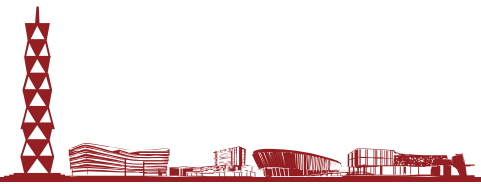
上海科技大学
ShanghaiTech University

CS286 AI for Science and Engineering

Lecture 0: Course Overview

Jie Zheng (郑杰)
PhD, Associate Professor
School of Information Science and Technology (SIST), ShanghaiTech University
Fall Semester, 2019



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)?

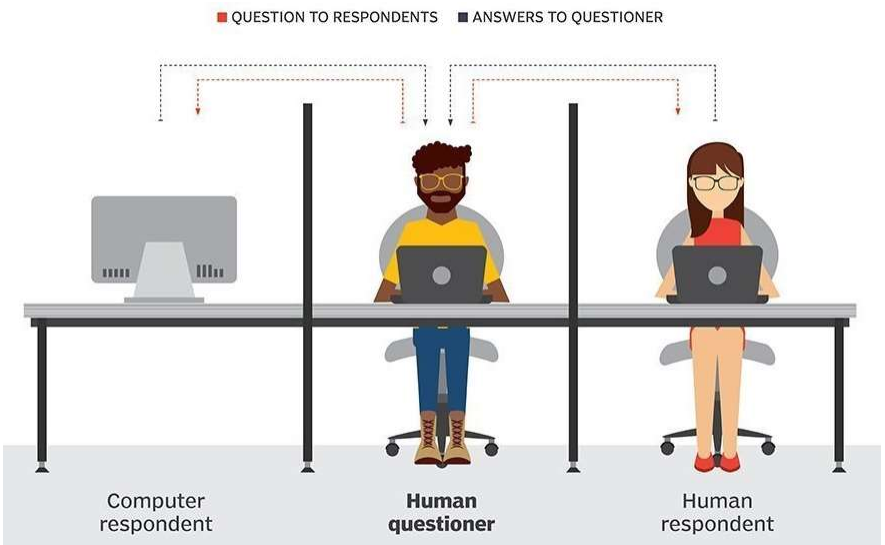
上海科技大学 ShanghaiTech University

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)

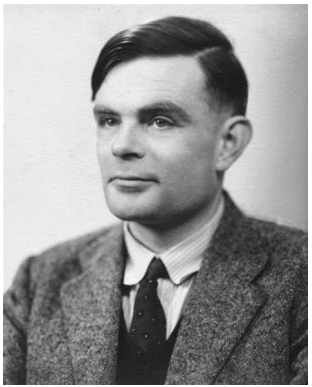
Source: Stuart Russell and Peter Norvig's book, "Artificial Intelligence: A Modern Approach" (3rd Ed.), 2010 (Fig. 1.1, page 2)

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

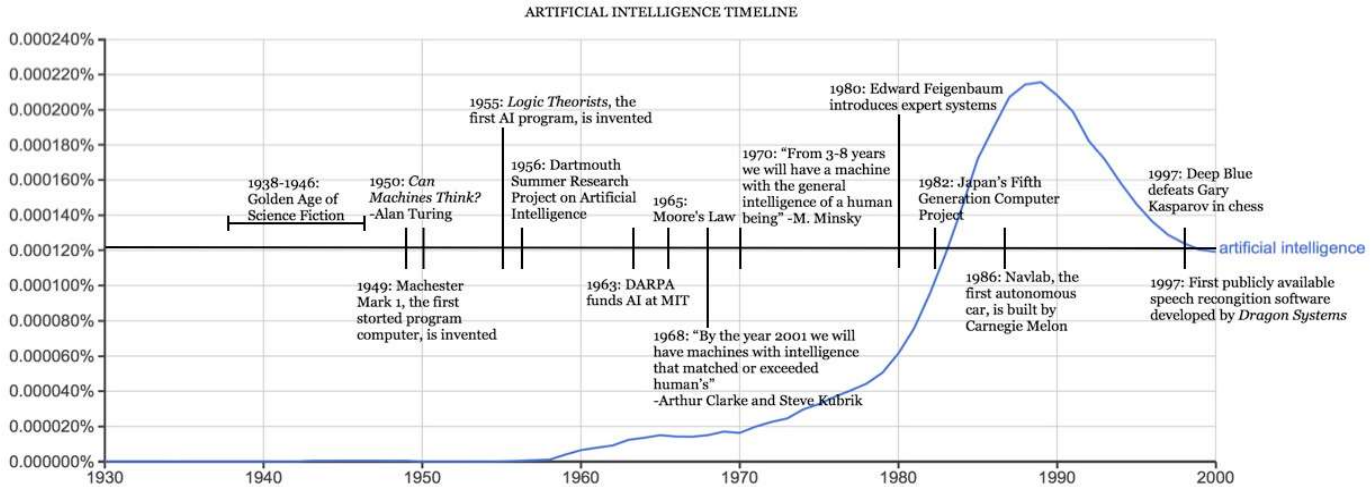
Birth of AI at Dartmouth College (1956)

