上海科技大学
ShanghaiTech University

CS286 AI for Science and Engineering

Lecture 1: Course Overview

Jie Zheng (郑杰)

PhD, Associate Professor

School of Information Science and Technology (SIST), ShanghaiTech University

Fall, 2023

立志成才 报国裕民

Outline

- A brief history of AI
- Learning objectives
- Instructors
- Prerequisites
- Learning plan
- Evaluation
- Computing platform
- Course materials and references
- Communication
- Academic integrity

What is Artificial Intelligence (AI)?

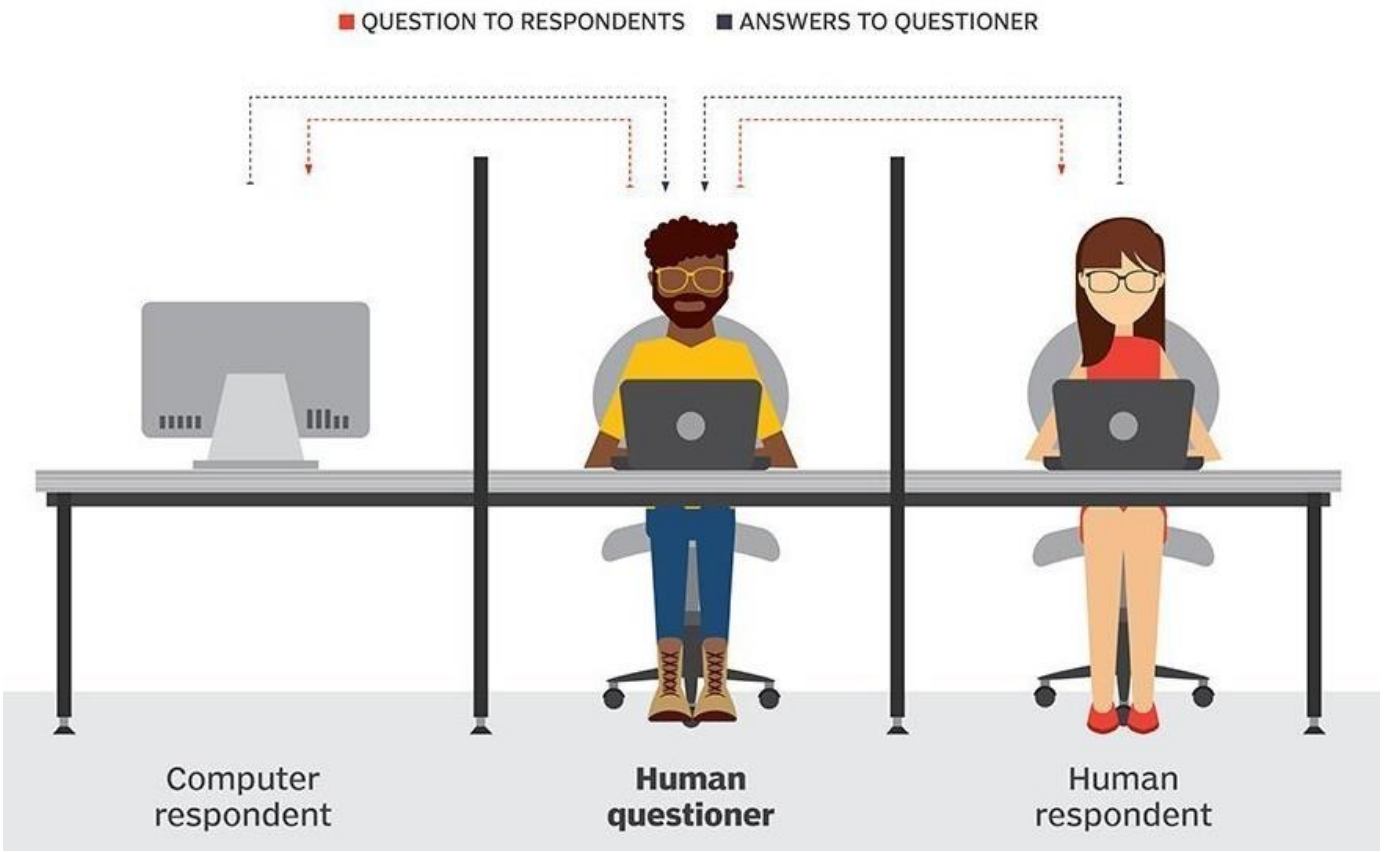
上海科技大学 ShanghaiTech University

Thinking Humanly	Thinking Rationally
"The exciting new effort to make computers think … *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning …" (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole et al., 1998) "AI … is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

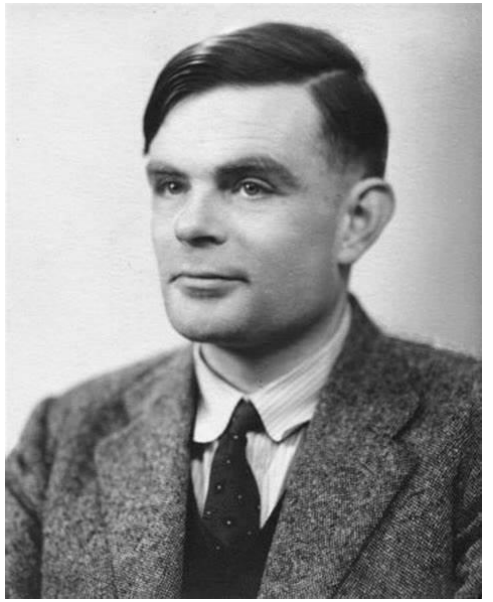
Source: Stuart Russell and Peter Norvig's book, "Artificial Intelligence: A Modern Approach" (3rd Ed.), 2010 (Fig. 1.1, page 2)

The Turing Test (1950)

- "Can machines think?" [Turing (1950), Computing Machinery and Intelligence]

