DS/AI

Self-Starter Handbook

BUILD YOUR OWN ROADMAP

Ankit Rathi

From a time around when DS/AI field started picking up, every other day I get at least 8–10 messages from DS/AI starters & enthusiasts on ‘How can I get into DS/AI field?’. Over a while, I have improvised my response based on the follow-up questions they ask like:

1. What is the difference between DS, ML, DL, AI, DM?

2. What are the roles in DS/AI, who does what?

3. What concepts, processes & tools they need to learn?

4. Which books, courses, etc they need to refer to?

5. How to build a DS/AI portfolio?

6. How to write a resume for DS/AI?

7. How to build a helpful network?

8. How to search for the job?

9. How to prepare for the interview?

10. How to stay up to date in this still-evolving field?

You can notice that these questions are not conceptual ones and there is no dedicated material to address these roadblocks. I thought why not to build a framework or a road-map for DS/AI starters and enthusiasts so that I need not to answer the same type of questions again and again. And that is when I started documenting what a starter or enthusiast need to do step by step in order to reach a level when he is ready to tackle any challenge thrown to him. My answer to the above questions in a structured way to help DS/AI starters & enthusiasts is this book. This book covers the framework to launch your DS/AI career in 8 chapters.

Ankit Rathi provides unique combination of Data Engineering (DB/ETL/DWH/BI)/Architecture (Data Management & Governance) & Data Science (ML/DL/AI) with more than a decade of demonstrated history of working in IT industry using Data & Analytics. His interest lies primarily in building end to end DS/AI applications/products following best practices of Data Engineering and Architecture.

In his free time, he blogs about various topics on DS/AI field & tries to simplify it for starters & enthusiasts.

ankitrathi.com

Rs 499/-

ISBN 978-1-64587-432-4

DS/AI Self-Starter Handbook

Build Your Own Roadmap

Ankit Rathi

To my wife, Divya, who's always accepted me the way I am and supported my hustle, drive & ambition.

To my children, Aarsh & Driti, who are the reason to wake up every morning and work as hard as I can.

DS/AI Self-Starter handbook is a great resource for aspirants starting in the space of Data Science. It covers approach and useful resources that can help in your learning journey and written by one who himself is an Data Science practitioner. I recommend this to anyone who are aspiring to get into Data Science and are looking for insights on how and where to get started.

Srivatsan Srinivasan
Chief Data Scientist (Cognizant)

Wow, this is very impressive! It has taken some time to review, but WOW!
I should have had you as a co-author next time!!!

T. Scott Clendaniel
Chief Data Scientist (Legg Mason)

To be great data scientist you should emphasis on skillset and mindset. Where a lot of book that give you skill set, this is the first book I read that dedicating to shape data scientist mindset.

Nabih Ibrahim Bawazir
Data Science Head (Datanest)

Extremely laudable & heroic attempt to put all your thoughts and experience together to help people.

Sumit Pal
Big Data Architect (Qcentive)

Ankit has done a great job summarizing what is possibly one of the toughest and most frequently asked questions, "How to get started with data science?". Packed with information, this book will definitely be helpful for people from both academia and industry looking to get started on their own Data Science and AI journey.

Dipanjan Sarkar
Data Scientist (Rad Hat)

I think it is a brilliant book for starting Career in Data Science as New Entrants to Data Science often deviate from Path to reach End Goal and this Book tries to solve that Problem in a easy way. I would really like to Congratulate Ankit for Providing Data Science Career Steps in this useful manner.

Yatin Bhatia
Data Scientist (RxLogix)

An indispensable guide and a valuable resource for anyone seeking to enter the field of Data Science. Replete with great advice directly from the author's personal experience.

Parul Pandey
Data Science Evangelist (H2O.ai)

This book kicks you into the right direction definitely worth reading for the beginners trying to break into DS/AI.

Avik Jain
Machine Learning Intern (EMA Solutions)

If you are one among people struggling to identify the right book for data science, this book would probably help to understand where to start, how to prepare, how to develop the habit of continuous learning.

