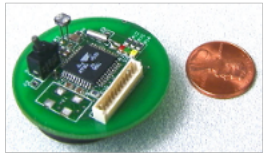
Data Science and Machine Learning

Dr. Mohammad Pourhomayoun
Assistant Professor of Computer Science
Director of Artificial Intelligence & Data Science Research Lab
California State University, Los Angeles

What is Data Science?

- **Data Science** is a new field of research that aims to develop automated techniques to **extract knowledge or insight from large-scale data** and use it for **future purposes** such as **prediction, decision making, recommendation, …** .
- It can be an integration of Machine Learning, Artificial Intelligence, Big Data Processing and Analytics, and Computing.

- **Question**: What is the difference between “knowledge” and “data”?



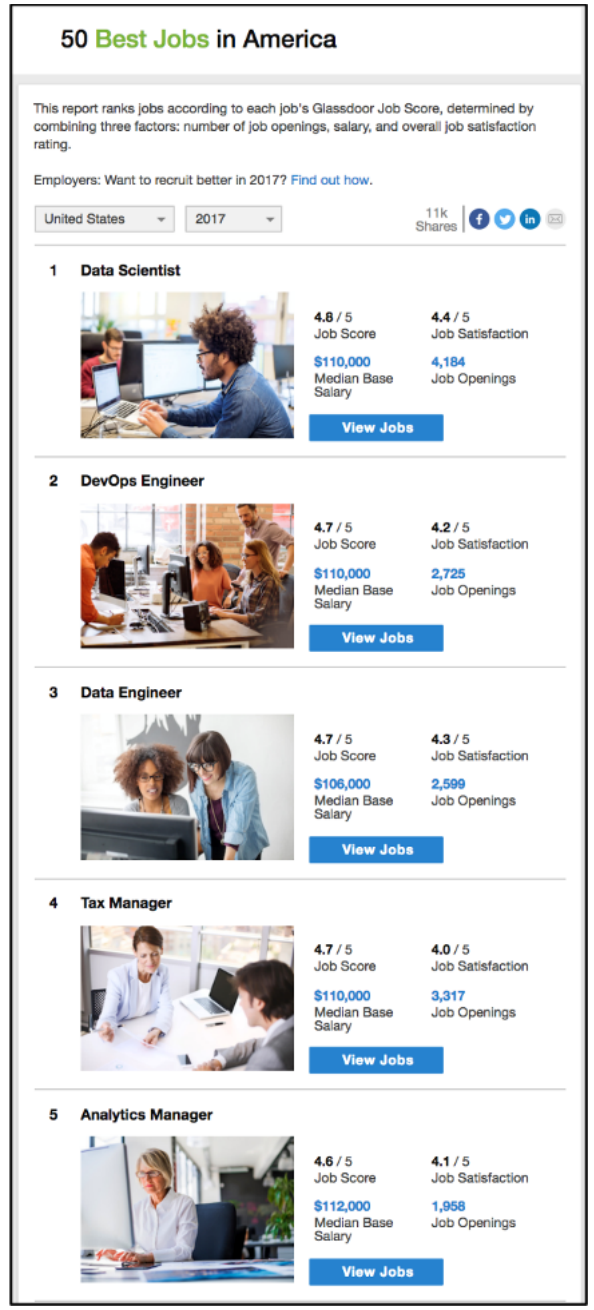
What is Data Science?

- Fortune Magazine:
  - “The Hot New Gig in Tech!”
- The New York Times:
  - "This hot new field promises to revolutionize industries, from business to government, healthcare to academia."
- Fortune Magazine
  - “Companies that want to make sense of all their bits and bytes are hiring so-called **data scientists** – **if they can find any!**”
- McKinsey Global Institute:
  - There will be 4 to 5 million jobs in the U.S. requiring data analysis skills.

Who is a Data Scientist?

- **Glassdoor:**
  - "**Data Scientist**" has been rated **#1 in the list of Best Jobs in America since 2016**
  - In this list, the jobs are determined by combining three key factors:
    - **number of job openings**
    - **salary**
    - **career opportunities rating**

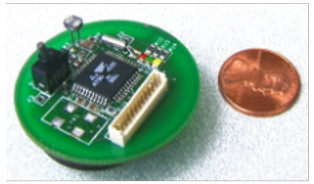
[Ref]: www.glassdoor.com/List/Best-Jobs-in-America-2019-LST_KQ0,25.htm



Why Now?

New Sources of Data

- Social Networks: Facebook, Twitter, …
- World Wide Web
- Online Activities: Amazon, ebay, …
- Smart Phone Activities
- IoT
- Electrical Health Records (EHR)
- Body and wearable sensors
- …



New Sources of Data

- "There was 5 exabytes of information ($5x10^{18}$ bytes) created between the **dawn of civilization through 2003**, but that much information is now created **every two days**."

