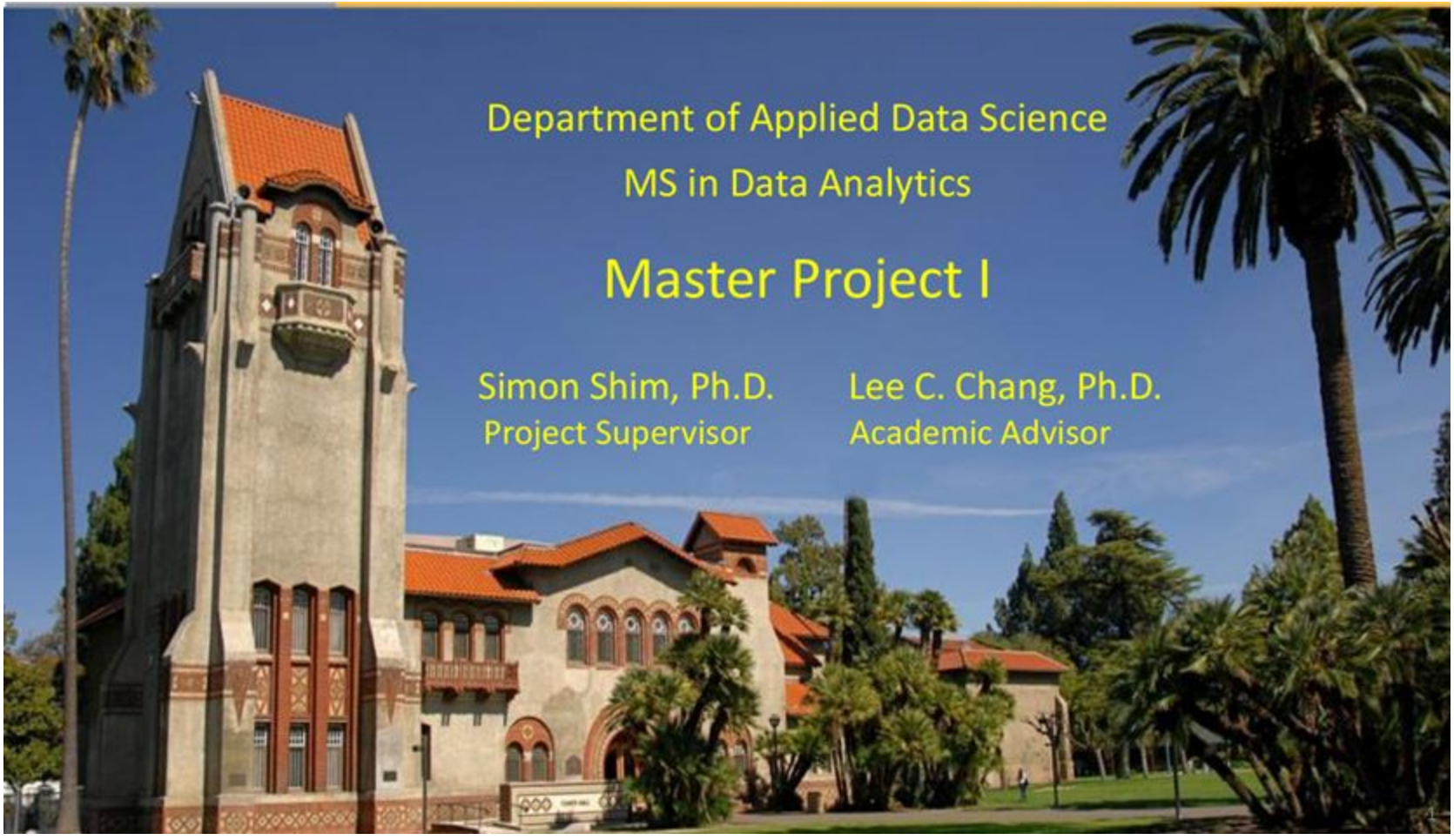
SJSU SAN JOSÉ STATE UNIVERSITY



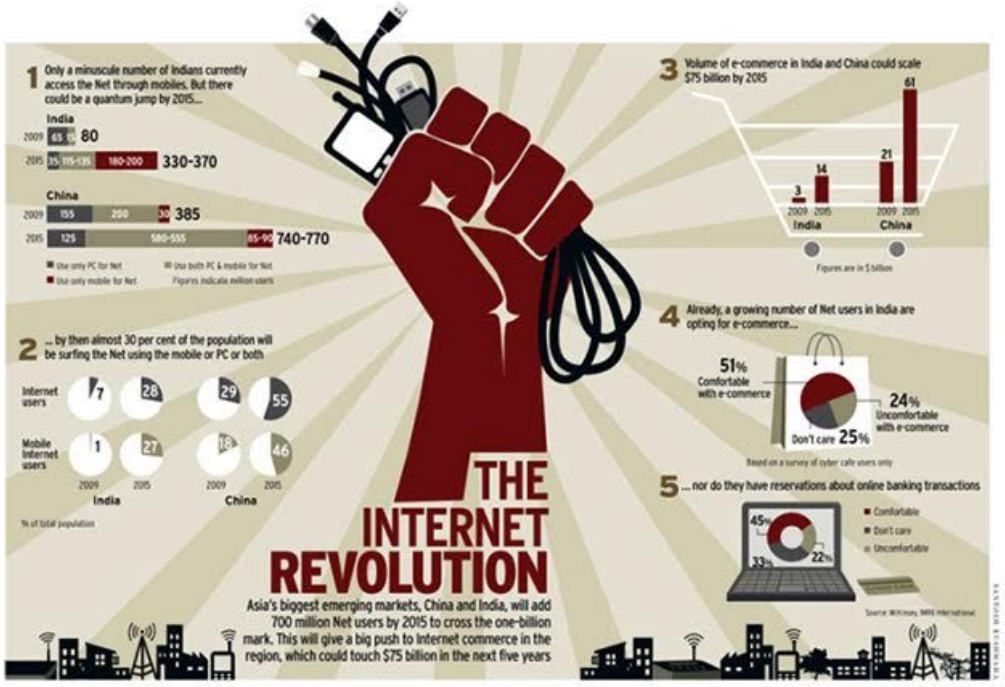
SJSU SAN JOSÉ STATE UNIVERSITY

Project Topics

- Each team decides its project topic in two ways
  - Selects from the list provided by the Project Supervisor
  - Proposes a topic to the Project Supervisor for approval
- Criteria of deciding the project topic
  - Sufficient academic contents
  - Adequate technical challenges
  - Capable to acquire domain knowledge
  - Necessary and sufficient data sets
  - Available environment and tools
  - Feasible to complete required deliverables in time

Internet Revolution

Between 1993 and 1995, the World Wide Web (www, or the Web), a user-friendly information-sharing network system, quietly came into being and began to spread.



LLM Revolution

The language revolution:
LLMs could transform the world

"If LLMs are humans, all the ideas are trivial: chain-of-thought prompting ('explain your answer'), self-consistency ('double check your answer'), least-to-most prompting ('decompose to easy subproblems'). The shocking thing is that LLMs are not humans but these still work!"

1-1 Build RT-2 VLA powered Application

