

ECE 20875

Python for Data Science

Chris Brinton, Qiang Qiu, and Mahsa Ghasemi

**(Adapted from material developed by Profs. Milind Kulkarni,
Stanley Chan, Chris Brinton, David Inouye, Qiang Qiu)**

concluding remarks

data science is a lot of things

making predictions
from data

identifying patterns in data

visualizing data

building systems
for data analysis

dealing with
privacy concerns

collecting/organizing data

interpreting data

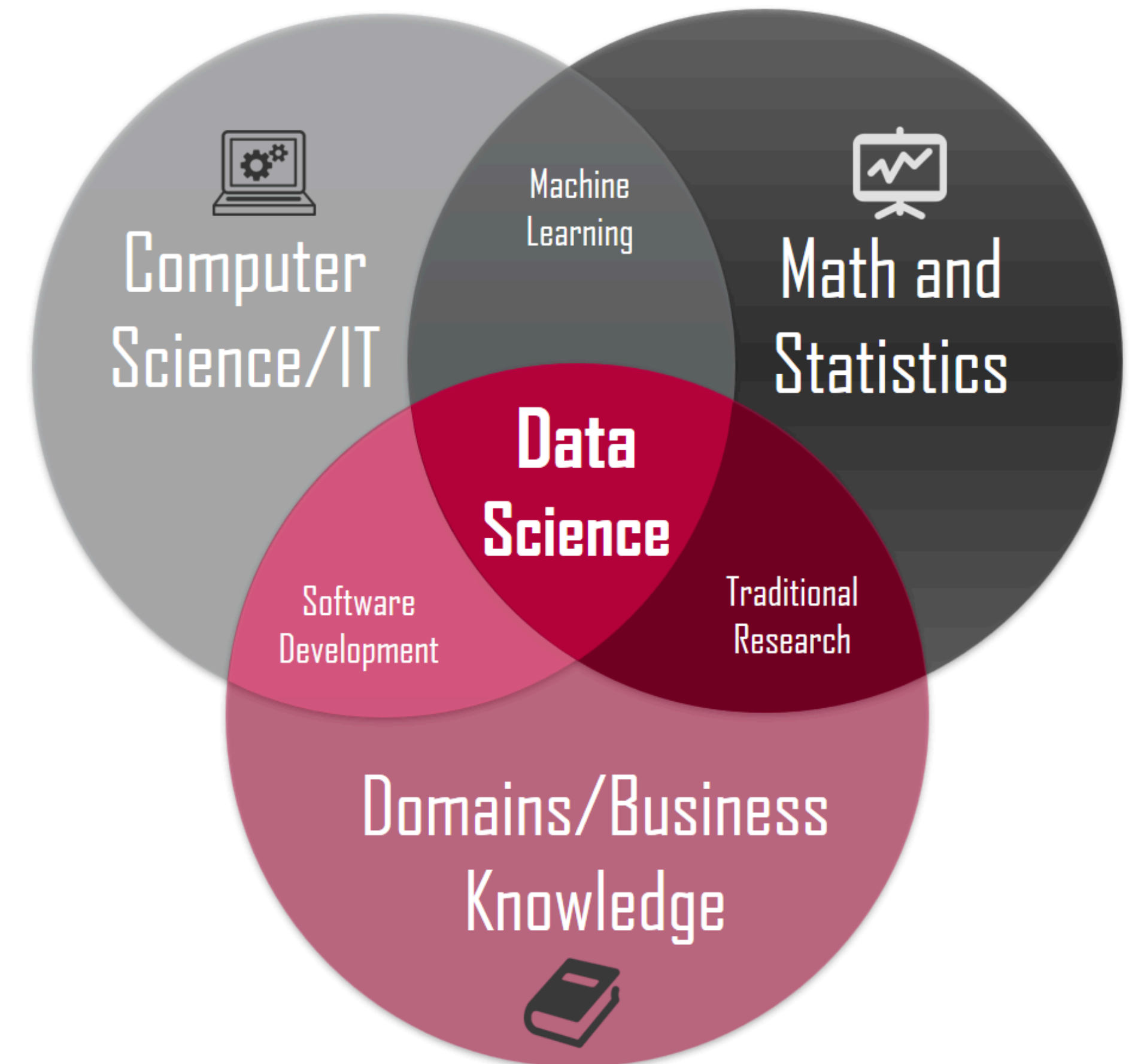