- Eric Schmidt, Google, Alphabet

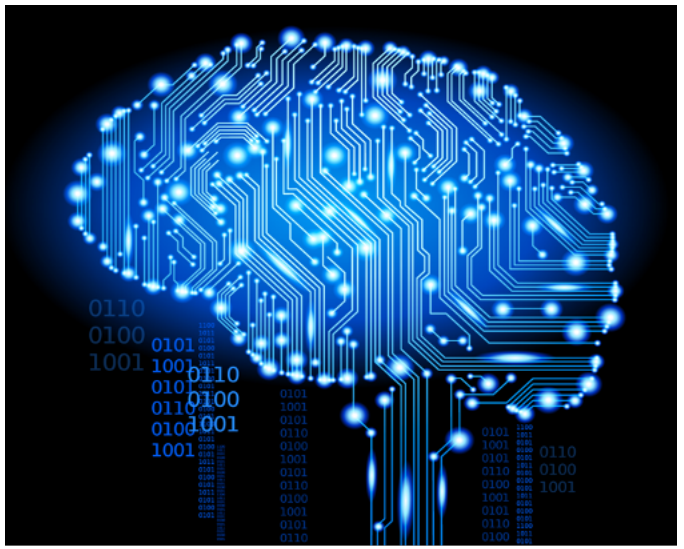
What is Artificial Intelligence (AI)?

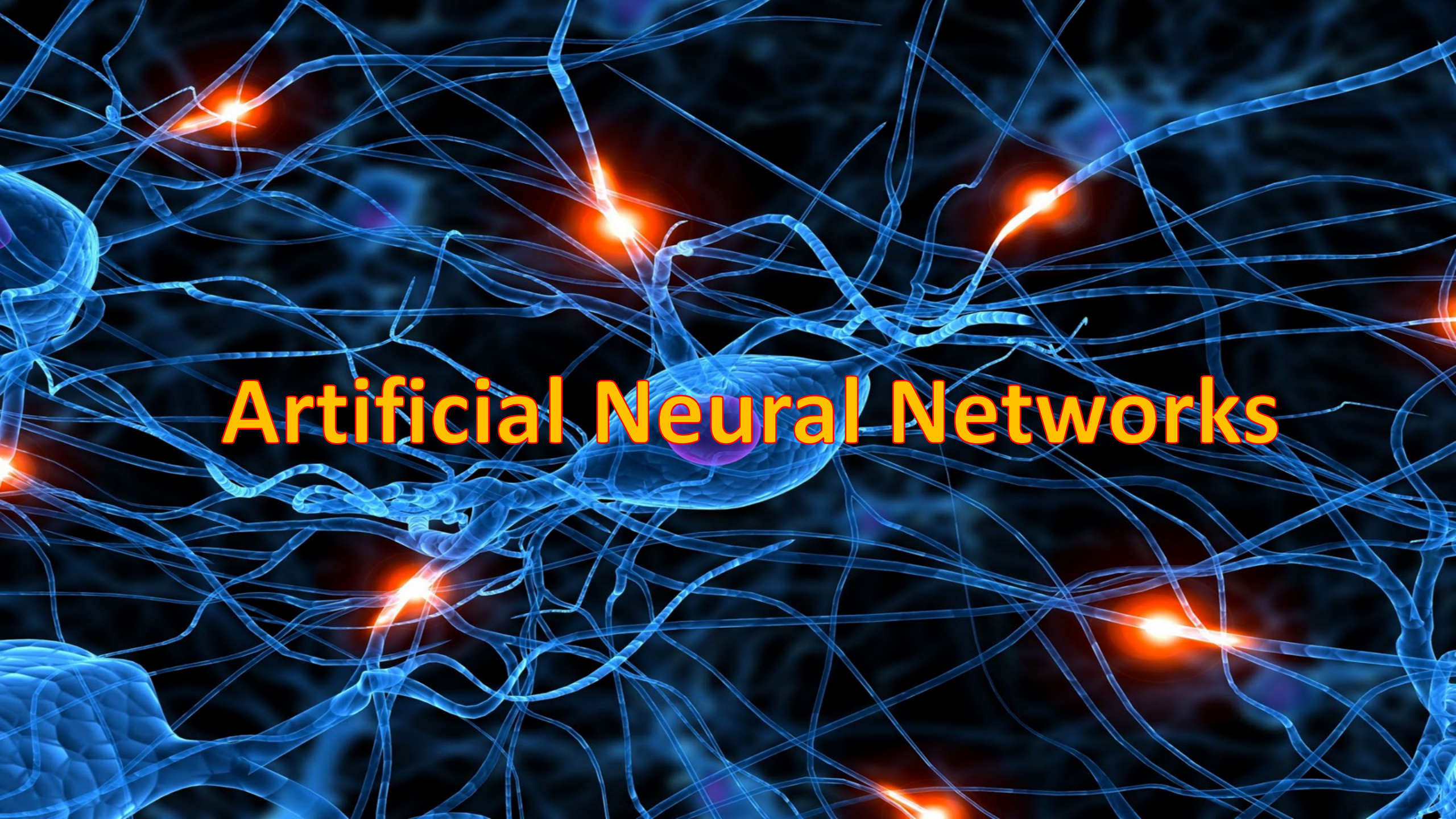
- **Artificial Intelligence (AI)** is a branch of computer science and Data Science that tries to build machines (computers) that can ***mimic human intelligence and functions***, such as "***learning***" , "***decision making***" , "***prediction***", and "***problem solving***".

What is Machine Learning?

- **Definition:** Designing and constructing algorithms or methods that give computers the ability to learn from past data, without being explicitly programmed, and then make predictions on future data.