Vishnu Durgha Prasaad
Data Science Practitioner

About the Author

Ankit Rathi is currently working as a Lead Architect-DS/AI at SITA aero. He is a Data Science (ML/DL/AI) practitioner with more than a decade of demonstrated history of working in IT industry using Data & Analytics. His interest lies primarily in the theory & application of artificial intelligence, particularly in developing business applications for machine learning and deep learning. Ankit's work at SITA aero has revolved around designing FlightPredictor product & building the CoE capability. During his tenure as a Principal Consultant at Genpact HCM, Ankit architected and deployed machine learning pipelines for various clients across different industries like Insurance, F&A. He was previously a Tech Lead at RBS IDC where he designed and developed various data intensive applications in AML & Mortgages area. Ankit is a well-known author for various publications (Towards Data Science, Analytics Vidhya etc) on Medium where he actively contributes by writing blog-posts on concepts & latest trends in Data Science. His blog-series on 'Probability & Statistics for Data Science' has been well received by Data Science community in 2018. He is followed by around 30K data science practitioners & enthusiasts on LinkedIn.

U0.1: Webpage: *https://www.ankitrathi.com/*

Table of Contents

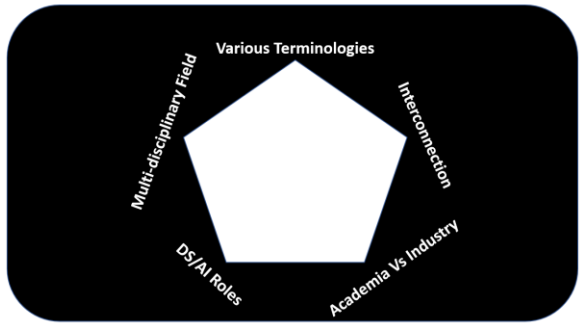
DS/AI: Self-Starter Kit

Build Your Own Roadmap

Chapter 2

Navigating the Landscape

DS/AI is a complex and evolving field. The first challenge a DS/AI aspirant faces is understanding the landscape and how he could navigate through it. Consider this, if you are travelling to a new city, and if you don't have the map, you will have trouble to navigate the city and you will need to ask a lot of random people during your travel without knowing how much they know about the place. Similarly, all the newcomers to data science have this trouble, and there are two ways to deal with this, arrange the map (or a guide) or travel yourself and learn with experience.



Navigating the Landscape

DS/AI: Self-Starter Kit

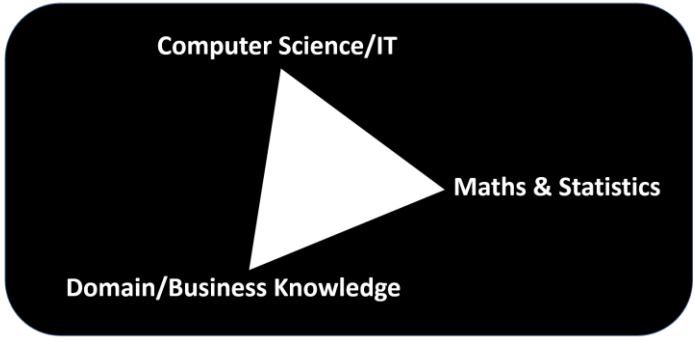
This chapter intends to serve as a map of DS/AI field.

You might have heard data science, machine learning, deep learning, artificial intelligence etc terminology but might not be fully aware of these terms, what to use when and how these topics are interconnected. After going through this chapter, you should be able to understand what is where in DS/AI field.

2.1 Multi-disciplinary field

DS/AI is a multidisciplinary field with sub-fields of study in Math/Statistics, CS/IT & Business/Domain knowledge.

Math/Statistics is required to understand the data and relationship between data elements. CS/IT skills are required to process the data to generate insights. And Business or domain knowledge is required to apply above to skills in the context of a business problem.



