

Becoming an AI Apprentice (Google Edition)

A Field Guide (**Version 4.0**)

12 - 18 months self-directed AI/ML learning journey



AI SINGAPORE

June 2021

Current Contributors

- Laurence Liew
- Najib Ninaba
- Kevin Oh
- Weng Jianshu
- Ng Kim Hock
- Ryzal Kamis
- Basil Han
- Azmi Mohamed Ridwan

Past Contributors

- Cheong Wei Yih
- Maurice Manning
- Jeanne Choo
- Daniel Ng

Preface

The AI Apprenticeship Programme (AIAP)[™] is a 9-month long programme run by AI Singapore (AISG) in collaboration with the Infocomm Media Development Authority (IMDA) to develop a pipeline of local AI engineers for the industry. It is a deep skilling programme and selection into it is rigorous. Applicants are expected to already have a solid foundational level of skills and knowledge in the field of AI and machine learning (ML), albeit with limited hands-on experience with and exposure to real-world data and production environments.

This field guide serves several needs. It offers a structured 12-month learning pathway to attain the requisite level for acceptance into AIAP[™]. New applicants can check themselves against the requirements outlined before they submit their applications. Unsuccessful applicants can also evaluate their skills set against them and be motivated to improve and try again. Finally, it defines the minimum standard in order to become an AI Associate Engineer in AISG's [AI Certification](#) pathway.

- *The Editor, Basil Han*

Table of Contents

[Preface](#)

[Introduction : Who is an AI engineer?](#)

[How to use this Field Guide](#)

[Suggested Learning Schedule](#)

[Python](#)

[The Python Tutorial](#)

[Crash Course on Python by Google](#)

[Video Tutorials](#)

[Software Engineering](#)

[Software development life cycle](#)

[Code Versioning](#)

[Object Oriented Programming \(OOP\)](#)

[Software Testing and Debugging](#)

[Data Structures and Algorithms](#)

[Databases](#)

[Computational Thinking](#)

[Mathematics](#)

[Data Science and Machine Learning](#)

[Machine Learning Crash Course by Google](#)

[The Data Science Design Manual](#)

[An Introduction to Statistical Learning \(ISL\)](#)

[Cloud Computing](#)

[Google Cloud Computing Foundations](#)

[Google Cloud: Machine learning and AI learning path](#)

[Ethics and Governance](#)

[Model AI Governance Framework](#)

[Data Science Ethics Course](#)

[Additional Readings](#)

[Practice](#)

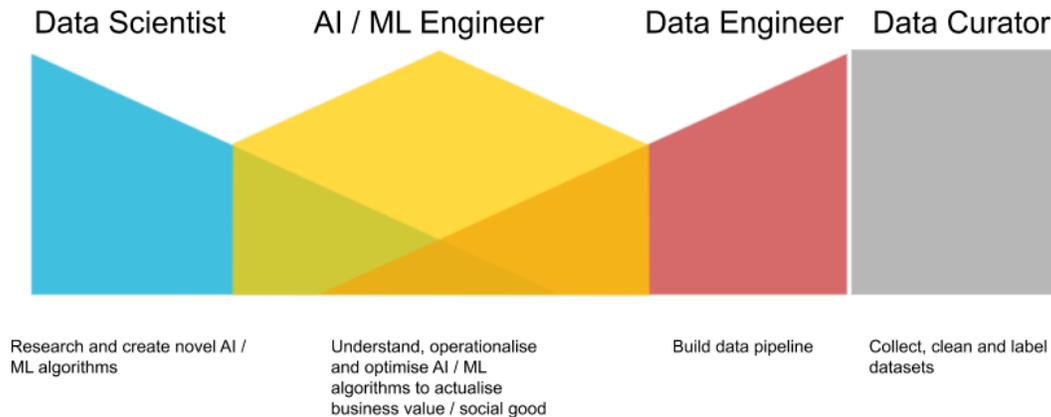
[Competitions](#)

[AI Bricks](#)

[Work with NGOs](#)

[Additional Resources](#)

Introduction : Who is an AI engineer?



The AI (or ML) engineer is a relatively recent specialisation with a skills set overlapping with those of the *data scientist* and *data engineer*¹. An engineer in general builds things to solve real-world problems. An AI engineer therefore harnesses AI technologies to build AI systems that solve real-world problems.

To be an AI engineer requires strong conceptual understanding of AI/ML algorithms as well as the requisite software engineering skills to operationalise and optimise AI systems in production. He or she makes intelligent use of the research of data scientists (or works in close collaboration with them) and, together with data engineers, builds systems that solve some business or social problem. During this process, the AI engineer builds an actual implementation which transforms a problem in the real world into a quantitative form which can be manipulated to deliver something of value to the organisation within ethical and regulatory boundaries. As can be seen, the AI engineering specialisation demands a wide range of skills and knowledge.

¹ [Data engineers vs data scientists](#)

How to use this Field Guide

The various sections in this guide serve as your curriculum in your journey to become an AI Engineer. With these, you will know what competencies are required in this domain.