Topic 1: ***Robotics Breakthrough***: Robotic Transformer 2 (RT-2) is a novel vision-language-action (**VLA**) model that learns from both web and robotics data, and translates this knowledge into generalised instructions for robotic control.Vision Language Models (**VML**), 1X Technologies, JESTON ISAC AMR, OMNIVERSE

Purpose

Build an industrial application that uses RT-2

Tasks

- Develop a task models for robot control using RT-2
- Build quality metric for model evaluation and improvements
- Collect data set for training and test

Outcome

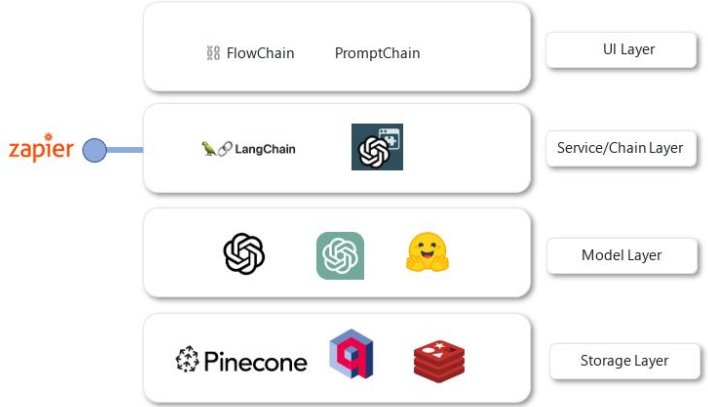
An intelligent Machine learning system that can translate text to robot commands

Q: What should the robot do to <task>?

Δ Translation = [0.1, -0.2, 0]
Δ Rotation = [10˚, 25˚, -7˚]

1-2 LLM Powered applications

Topic 4: Large Language Model (LLM) Powered Conversational Applications



Purpose

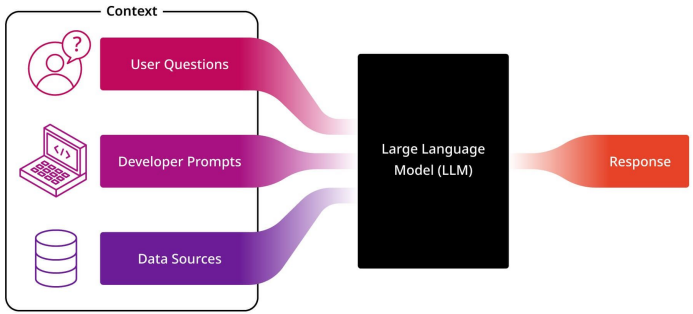
Build an LLM powered application in ***your expertise area***

Tasks

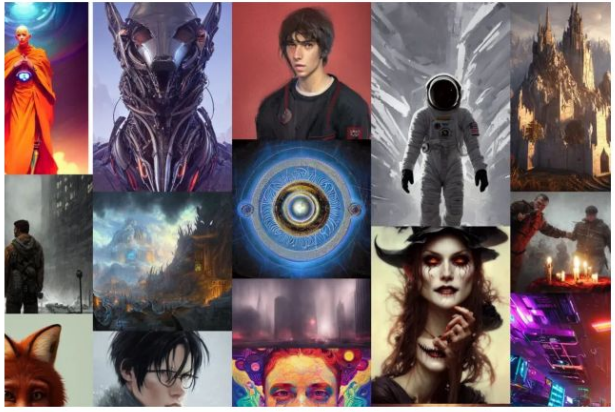
- Develop an application using LLM APIs
- Build quality metric for model evaluation and improvements
- Collect data set for training and test