analyzing data

ethics

writing data analyses

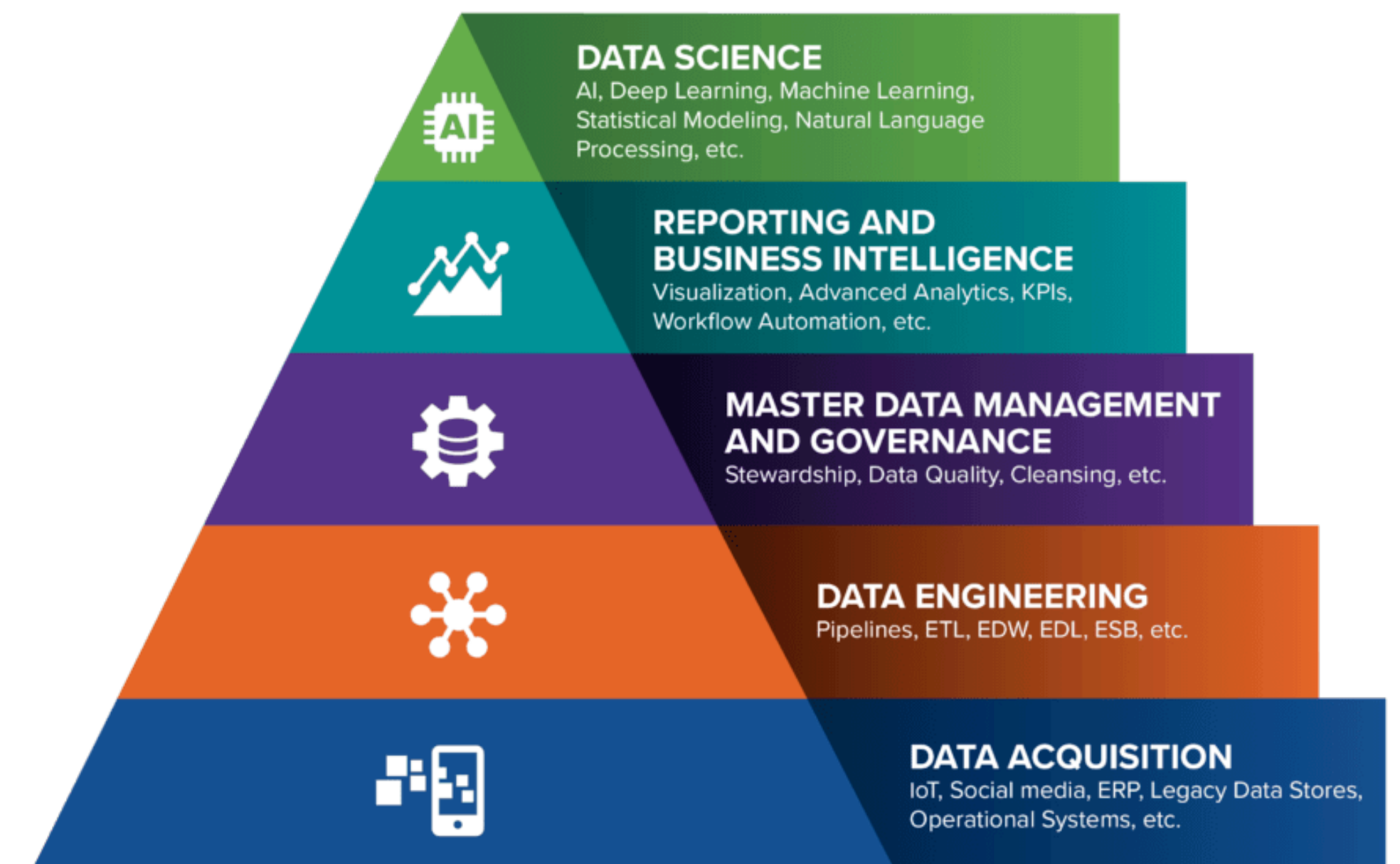
tackling data science problems

- *Here's a dataset, what can we learn from it?*
 - This is a very high-level yet important question
 - Everything we learned in this class is fundamentally about answering this question
 - You now have a lot of tools in your belt to help you answer it!
 - And more importantly, you know the process involved, which will help you