- 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

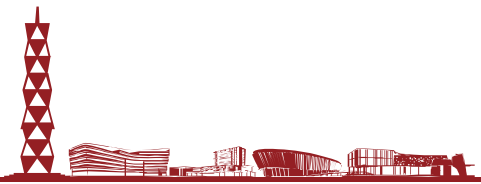
Recent news about AI for sciences

- There are many recent success stories about using AI techniques (often machine learning) to make scientific breakthrough

Case studies	Time	Domain	AI Technique	Research team
Mankind has the first photo of a black hole	April 2019	Astrophysics	CHIRP algorithm based on Bayesian statistical model	MIT etc.
AlphaFold won CASP13 protein-folding competition	Dec. 2018	Structural biology	Deep residual network (ResNet)	Google DeepMind
The AI program called **Atom2Vec** recreated the periodic table of chemical elements	June 2018	Chemistry, material science	Natural language processing, 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
Atomwise Inc. found two candidate drugs for controlling the Ebola virus in less than a day	March 2015	Drug design	A deep convolutional neural network named AtomNet	Atomwise Inc., Toronto University

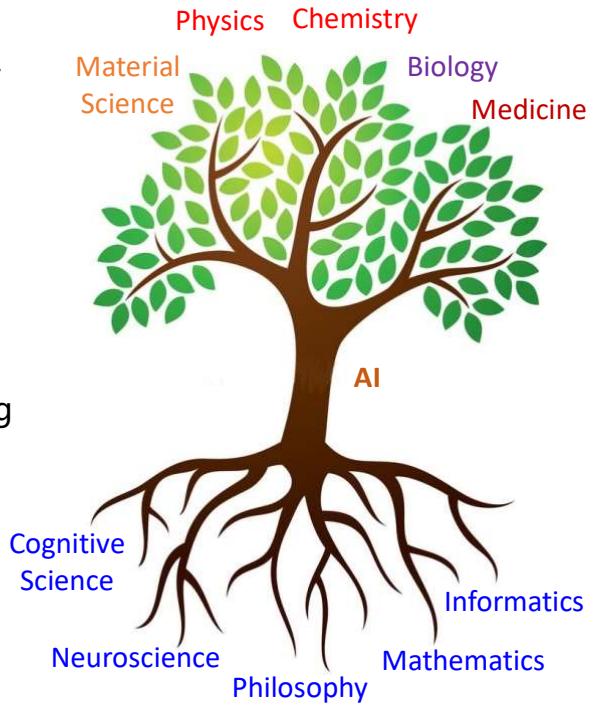
Outline

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



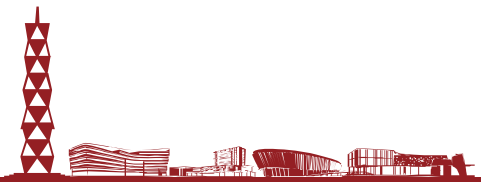
Learning objectives

- After taking this course, you should be able to:
  - Understand the basic ideas of machine learning (especially deep learning)
  - 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 method for a given research problem
  - Propose innovation to improve the performance of existing machine learning methods, in terms of accuracy, computational efficiency, etc.
- Teaching strategy:
  - Learning by doing
  - Problem-based 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



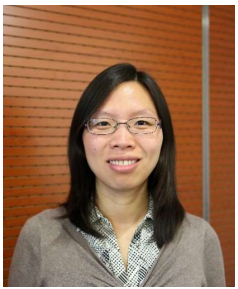
Lecturers (partial list)

郑杰，信息学院
Course coordinator
zhengjie@shanghaitech.edu.cn

何旭明，信息学院
hexm@shanghaitech.edu.cn

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

武丽杰，iHuman 研究所
wulj@shanghaitech.edu.cn



赵素文
iHuman 研究所, 生命学院
zhaosw@shanghaitech.edu.cn

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

Yongjin Lee，物质学院
leeyj@shanghaitech.edu.cn

杨波，物质学院
yangbo1@shanghaitech.edu.cn

Teaching Assistants (TAs)