Outcome

An intelligent Machine learning system that can utilize LLM on backends



1-3 Build high-quality AI text to Image Generator



Purpose

To develop a **high-quality AI Text to Image Generators**

Tasks

- AI image generator should be able to take text and turn it into an image
- Build GANs (Generative Adversarial Networks)/Diffusion model to generate an image

Outcomes

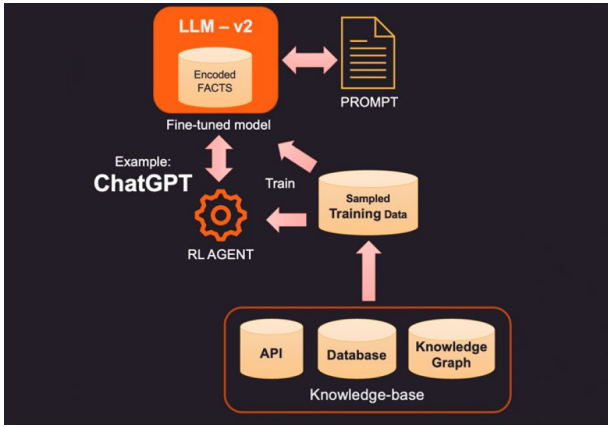
- Innovative **high-quality AI Text to Image Generators**
- A web-based portal for generating high quality images

Applications:

Generating photo-realistic images from text has tremendous applications, including photo-editing, computer-aided design

Examples are Open AI Dall-E 2, Midjouney, DreamStudio.

1-4 Customize/Fine-tune LLM



Purpose

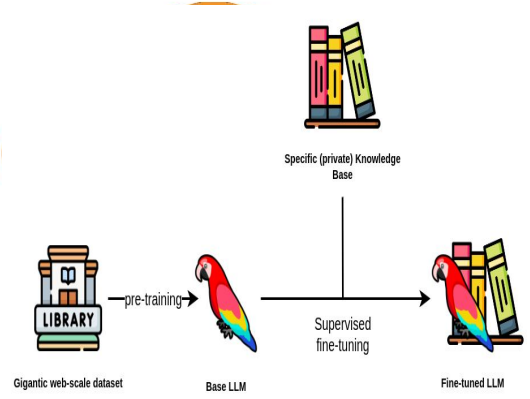
To study and develop a clone of ChatGTP system, called SpartaChatGTP, for NLP applications

Tasks

- Develop a transformer based ChatGTP language generation similar to ChatGTP.
- Develop built-in quality evaluation model and metrics to support continuous quality evaluation.
- Collect and prepare transformer based ChatGTP training and test data for chatbot development.

Outcomes

- An intelligent SpartaChatGTP system with testability
- Big training and test datasets with demo examples
- A final SpartaChatGTP system demo as an intelligent NLP generation system



Applications

Intelligent customer support for stationary service robots, for example soft drinking service robots in the dining/travel industry.

1-5 LLM Powered Personalized Chatbot

Purpose

To study and develop a rich-media intelligent chatbot system for customer support of stationary service robots in a restaurant or travel agent planning/reservation.

Tasks

- Develop a rich-media intelligent chatbot for stationary service robots in a restraint with a deep learning machine model and NLP models.
- Develop built-in quality evaluation model and metrics to support continuous quality evaluation.
- Collect and prepare rich-media chatbot training and test data for chatbot development.

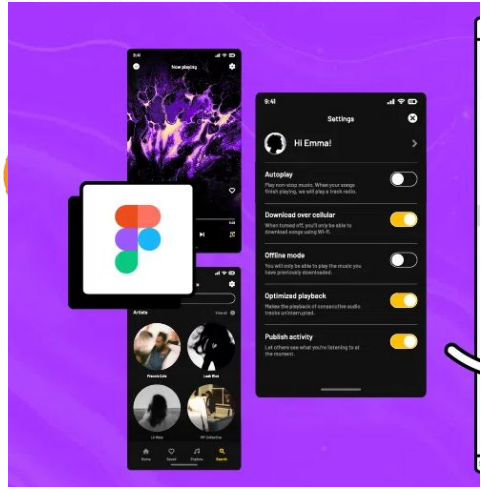
Outcomes

- An intelligent rich-media-based online chatbot system with testability
- Big data training and test datasets with demo examples
- A final chatbot system demo as an intelligent customer support for a restaurant robot/travel planning

Applications

Intelligent customer support for stationary service robots, for example soft drinking service robots in the dining/travel industry.

1-6 LLM Powered Figma Development



Purpose

To study and develop a rich-media intelligent chatbot system for customer support of stationary service robots in a restaurant or travel agent planning/reservation.

Tasks

- Develop a rich-media intelligent chatbot for stationary service robots in a restraint with a deep learning machine model and NLP models.
- Develop built-in quality evaluation model and metrics to support continuous quality evaluation.
- Collect and prepare rich-media chatbot training and test data for chatbot development.

Outcomes

- An intelligent rich-media-based online chatbot system with testability
- Big data training and test datasets with demo examples
- A final chatbot system demo as an intelligent customer support for a restaurant robot/travel planning

Applications

Intelligent customer support for stationary service robots, for example soft drinking service robots in the dining/travel industry.