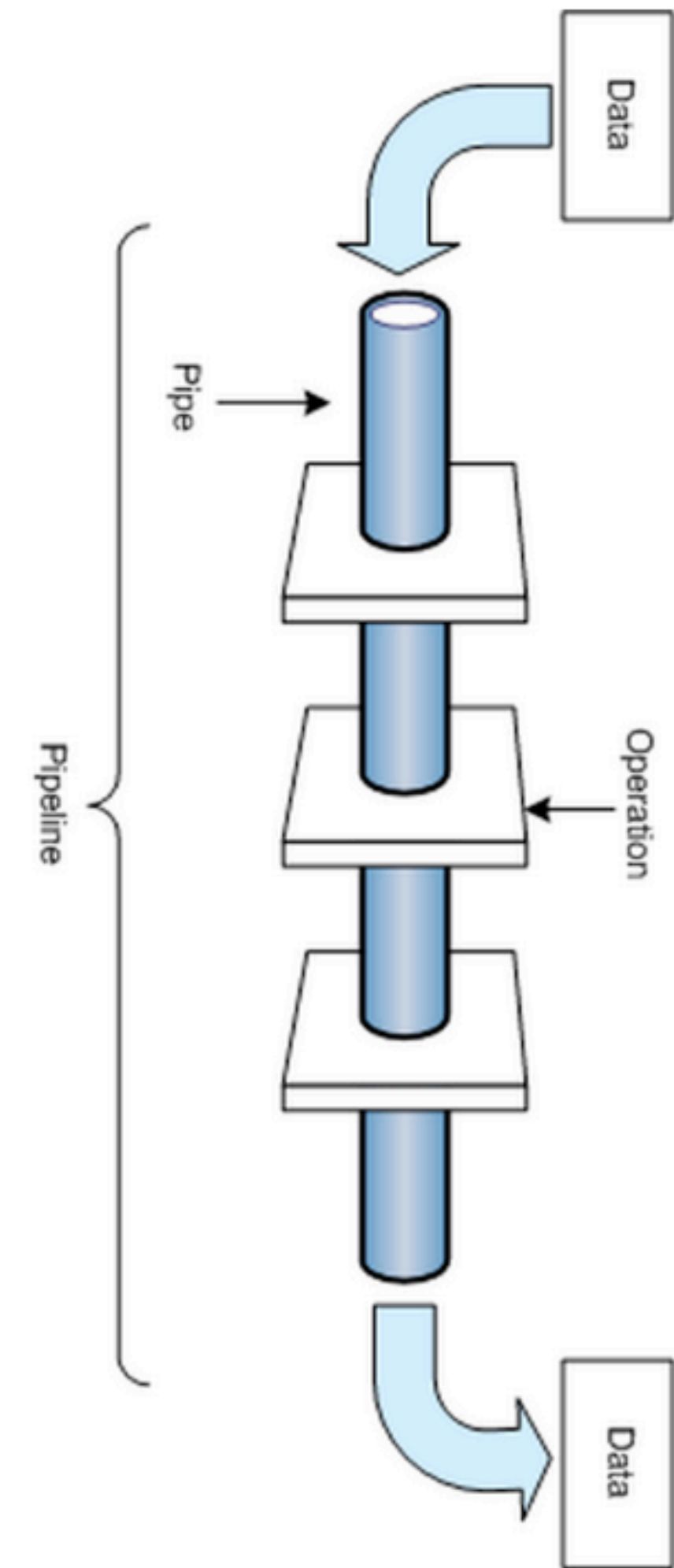
classes of methods

- The methods we have covered largely fall into two categories:
 - Data engineering
 - Statistics and machine learning
- There are full careers available in both of these disciplines at most companies today
- But the distinction between the two is beginning to blur more and more