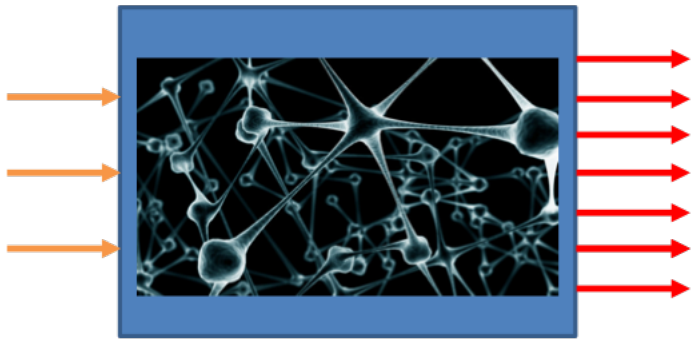
- Similar to our Brain!

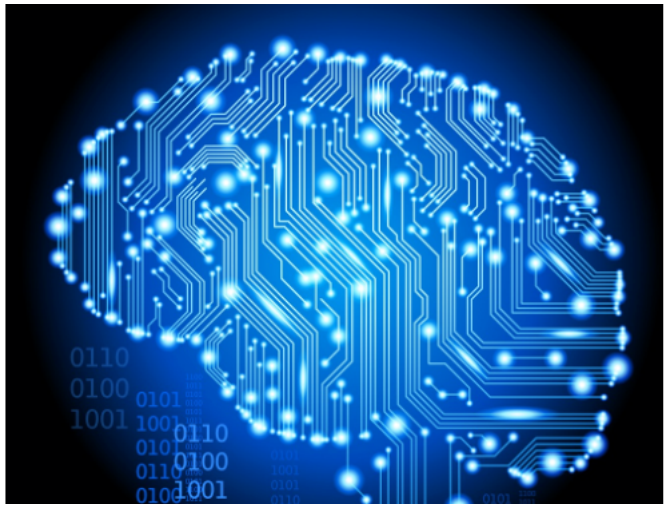



Artificial Neural Networks

Artificial Neural Networks

- **Artificial Neural Networks (ANN)**, or simply called **Neural Networks** are a family of ML models/methods inspired by the human's nervous system, particularly the brain.
- **Artificial Neural Network algorithms try to mimic the brain!**





Only Some of the Applications!

Online Shopping and Advertisements

Speech Recognition and Natural Language Processing

Marketing

Computer Vision

Recommendation Systems

Self-Driving Cars

Healthcare

Social Media
Search Engines
Email

Example: Recommender System

amazon

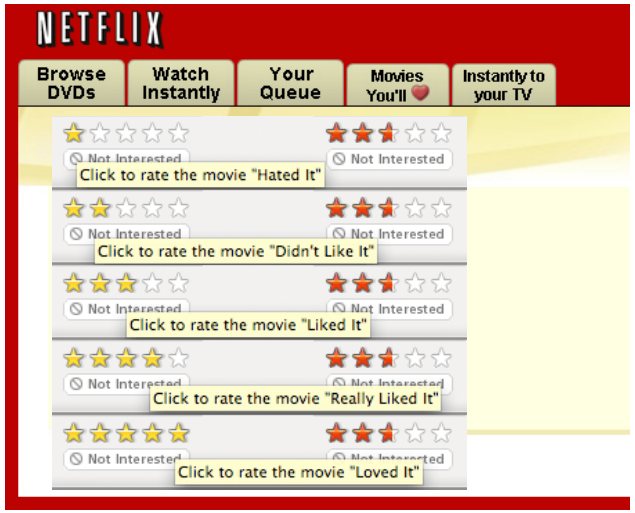
- **Customers Who Bought This Item Also Bought:**

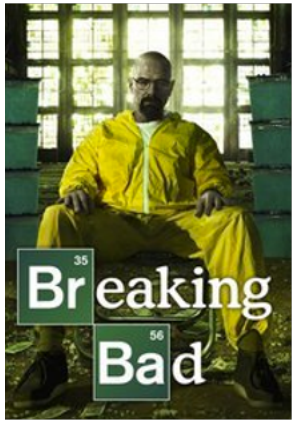
+

+

Example: Recommender System

Netflix Prize: $1,000,000 in an open competition for the best algorithm to predict user ratings for films, based on previous ratings without any other information about the users or films.





+

+

?????

