ARTIFICIAL INTELLIGENCE 501

Lesson 3
AI in the Enterprise

Legal Disclaimers

Learning Objectives

You will be able to:

- Identify the steps in the data science workflow
- Identify the key roles and skill sets within the field of AI
- Describe the different ways to structure an AI team
- Identify common data science misconceptions
- Identify the components of model maintenance after deployment

DATA SCIENCE WORKFLOW

Data Science Workflow

Problem Statement	What problem are you trying to solve?
Data Collection	What data do you need to solve it?
Data Exploration & Preprocessing	How should you clean your data so your model can use it?
Modeling	Build a model to solve your problem?
Validation	Did I solve the problem?
Decision Making & Deployment	Communicate to stakeholders or put into production?

Problem Statement

What problem are you trying to solve?

- Data scientists first need to identify the problem to solve.
- Knowledge of the business is needed to identify impactful opportunities.
- Technical knowledge is needed to ask the right questions, and to know what is possible.

Data Collection

What data do you need to solve your problem?

- The data required to solve the problem needs to be identified and collected.
- Data and engineering skills are needed to collect and consolidate data from multiple sources.

Data Exploration and Preprocessing

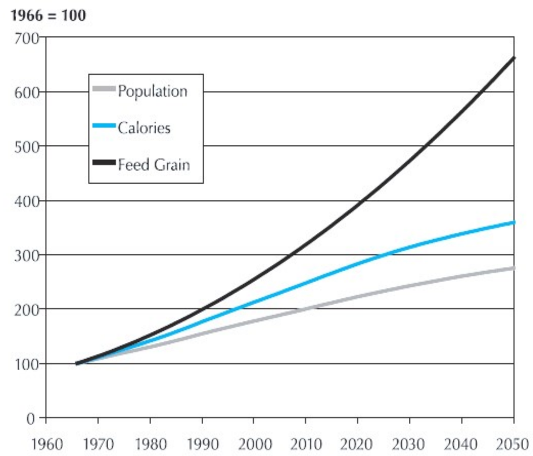
How should you clean your data so your model can use it?

- Data needs to be cleaned and processed so that it's in a usable format for modeling.
- Exploration is required to identify important elements within the data and to identify any data quality issues.
- Data, engineering, and statistics skills are needed to appropriately process the data and make inferences.

Modeling

What model needs to be built to solve your problem?

- Several factors go into modeling such as complexity, required data, speed, and performance.
- This step requires skills in engineering, modeling, and statistics.



Validation

Did the problem get sufficiently solved?

- Validation is required to ensure the original problem was solved.
- Model performance needs to be accurately measured.
- Statistics and modeling skills, as well as domain knowledge, are needed to make sure the results align with the business problem.

Decision Making and Deployment

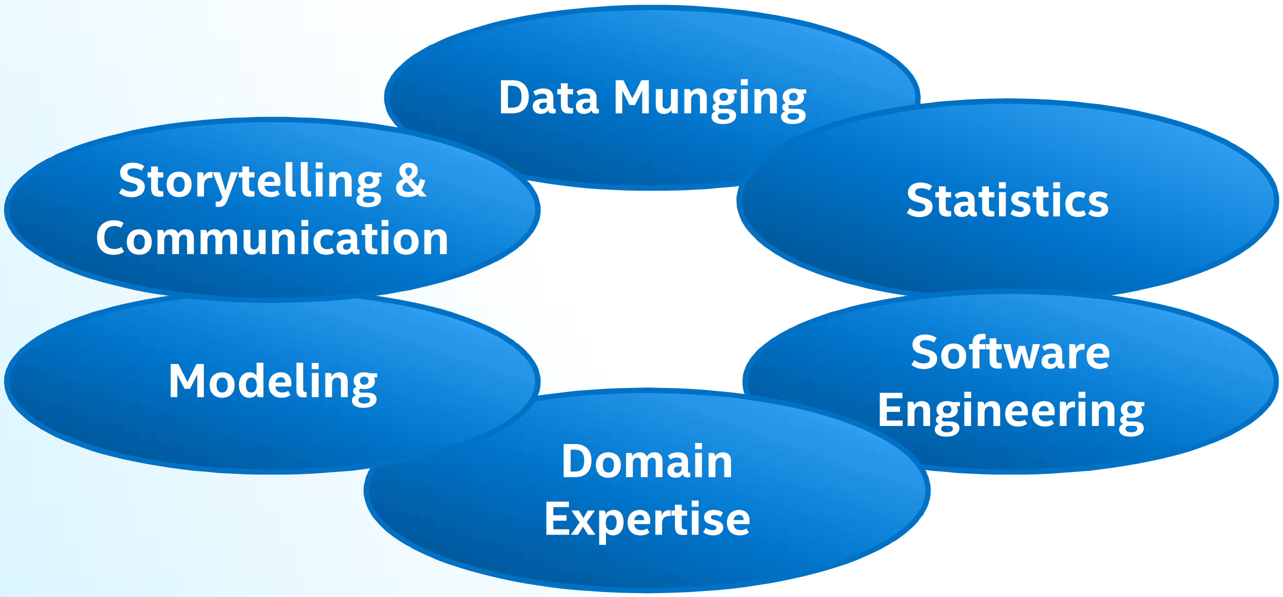
Communicate to stakeholders or put into production.