上海科技大学 ShanghaiTech University

Name	Affiliation	Position	Email
宋柯	SIAIS （免疫化学所）	博士，平台工程师	songk@shanghaitech.edu.cn
刘旭阳	iHuman 研究所	博士研究生	liuxy6@shanghaitech.edu.cn
苏欣	iHuman 研究所	硕士研究生	suxin@shanghaitech.edu.cn
熊招平	SLST （生命学院）	博士研究生	xiongzhp@shanghaitech.edu.cn
刘永飞	SIST （信息学院）	硕士研究生	liuyf3@shanghaitech.edu.cn
张可	SIST（信息学院）	博士研究生	zhangke1@shanghaitech.edu.cn
何彦男	SIST（信息学院）	硕士研究生	heyn@shanghaitech.edu.cn
罗海旻	SIST（信息学院）	硕士研究生	luohm@shanghaitech.edu.cn
张晓峰	SPST（物质学院）	博士，助理研究员	zhangxf2@shanghaitech.edu.cn
王宝川	SPST（物质学院）	硕士研究生	wangbch@shanghaitech.edu.cn

Outline

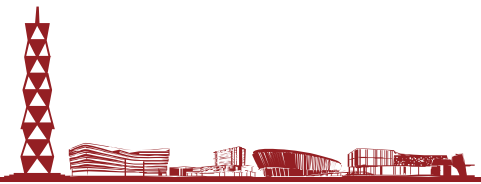
- A brief history of AI
- Learning objectives
- Instructors
- **Prerequisites**
- Learning plan
- Evaluation
- Computing platform
- Course materials and references
- Communication
- Academic integrity

Prerequisites

- Required:
  - Basic programming skills in Python (to be tested on next Tuesday)
  - Linear algebra
  - Probability and statistics
  - Calculus
- Preferred:
  - If your major is in information science, strong interest and knowledge in physics, chemistry or molecular biology would help
  - If your major is in life sciences or physical sciences, interest and experiences in coding would be useful
  - Knowledge about algorithms

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

上海科技大学 ShanghaiTech University

- Lectures:
  - 15 weeks in total
  - 4 hours / week
    - Tuesday 8:15 – 9:55 am
    - Thursday 8:15 – 9:55 am
  - Location: Teaching Center 204
- Tutorials (Recitations):
  - 2 hours / week
  - To start in week 2
  - TAs demonstrate hands-on examples, and students try to solve small problems by themselves
  - Project discussion: Students report progress to TAs to get feedback
  - Tutorial attendance is mandatory!
- Office hours: TBA

Tentative course schedule (part 1)

上海科技大学 ShanghaiTech University

Week	Day	Lecture Topics	Lecturers	Tutorials	Homework and Project
Week 1	Tuesday	Course overview, machine learning landscape	Jie Zheng	N.A.	N.A.
	Thursday	Machine learning algorithms	Jie Zheng		
Week 2	Tuesday	Introduction to bioinformatics and genomics, Python test	Lichun Jiang, Jie Zheng	Set up computing platform, install software	Project orientation
	Thursday	Introduction to imaging (cryo-EM and SHINE)	Lijie Wu, Xuming He, and Ping Huai		
Week 3	Tuesday	Introduction to structural biology and drug design	Suwen Zhao, Zhaoping Xiong	Use Scikit-Learn and TensorFlow, etc.	Project orientation
	Thursday	Introduction to physical sciences	Bo Yang		
Week 4	National Day Holidays				Choose projects, form groups
Week 5	Tuesday	Artificial neural networks, training deep neural nets	Jie Zheng	Exercises on machine learning	Projects start
	Thursday	Convolutional neural networks (CNN)	Jie Zheng		
Week 6	Tuesday	Recurrent neural networks (RNN)	Jie Zheng	Exercises on deep neural nets, CNN	Homework 1 starts
	Thursday	Autoencoders	Jie Zheng		

Tentative course schedule (part 2)

