
 - 新建文本文件, 改后者名为“.py”
 - 编辑文件内容为print(“hello world”), 通过右键菜单运行

- Pip指令
 - 安装指令: `pip install [-i 安装源] 包名称`
 - “-i 安装源”可指定国内源以提升下载速度
 - 清华大学安装源<https://pypi.tuna.tsinghua.edu.cn/simple>
 - 阿里云安装源<http://mirrors.aliyun.com/pypi/simple>
 - 豆瓣安装源: <http://pypi.douban.com/simple>
- 本教材需要用到的包:
 - NumPy、Matplotlib、Scikit-learn、Pytorch、Tensorflow
 - 使用pip指令一一安装演示

Python人工智能库安装



- 显示安装包信息
 - pip show 包名称
 - 如：pip show numpy
- 卸载包
 - pip uninstall 包名称
 - pip uninstall numpy

Python集成开发环境

- 集成开发环境
 - 简称IDE，一般包括代码编辑器、编译器、调试器和图形用户界面等工具，能够提升开发效率
- Python常用IDE
 - IDLE: Python自带，演示print(“hello world”)
 - JupyterLab: 基于Web的Python集成开发环境
 - 安装指令: `pip install jupyterlab`
 - 运行指令: `jupyter lab`
 - PyCharm:
 - 下载: <https://www.jetbrains.com/pycharm-edu/>
 - Vscode+命令行

- 安装Python人工智能开发环境需要先安装Python，再通过pip指令下载安装相应的库，最好再安装一款好用的集成开发环境
- Anaconda是一个Python的集成库，是一个开源的Python发行版本，包含了conda、Python等180多个科学包及其依赖项
- 下载地址： <https://www.anaconda.com/>

- 数据挖掘

- 从复杂网络（社交网络、分子网络）等挖掘关键的节点、信息
 - 细胞分类
 - 社交网络信息传播分析
 - 网络关键节点分析
- 技术手段
 - 图神经网络

- 强化学习

- 离线强化学习实现低代价模型训练，多智能体博弈强化学习
 - 机械臂、无人车
 - 动态定价
 - 多智能体博弈
- 技术手段
 - 强化学习
 - 博弈论

谢谢！