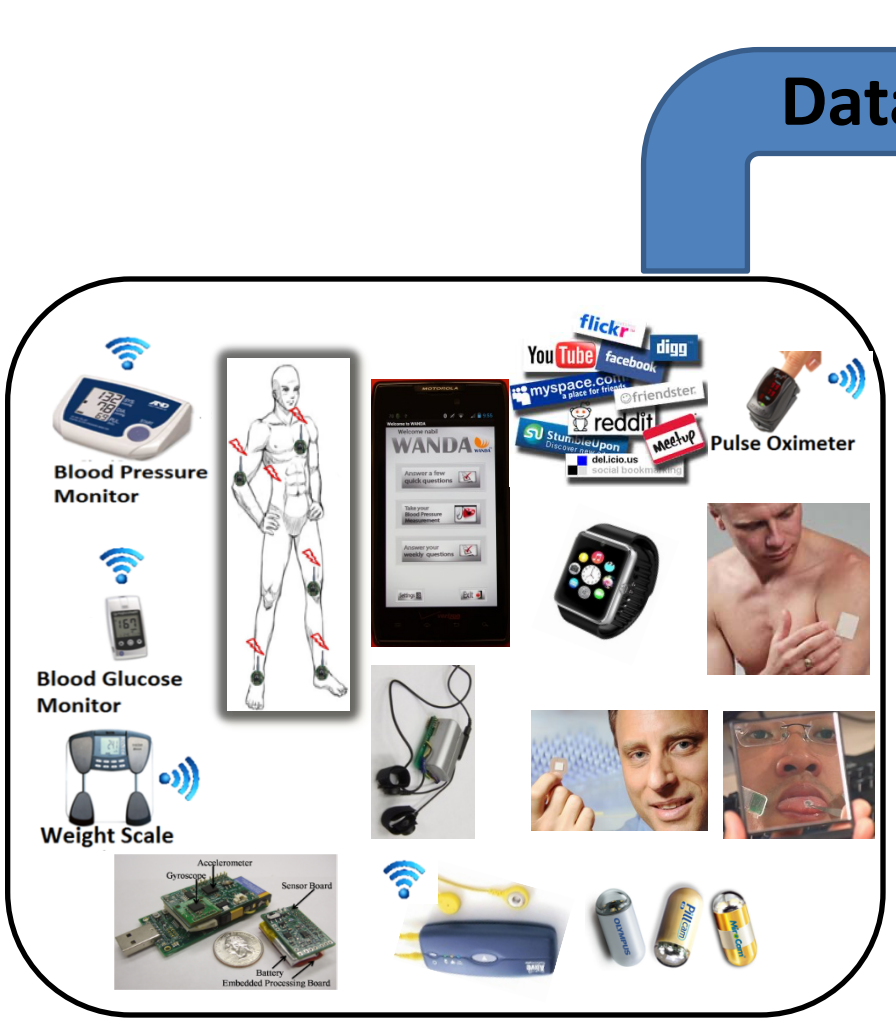
Example: Speech Recognition and Natural Language Processing (NLP)

- Apple Siri
- Amazon Alexa
- Google Home
- Google Duplex

Example: Self-Driving Cars

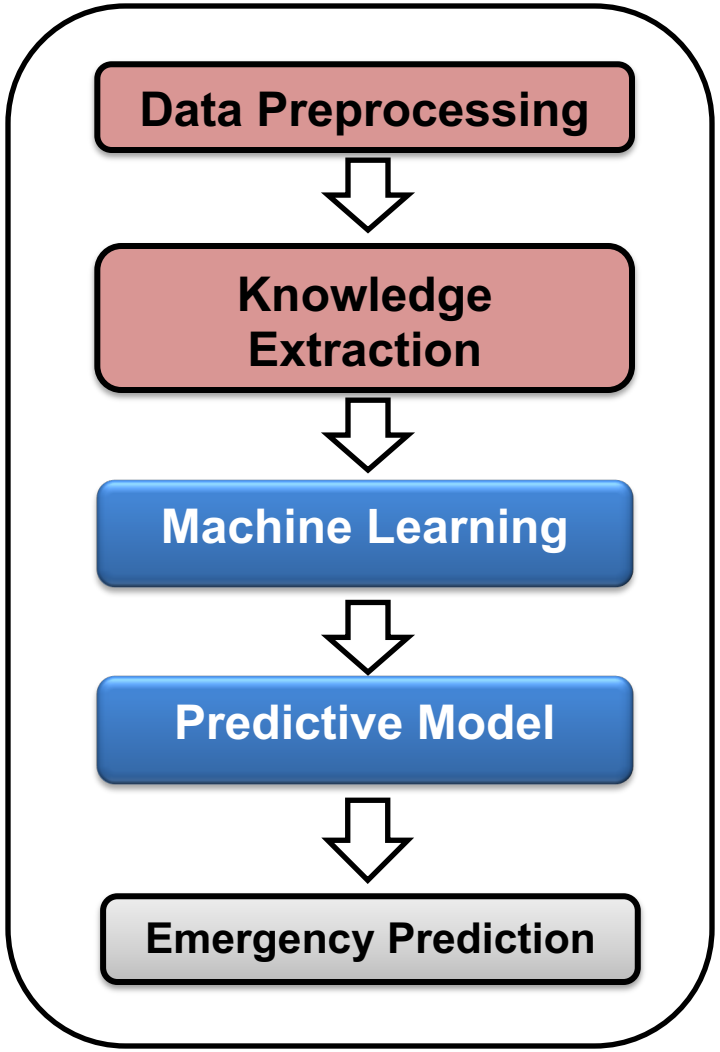
Watch this:
www.youtube.com/watch?v=B8R148hFxPw