As for the learning materials, we have tried our best to find high quality and free resources for you. Some of the topics even have a few alternatives for you to explore. However, these are not the only ones available. We are fortunate that there are many other resources available and many others will be created in the future. You can find resources and materials that can fit your specific needs; based on your background, current knowledge, learning styles etc. All you have to do is a bit of research on the Internet.

Good luck on your journey and happy learning!

Suggested Learning Schedule

To better prepare AIAP™ aspirants for selection into the programme, AISG has charted out a 12-month self-learning roadmap. It is a curation of online courses and resources which provide the essential foundational knowledge grouped into 10 topics. Many of these resources are free to access. Refer to the chart on the following page.

		Months											
SN	Topics	1	2	3	4	5	6	7	8	9	10	11	12
1	Python Programming	█											
2	SDLC, Code versioning, OOP, Software Testing		█	█									
3	Data Structures and Algorithms			█	█								
4	Databases (SQL)				█	█							
5	Computational Thinking					█							
6	Mathematics						█	█	█				
7	Data Science and Machine Learning							█	█	█	█		
8	Cloud Computing									█	█	█	█
9	Ethics and Governance												█
10	Practise								█	█	█	█	█

Python

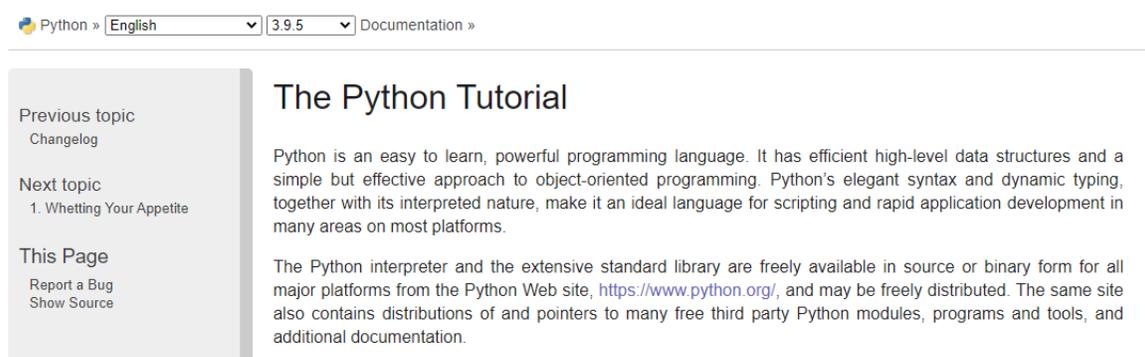
Code is what animates computers. An AI engineer must be able to write, execute and debug code as it is the means to translate concepts into real-world actions.

While there are a few options available in selecting a programming language for AI/ML development, Python remains the first choice for many developers ². Python is known for its simple syntax and having a very strong support community allowing new learners to pick it up easily. It has a rich and extensive set of libraries which allows software developers to build any product that they wish. These libraries include those required for building AI/ML models. Python is also being very actively developed allowing the language to be improved and evolved constantly. Finally, Python is also mature enough that its presence can be found in many IT systems and software.

There is no shortage of online learning materials available to learn Python, with many of them freely accessible. You can always find one that will be best suited to your needs, regardless of whether you have no programming experience or if you are already an expert in another programming language. Here, we will highlight a couple of resources for the beginner:

[The Python Tutorial](#)

This is the official Python tutorial. This tutorial does not attempt to be comprehensive and cover every single feature, or even every commonly used feature. Instead, it introduces many of Python's most noteworthy features, and will give you a good idea of the language's flavor and style.



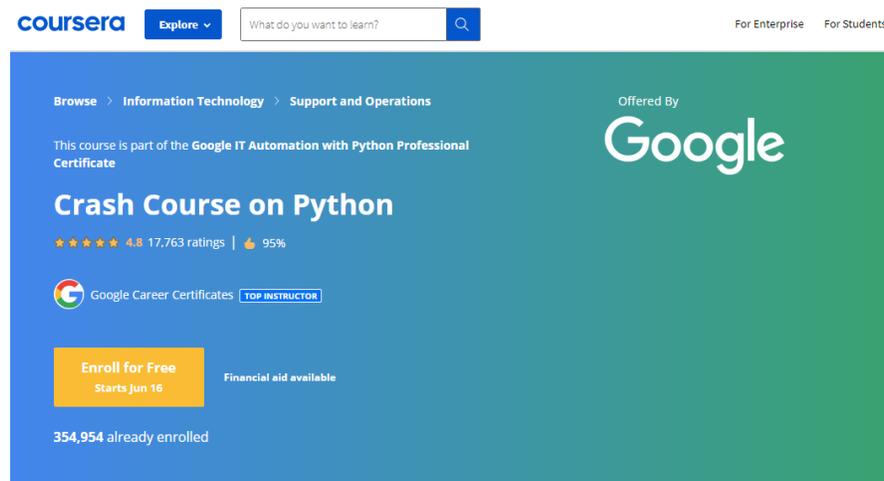
The screenshot shows the top of the Python documentation page. At the top, there is a navigation bar with "Python »", a language dropdown set to "English", a version dropdown set to "3.9.5", and "Documentation »". Below this is a sidebar on the left with the following links: "Previous topic: Changelog", "Next topic: 1. Whetting Your Appetite", and "This Page: Report a Bug, Show Source". The main content area has the title "The Python Tutorial" and two paragraphs of text. The first paragraph describes Python as an easy-to-learn, powerful programming language with efficient high-level data structures and a simple but effective approach to object-oriented programming. The second paragraph states that the Python interpreter and its extensive standard library are freely available in source or binary form for all major platforms from the Python Web site (<https://www.python.org/>), and may be freely distributed. The same site also contains distributions of and pointers to many free third-party Python modules, programs and tools, and additional documentation.

