

# Mu Sigma Data Science Series

## Part 1: A Beginner's Guide to Data Science

February 10, 2022



# Key Messages

- Rapidly evolving field
- Vast ocean of techniques and technologies
- Human judgment and ingenuity are essential for their success: the key lies in applying meaningfully
  - Fanciest or most complex is not the best
  - Model the operational set-up / mechanics of a phenomenon
- Confluence of these has resulted in blurring of boundaries between roles
- There are layers of problem-solving, and implementation: technology, orgs / people, monetary
- 80%+ of Model-building effort goes into data prep

# Definitions

**Machine Learning** constructs algorithms that can learn from data.

**Statistical Learning** is a branch of applied statistics that emerged in response to machine learning, emphasizing statistical models and assessment of uncertainty.

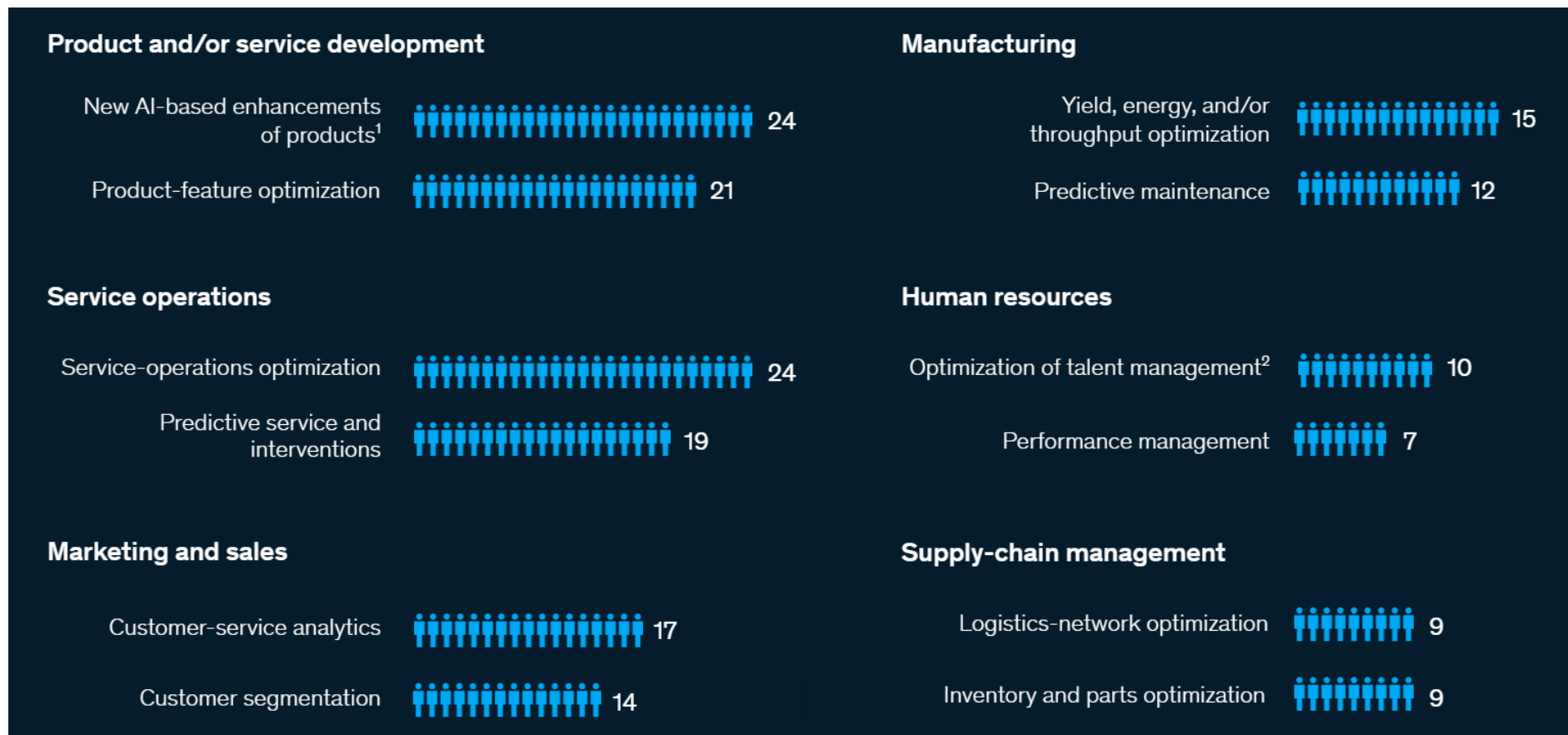
**Data Science** is the extraction of knowledge from data, using ideas from mathematics, statistics, machine learning, computer science, engineering. . .

**Artificial Narrow Intelligence (ANI)** refers to any AI that can outperform a human in a narrowly defined and structured task. It is designed to perform a single function like face recognition, or speech detection under various constraints and limitations. It is the constraints that lead people to refer to these functions as 'narrow' or 'weak'.

All of these are very similar *with different emphases.*

# AI adoption is highest within the product- or service-development and service-operations functions

AI use cases most commonly adopted within each business function, %



Source: McKinsey & Company

# Artificial Intelligence:: Real-time Decision-making

## Neural Networks

“Apply cutting-edge research to train deep neural networks on problems ranging from **perception to control**. Our per-camera networks analyze raw images to perform **semantic segmentation**, **object detection** and **monocular depth estimation**. Our birds-eye-view networks take video from all cameras to output the road layout, static infrastructure and 3D objects directly in the top-down view. Our networks learn from the most complicated and diverse scenarios in the world, **iteratively sourced from our fleet of nearly 1M vehicles in real time**. A full build of Autopilot neural networks involves **48 networks that take 70,000 GPU hours to train**.