Example: Medical Emergency Prediction



DATA COLLECTION SYSTEM

Data



ANALYTICS ENGINE

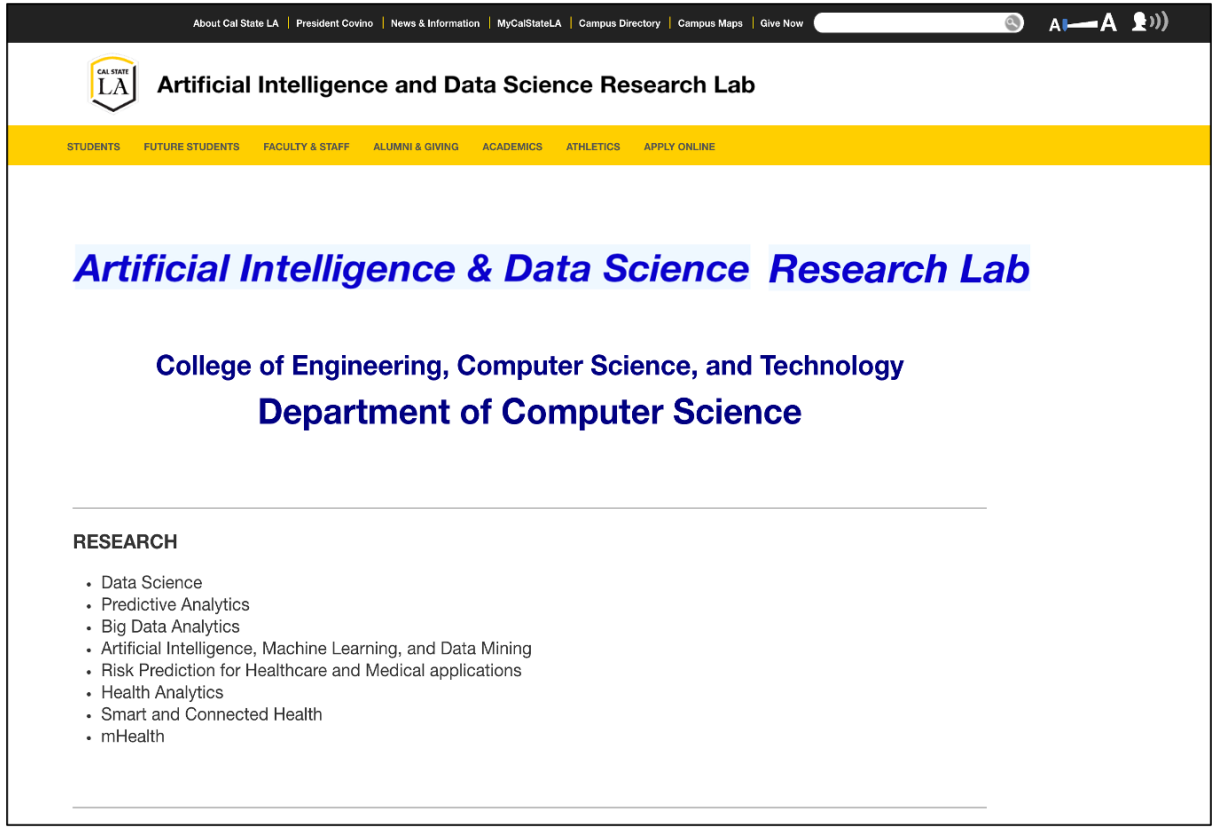
Artificial Intelligence and Data Science at Cal State LA

Artificial Intelligence & Data Science Research at Cal State LA

Visit our AI & Data Science Research Lab at:

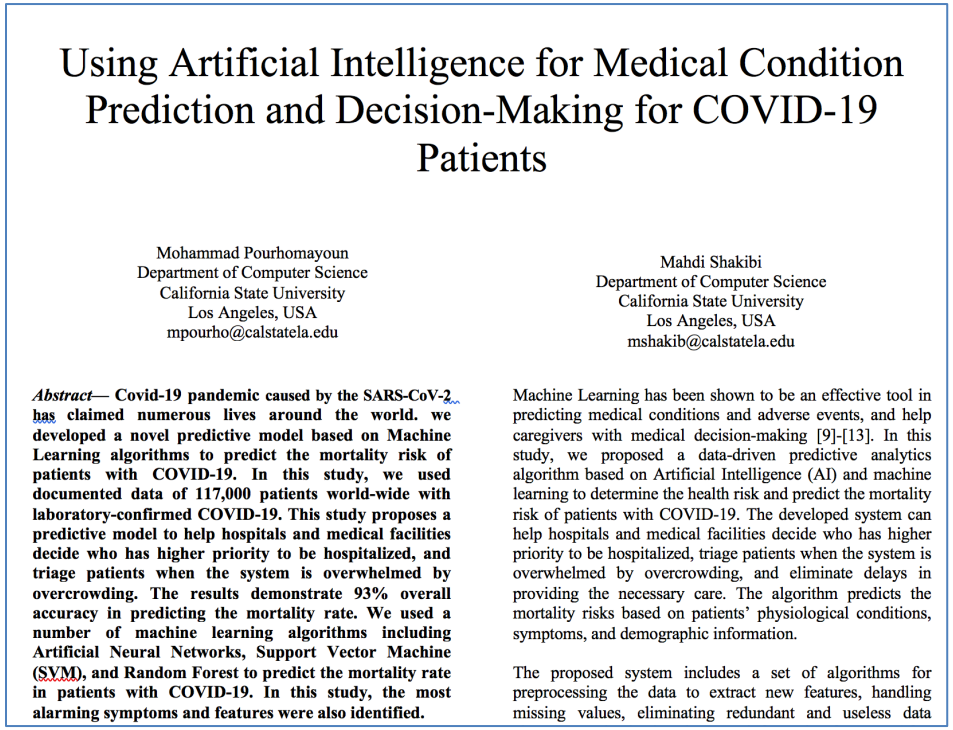
www.calstatela.edu/research/data-science