- A business decision needs to be made, or a product needs to be put into production, so the business can see value from the project.
- This requires domain knowledge, as well as communication and storytelling skills.
- Engineering skills are needed to integrate code into back-end software systems.

DATA SCIENCE SKILL SET

Data Science Skill Sets

Data science teams need a variety of skills to be successful.



Data Science Team Skills

Data science teams need a variety of skills to be successful.

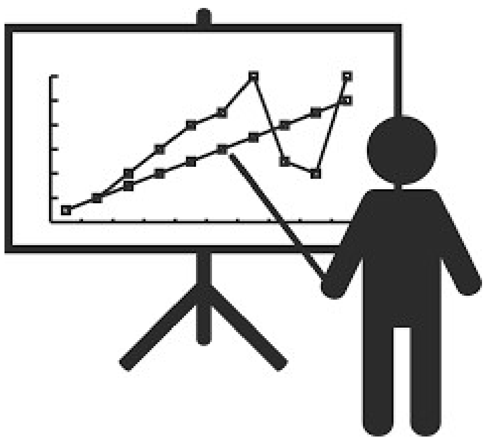
Communicate with business leaders ⟷ Integrate code into software systems

Communication/ Storytelling	**Modeling**	**Data Munging**
Domain Expertise	**Statistics**	**Software Engineering**

Communication, Storytelling, Domain Expertise

Understand the business needs and communicate how to address them.

- Domain expertise to understand the process and business problem to help their business.
- Persuade decision makers to support their idea.
- Communicate complicated concepts clearly, and tell stories.



Modeling and Statistics

Use data to make predictions via models, and using statistics to assess the validity of those predictions.

- Work with a variety of modeling techniques, from regression to DL.
- Use statistics to assess the performance of one model vs. another.
- Design experiments and perform A/B testing.

Data Munging/Software Engineering

Transform messy data into clean, usable data, as well as building software systems to deploy their models.

- Raw data can be messy and unstructured.
- Data must be manipulated and stored in databases before it can be used.
- Models need to be deployed.

DATA SCIENCE ROLES

Roles on Data Science Teams

Roles have evolved that fit on different places on this spectrum.

Communicate with business leaders ⟷ Integrate code into software systems

- Product Manager
- Machine Learning Engineer
- Data Engineer
- Business Analyst
- Statistician
- Research Scientist
- Software Engineering

Business Analysts

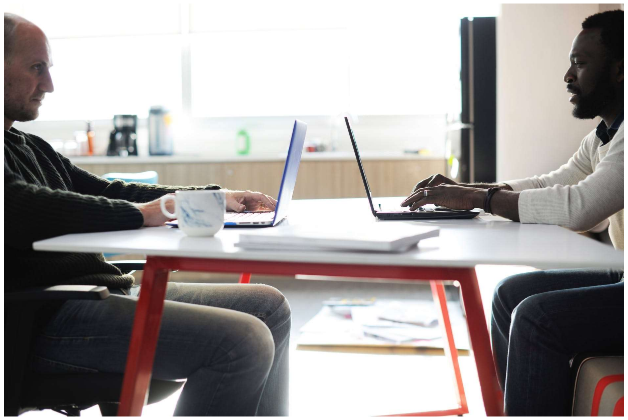
Business analysts interact with decision-makers.

- Create reports and provide insights.
- Create dashboards displaying key product KPIs.
- Perform analysis to determine business impact of a new product/feature.
- Excel*, PowerBI* and Tableau* are examples of tools used.

Product Managers

Product Managers get requirements from business.