Multi-disciplinary Field
Navigating the Landscape

Computer Science/IT

Programming is an essential skill to become a data scientist but one needs not be a hard-core programmer to learn DS/AI. Having familiarity with basic concepts of programming will ease the process of learning data science programming tools like Python/R. These basic concepts of

programming should help a candidate get a long way on the journey to pursue a career in DS/AI as it is all about writing efficient code to analyse big data and not being a master of programming. Individuals should learn the basics of programming in Python/R (or any relevant language) before they begin to work on DS/AI problems/projects.

Maths & Statistics

Data science teams have people from diverse backgrounds like chemical engineering, physics, economics, statistics, mathematics, operations research, computer science, etc. You will find many data scientists with a bachelor's degree in statistics and machine learning but it is not a requirement to learn DS/AI. However, having familiarity with the basic concepts of Math and Statistics like Linear Algebra, Calculus, Probability, etc. is important to learn DS/AI.

Domain/Business Knowledge

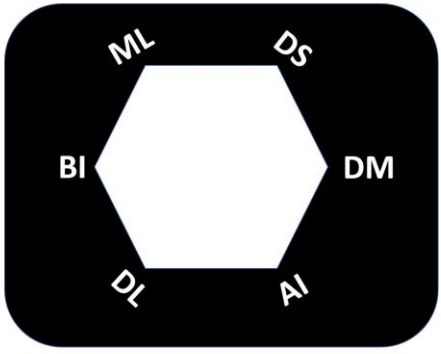
Subsequently, the business knowledge that the data scientists would need to have would be related to the domain that the project/analysis is in. For instance, if the data scientist is working for a credit card department in a bank, it will need to understand the specific business definitions, regulations, accounting policies & international standards, processes etc. This is the part that is more specific to the organization the data scientist is deployed in.

In my view, one thing to take care while the hiring data scientists is not to give huge preference to domain knowledge. This may severely limit the supply of data science talents to the organization. You would have a better chance of getting more value from data science by looking for those that are strong in math & programming, being able to convert business objectives to mathematical models. Based on my observation, this is a much more difficult skill to find or train, as compared to domain knowledge.

U02.1: Multi-disciplinary field: *https://datajobs.com/what-is-data-science*

2.2 Various Terminologies

As a DS/AI starter, you will come across many similar terminologies. First thing you need to do is to understand what each term means and where each fits in the bigger picture. Data Science, Business Intelligence, Data Mining, Machine Learning, Deep Learning, Artificial Intelligence; let's have a look at Wikipedia definition for each term & later see how these are interconnected.



Various Terminologies
Navigating the Landscape

Data Science

Data Science is an interdisciplinary field, where we try to solve a business problem by getting insights from available data after applying scientific methods on it.

Business Intelligence

Business intelligence is a technology-driven process of converting raw data into meaningful information which can be used by business.

Data Mining

Data mining is a process of collecting & discovering patterns in the data.

Machine Learning

Machine learning is a field where we develop predictive models that are generic and can be applied to any domain related data problem.

Deep Learning

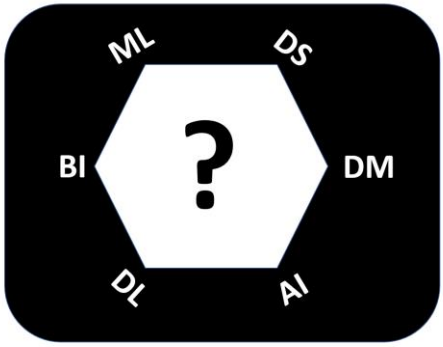
Deep learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms.

Artificial Intelligence

Artificial intelligence is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals.

U02.2: Various Terminologies: *https://www.dezyre.com/article/data-science-compared-with-different-analytics-disciplines/175*

2.3 Interconnection



Interconnection
Navigating the Landscape

Business Intelligence is a plain technology-driven process to analyse and provide actionable intelligence to executives/managers.

Data mining uses statistics and other programming languages to find hidden patterns in the data to explain a certain phenomenon. It helps in building a perception about the data using both math and programming.

Machine Learning deploys data mining techniques as well as other algorithms to develop models of what is happening behind some data to forecast future outcomes.