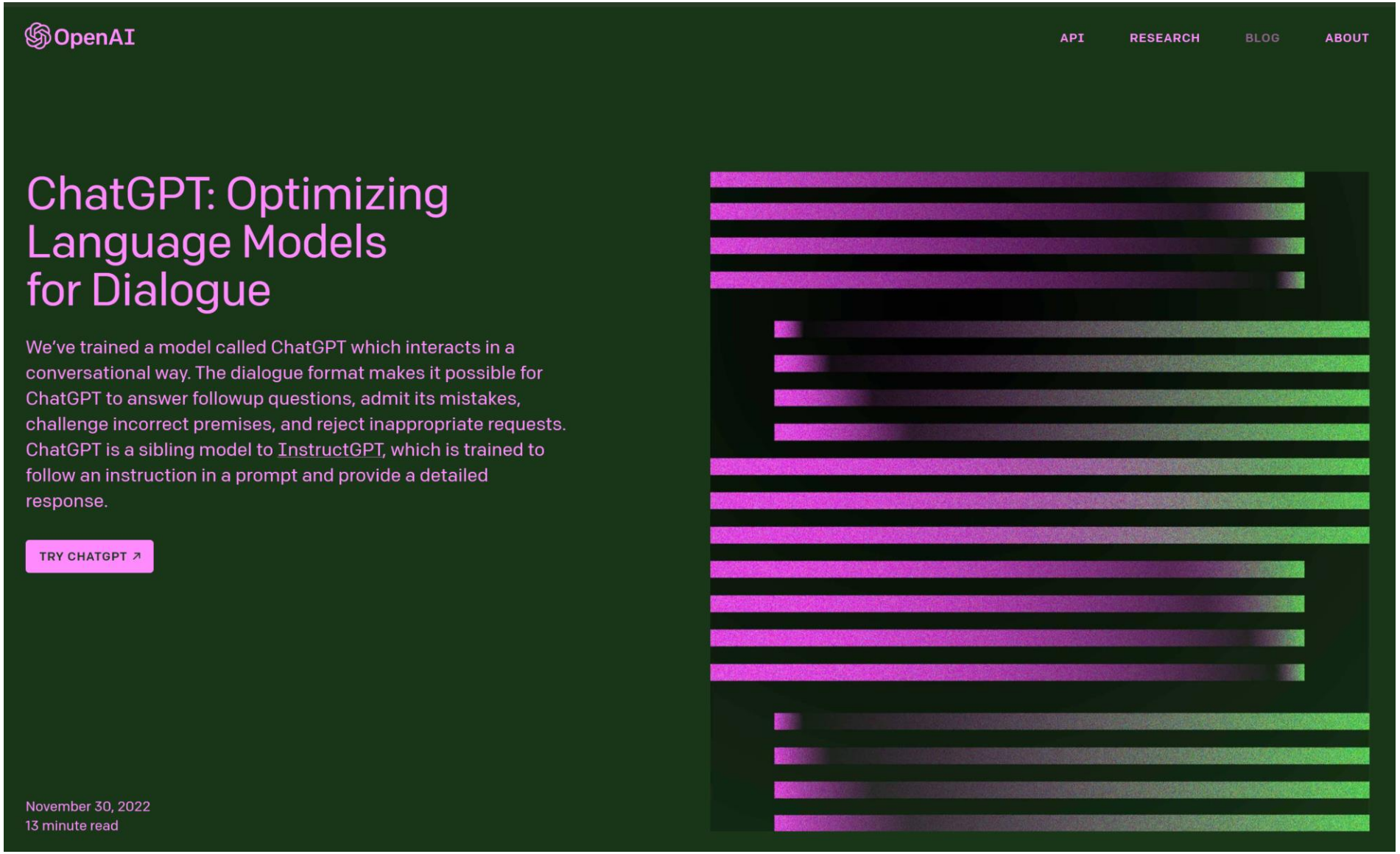


Source: WhatIs.com



Alan M. Turing
(1912 – 1954)
English mathematician, computer scientist, logician, cryptanalyst, philosopher and theoretical biologist

Has ChatGPT passed the Turing Test?



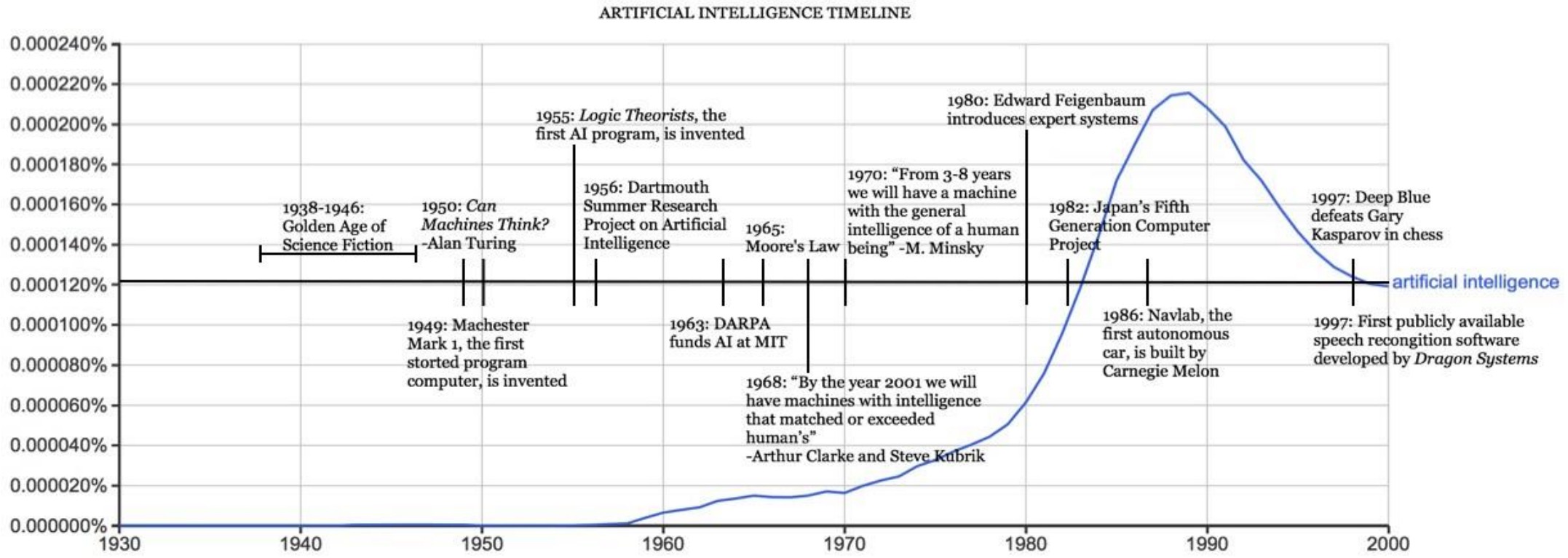
Birth of AI at Dartmouth College (1956)