² <https://towardsdatascience.com/top-programming-languages-for-ai-engineers-in-2020-33a9f16a80b0>

[Crash Course on Python](#) by Google

Targeted as novices, this course is designed to teach you the foundations in order to write simple programs in Python using the most common structures. By the end of this course, you'll understand the benefits of programming in IT roles; be able to write simple programs using Python; figure out how the building blocks of programming fit together; and combine all of this knowledge to solve a complex programming problem.

Note: This course is offered on the [Coursera](#) platform. For a large number of courses, you can access most of the learning materials for free using the 'audit' mode. If you are interested in accessing the graded assignments and earning a certificate (from Google in this case), you can purchase the Certificate Experience either before or after the audit.



Video Tutorials

Youtube provides another source of excellent learning videos. Some creators of these content may have accompanying websites or code repositories that the learner can use to follow along. The channels and playlists listed below (in no particular order) are just some of the more popular ones for learning Python.

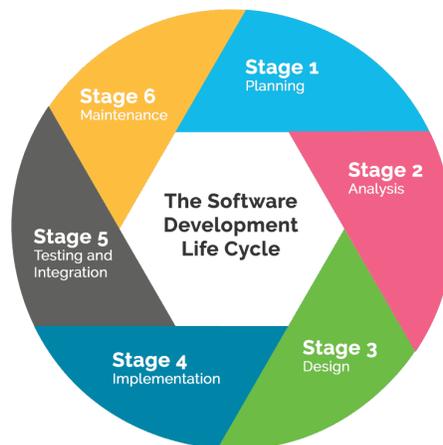
- [Python Tutorial for Beginners](#) by Telusko
- [Python Tutorials for Absolute Beginners](#) by CS Dojo
- [Python Programming Beginner Tutorials](#) by Corey Schafer

Software Engineering

Once you have built a strong foundation of a programming language, you should then learn about the various aspects of building a program. As an AI Engineer, you are not only expected to build accurate, reproducible ML models but also good quality, fault tolerant and well designed applications. The following learning resources will help you build those skills in some of the most key aspects of software development.

Software development life cycle

A systems development life cycle (SDLC) is composed of a number of clearly defined and distinct work phases which are used by systems engineers and systems developers to plan for, design, build, test, and deliver information systems³. Knowing SDLC allows you to integrate well into an existing development team or help you plan and structure the project you are working on.



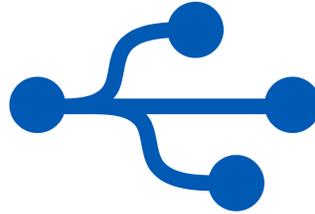
[Software Development Processes and Methodologies](#) by the University of Minnesota.

This course is offered on the Coursera Platform. In this course, you will get an overview of how software teams work? What processes do they use? What are some of the industry standard methodologies? What are the pros and cons of each? You will learn enough to have meaningful conversation around software development processes. If you would like to learn more about Software Development, this course is part of the [Software Development Lifecycle Specialization](#) offered by this university.

³ https://en.wikipedia.org/wiki/Systems_development_life_cycle

Code Versioning

Version control, also known as source control, is the practice of tracking and managing changes to software code. Version control systems are software tools that help software teams manage changes to source code over time. As development environments have accelerated, version control systems help software teams work faster and smarter⁴.



Two options are presented here for you to choose, both targeted at beginners.

- [Introduction to Git and GitHub](#) by Google (a Coursera course)
- [Version Control with Git](#) on Udacity

Object Oriented Programming (OOP)

Python is a flexible programming language that gives you the option of using one of many programming paradigms. OOP is one such paradigm commonly used. It is based on the concept of “objects” that contain both data and code. At this beginning part of your learning journey, it is not necessary to become highly competent in OOP. However, you may come across examples and projects written in this paradigm. Here, we have presented 2 short tutorials covering the basic concepts of OOP in Python.

- [Object-Oriented Programming \(OOP\) in Python 3](#) by Real Python
- [Learn Object Oriented Programming Basics in 30 Minutes: A Free Crash Course](#) by FreeCodeCamp

Software Testing and Debugging

Software that does not behave consistently or correctly when faced with different scenarios can cause significant issues when deployed. A primary purpose of testing is to detect software failures so that defects may be discovered and corrected. Two resources are provided with different depths of coverage. The first is a tutorial which can serve as a good overview of this

⁴ <https://www.atlassian.com/git/tutorials/what-is-version-control>

topic. The second is a website by Brian Okken, author of “[Python Testing with pytest](#)”. He has listed a number of materials for various different Python testing frameworks for you to explore.