上海科技大学 ShanghaiTech University

Week	Day	Lecture Topics	Lecturers	Tutorials	Homework and Project
Week 7	Tuesday	Deep learning for bioinformatics and genomics	Lichun Jiang	Exercises on RNN and Autoencoders	
	Thursday	Deep learning for single-cell RNA-seq data analysis	Jie Zheng		
Week 8	Tuesday	Deep learning for 3D genomics (e.g. Hi-C data)	Lichun Jiang	Exercises on Bioinformatics	Homework 1 due, Homework 2 starts
	Thursday	Selected areas of bioinformatics and genomics	Lichun Jiang / Jie Zheng		
Week 9	Tuesday	Molecular imaging (Cryo-EM etc.)	Lijie Wu / Xuming He	Exercises on Genomics	
	Thursday	Molecular imaging (Cryo-EM etc.)	Lijie Wu / Xuming He		
Week 10	Tuesday	Molecular imaging (Cryo-EM etc.)	Lijie Wu / Xuming He	Exercises on imaging	Homework 2 due, Homework 3 starts
	Thursday	Machine learning in SHINE and related research	Ping Huai		
Week 11	Tuesday	Computational methods in structural biology	Suwen Zhao	Exercises on imaging	**Project interim reports due**
	Thursday	Machine learning in structural biology (with focus on protein folding)	Suwen Zhao / Sheng Wang		
Week 12	Tuesday	Deep learning in structural biology	Suwen Zhao / Sheng Wang	Exercises on structural biology	Homework 3 due, Homework 4 starts
	Thursday	Deep learning in structural biology	Suwen Zhao / Sheng Wang		

Tentative course schedule (part 3)

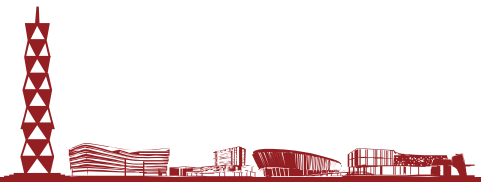
上海科技大学 ShanghaiTech University

Week	Day	Lecture Topics	Lecturers	Tutorials	Homework and Project
Week 13	Tuesday	AI for drug design and healthcare industry	TBA	Exercises on structural biology	
	Thursday	AI for drug design and healthcare industry	TBA		
Week 14	Tuesday	AI for physical sciences (e.g. material genomics)	Faculty from SPST (TBA)	Exercises on drug design	Homework 4 due
	Thursday	AI for physical sciences	Faculty from SPST (TBA)		
Week 15	Tuesday	AI for physical sciences	Faculty from SPST (TBA)	Exercises on physical sciences	
	Thursday	AI for physical sciences	Faculty from SPST (TBA)		
Week 16	Tuesday	Project presentations by students	Instructors as judges	N.A.	**Project final reports due**
	Thursday	Project presentations by students	Instructors as judges		

Note: For each homework, a student is expected to submit his/her solution in **about 2 weeks** (usually announced online on Friday, and due by mid-night of Sunday)

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 (on Tuesday, Sept. 17th): **4%**
- Class attendance: **6%** (2 quizzes, date unknown in advance)
- Homework assignments: **30%**
  - Individual based
  - 4 homework assignments
    - Homework 1: **6%**
    - Homework 2, 3, 4: **8%** each
  - Submit solutions in form of Jupyter Notebook in GitHub Classroom (specific instructions to be given by TAs in tutorials)
- Project: **60%**
  - Group based (~3 students per group)
  - To choose projects and grouping during the National Day holidays
  - Interim reports (due in Week 11)
  - Final reports (including data and code) and presentation (due in Week 16, end of December)
  - **Specific instructions to be announced in Week 3**

Policies for attendance and homework submission

- Attendance will be checked by 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

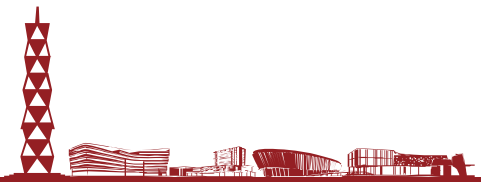
Perseverance is needed

- At the beginning of a semester:

- Toward the end of the semester:

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) has set up computational environment for this course, based on Tesla P40 cluster (GPU servers) to support students' homework and projects
- Hardware (10 nodes):
  - 60 virtual machine accounts (5 nodes)
  - 20 docker environment (5 nodes)
- Software:
  - Ubuntu Linux
  - Python 3.6.8
  - TensorFlow and Torch
  - CUDA
  - Additional software (e.g. Python scientific libraries like 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

上海科技大学 ShanghaiTech University

- Main course page (on Piazza):
  - piazza.com/shanghaitech.edu.cn/fall2019/cs286/home
  - Content includes:
    - Lecture slides (in PDF)
    - Tutorial information
    - Homework assignments
    - Project description
    - Q&A forum
  - Please register for the course on Piazza via this Signup link: piazza.com/shanghaitech.edu.cn/fall2019/cs286

- **Note**: Please do **not** share the course materials outside the class, to avoid potential copyright issue