- The Dartmouth workshop in the summer of 1956 gave birth to the field of AI
  - Attendees are founding fathers of AI: John McCarthy, Marvin Minsky, Claude Shannon, Herbert Simon, etc.
  - John McCarthy coined the term "Artificial Intelligence"
- The Dartmouth proposal:

  "**Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.**"

Dartmouth Hall
The birthplace of AI

AI timeline

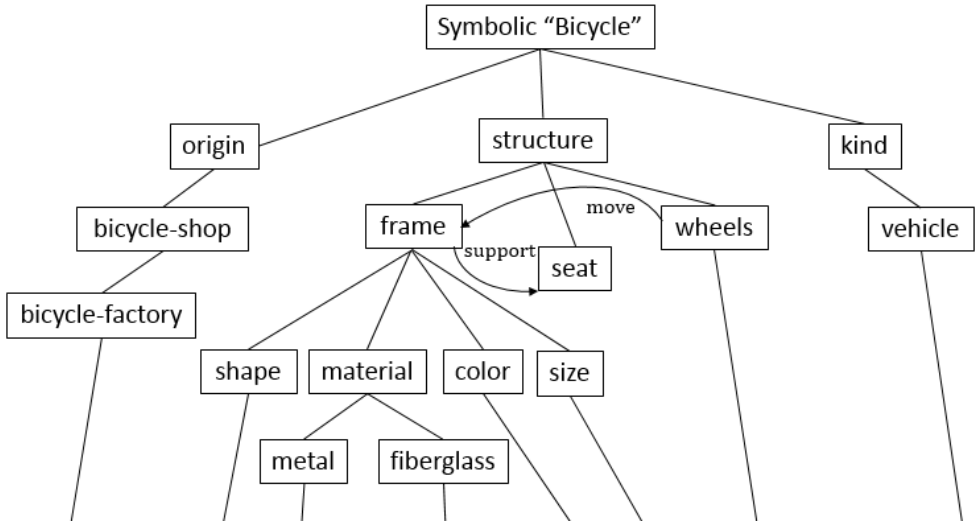


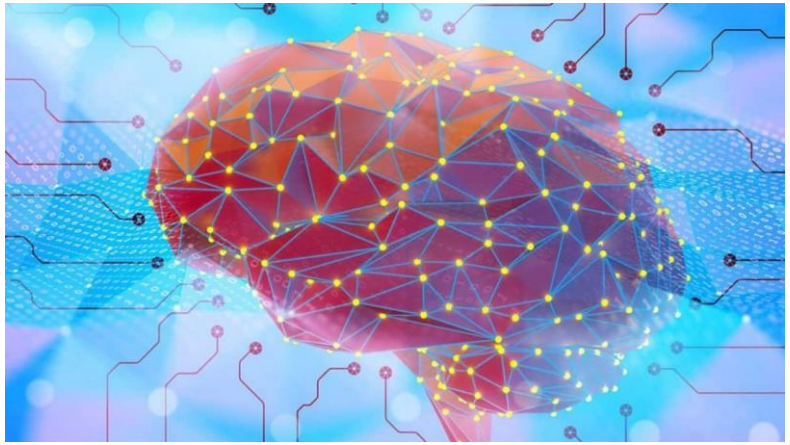
http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

- 1955: Logic Theorist, the first AI program was invented
- 1956: Dartmouth workshop
- 1966 – 1975: Funding cut
- 1970-80s: Expert systems created and proliferated
- 1980s: Fifth-Generation Computer Project (Japan), Strategic Computing Initiative (DARPA)
- 1987: Collapse of LISP market, funding cut
- 1990s: Rise of machine learning
- 2010s: Heavy investment in deep learning