- [Getting Started With Testing in Python](#) by Real Python
- [Python Testing](#) website

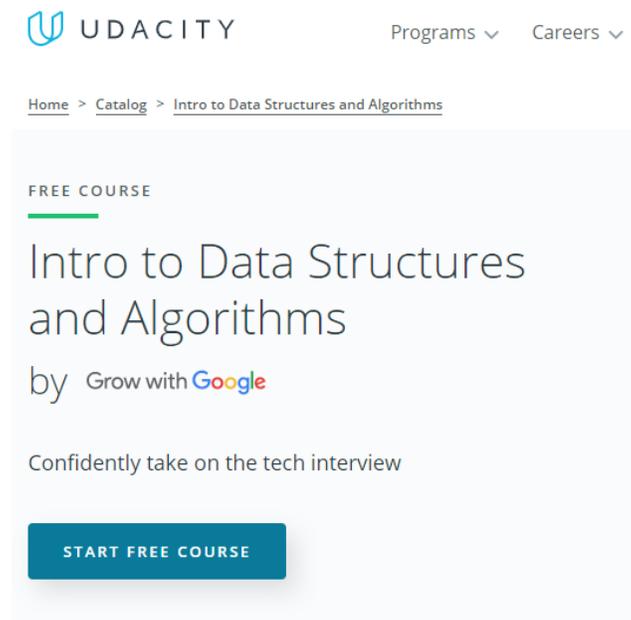
Data Structures and Algorithms

A computer program is a collection of instructions to perform a specific task. For this, a computer program may need to store data, retrieve data, and perform computations on the data.

Data Structures are the programmatic way of storing data so that data can be used efficiently. Algorithm is a step-by-step procedure, which defines a set of instructions to be executed in a certain order to get the desired output. Learning data structures and algorithms allow us to write efficient and optimized computer programs.

[Intro to Data Structures and Algorithms](#) by Google (a Udacity course)

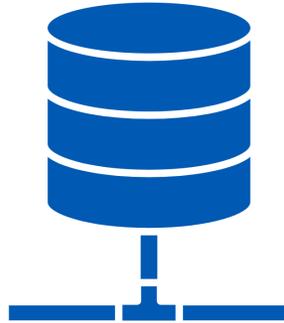
This learning resource covers commonly used data structures found in most programming languages as well basic search algorithms.



The screenshot shows the Udacity website header with the logo and navigation links for 'Programs' and 'Careers'. Below the header is a breadcrumb trail: 'Home > Catalog > Intro to Data Structures and Algorithms'. The main content area features the text 'FREE COURSE' with a green underline, followed by the course title 'Intro to Data Structures and Algorithms' in a large font. Below the title, it says 'by Grow with Google'. A sub-headline reads 'Confidently take on the tech interview'. At the bottom of the card is a blue button with the text 'START FREE COURSE'.

Databases

A database is an organized collection of data, generally stored and accessed electronically from a computer system⁵. With the enormous amount of data that is available and being constantly generated, an effective, fast and reliable database system is the key that makes AI possible.



If a database is the “Housing complex” where data lives, then SQL (Structured Query Language) is like the address book that allows you to find someone quickly.

3 learning resource options are listed for your selection. The first resource gives you a more in depth understanding about database systems. If you are more interested in Data Engineering, this would be a good place to start as you will need to interact with Databases often.

The second resource focuses on SQL, the ‘language of data’. This course is hosted on [Khan Academy](#), which is a well known, popular online learning platform that focuses on students of various levels. For some, this could be a very good, gentle introduction to the topic.

The third resource also focuses on SQL but using the perspective of querying data for data analysis.

- [Database Management Essentials](#) by University of Minnesota (Coursera)
- [Intro to SQL: Querying and managing data](#) on Khan Academy
- [SQL for Data Analysis](#) by Mode (Udacity)