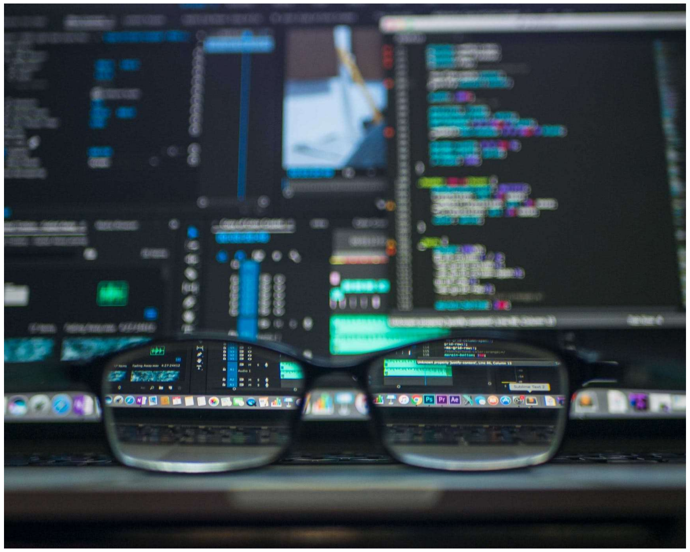
- Translate business ideas into product ideas.
- Determine feasibility of solving business problems.
- Consider impact of new product or model on key business metrics.
- Prioritize projects and tasks.
- Examples of tools used: Microsoft Project*, Trello*, and JIRA*.



Statisticians

Statisticians determine the validity of models.

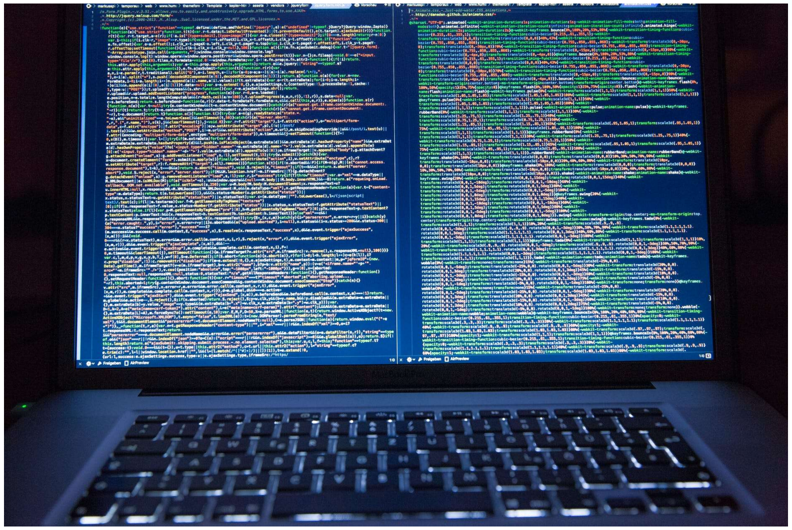
- Applies statistical concepts to determine amount of data required.
- Explores outliers and trends.
- Determines if results are statistically significant.
- Examples of tools used: R*, SAS*, Python*.



Machine Learning Engineers

Machine learning engineers solve problems involving large amounts of high-dimensional data.

- Apply machine learning techniques.
- Focused on the engineering that makes models accurate and fast.
- Examples of tools used: Python*, R*, and MATLAB*.



Research Scientists

Research scientists work on problems in bleeding-edge fields.

- Work on the toughest problems in big data and machine learning.
- Expert at a particular sub-discipline.
- Understand how algorithms work under the hood.
- Can be part of a separate research team, interfacing with data science team when necessary.
- Examples of tools used: Python*, MATLAB*, R*, C++, and Java*.

*Other names and brands may be claimed as the property of others.

Data Engineers

Data engineers build data ingestion, storage, and infrastructure.

- Databases experts.
- Know the tradeoffs between speed, reliability, and size.
- Automate data cleaning.
- Build ETL (extract, transform, load) pipelines to make data available on a regular cadence.
- Examples of tools used: Java*, SQL, and noSQL.

Software Engineers

Software engineers are responsible for optimizing code and deploying.

- Get code into production.
- Write tests to detect code breaking and bugs.
- Ensure model code is maintainable.
- Examples of tools used: Python*, Ruby*, C++, and Java*.