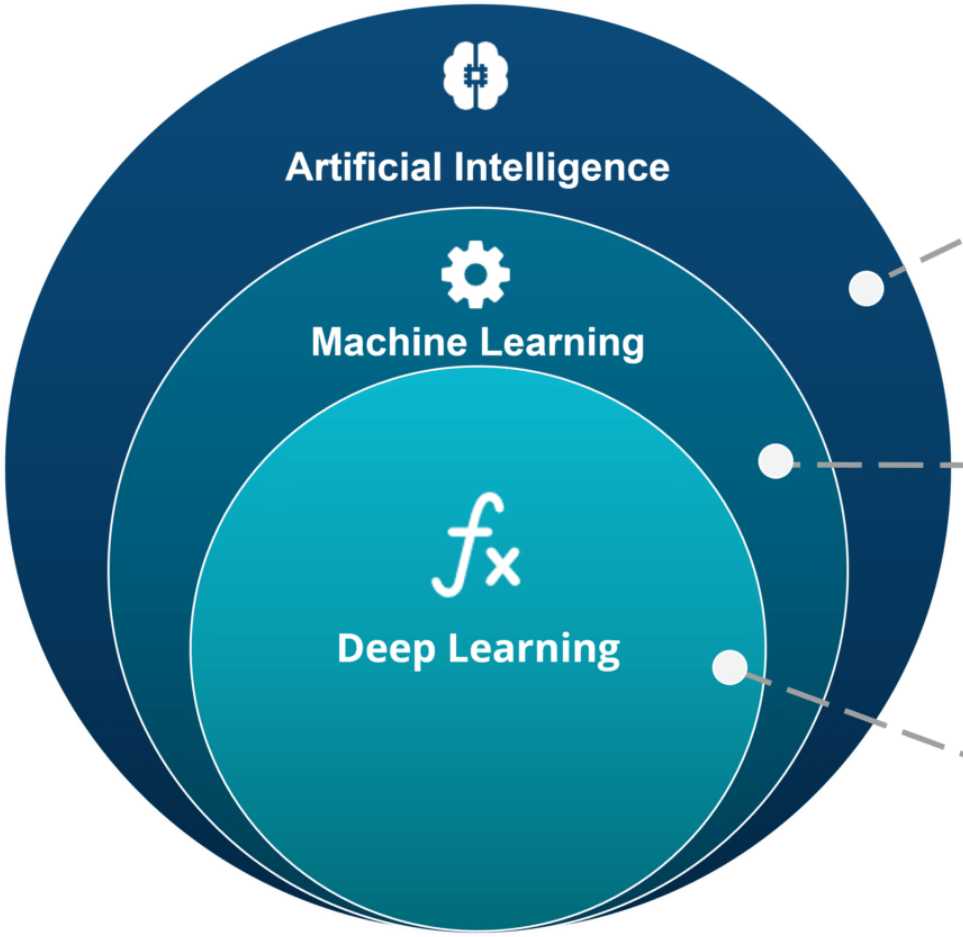
AI Paradigms

- **Symbolism (符号主义)**: Symbolic AI represents information through symbols and their relationships. Specific algorithms are used to process these symbols to solve problems or deduce new knowledge
- **Connectionism (连接主义)**: Connectionist AI represents information in a distributed, less explicit form within a network. It imitates biological processes underlying learning, task performance and problem solving
- Other paradigms:
  - Actionism (行为主义)
  - Statistical approach





Machine learning as a subfield of AI



ARTIFICIAL INTELLIGENCE
A technique which enables machines to mimic human behaviour

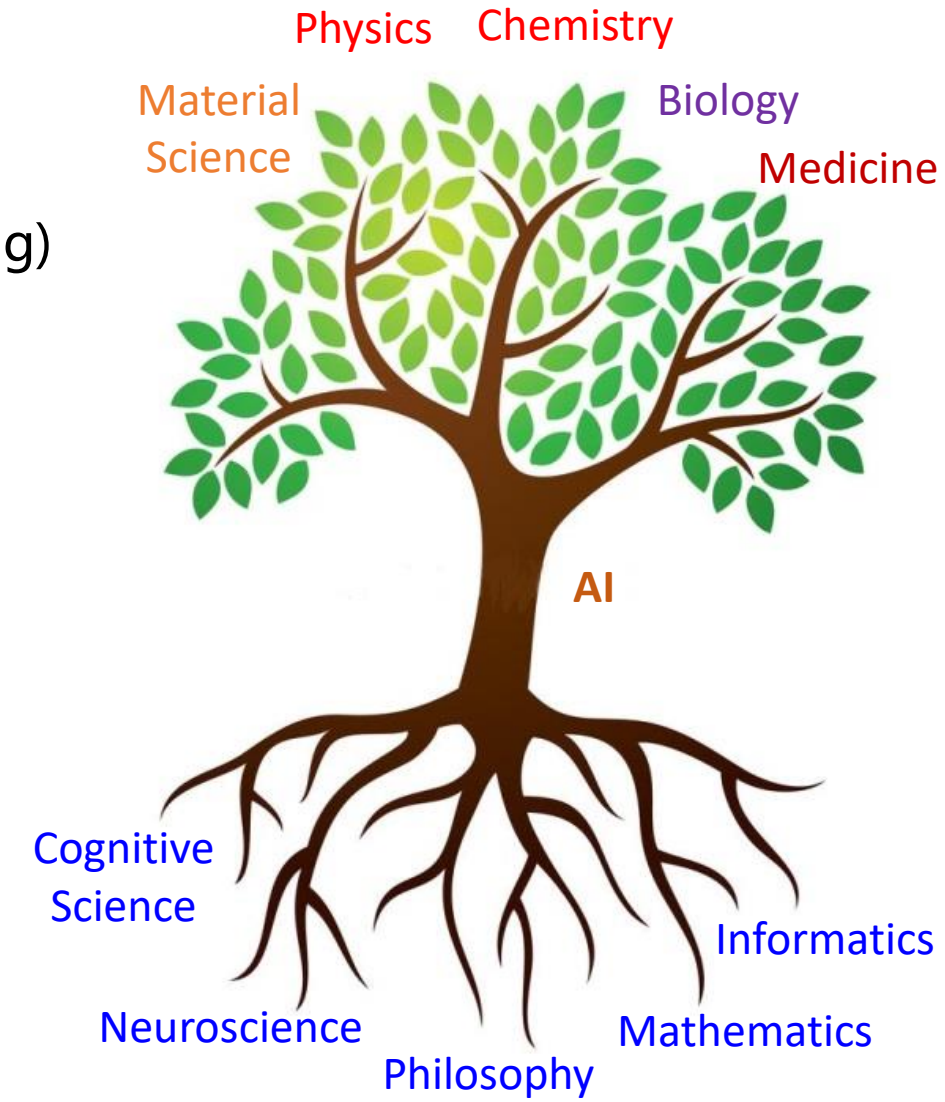
MACHINE LEARNING
Subset of AI technique which use statistical methods to enable machines to improve with experience

DEEP LEARNING
Subset of ML which make the computation of multi-layer neural network feasible

Figure credit: https://www.edureka.co/blog/ai-vs-machine-learning-vs-deep-learning/

Why does Science need AI?