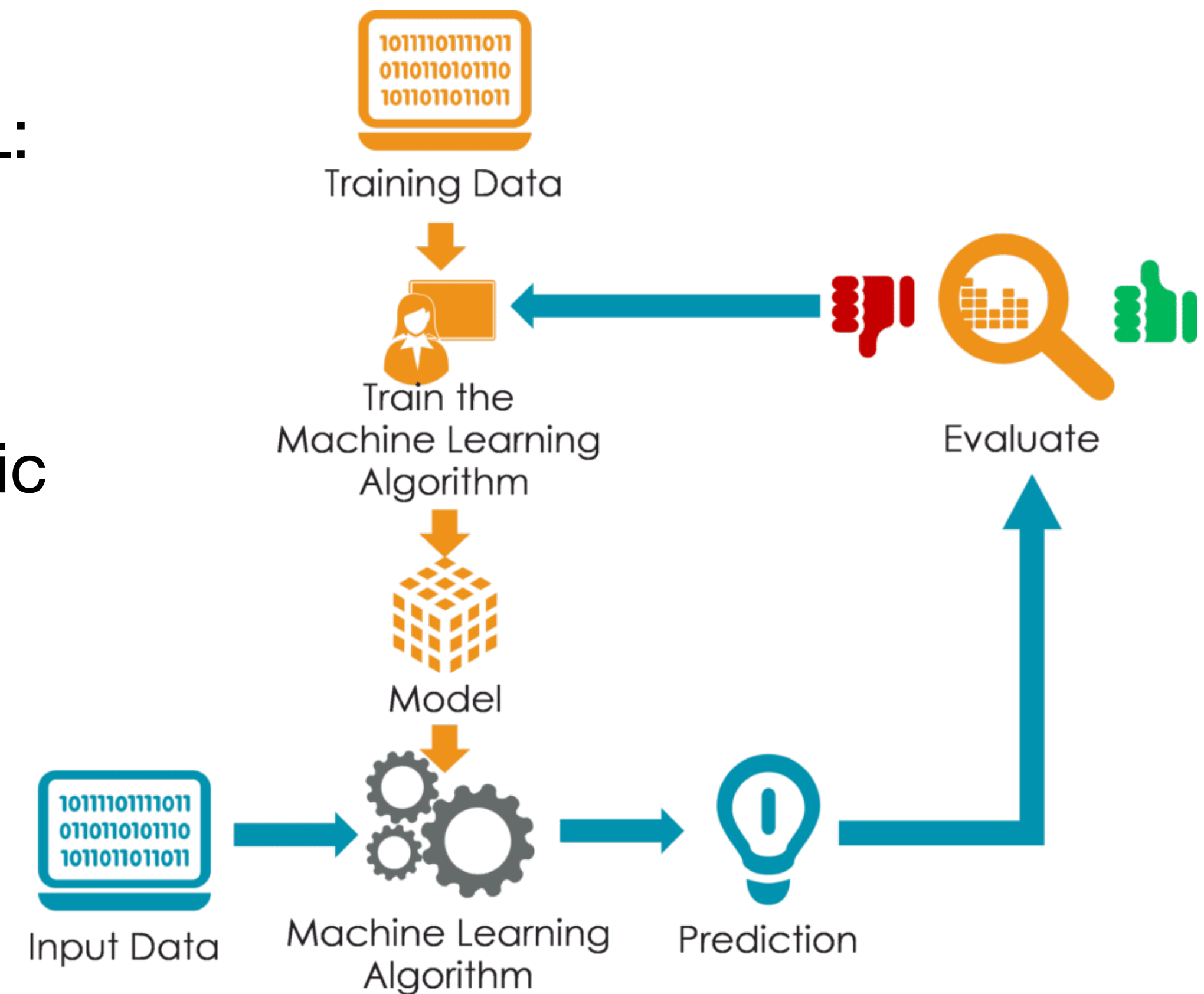
data engineering

- First, we need to set up methods to process the data in an efficient and scalable manner
- We learned how to do this, mainly in Python:
 - Data structures: Storing data in different formats
 - Functions/methods: Manipulating, translating, and aggregating data efficiently
 - Classes/objects: Containers for data fields
 - Iterators/generators: Finding specific data elements
 - Bash: Redirecting data between different files