- 25 grad students theses in past 4 years
- More than 80 peer-reviewed paper publications
- PI on more than $1M in past 2 years



Artificial Intelligence and Data Science to Address COVID-19

Dr. Mohammad Pourhomayoun



Using Artificial Intelligence for Medical Condition Prediction and Decision-Making for COVID-19 Patients

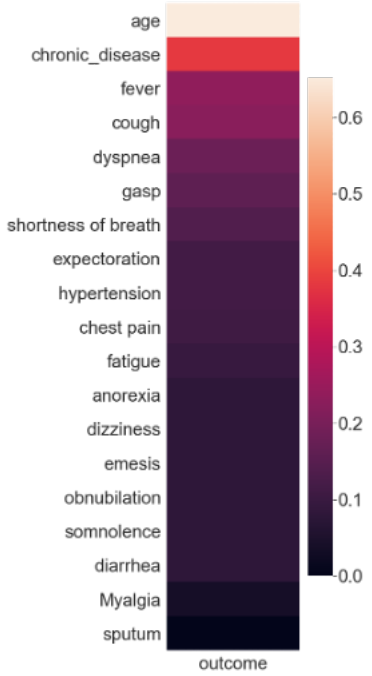
Mohammad Pourhomayoun
Department of Computer Science
California State University
Los Angeles, USA
mpourho@calstatela.edu

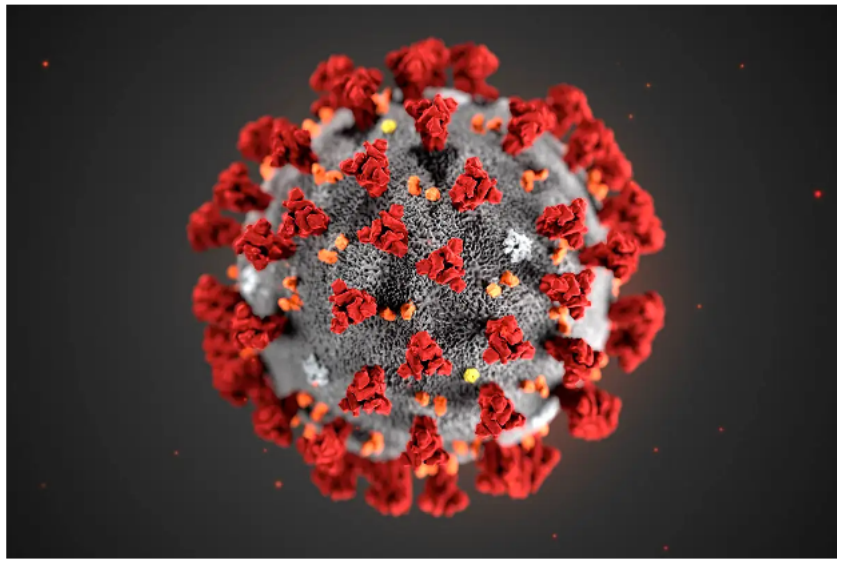
Mahdi Shakibi
Department of Computer Science
California State University
Los Angeles, USA
mshakib@calstatela.edu

***Abstract*— Covid-19 pandemic caused by the SARS-CoV-2 has claimed numerous lives around the world. we developed a novel predictive model based on Machine Learning algorithms to predict the mortality risk of patients with COVID-19. In this study, we used documented data of 117,000 patients world-wide with laboratory-confirmed COVID-19. This study proposes a predictive model to help hospitals and medical facilities decide who has higher priority to be hospitalized, and triage patients when the system is overwhelmed by overcrowding. The results demonstrate 93% overall accuracy in predicting the mortality rate. We used a number of machine learning algorithms including Artificial Neural Networks, Support Vector Machine (SVM), and Random Forest to predict the mortality rate in patients with COVID-19. In this study, the most alarming symptoms and features were also identified.**

Machine Learning has been shown to be an effective tool in predicting medical conditions and adverse events, and help caregivers with medical decision-making [9]-[13]. In this study, we proposed a data-driven predictive analytics algorithm based on Artificial Intelligence (AI) and machine learning to determine the health risk and predict the mortality risk of patients with COVID-19. The developed system can help hospitals and medical facilities decide who has higher priority to be hospitalized, triage patients when the system is overwhelmed by overcrowding, and eliminate delays in providing the necessary care. The algorithm predicts the mortality risks based on patients' physiological conditions, symptoms, and demographic information.