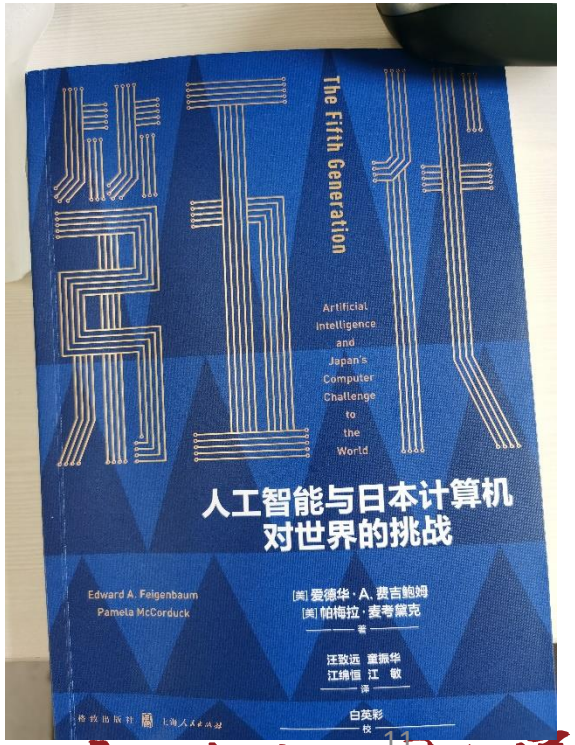
- Massive amounts of **data** have been generated in all different fields of natural science (e.g. astrophysics, genomics, material science)
  - The mainstream techniques of AI (e.g. deep learning) are **data-driven**
  - The overwhelming volume and complexity of scientific data can’ t be handled by traditional computer software that lacks intelligence
- Lots of daily works in scientific research are repetitive routines, which should be done by machines instead of humans
- AI can help humans better understand the natural world:
  - Things too small: atoms
  - Processes too fast: protein folding
  - Phenomena too complex: cancer



Expert systems for science and engineering

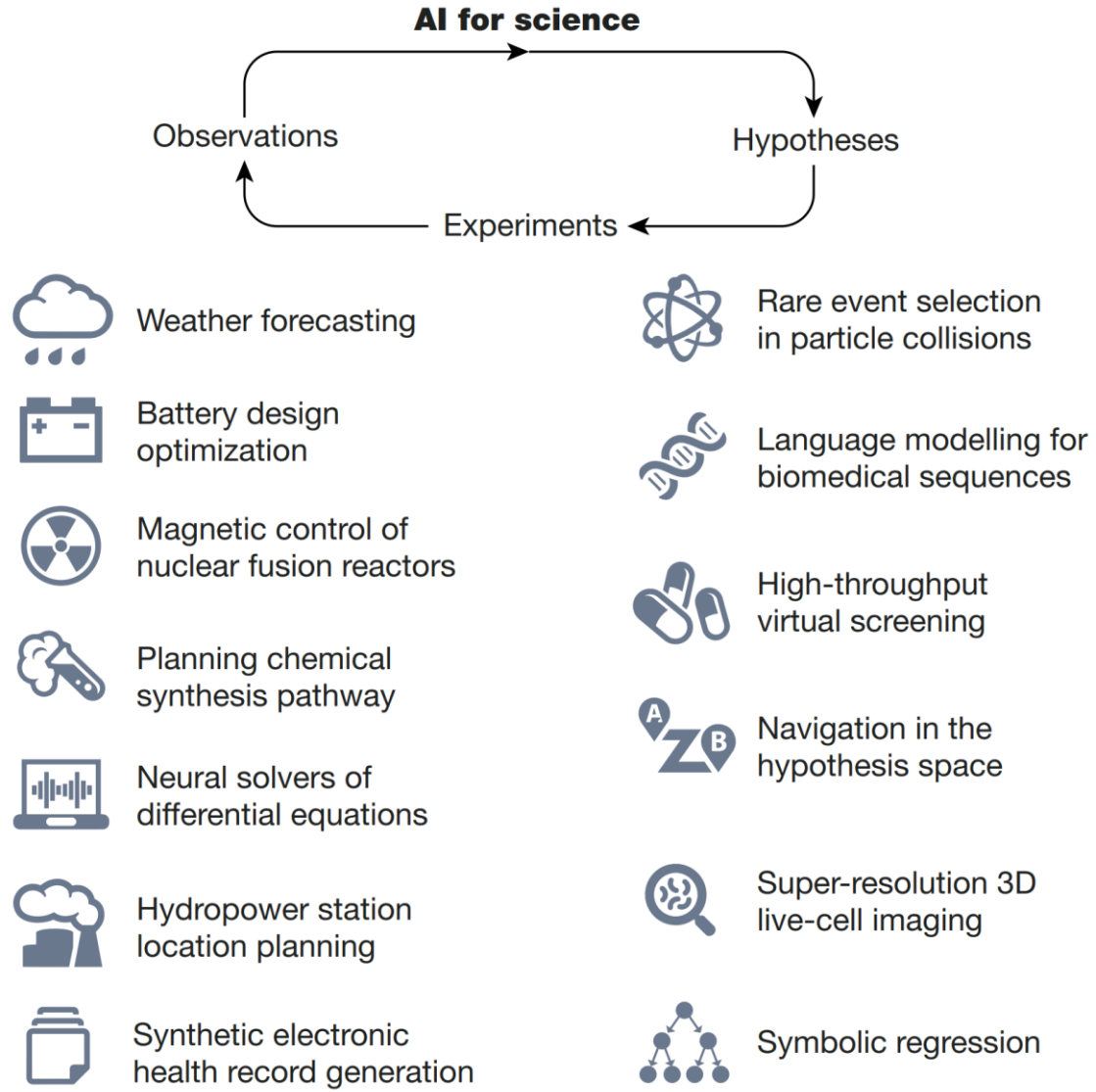
- An **expert system** is a computer program that uses AI methods to solve problems within a specialized domain that ordinarily requires human expertise (from Britannica)
  - Designed to solve complex problems by reasoning through bodies of knowledge, represented mainly as if-then rules (from Wikipedia)
  - Consisting of knowledge base (representing facts and rules) and inference engine (applying rules to known facts to deduce new facts)
- **DENDRAL**, an early expert system, developed from 1965 by AI researcher **Edward Feigenbaum** and geneticist (Nobel Laureate) **Joshua Lederberg**, both at Stanford University
  - To identify the structures of chemical compounds
  - Starting from spectrographic data obtained from substance (e.g. a compound of carbon, hydrogen and nitrogen), it would hypothesize the substance's molecular structure
  - Its performance rivaled that of human chemists at this task
- Many expert systems were derived from DENDRAL, such as:
  - **MYCIN**: Medical system for diagnosing blood disorders, first used in 1979
  - **MOLGEN**: A system that can plan or design laboratory experiments in molecular genetics (e.g. gene cloning)