Skills and Roles

	Business Analyst	Product Manager	Statistician	Machine Learning Engineer	Research Scientist	Data Engineer	Software Engineer
Communication/Story	X	X					
Domain Expertise	X	X					
Modeling		X	X	X	X		
Statistics			X	X	X	X	
Data Munging				X	X	X	X
Software Engineering							X

STAFFING STRUCTURE

Data Science Organizational Structure

There are multiple ways to organize data science teams.

- Centralized teams where are all data scientists report to the same head.
- Distributed teams where individual data scientists work with a business team.
- Teams embedded within functional business units.

Centralized Teams

All the data scientists report to the same group head.

- Usually within a technology or IT team.
- Pros:
  - Standardization of skills and tools
  - Reduction of redundant roles
  - Closer collaboration amongst scientists
- Cons:
  - Further removed from the business units
  - Tendency to be more reactive to problems

Distributed Teams

Distributed teams are where individual data scientists sit within a particular team within a business unit.

- Pros:
  - Closer access to the business, domain experts, and end users
  - More likely to come up with solutions to immediate business problems
- Cons:
  - Destandardization of tools and skills throughout the organization
  - Less communication and collaboration between data scientists

Additional Team Structures

There are multiple hybrid ways to organize data science teams.

- Some examples include:
  - Full data science teams can be embedded within a business function.
  - Data science centers of excellence can be created for the company.

COMMON MISCONCEPTIONS

Misconception #1: Data Science “Unicorn”

Data Scientists who are experts in all areas are called “unicorns”.

- Successful teams contain people with a diverse array of skillsets and backgrounds.
- Some excel at communication, while others may excel at statistics.
- Successful teams have experts in the three main areas: business, science, and engineering.

Misconception #2: Research and Algorithms Focus

Data science teams cannot just focus on research and algorithms.

- Effective teams have mechanisms to:
  - Identify problems
  - Communicate findings
  - Work with engineering to understand how to put their models into production

Misconception #3: Complex and Advanced Systems

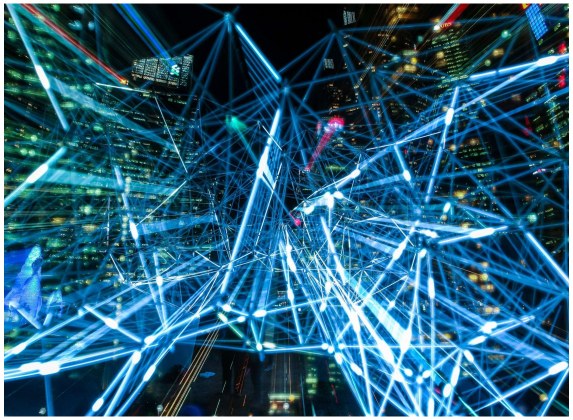
The most complicated solution isn't always the best.

- Teams tend to be more successful when they start simple and then move on to more complex modeling techniques.
- Complex models may be more accurate, but are less interpretable, more likely to fail in unpredictable ways, and harder to maintain.
- Starting simple also ensures that what the team is building aligns with business needs.

Misconception #4: Industry Differences

The underlying modeling and data analysis techniques are largely transferable between industries.

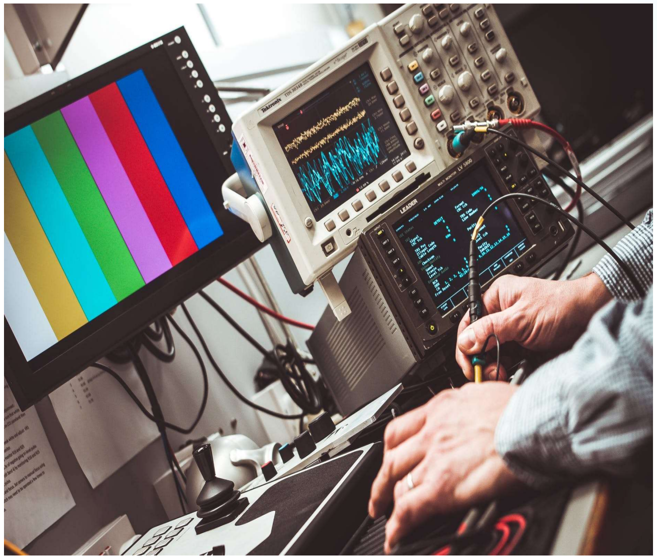
- Domain expertise is required to understand which data is relevant and which problems are most important to solve.
- The techniques used to clean data, store it, and extract useful insights and modeling remain very similar.



Misconception #5: Projects Begin Well-Defined

Data Science projects are often exploratory and experimental in nature.