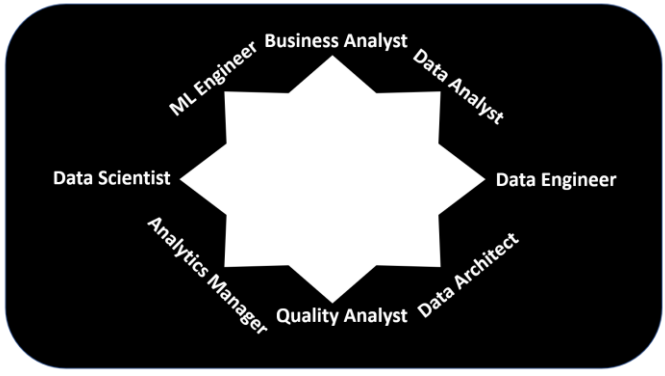
Artificial Intelligence uses models developed by Machine Learning and other algorithms to lead to intelligent behaviour. AI is very much programming based.

- Business Intelligence provide actionable insights
- Data Mining demonstrates patterns
- Machine Learning forecasts with models
- Artificial Intelligence shapes behaviours

So you see that these terms are different but still inter-connected.

U02.3: Interconnection: *https://www.quora.com/What-is-the-difference-between-artificial-intelligence-machine-learning-data-mining-and-business-intelligence-How-they-are-related*

2.4 DS/AI Roles



DS/AI Roles
Navigating the Landscape

Before looking into the skill-set of a data scientist, let's have a look at various roles required to work and deliver a data science project, after all, it's a teamwork.

Every role has its own skills that are critical to data science projects at various stages.

Data Analyst

Data analysts translate numbers into plain English. Every business collects data, whether it's sales figures, market research, logistics, or transportation costs. A data analyst's job is to take that data and use it to help companies make better business decisions. There are many different types of data analysts in the field, including operations analysts, marketing analysts, financial analysts, etc.

Data Scientist

A data scientist is someone who knows how to extract meaning from and interpret data, which requires both tools and methods from statistics and machine learning. She spends a lot of time in the process of collecting, cleaning, and munging data. Domain knowledge is also an integral part of the skill.

Data Engineer

Data Engineers are responsible for the creation and maintenance of analytics infrastructure that enables almost every other function in the data world. They are responsible for the development, construction, maintenance and testing of architectures, such as databases and large-scale processing systems.

Data Architect

Data architects build complex computer database systems for companies, either for the general public or for individual companies. They work with a team that looks at the needs of the database, the data that is available and creates a blueprint for creating, testing and maintaining that data architecture.

Analytics Manager

The data science manager coordinates the different tasks that must be completed by their team for a DS/AI project. Tasks may include researching and creating effective methods to collect data, analyzing information, and recommending solutions to business.

Business Analyst

Data science business analyst converts the business problem statement to a DS/AI problem statement which means what data needs to be analyzed to arrive at the insights. The data would then be reviewed with the technology team and results would be delivered to the business team in the form of insights and data patterns. The business analyst should also be knowledgeable enough to apply various predictive modelling techniques and right model selection for generating insights for the problem at hand.

Quality Analyst

The job of quality analyst includes checking the quality of the training data-set, preparing data-sets for testing, running statistics on human-labelled data-sets, evaluating precision and recall on the resulting ML model, reporting on unexpected patterns in outputs, and implementing necessary tools to automate repetitive parts of the work. Experience in software testing with data quality or DS/ML focus, understanding of

statistics, exposure to Data Science / Machine Learning techniques and coding proficiency in Python, are some of the skills required for the job.

To work on DS/AI projects in any of the above mentioned roles, one needs to have an understanding of the core concepts at a high level but depth is required in the specific area you would be working in.