Edward A. Feigenbaum (1936 -), American Computer Scientist, Turing Award (1994)



AI is reshaping scientific discovery

- Data collection & generation
  - Automation in data preparation
- Representation learning
  - Data-driven and knowledge-driven methods
- Hypothesis generation & search
  - Efficient strategies
- Experimentation & simulation
  - Automatic testing + efficient solver



Cases of AI for Science (AI4S)

- Many success stories of using AI techniques (often deep learning) to make scientific discovery

Case studies	Time	Domain	AI Technique	Research team
AlphaFold 2 won CASP14 protein-folding competition	Dec. 2020	Structural biology	NLP modeling (Transformer)	Google DeepMind
Mankind has the first photo of a black hole	Apr. 2019	Astrophysics	CHIRP algorithm based on Bayesian statistical model	MIT etc.
The AI program called **Atom2Vec** recreated the periodic table of chemical elements	Jun. 2018	Chemistry, material science	Natural language processing (NLP), knowledge representation	Stanford University
An AI system speeded up the discovery of metallic glass by 200 times	May 2018	Material science & engineering	Supervised machine learning	SLAC, NIST, Northwest University

Cover image (Aug. 2021): DeepMind's AlphaFold2

Outline

- A brief history of AI
- **Learning objectives**
- Instructors
- Prerequisites
- Learning plan
- Evaluation
- Computing platform
- Course materials and references
- Communication
- Academic integrity

Learning objectives

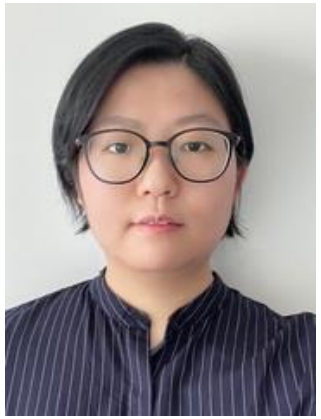
- After taking this course, you should be able to:
  - Understand basic ideas of machine learning (especially deep learning) and other techniques of AI
  - Get experiences of solving scientific or engineering problems using AI techniques
  - Identify research problems that can be solved using AI techniques
  - Select the most suitable machine learning methods for given research problems
  - Propose innovation to improve the performance of existing machine learning methods, in terms of accuracy, computational efficiency, etc.
- Teaching strategy:
  - Learning by doing
  - Problem-oriented learning
  - Project-based learning

Outline

- A brief history of AI
- Learning objectives
- **Instructors**
- Prerequisites
- Learning plan
- Evaluation
- Computing platform
- Course materials and references
- Communication
- Academic integrity

Lecturers (SIST)

Jie Zheng (郑杰)，SIST
Course coordinator
zhengjie@shanghaitech.edu.cn



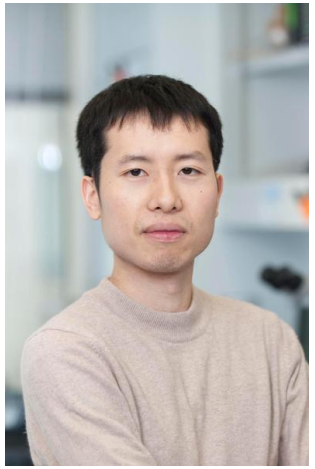
Sibei Yang (杨思蓓)，SIST
yangsb@shanghaitech.edu.cn

Xuming He (何旭明)，SIST
hexm@shanghaitech.edu.cn

Quan Li (李权)，SIST
liquan@shanghaitech.edu.cn

Lecturers (SLST)

王权，免疫化学所，生命学院
wangq@shanghaitech.edu.cn



徐一驰，生命学院
xuych2@shanghaitech.edu.cn

蒋立春，免疫化学所
jianglch@shanghaitech.edu.cn

刘雪松，生命学院
liuxs@shanghaitech.edu.cn

Lecturers (SPST)

孙兆茹，物质学院
sunzhr@shanghaitech.edu.cn

怀平，物质学院，高研院
huaiping@shanghaitech.edu.cn

姜珊，物质学院
jiangshan@shanghaitech.edu.cn

胡培君，物质学院
hupj@shanghaitech.edu.cn

Teaching Assistants (TAs)

Name	Affiliation	Email
Tao Zhang (张涛)	SIST	zhangtao4@shanghaitech.edu.cn
Yang Yang (杨扬)	SIST	yangyang2022@shanghaitech.edu.cn

- There will be additional “project TAs” (graduate students or postdoc scientists who help lecturers supervise the projects)

Outline

- A brief history of AI
- Learning objectives
- Instructors
- **Prerequisites**
- Learning plan
- Evaluation
- Computing platform
- Course materials and references
- Communication
- Academic integrity

Prerequisites

- Required:
  - Programming skills in Python
  - Linear algebra
  - Probability and statistics
  - Calculus
- Preferred:
  - Knowledge and experience in machine learning
  - If your major is in information science, strong interest and knowledge in physics, chemistry or biology, etc. would help
  - If your major is in physical or life science, strong interest and experience in computing would be useful

Outline

- A brief history of AI
- Learning objectives
- Instructors
- Prerequisites
- **Learning plan**
- Evaluation
- Computing platform
- Course materials and references
- Communication
- Academic integrity