The proposed system includes a set of algorithms for preprocessing the data to extract new features, handling missing values, eliminating redundant and useless data



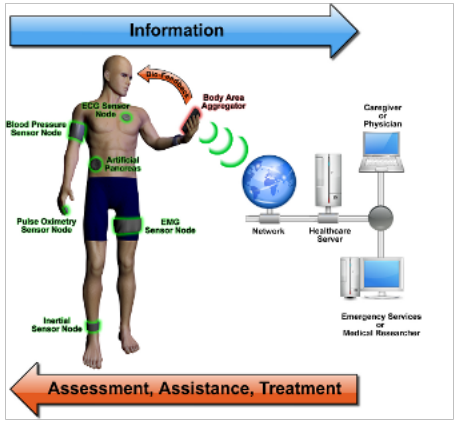


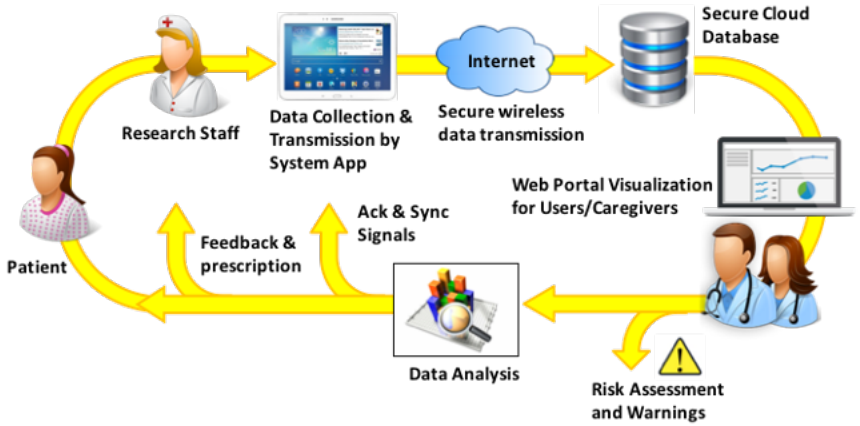
www.medrxiv.org/content/10.1101/2020.03.30.20047308v1

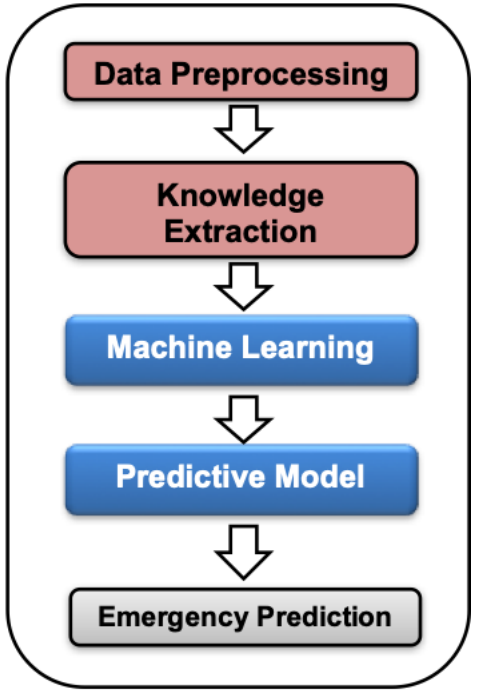
Developed a predictive model based on AI and ML to ***determine the health risk and predict the mortality risk of patients with COVID-19***. This model helps hospitals and medical facilities decide who needs to get attention first, who has higher priority to be hospitalized when the system is overwhelmed, and eliminate delays in providing the necessary care.

Predictive Analytics and Machine Learning for Chronic Disease Monitoring & Management, and Predicting Medical Conditions

Dr. Mohammad Pourhomayoun







Case Studies in: Cardiovascular Disease, Diabetes, Cancer, AIDS, Heart Failure, Osteoporosis, Kidney failure, Liver disease, Hospital Readmission

Ref: [6][12][14][18]-[65]
(www.calstatela.edu/research/data-science)

AI and ML in Cytopathology for Early Detection of Cancer

Dr. Mohammad Pourhomayoun

- Although the Pap test is an effective method for early detection of cervical cancer, it is very expensive and time-consuming because the Pap test method requires visual examination of thousands of cells by a trained pathologist.

Ref: [1][8][20]
(www.calstatela.edu/research/data-science)

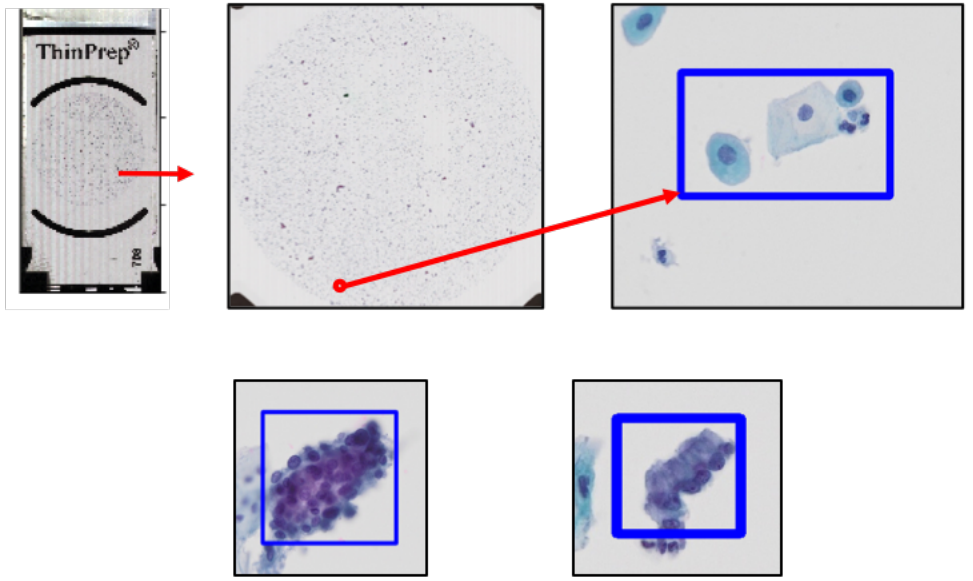


Figure2: Abnormal cell (in this case a malignant cancerous cell) (left), versus normal cell (right).