1-7 LLM Ontology-prompting for Knowledge Graph Extraction

Purpose

Prompting an LLM with an ontology to drive Knowledge Graph extraction from unstructured documents

Tasks

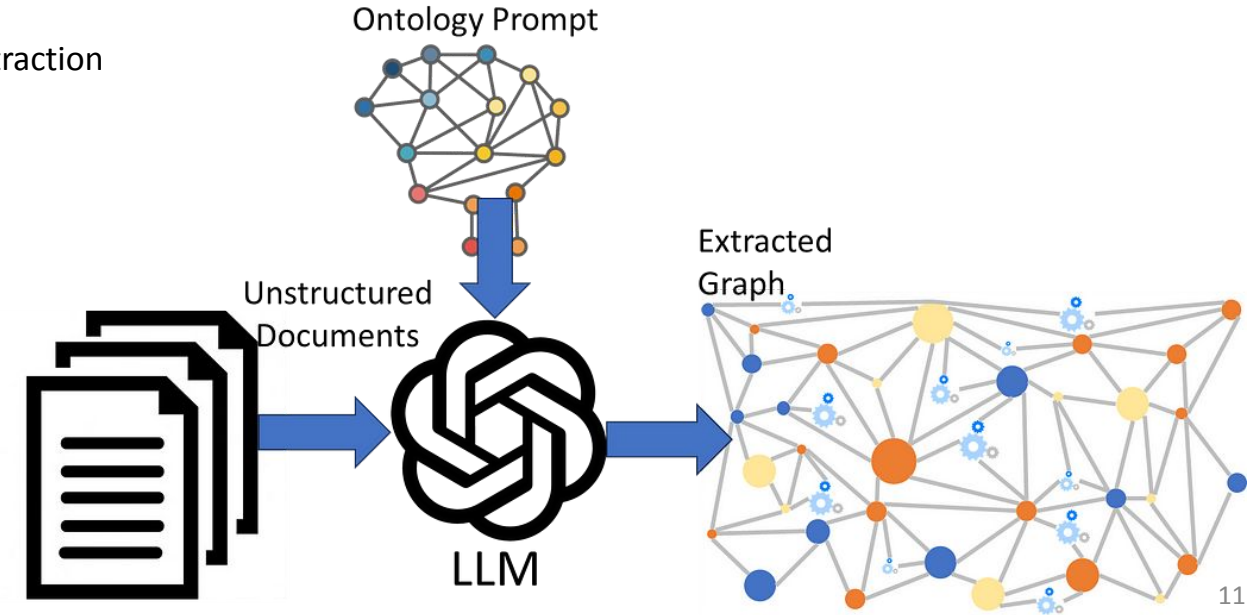
- *Process unstructured document*
- *Extract graphs to a specific ontology*
- Ontology Prompting

Outcome

Develop LLM Powered Knowledge graph extraction

Application

Text Analysis for Content Management
Knowledge Graph Applications



1-8 Object Tracking with Detectron2

Purpose

Apply Detectron2 for object detection and tracking from satellite video/series of images

Tasks

- Collect moving object video/images
- Design metrics for model validation
- Build a object tracking system using Detectron2
- Adapt Detectron2 and enhance it (by re-training or other techniques) for object tracking.

Outcomes

Develop object identification and tracking

Measure the improvement (or degradation) of object detection and tracking over satellite images

Applications

Satellite Images Help Track a Vehicle

assessing building damages from satellite images

1-9 Amenity Detection and Inventory Tracking with Detectron2

Purpose

Apply Detectron2 for amenity detection

Tasks

- Collect image dataset, taxonomy, creating annotations
- Model training and model validation
- Model deployment and online serving

Outcomes

Develop amenity detection
Measure the improvement (or degradation) of amenity detection

Applications

Amenity Detection, Broad-scope Object Detection and Image Quality Control help Airbnb become a smarter and safer home-sharing platform for our hosts and guests

1-9 Sign language generation

Text to Sign Language

Purpose

Generating continuous sign language videos from text input

Tasks

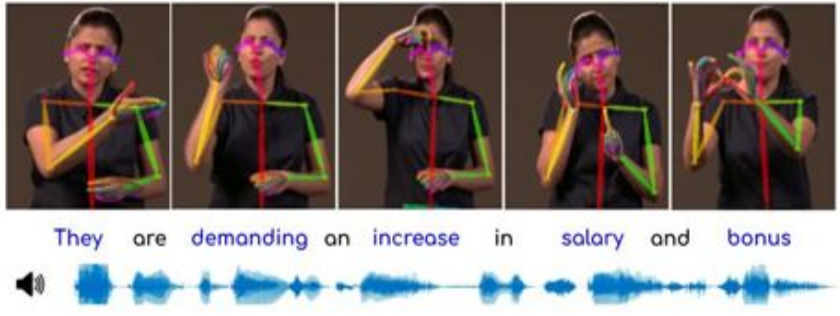
- *Text to Sign language generation*
- *Natural and continuous motion video*
- GAN/Stable Diffusion model training