Basic course structure

- Lectures:
  - 8 weeks
  - **Time**:
    - Tuesday 1 – 2:40 pm
    - Thursday 1 – 2:40 pm
  - **Venue**: Teaching Center 303
- Tutorials (Recitation):
  - Around 1 hour / week
  - To start in week 2
  - TAs demonstrate hands-on examples, and students try to solve small problems by themselves
- Attendance of **both** lectures **and** tutorials is mandatory!
- Office hour: By email appointment (SIST Building 3, Room 432)
- Project meeting: Students should meet regularly with instructors supervising the projects

Tentative course schedule

上海科技大学 ShanghaiTech University

Week	Lecture Topics	Contact hours per week	Teaching modes
Week 1	1. Overview of course (history of AI, instructors, evaluation, etc.) 2. 10th anniversary celebration of ShanghaiTech	2	Lecture
Week 2	1. National Holidays 2. Landscape of machine learning	2	Lecture & Tutorial
Week 3	1. Python test 2. Basics of deep learning: Introduction to artificial neural networks, training deep neural networks, etc.	4	Lecture & Tutorial
Week 4	Classic topics of deep learning: Convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoders, graph neural network, etc.	4	Lecture & Tutorial
Week 5	Selected topics of AI: Transformer, large language models, generative AI, explainable AI, etc.	4	Lecture & Tutorial
Weeks 6-7	Case studies of AI in Physical Sciences	4	Lecture & Tutorial
Weeks 8-9	Case studies of AI in Life Sciences	4	Lecture & Tutorial
Weeks 10-16	working on projects (group meetings, mid-term presentations, final presentations)	4	Practical training

Outline

- A brief history of AI
- Learning objectives
- Instructors
- Prerequisites
- Learning plan
- **Evaluation**
- Computing platform
- Course materials and references
- Communication
- Academic integrity

Evaluation methods

- Python Test: **10%**
- Class Attendance (including quizzes at random dates): **10%**
- Homework Assignments: **30%**
  - Individual-based
  - 3 homework assignments
  - Submit solutions on our Blackboard course site (specific instructions to be given by TAs in tutorials)
- Project: **50%**
  - Group-based (2-3 students per group)
  - To choose projects and grouping in about **Week 5**
  - Mid-term presentation (**around Week 12**)
  - Final presentations and submission of reports (also including PPT, code and data) (**due in Week 16**)
  - Specific instructions to be announced **around week 6**

Policies for attendance and homework submission

- Attendance will be checked by calling the roll or in-class quizzes (at random dates)
- If unable to attend a class for special reasons, you must:
  - Apply for approval of leave officially (in Egate and by email to course coordinator) **before** the class begins
  - Provide an official medical certificate (MC) from a clinician or similar official document to justify your leave
- Penalty for late submission of homework solutions:
  - **50% deduction** within the first 24 hours after deadline
  - **100% deduction (0 score)** more than 24 hours after the deadline

Outline

- A brief history of AI
- Learning objectives
- Instructors
- Prerequisites
- Learning plan
- Evaluation
- **Computing platform**
- Course materials and references
- Communication
- Academic integrity

Computational resources

- University IT center (学校图书信息中心) will set up computational environment GPU servers for this course, to support students' homeworks and projects

- Software (e.g. PyTorch, SciKit-Learn, NumPy, Pandas, Matplotlib) can be installed with help of TAs

Outline

- A brief history of AI
- Learning objectives
- Instructors
- Prerequisites
- Learning plan
- Evaluation
- Computing platform
- **Course materials and references**
- Communication
- Academic integrity

Course materials

- Main course page (on Blackboard system, 教学互动平台):
  - Content includes:
    - Lecture slides (in PDF)
    - Tutorial information
    - Homework assignments
    - Project description
    - Q&A forum
  - Please sign it up in Egate
- Note: Please do **not** share the course materials outside the class, to avoid potential copyright issue

The main reference book

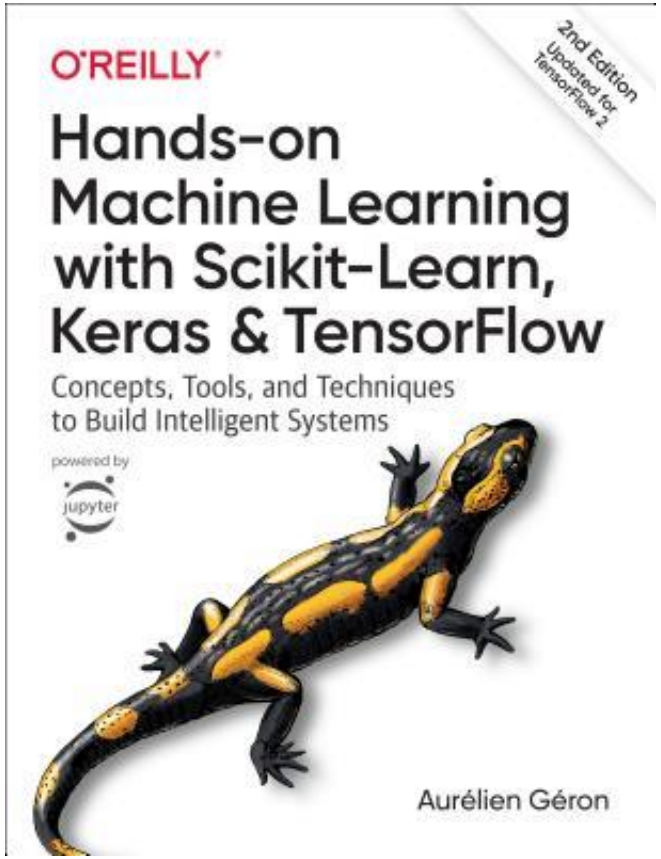
Book title: Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems