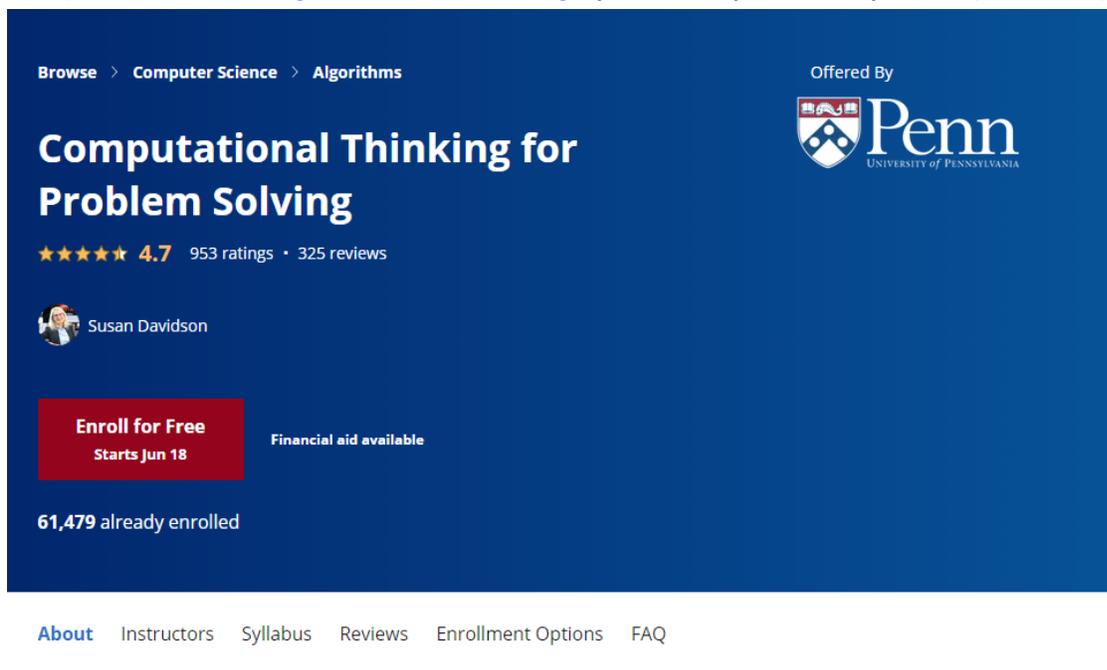
⁵ <https://en.wikipedia.org/wiki/Database>

Computational Thinking

Computational thinking (CT) is a set of problem-solving methods that involve expressing problems and their solutions in ways that a computer could also execute. It involves the mental skills and practices for designing computations that get computers to do jobs for people, and explaining and interpreting the world as a complex of information processes⁶.

Computational thinking is critical if you want to become a good AI Engineer as you will be asked to solve problems that do not have obvious solutions.

[Computational Thinking for Problem Solving](#) by University of Pennsylvania (Coursera)



Browse > Computer Science > Algorithms

Offered By



Computational Thinking for Problem Solving

★★★★★ 4.7 953 ratings • 325 reviews

 Susan Davidson

Enroll for Free
Starts Jun 18

Financial aid available

61,479 already enrolled

[About](#) [Instructors](#) [Syllabus](#) [Reviews](#) [Enrollment Options](#) [FAQ](#)

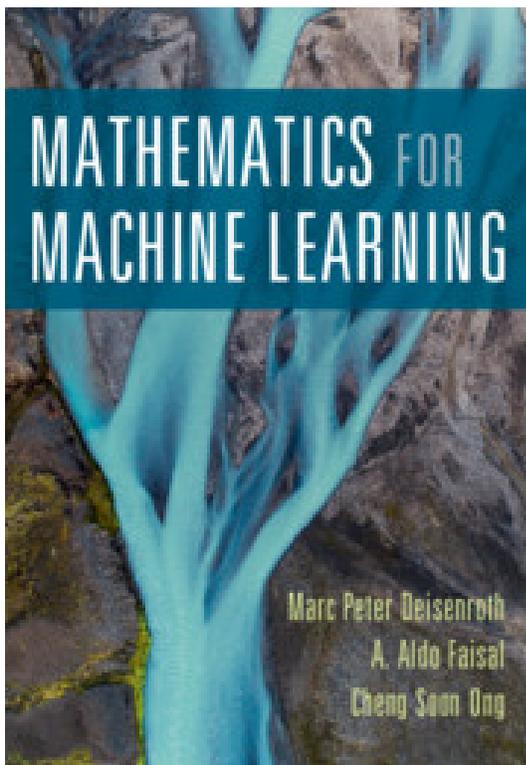
⁶ https://en.wikipedia.org/wiki/Computational_thinking

Mathematics

Mathematics is needed if you want to truly understand how various Machine Learning algorithms work. Fortunately, you do not need advanced mathematics skills to do so. Use these resources to brush up on key concepts in areas such as Linear Algebra and Calculus. Additionally, these resources have been developed specifically with Machine Learning in mind rather than being a generic course in Mathematics.

The first resource is similar to previous recommendations in that it is an online course. The second resource is an alternative where the authors, who are respected academics in their field, have made their book available for free. You can download a PDF copy from their website which also contains additional resources to support your learning.

- [Mathematics for Machine Learning Specialization](#) by Imperial College London (Coursera). You should focus on the first 2 courses.
- [Mathematics for Machine Learning](#) free eBook



Data Science and Machine Learning

[Machine Learning Crash Course](#) by Google

Developed by Google, this free online course features a series of lessons with video lectures, real-world case studies, and hands-on practice exercises. The course does not presume or require any prior knowledge in machine learning. However, to understand the concepts presented and complete the exercises, the learning and preparation that you have done up to this point will very useful.