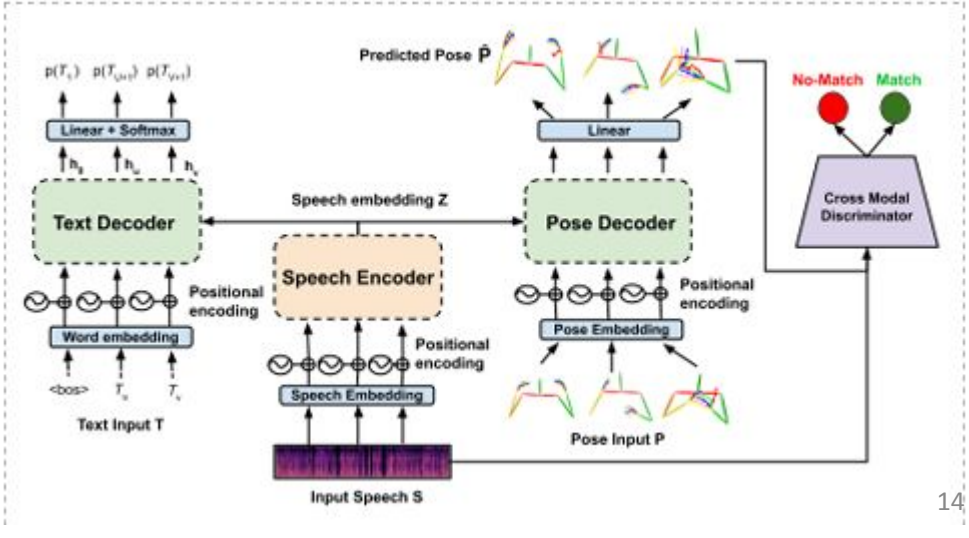
Outcome

Continuous sign language videos

Application

Sign language generator/translator





Topics from Previous Semester

2-1 Identification Fraud Detection Using Machine Learning

Purpose

Apply machine learning to detect potential identification fraud using physiological and audio data

Task

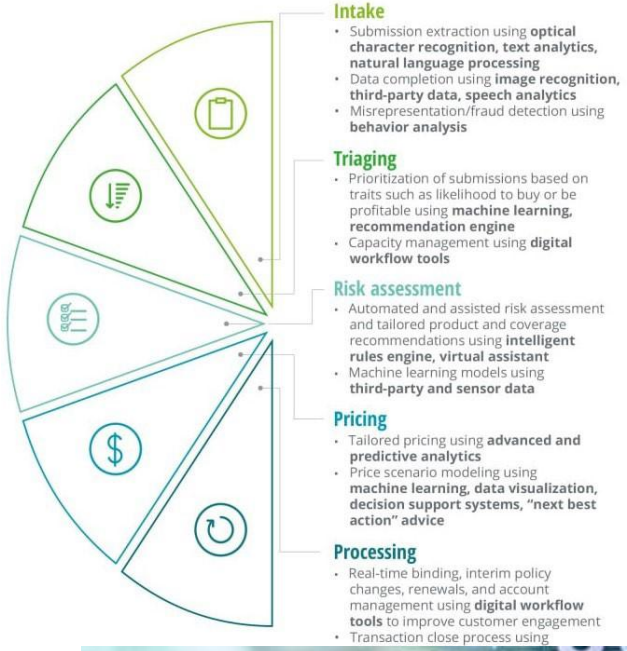
- Collect personal vital and audio data to build a data-driven machine learning model to detect distinct identities
- Develop a model capable of detecting fraud in identity by using vital & audio data.
- Setup an online web portal for demos.

Outcomes

- A data-driven machine learning model able to detect potential fraud in identity using per
- An interactive web portal to demonstrate the detection of identification fraud

Applications

Insurance underwriting and other commercial applications



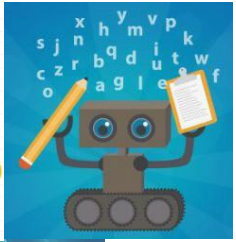
https://www2.deloitte.com/us/en/insights/industry/financial-services/future-of-insurance-underwriting.html

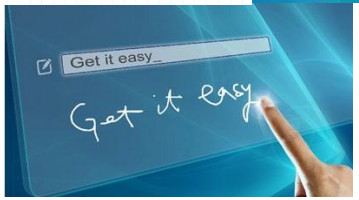
SJSU SAN JOSÉ STATE UNIVERSITY

2-2: An Intelligent Multi-language Question Answering System Based on Handwriting Recognition

Purpose

Develop a smart, self-learning multi-language question answering system for customer support in a selected application domain



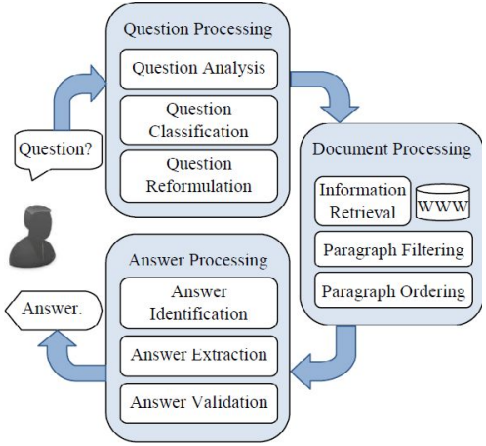


Task

- Develop a handwritten recognition system with the ability to automatically recognize at least 2 languages.
- Develop an answering system with the ability to search in Big Data knowledge base.
- Provide rich-media answers (e.g. image + text) in the corresponding question asked language.