Author(s): Aurélien Géron

Publisher: O’ Reilly

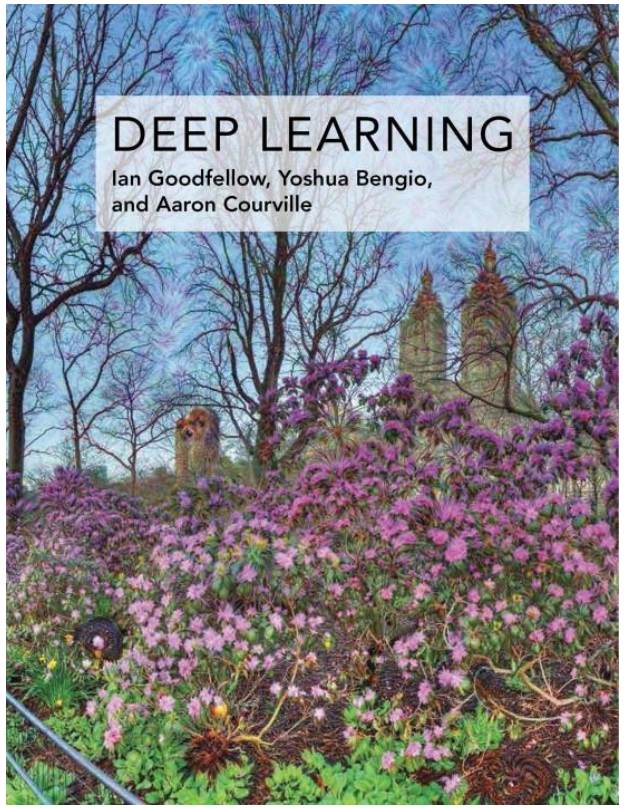
Publication Year: 2019

ISBN: 978-1492032649

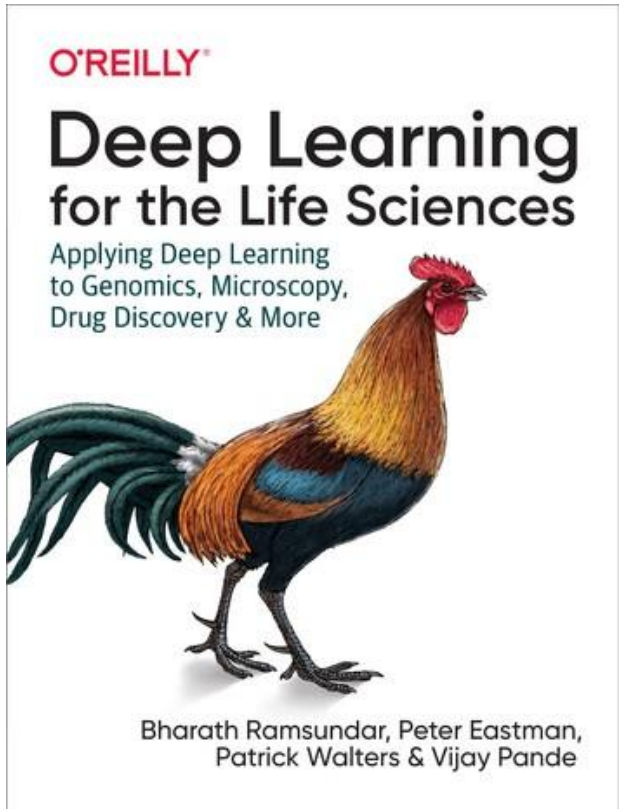


Other references

Book title: Deep Learning
Author(s): Ian Goodfellow, Yoshua Bengio, and Aaron Courville
Publisher: MIT Press
Publication Year: 2016
ISBN: 978-0262035613



Book title: Deep Learning for the Life Sciences：Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More
Author(s): Bharath Ramsundar, Peter Eastman, Patrick Walters, and Vijay Pande
Publisher: O’ Reilly
Publication Year: 2019
ISBN: 978-1492039839



Note:

- A reading list will be provided for each project
- Search internet to find related papers in journals and conference proceedings

Outline

- A brief history of AI
- Learning objectives
- Instructors
- Prerequisites
- Learning plan
- Evaluation
- Computing platform
- Course materials and references
- **Communication**
- Academic integrity

Communication

- Each enrolled student is required to check his/her ShanghaiTech email (as used to enroll in the course) **at least once every 24 hours**. If missing any important email about the course, the student should bear the consequence by himself/herself
- For technical discussions:
  - Face-to-face meetings during tutorials or office hours
  - Emails should be used **only for short communication** (e.g. making an appointment), but **not** for technical discussions
  - Use Blackboard forum
- Lecturers or TAs will **not** help any student debug his/her code

Outline

- A brief history of AI
- Learning objectives
- Instructors
- Prerequisites
- Learning plan
- Evaluation
- Computing platform
- Course materials and references
- Communication
- **Academic integrity**

Academic integrity rules (part 1)

- Unless explicitly noted, your solutions submitted should be done **all by yourself** and reflect your **independent** thinking and working.
- **No cheating**:
  - **No plagiarism** (i.e. you should not copy any part of solution from others): We will scan your solutions for similarity both automatically and manually.
  - Do not ask anyone else (whether inside or outside the university) to do the homework or project on your behalf. **The TAs can give advice only for the projects, but their contributions should be openly and precisely declared in writing.**
  - Do not show, explain or share your homework or project solutions to any other student under any circumstance (except for discussion about project with your groupmates). In case of plagiarism, a student being copied shall be punished as severely as a student copying others.
  - If some sentences in your answer or report are from books, papers or web (e.g. Wikipedia), do cite the sources as references and add **double quotations**. Otherwise, it is counted as plagiarism.
  - If a figure, table or piece of code in your project report/presentation is borrowed or adapted from the web (e.g. GitHub), you should acknowledge the source, and state clearly in the code (as comments) and/or in documentation about how you have borrowed it.
  - Do not violate academic integrity rules for test, quiz or exam (further specific instructions will be given in due time).

Academic integrity rules (part 2)

- **Consequences of cheating**:
  - 0 score for the course component involved,
  - Fail the course,
  - Record the cheating in personal file,
  - Expulsion from the university.
- **Reproducibility**:
  - Solutions that you submit should be **reproducible** independently by yourself as well as others.
  - Your code is expected to run again to output results that are consistent with you report.
  - An answer that is not reproducible is considered a "**fake solution**", and counted as **cheating**.

End of Lecture 1