The main reference book

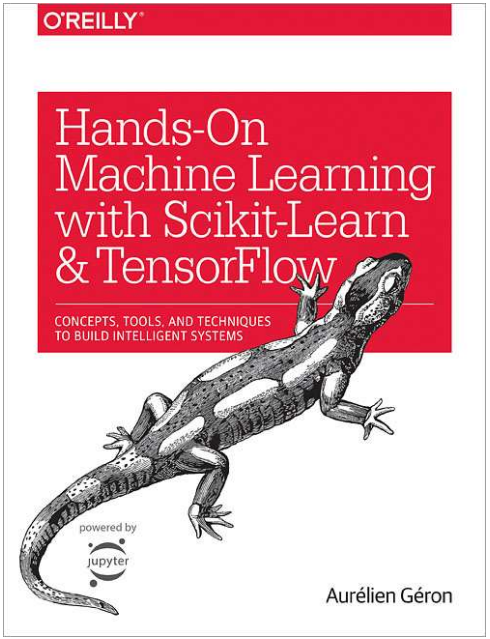
Book title: Hands-On Machine Learning with Scikit-Learn & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems

Author(s): Aurélien Géron

Publisher: O' Reilly

Publication Year: 2017

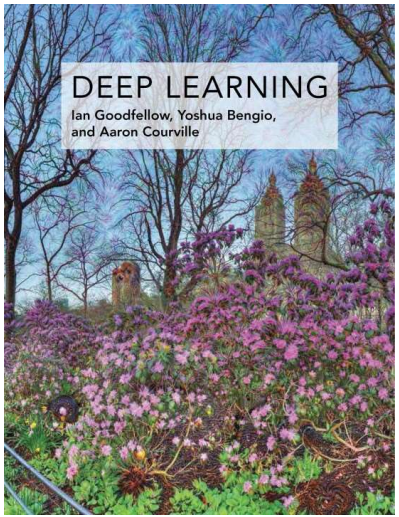
ISBN: 978-7-5641-7371-5



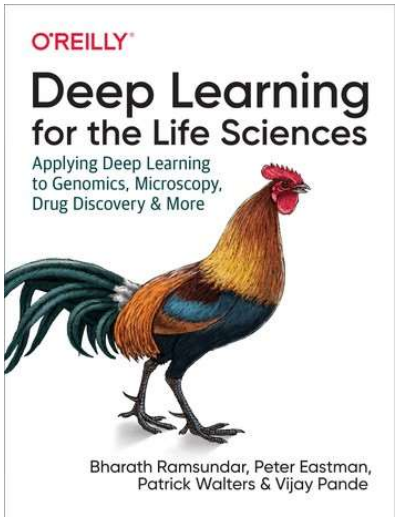
Note: A new edition titled "Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems" is going to be published by O' Reilly in Oct. 2019

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

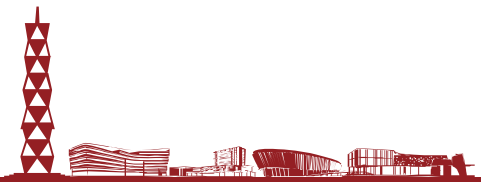


Note:

- A reading list will be provided for each project
- Search the literature to find related publications

Outline

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

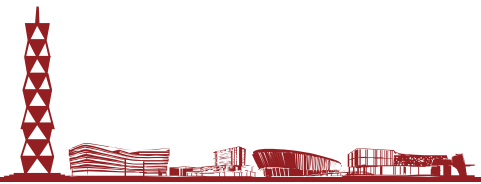


Communication

- 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:
  - Use Piazza forum
  - Face-to-face meetings during office hours
  - Emails should be used *only for short communication* (e.g. making an appointment), but **not** for technical discussions
- Lecturers would **not** communicate with students via WeChat or QQ (emails would be enough)
- Lecturers or TAs would **not** help any student debug his/her code

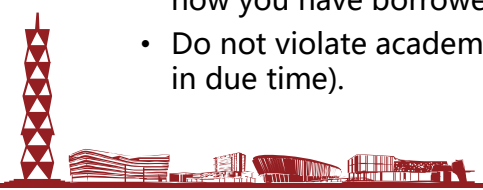
Outline

- 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 work submitted should be done **all by yourself** and reflect your **independent** capabilities.
- **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 within your own project group). 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 state clearly as comments in the code and/or in documentation how you have borrowed it.
  - Do not violate academic integrity rules for test or quizzes (further specific instructions will be given in due time).



Academic integrity rules (part 2)

上海科技大学 ShanghaiTech University

- **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 not reproducible is considered a "**fake solution**", and will get 0 score.

End of Lecture 0