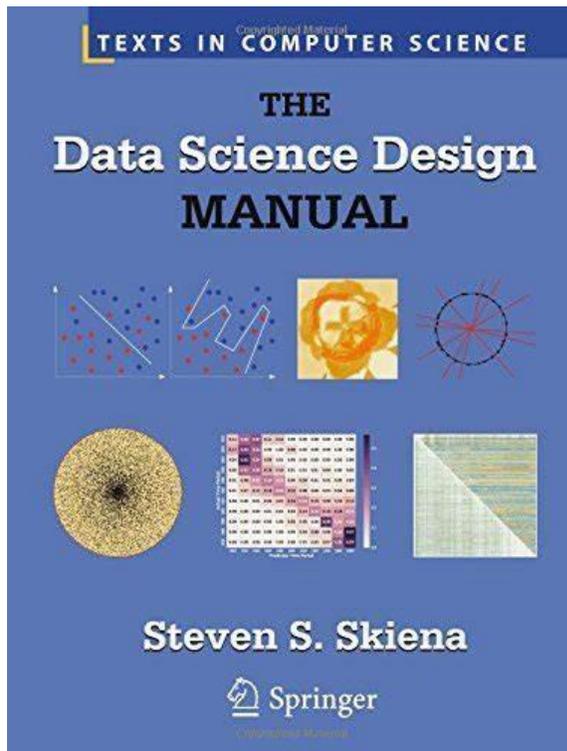


A self-study guide for aspiring machine learning practitioners

Machine Learning Crash Course features a series of lessons with video lectures, real-world case studies, and hands-on practice exercises.

The Data Science Design Manual

An **excellent** introduction and foundation course in Data Science and Machine Learning by Professor [Steven Skiena](#) who is the Distinguished Teaching Professor of Computer Science at Stony Brook University. His course and book [The Data Science Design Manual](#) provides an excellent introduction with interesting war stories. Additional resources including data sets for projects and assignments can be found at the [website](#) of the book.



With the kind permission of Prof Skiena, AI Singapore has mounted his lectures as a course module [here](#).

Table of Content :

1. What is Data Science
2. Mathematical Preliminaries
3. Data Munging
4. Scores and Ranking
5. Statistical Analysis
6. Visualizing Data
7. Mathematical Models
8. Linear Algebra
9. Linear and Logistic Regression
10. Distance and Network Methods
11. Machine Learning
12. Big Data

An Introduction to Statistical Learning (ISL)

Written by well respected academics, this is one of the most known books in Machine Learning. It focuses on explaining algorithms and techniques which are statistical based. Similar to the Mathematics learning resource earlier, the authors have made their book available for [free download](#). In addition, video recording of a course taught at Stanford by two of the authors (together with various course materials) is also available [here](#).

Note that the book uses R. However, the equivalent Python notebooks kindly contributed to the community by Jordi Warmenhoven can be found [here](#).

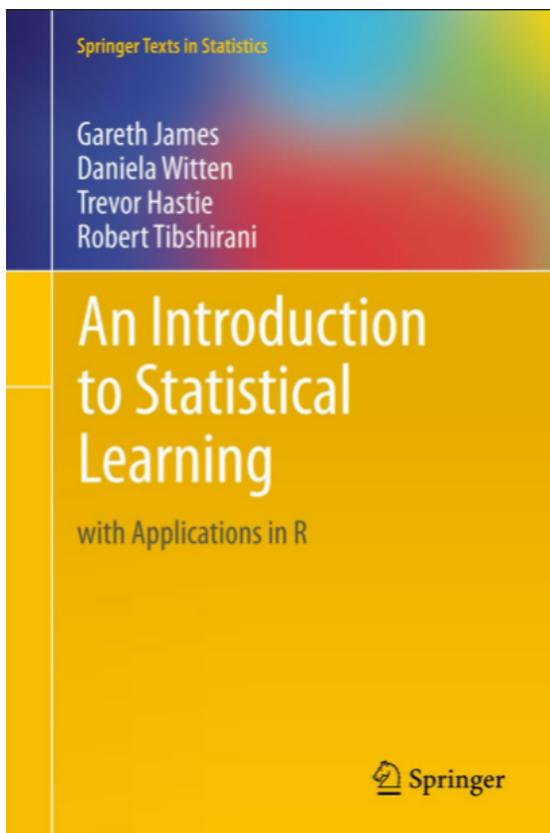


Table of Content :

1. Introduction
2. Statistical Learning
3. Linear Regression
4. Classification
5. Resampling Methods
6. Linear Model Selection and Regularization
7. Moving Beyond Linearity
8. Tree-based Methods
9. Support Vector Machines
10. Unsupervised Learning

Cloud Computing

Cloud computing is the delivery of on-demand computing services -- from applications to storage and processing power -- typically over the internet and on a pay-as-you-go basis. Rather than owning their own computing infrastructure or data centers, companies can rent access to anything from applications to storage from a cloud service provider.