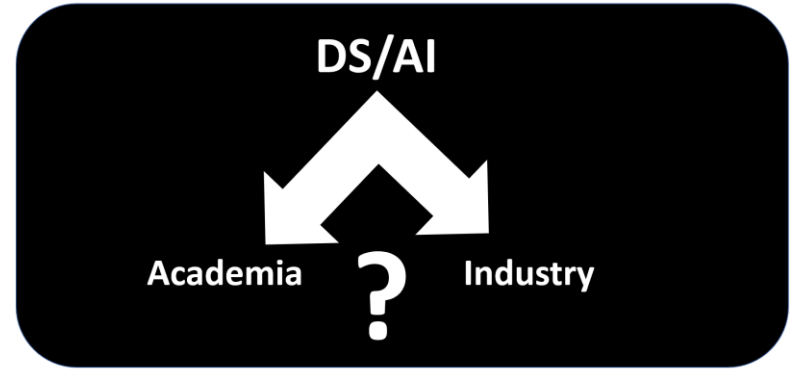
U02.4: DS/AI Roles: *https://hackernoon.com/top-10-roles-for-your-data-science-team-e7f05d90d961*

2.5 Academics Vs Industry

Academia and Industry are different fields with different people and culture. People working in Academia for longer tenure may find it difficult to adjust to industry culture and vice versa.

There is also an academic trap when your career trajectory is so specialized for academia that you're unprepared for a job outside of it.

The academic trap happens in all areas of study, but for this chapter, we focus only on DS/AI students who want to leave academia for data science positions.