Together, they output **1,000 distinct tensors (predictions) at each timestep.**”





# Machine Learning::

## Recommender Systems | Collaborative Filtering

Collaborative filtering is based on (user, item, rating) tuples. So, unlike content-based filtering, it leverages other users' experiences.

Your recently viewed items and featured recommendations

Sponsored products related to this search [What's this?](#)

Page 1 of 3

<



All-new Echo Show (2nd Gen) + Ring Video Doorbell 2- Charcoal  
1 offer from \$428.99



AmazonBasics Microwave, Small, 0.7 Cu. Ft, 700W, Works with Alexa  
★★★★☆ 1,375  
\$59.99 ✓prime



Echo Look | Hands-Free Camera and Style Assistant with Alexa—includes Style Check to...  
★★★★☆ 413  
\$99.99 ✓prime



Sonos Beam - Smart TV Sound Bar with Amazon Alexa Built-in - Black  
★★★★☆ 474  
\$399.00 ✓prime



Echo Wall Clock - see timers at a glance - requires compatible Echo device  
★★★★☆ 1,231  
\$29.99 ✓prime



Echo Spot Adjustable Stand - Black  
★★★★☆ 933  
\$19.99 ✓prime



AHASTYLE Wall Mount Hanger Holder ABS for New Dot 3rd Generation Smart Home Speakers...  
★★★★☆ 12  
\$10.99 ✓prime



Angel Statue Crafted Stand Holder for Amazon Echo Dot 3rd Generation, Alexa Smart...  
★★★★☆ 57  
\$25.99 ✓prime

>

Explore more from across the store

Page 1 of 6

<



Actionable Gamification: Beyond Points, Badges...  
› Yu-kai Chou



The Model Thinker: What You Need to Know to...  
› Scott E. Page



Don't Make Me Think, Revisited: A Common...  
› Steve Krug



Hooked: How to Build Habit-Forming Products  
› Nir Eyal



Microservices Patterns: With examples in Java  
› Chris Richardson



Solving Product Design Exercises: Questions &...  
› Artiom Dashinsky



100 Things Every Designer Needs to Know About...  
Susan Weinschenk



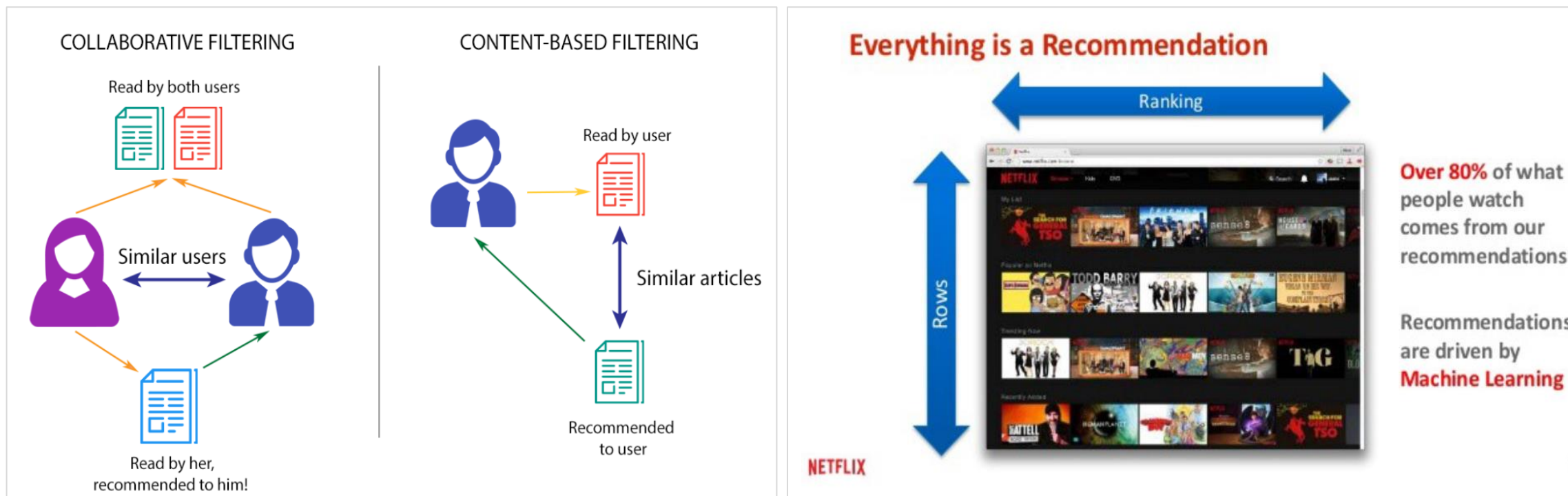
Infinity  
› Jonathan Hickman  
★★★★☆ 182

>

# Machine Learning::

## Recommender Systems | Content-based Collaborative Filtering

Collaborative filtering is based on (user, item, rating) tuples. So, unlike content-based filtering, it leverages other users' experiences.



# Predictive Analytics::

## Marketing Sciences

- DTC targeting
- Marketing-Mix Modeling
- Lift Analysis
- RFM Analysis
- CLTV Modeling
- Customer Acquisition / Churn





# Self-training

## Python libraries:

Data Processing: pandas; re; numpy; os; datetime

Visualization: matplotlib; seaborn; bokeh;

Predictive Analytics: scikit-learn

## R libraries:

Data Processing: dplyr, data.table, tidyr, magrittr

Visualization: ggplot2

## Frameworks:

Spark

PyTorch

TensorFlow

H2O

## Learning Resources:

<https://pandas.pydata.org/>



<https://stackoverflow.com>



<https://github.com>



<https://Kaggle.com>



<https://youtube.com/>