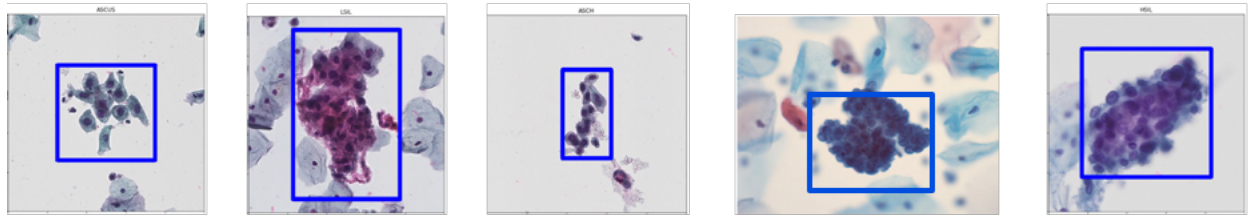


Figure3: Different types of abnormalities: from left to right: Atypical squamous cells of undetermined significance (ASCUS), Low–grade squamous intraepithelial lesion (LSIL) , Endometrial, Atypical squamous cells HSIL(ASCH).

AI, ML, DS for Traffic Monitoring, Prediction, Management

- Results on 5x12 hours of video streams captured from actual traffic cameras in the city of Los Angeles.

Automated Counted by Developed System	Ground Truth Counted by Human
2268	2230
Overall Percent Error = 1.7%	

Let’s use Data Science for social good!

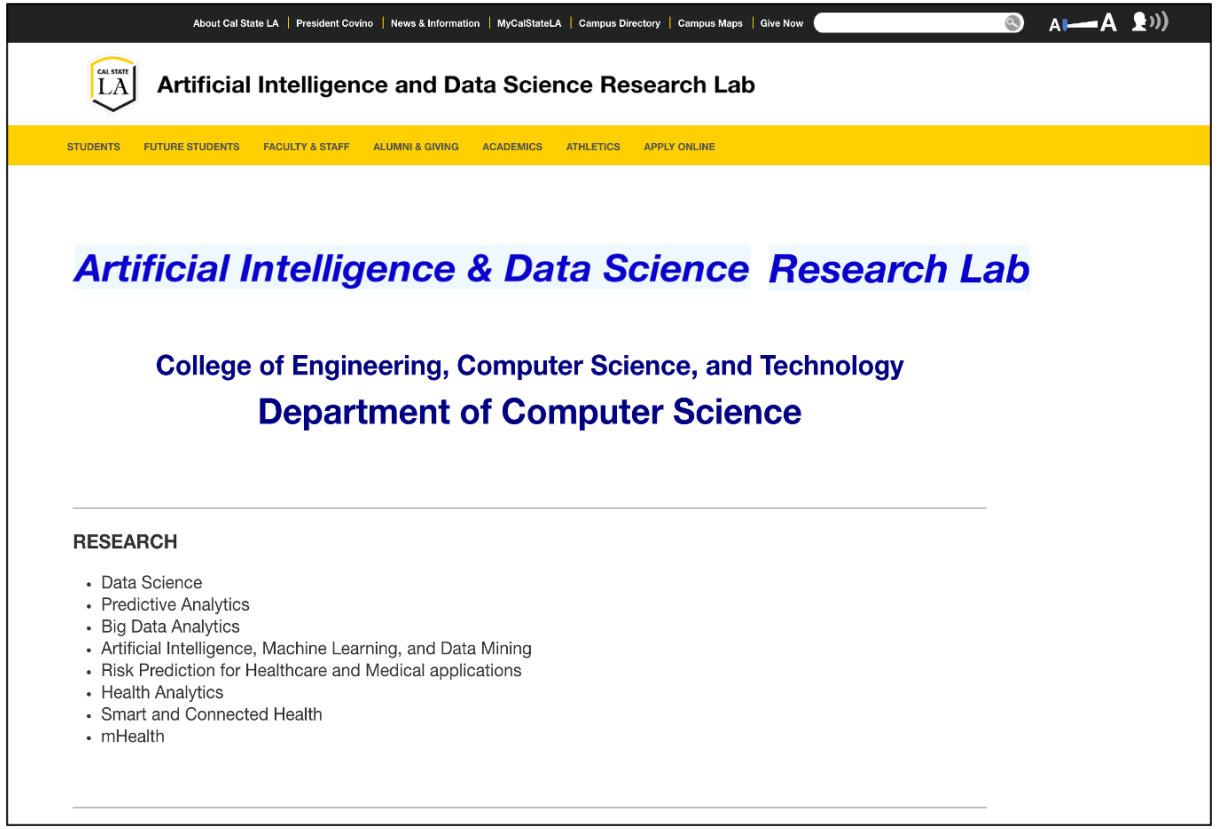
[Figure Ref]: UN.

Artificial Intelligence & Data Science Research at Cal State LA

Visit our AI & Data Science Research Lab at:

www.calstatela.edu/research/data-science

- 25 grad students theses in past 4 years
- More than 80 peer-reviewed paper publications
- PI on more than $1M in past 2 years



Thank You!