Outcomes

- A multilingual smart question and answering system which can automatically recognized the user's handwritten language and provide valuable answer based on data stored or web-search for the question asked.



Applications

Public services, amusement park, hotels, shopping malls, and visitor centers, etc.

2-3: Climate Change Impact Prediction Using Lidar Data and Machine Learning

Purpose

Apply machine learning to predict major effects of climate changes using Lidar data such as glacier melting, rising sea-level, potential coastal flooding.

Task

- Collect relevant Lidar dataset to build a data-driven machine learning model to forecast the potential climate change impacts – glacier melting, rising sea-level, coastal flooding risk.
- Develop a model able to predict and project changes on map.
- Setup an online web portal for demos.

Outcomes

- A data-driven machine learning model able to forecast impacts from climate change and project results on map.
- An interactive web portal to demonstrate impacts from climate change.

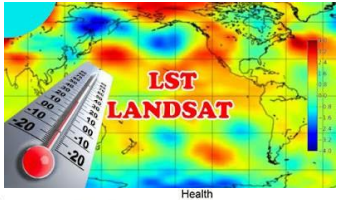
Applications

Environmental groups, geological survey groups, and climate change surveillance.

2-4: Crops Yield And Price Prediction Using Remote Sensing Images

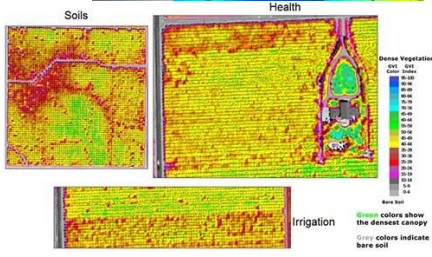
Purpose

Apply machine learning to predict land surface temperature and soil health using satellite images and their effect on crops yield and price.



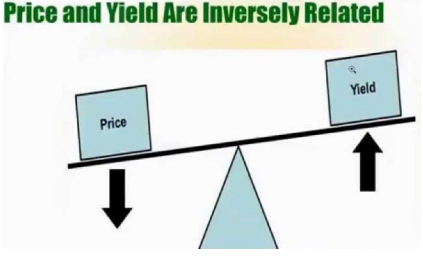
Task

- Collect data from NOAA and generate a dataset from remote sensing image to predict land surface temperature of a specific area (regional, can be county or state).
- Build a model to predict crops yield using land surface temperature and the corresponding soil health data.
- Predict price changes from the amount of crops yield.
- Setup an online web portal for demos.



Outcomes

- A data-driven machine learning model able to predict crops yield and price using land surface temperature and soil health.
- An online web portal supporting regional agricultural community to predict and analyze crops yield and price.



Applications

Agriculture industry marketing analysis, production forecast, planning & price prediction.

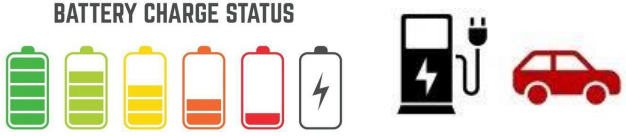
2-5: EV Battery Optimization – Real Time SOC and Plug-In Duration Forecasting Using Machine Learning

Purpose

Real time estimation of the state-of-charge (SOC) of electric vehicle (EV) under dynamic operating conditions and forecast the full charge plug-in duration.

Tasks

- Develop an online system to estimate the SOC of the battery based on different driving and operating conditions.
- Forecast the charging time needed to fully-charge the battery.
- Create a dashboard for the battery management system (BMS).



Outcomes

- A data-driven machine learning battery management system (BMS) for electric vehicles.

Applications

Battery management system for EV or other application (e.g. cellular phone).

Topics from
DATA298A-21

3-1: Smart Mirror-Based Fashion AI Based on Machine Learning

Purpose

Apply and develop machine learning techniques to Fashion AI for smart home mirrors

Tasks

- Develop and train machine learning models to support customized fashion AI based on smart home mirrors.
- Prepare necessary training and test datasets by collaborating with selected fashion companies in fashion lines.
- Develop a web-based solution based on a given smart mirror to show transforming digital models to customers in latest fashion lines.

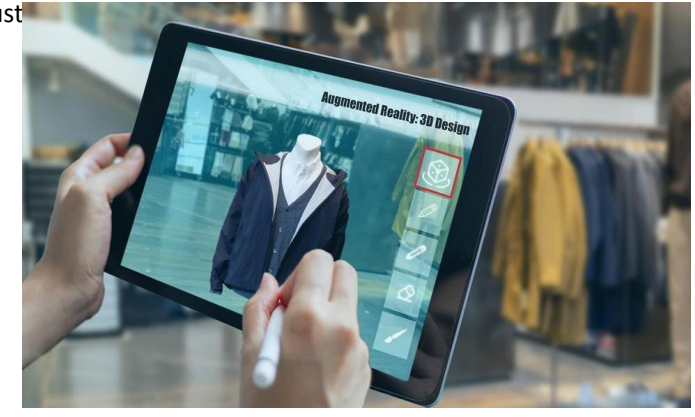
Outcomes

- Developed comparative models and innovative fashion AI models
- Completed training and test datasets in fashion AI for customers.
- Developed a web-based interactive solution to show prediction results comparative results and demos based on given inputs.