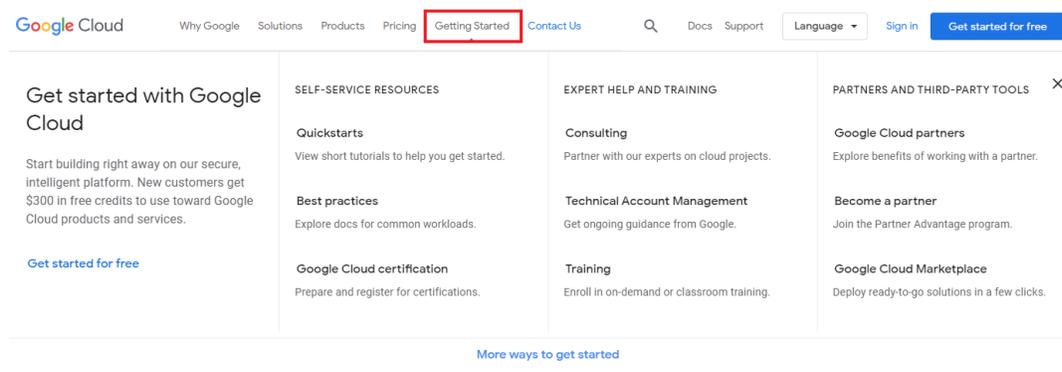
Some of the more well known cloud service providers are:

- [Google Cloud](#)
- [Amazon Web Services](#)
- [Microsoft Azure](#)
- [Alibaba Cloud](#)

This [report](#) by Gartner describes the state of cloud computing as of 2020 if you are interested in understanding the landscape.

For this Field Guide, we will focus on Google Cloud as it is user friendly and has lots of learning resources available. Alternatively, you may want to learn about the provider used by your organization instead.

To begin learning, you will need to create an account on [Google Cloud](#). To encourage new users, Google gives USD300 free credits which allows you to explore the platform. In addition, you can use more than 20 products and services for free (up to a monthly limit)⁷.



The screenshot shows the Google Cloud website's 'Getting Started' page. The navigation bar at the top includes 'Why Google', 'Solutions', 'Products', 'Pricing', 'Getting Started' (highlighted with a red box), and 'Contact Us'. Below the navigation bar, the page is organized into four main sections:

- Get started with Google Cloud:** Includes a 'Get started for free' button and text about new customer credits.
- SELF-SERVICE RESOURCES:** Contains 'Quickstarts' (short tutorials), 'Best practices' (docs for common workloads), and 'Google Cloud certification' (prepare and register).
- EXPERT HELP AND TRAINING:** Contains 'Consulting' (partner with experts), 'Technical Account Management' (ongoing guidance), and 'Training' (on-demand or classroom).
- PARTNERS AND THIRD-PARTY TOOLS:** Contains 'Google Cloud partners' (benefits of working with a partner), 'Become a partner' (Partner Advantage program), and 'Google Cloud Marketplace' (ready-to-go solutions).

A 'More ways to get started' link is located at the bottom of the page.

You can actually learn about Google Cloud from its various documentation (shown above). However, if you prefer a more structured learning experience, we recommended the following resources.

⁷ <https://cloud.google.com/free> Information accurate as of June 2021

Google Cloud Computing Foundations

This is a set of four courses that will give an overview of concepts important to Cloud Computing and how Google Cloud fits in. After this learning, you will have a good understanding of all the various products and services available.

Note: The learning resources in these sections are delivered on the [Qwiklabs](#) platform. The hands-on lab exercises may require the purchase of credits. However as a new learner, you may be able to access them for a limited time.

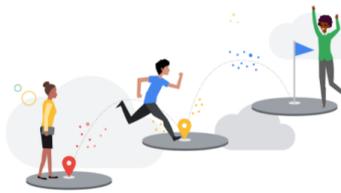


Build Google Cloud skills Cloud learning from virtually anywhere.

Ready to start learning the skills needed for high demand fields like cloud computing? Get one month of free access to introductory training across cloud computing, big data, and machine learning. Sign Up for your free account or sign in to your existing account, and get one month access when you enroll:

1m-GwG-GCCF-967 [Copy code](#)

[New to Qwiklabs? Take the tour](#) →



Google Cloud Computing Foundations

Complete this four-course series in order. Having trouble with the code? We can help. Watch this [4-step how-to video](#).

Google Cloud: Machine learning and AI learning path

This learning path from Google consists of 5 courses which you can take on Coursera and 21 labs on the Qwiklabs platform. This path allows you to go deeper into the Google products such as data storage, pipelines, computes etc. that are typically used for building AI/ML products.

Machine learning and artificial intelligence

Did you know that the adoption of machine learning results in 2x more data-driven decisions, 5x faster decision-making, and 3x faster execution?¹

Learn how to implement the latest machine learning and artificial intelligence technology by exploring training on BigQuery, Datalab, TensorFlow, Cloud Vision, Natural Language API, and more.

[Pick your path](#) ▾



Ethics and Governance

It is important that AI be built and used ethically, fairly and responsibly. An AI engineer should be aware of these principles.

Model AI Governance Framework

Singapore released the first edition of the Model AI Governance Framework in 2019 to provide guidance on key ethical and governance issues when deploying AI solutions. The second and latest edition was released on 21 January, 2020.