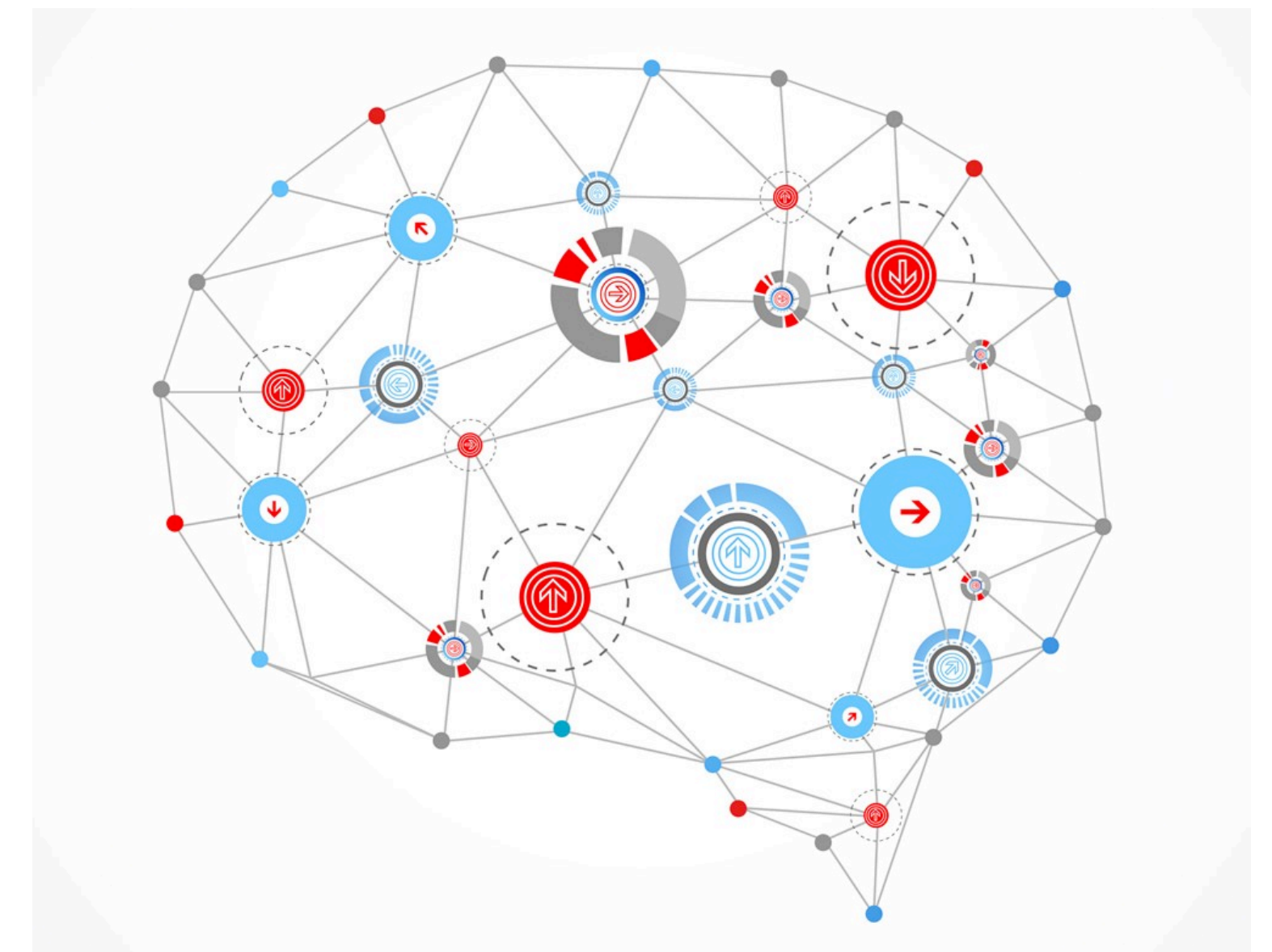
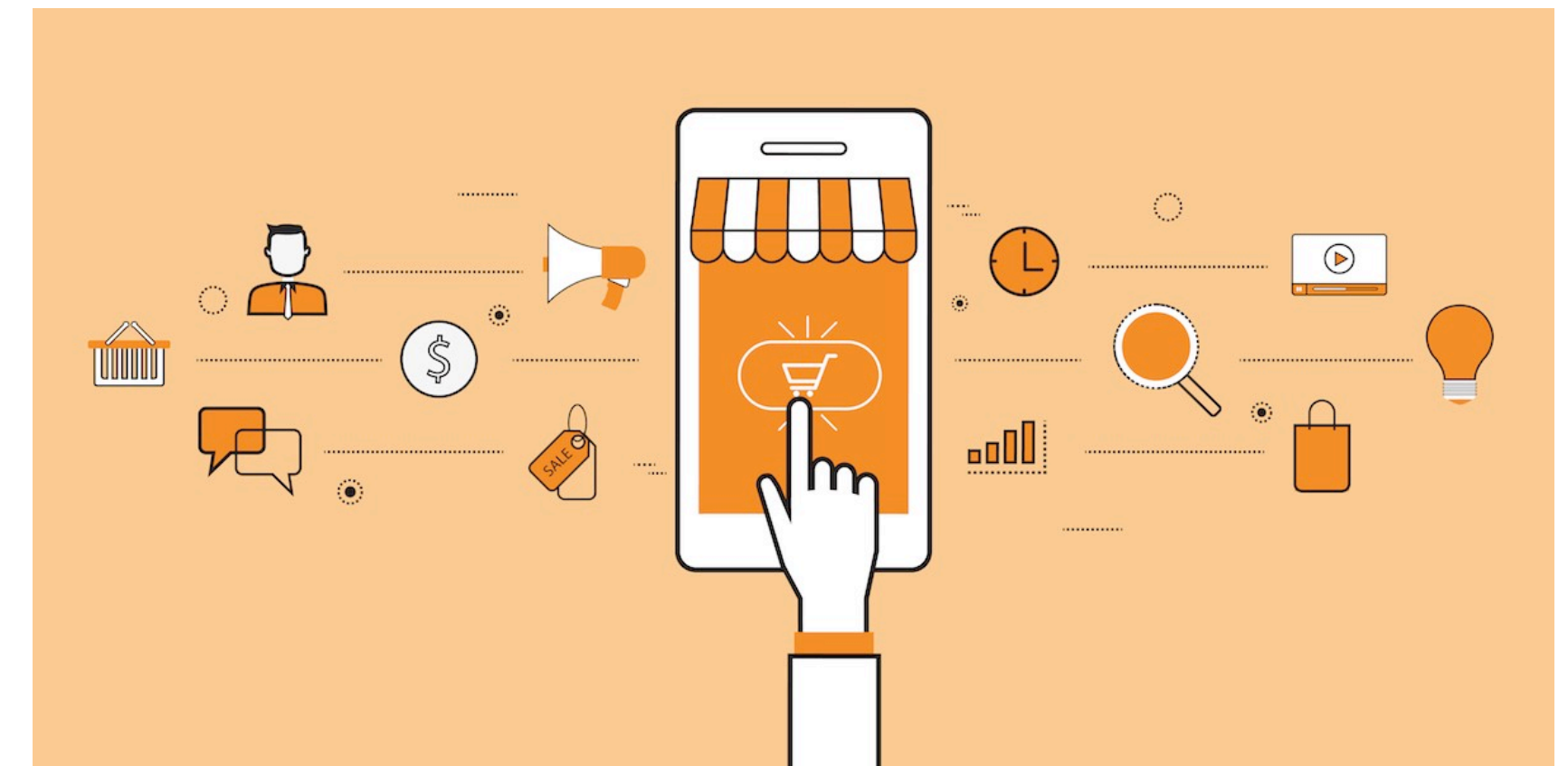
statistics and machine learning

- Second, we need to model the data, make predictions about the future, and interpret results
- We covered several classes of statistics and ML:
 - Confidence intervals and hypothesis testing
 - Supervised learning: Linear regression, logistic regression, naïve Bayes, kNN
 - Unsupervised learning: K-Means, GMMs
 - Natural language processing: tf-idf
 - Deep learning: Neural networks



what's next?

- We are in the middle of another AI “revolution”
 - Key challenges in predictability and interpretability
 - The field will continue to evolve over the next decade
- Other courses to take:
 - ECE 302 - Probabilistic Methods in ECE
 - ECE 473 - Intro. to Artificial Intelligence (some classical AI reasoning)
 - ECE 495 - (New) Data Mining: Basic Concepts and Techniques
 - ECE 570 - Artificial Intelligence (probabilistic models, project-based)
 - ECE 595 - (New) Machine Learning I
 - ECE 595 - (New) Deep Learning for Computer Vision
 - More being created as well. And these are just in ECE!



thank you for a great semester!

- Please review the course :)
- We hope to see you in other classes down the line!