Applications

- Apply machine learning models to fashion industry for retail customers based on smart mirror technologies.

https://www.forbes.com/sites/bernardmarr/2021/03/26/three-ai-and-tech-trends-that-will-transform-fashion-indust



3-2: Detecting and Classifying Defects In Chips With Machine Learning

Purpose

Apply machine learning techniques to detect and classify diverse detects in chip manufacturing in semiconductor industry.

Tasks

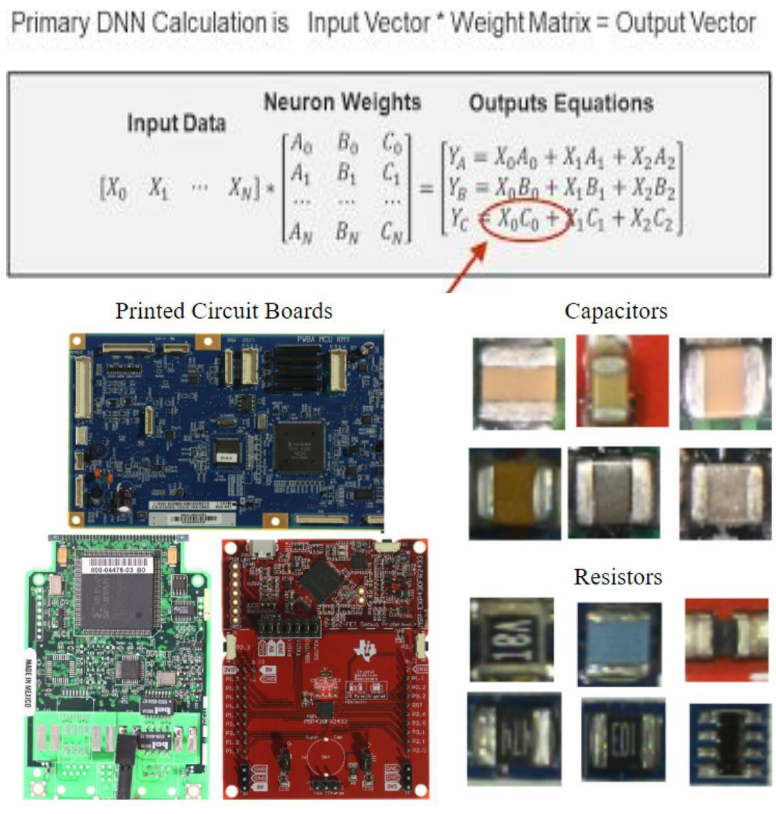
- Develop and train machine learning models that detect and classify different types of chip detects.
- Prepare necessary training and test datasets based on the industry provided datasets.
- Develop a web-based solution to show interactive comparative results and demos.

Outcomes

- Developed comparative models and innovative models
- Completed training and test datasets.
- Developed a web-based interactive solution to show prediction results comparative results and demos based on given inputs.

Applications

- Apply the solutions to improve chip detection and classification in semiconductor industry.



https://www.mdpi.com/2410-387X/5/1/9

3-3: Data-Driven Soil Analysis and Evaluation for Agriculture Using Machine Learning Approaches

Purpose

Using machine learning approaches to conduct soli analysis and quality evaluation based on big data

Tasks

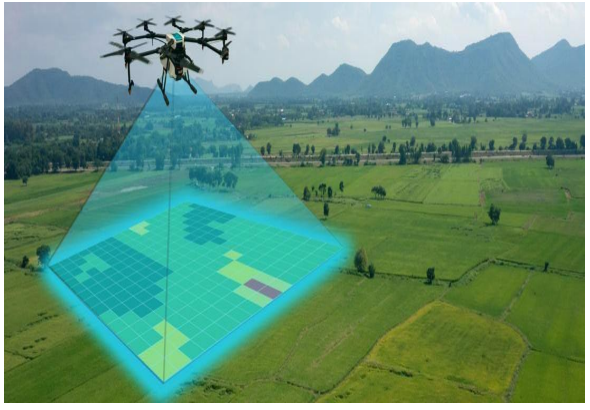
- Develop data-driven machine learning models to support farm land soil analysis and quality evaluation for selected farm lands based on big data sets, including satellite and remote sensing datasets, and drone collected data.
- Prepare soil big data for farm lands, including diverse big data types.
- Develop a web portal to support farm land oriented soil analysis and quality evaluation based on big data and machine learning models learning models.

Outcomes

- Innovative soil analysis and evaluation models using machine learning approaches
- Prepared diverse soil data for farm lands.
- A web portal system for project demonstration

Applications

Smart agriculture and intelligent platforms



3-4: Data-Driven Air Quality and Environment Analysis and Evaluation for Cattle Farms

Purpose

To learn, analyze, and evaluate air quality and environmental impacts at selected livestock farms (such as cattle farms) based on machine learning models and big data.