- <https://www.imda.gov.sg/infocomm-media-landscape/SGDigital/tech-pillars/Artificial-Intelligence>
- <https://www.pdpc.gov.sg/Resources/Model-AI-Gov>

AISG is leveraging this framework as part of its engineering best practices for the 100E projects. The real-world experience from the projects delivered will be used to refine the framework further.

Data Science Ethics Course

Enroll in an online course to learn about issues of data privacy and algorithmic bias and fairness.

- <https://www.edx.org/course/data-science-ethics>

Additional Readings

- Microsoft: [AI Principles & Approach from Microsoft](#)
- Microsoft: [FATE: Fairness, Accountability, Transparency, and Ethics in AI](#)
- Microsoft: [Responsible bots: 10 guidelines for developers of conversational AI](#)
- Microsoft: [The Future Computed](#) (eBook)
- Google: [Responsible AI Practices](#)
- DeepMind: [Safety & Ethics](#)
- [Interpretable Machine Learning](#)
- [Fairness and machine learning](#)

Practice

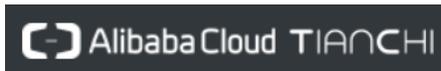
Put learning to practice by building an actual AI project. This can be through joining a competition platform or, even better, building your own real-world application.

Competitions



[Kaggle](#) is the most popular AI/ML competition in the industry, and has served as a starting point for many practitioners. It also allows you to discuss with, learn from, and benchmark against other aspiring and experienced AI engineers and data scientists.

Below is a snapshot of other popular competitive platforms.

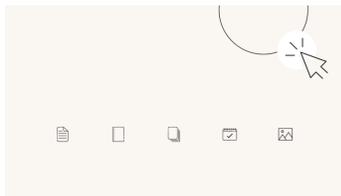


AI Bricks



AI Singapore's very own [AI Bricks](#) platform contains curated tools and resources for you to get started on solving your own real-world AI problems.

As of June 2021, we have the following collections:



TagUI

TagUI is our open source full-featured desktop RPA tool. It helps you automate your repetitive tasks, such as data acquisition and testing of web apps.



Speech Lab

Speech recognition helps you convert audio to text. This is our uniquely developed code switching speech engine which can recognise English, Mandarin and Singlish



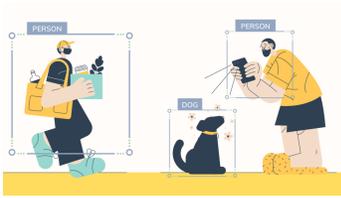
Fine Pose

Social distancing app that utilises human pose estimation.



CUDO

CUDO (Collaborative Urban Delivery Optimisation) is our resource planning and scheduling tool for logistics service providers.



Computer Vision Hub

Our open source tools for Computer Vision



AI-Ready Bricks

Plug-and-play tools built on machine learning (ML) platforms.



Natural Language Processing Hub

Our open source tools for Natural Language Processing



Synergos

Our open-source platform for Federated Learning

[Contact](#) AI Singapore if you would like to use any of these Bricks to build your project.

Work with NGOs



[DataKind](#) brings together data scientists and local nonprofit organisations who require technical help. You can do good while honing your skills and meeting like-minded enthusiasts in person.

Additional Resources

To supplement the journey, the following materials are also recommended. Note that some of the books do get updated regularly.

Learning Python

- [Python Crash Course, A Hands-On, Project-Based Introduction to Programming](#) by Eric Matthes
- [Learning Python](#) by Mark Lutz
- [Fluent Python: Clear, Concise, and Effective Programming](#) by Luciano Ramalho
- [Learn Python the Hard Way](#) by Zed Shaw, New York: Addison-Wesley (<https://learnpythonthehardway.org/>)

Python for Data Analysis

- [Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython](#) by Wes McKinney
- [Pandas for Everyone: Python Data Analysis](#) by D Chen D (2017). New York: Addison-Wesley.
- [Python Data Science Handbook: Essential Tools for Working with Data](#) by Jake VanderPlas

Machine Learning

- [The Hundred-Page Machine Learning Book](#) by Andriy Burkov
- [The Hundred-Page Machine Learning Engineering Book](#) by Andriy Burkov
- [Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition](#) by Aurelien Geron
- [Deep Learning with Python](#) by Francois Chollet
- [Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence](#) by Jon Krohn , Beyleveld Grant
- [Deep Learning Book](#) at <http://www.deeplearningbook.org/> (Advanced)

AI (Beyond Machine Learning)

- [Artificial Intelligence: A Modern Approach](#) by Stuart Russell

General AI

- [On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines](#) by Jeff Hawkins
- [Weapons of Math Destruction](#) by Cathy O'Neil
- [Rise of the Robots: Technology and the Threat of a Jobless Future](#) by Martin Ford
- [Human Compatible](#) by Stuart Russell
- [Architects of Intelligence](#) by Martin Ford

- *Rebooting AI – Building Artificial Intelligence We Can Trust* by Gary Marcus, Ernest Davis
- *Automating Inequality* by Virginia Eubanks