Academia Vs Industry

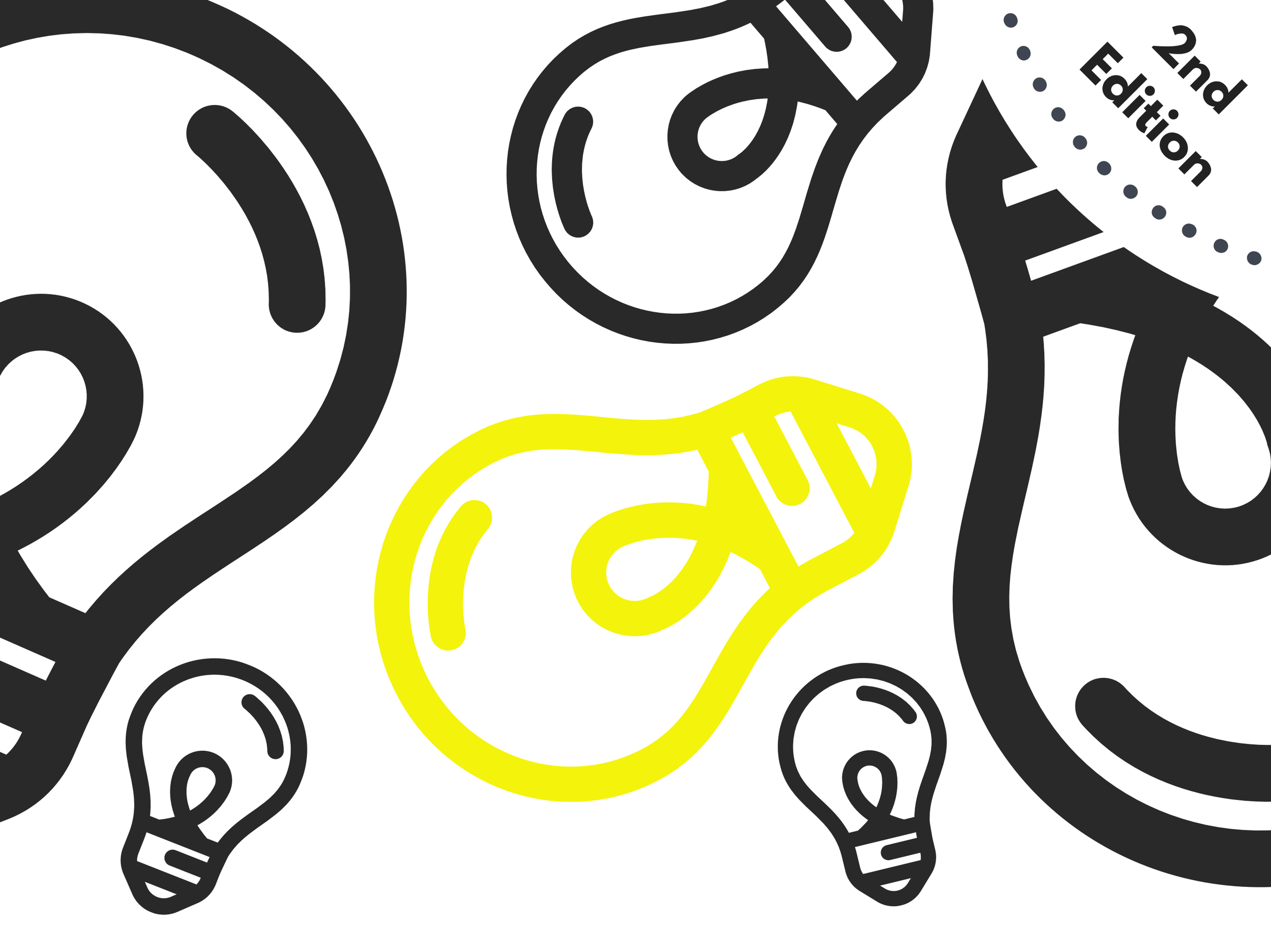
Navigating the Landscape

Further, companies are often hesitant to hire people coming straight from academia for various reasons like:

- In academia, individuals prefer writing papers over internships, making grants over learning programming languages, and not doing the things that could help you in the industry but not academia. The things that are important for academic hirings, such as papers, talks, and grants, are not as important in the industry.
- Working as a data scientist within a corporation requires an understanding of how the business world works, including how quickly deliverable need to be made, how to craft a good presentation, and how to word an email to make a request.
- In academia, you are encouraged to find the most innovative and elegant solution. In industry, you are encouraged to spend as little time as possible to find an analytical solution that just fits the need.
- Salary expectations for advanced degree holders are higher than someone with only an undergraduate degree. This also pushes away recruiters as the industry works in a different way, culture is simply different than the academic one. People coming from academia

need to learn these lessons at their first job, which means that there is a lot of risk for the hiring company.

U02.5: Academia Vs Industry: *https://www.northeastern.edu/graduate/blog/working-in-industry-vs-academia/*



Artificial Intelligence

Self-Starter Handbook

BUILD YOUR OWN ROADMAP

Ankit Rathi

Coming Soon… 2nd Edition

with revised content & 3 more chapters…

ankitrathi.com

From a time around when AI field started picking up, every other day I get many questions from AI starters & enthusiasts on ‘How can I get into AI field?’. Over a while, I have improvised my response based on the follow-up questions they ask like:

- What is AI and why is it important?
- What is the difference between AI, ML, DL, DS, DM, BI?
- What an end-to-end AI project looks like?
- What are the roles in AI projects, who does what?
- What AI concepts & tools you need to learn?
- Which books, courses, channels etc you need to refer to?
- How to practice & build an AI portfolio?
- How to write a resume for an AI role?
- How to build a helpful network?
- How to search for the job?
- How to prepare for the interview?
- How to switch into an AI role (inside or outside)?
- How to lead an AI initiative in your organization?
- How to stay up-to-date in this ever-evolving field?

You can notice that these questions are not conceptual ones and there is no dedicated material to address these roadblocks. I thought why not to build a framework or a road-map for AI starters and enthusiasts so that I need not answer the same type of questions again and again. And that is when I started documenting what a starter or enthusiast need to do step by step in order to reach a level when he is ready to tackle any challenge thrown to him. My answer to the above questions in a structured way to help AI starters & enthusiasts is this book. This book covers the framework to launch your AI career in 11 chapters.

Ankit Rathi is a data & AI architect, published author & well-known speaker. His interest lies primarily in building end to end AI applications/products following best practices of Data Engineering and Architecture.

In his free time, he blogs about various topics on Data & AI field & tries to simplify it for starters & enthusiasts.

ankitrathi.com

₹ 499 | $ 9.99 | € 8.99

ISBN 978-1-64587-432-4