Tasks

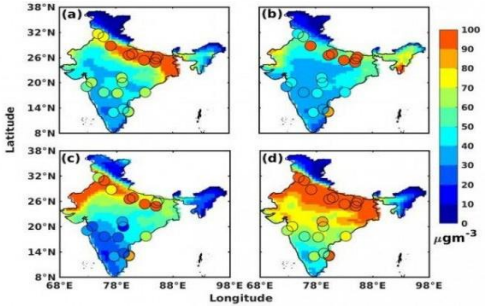
- Develop data-driven machine learning models supporting air quality analysis and environment impacts for selected cattle farms based on diverse big data.
- Typical data sets includes satellite and remote sensing data, cattle farm data as well as weather big data sets.
- Develop a web portal to support air quality analysis and environmental impact evaluation for selected cattle farms based on developed machine learning models and diverse cattle farm big data, including remote sensing and satellite image data, weather data, and cattle farm data.

Outcomes

- Machine learning models for cattle farm air quality analysis and environment impacts.
- A web portal supporting the project demonstration

Applications

- Smart livestock framing and platforms.

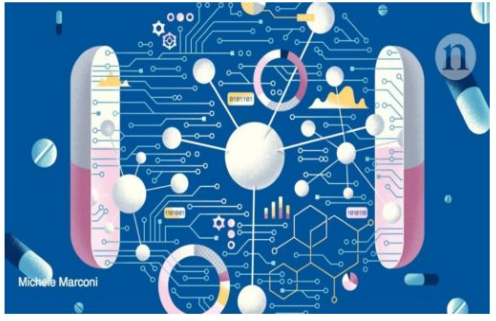


[Outstanding projects: http://cs230.stanford.edu/past-projects/](http://cs230.stanford.edu/past-projects/)

Propose A Topic

https://cs230.stanford.edu/past-projects/#fall-2021

Everybody Compose: Deep Beats To Music by Yixin Liu, Tom Shen, Violet Yao: report
Using pre-Q sequences of Reddit posts to predict user-level QAanon participation by Lillian Ma, Stephanie Vezich: report
Semantic Segmentation of Extreme Climate Events by Hannah Grossman, Lucas Hendren, Romain Lacombe, David Ludeke: report
Style transfer for rooms by Marie Chu, Warren Xia: report
DARE-Net: Speech Dereverberation And Room Impulse Response Estimation by Paul Calamia, Jacob Donley: report
A Generative Deep Learning Approach for Alzheimer's Disease Drug Discovery by Adrian Gamarra Lafuente, Akash Gupta, Justin Shen: report



Other Topics

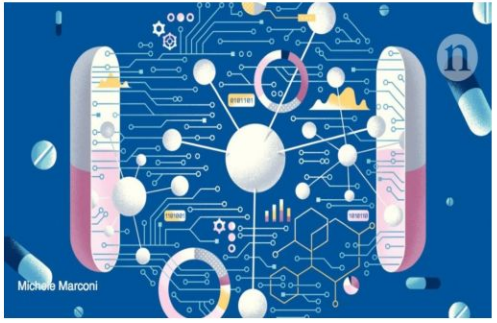
Project Proposal Approval: email to simon.shim@sjsu.edu
subject line: DATA 298A project approval, Team#

What should include in the proposal:

- One paragraph abstract
- Cite the source and how to improve the existing projects.
- Clearly state **what/how you improve/change** from the original
- link to datasets (different dataset from the source) or prepare for your own training and test data for development.
- application areas

Outcomes

- An intelligent system with testability
- Big data training and test datasets with demo examples



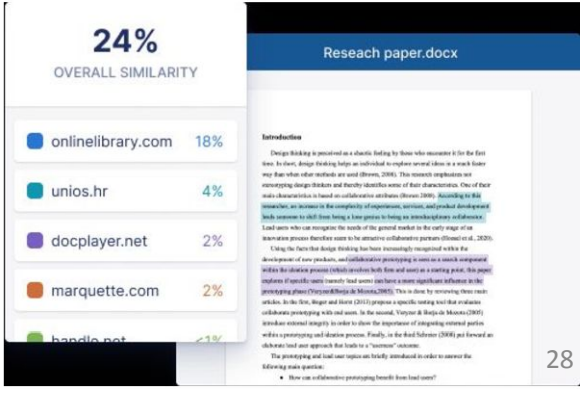
plagiarism

ChatGPT is not allowed in writing reports.

- ACADEMIC INTEGRITY POLICY
- https://www.sjsu.edu/senate/docs/S07-2.pdf

- San José State University defines plagiarism as the act of representing the work of another as one's own without giving appropriate credit, regardless of how that work was obtained, and submitting it to fulfill academic requirements.

- Every assignment will be checked via turnitin plagiarism checker
    - Report to the university for dishonesty if Turnitin score > 40%



SJSU SAN JOSÉ STATE UNIVERSITY

Report Grading

- You will see "poor", "good", "excellent" in assignment rubrics.
- "**poor**' means it was not meeting the general expectations and had some minor flaws/missing parts,
- "**good**" means the section was meeting the expectations and no major flaws.
- "**excellent**" means it went beyond the expectations and much better than other teams' reports.
- I am expecting all teams to get "good" rating and only few teams will get "excellent" ratings.

Team Work


“CONSTANTLY THINK ABOUT HOW YOU COULD BE DOING THINGS BETTER AND QUESTIONING YOURSELF.”
DANSILVESTRE.COM


"I think it's important to take as much feedback as you can from as many people as you can about whatever idea you have. You should seek negative feedback, especially from friends."
- Elon Musk
Abbott Media

SJSU SAN JOSÉ STATE UNIVERSITY

Team Work