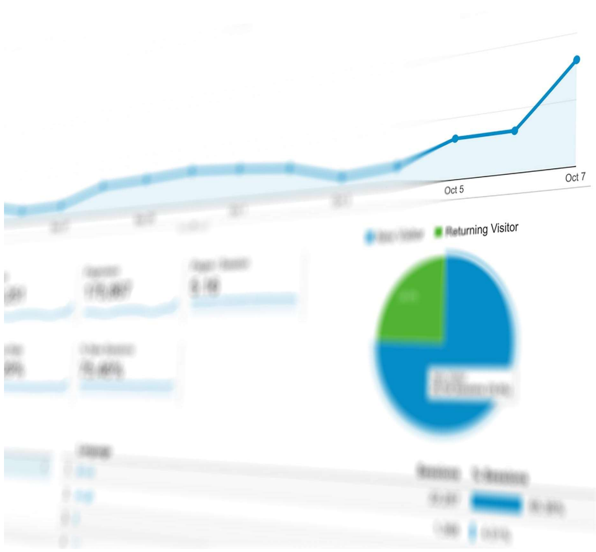
- It may not be clear how hard the problem is to solve until investing time exploring the data.
- Product Managers must actively work with both the team and the business stakeholders to manage expectations.



Misconception #6: Best Prediction Models are Best

There are more challenges involved when selecting a model than its predictive abilities.

- Some models may be too slow or complicated to include in production.
- Some models may not be interpretable, and would have a tough sell with decision makers.



AFTER DEPLOYMENT

After Deploying a Model

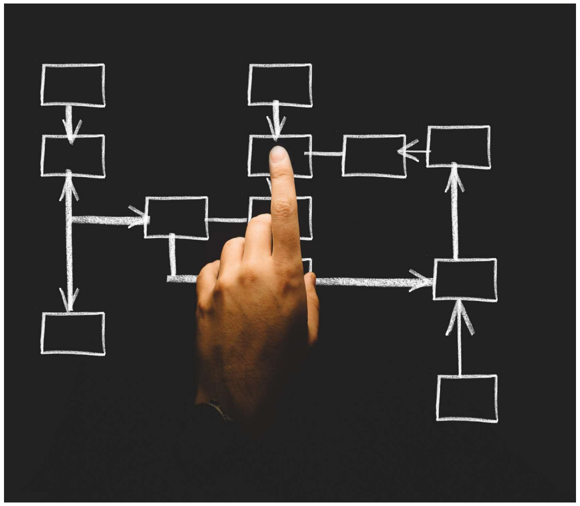
Once a model is deployed, relevant teams must monitor and manage the model for it to be useful.

- **Business Intelligence teams** should build reports/dashboards displaying model results.
- **Business teams** (e.g. operations, merchandising) or customers should use model outputs.
- **Data Science teams** must update model at appropriate cadence (for example, monthly).

Business Intelligence Team

The business intelligence team monitors how model predictions are changing over time.

- Monitor model usage and output.
- Develop key performance indicators and dashboards.
- Visualizations should align with the needs of the functional teams (for example, operations) that will be using the model.



Functional Business Team

Use dashboards to improve decision making.

- For example: marketing team could use churn model to decide when to reach out to groups of customers.
- For example: operations team could use model to predict shipping/logistics times.

Data Science Team

Monitor model to ensure continued validity and usefulness.

- For example: changes in marketing may lead to different types of customers coming in the door, making the old model of customer behavior less accurate.
- Depending on how fast the business conditions are changing, the data science team should update model with appropriate frequency.
- Retraining models with new data.

Learning Objectives Recap

In this lesson, we worked to:

- Identify the steps in the data science workflow
- Identify the key roles and skill sets within the field of AI
- Describe the different ways to structure an AI team
- Identify common data science misconceptions
- Identify the components of model maintenance after deployment

intel®

Sources for images used in this presentation (listed by slide number)

20 https://www.pexels.com/photo/adult-brainstorming-business-chair-515167/

22 https://www.pexels.com/photo/blue-screen-of-death-in-silver-black-laptop-177598/

31 https://www.pexels.com/photo/ground-group-growth-hands-461049/

33 https://www.pexels.com/photo/yellow-blue-and-red-plastic-gameboard-toy-209712/

34 https://www.pexels.com/photo/adult-agreement-beard-beverage-618550/

38 https://www.pexels.com/photo/blue-and-green-pie-chart-97080/

39 https://www.pexels.com/photo/arrows-box-business-chalk-533189/