Omsairam

-- Understandable

-- Accessible

-- Structured

-- Write in purpose evidence-based

2nd floor 237B TMH -> prof office

[gregory.tomlinson@sjsu.edu](mailto:gregory.tomlinson@sjsu.edu)

assignments

-- diagnostic -> how u approach writing

-- resume/cover letter

-- technical docs

-- technical

-- simplified

-- research project proposal

-- peer-reviewed literature review

-- oral ppt

Grading criteria

-- argument

-- clarity

-- grammar spelling syntax

-- structure

-- Know your audience

peer and faculty feedback

* revise and resubmit

with 3 other people's feedback (peer)

* submit printed copy in the class

2nd floor MLK lbrary

[www.sjsu.edu/writingcenter/](http://www.sjsu.edu/writingcenter/)

Admin building 110

* Assignment 1

Write a 2-3 paragraph composition telling me

Who you are

Where you went to school

Why you choose sjsu for graduate school

What project you’d like to work on

Omsairam

Lec-02

Active vs passive

Past perfect

Acronym

Spelling, grammar, syntax

Proper spelling and spell check

Grammar

Active voice -> subject doing it is the action doer.

Subject-object language

Sentence length > avoid run -ons.

Comma splices -> avoid

Colons and semicolons -> separate sentence , avoid them by using short sentences

Idiomatic expressions vs formal writing ->

Editing and revision

Example : write / share about mac vs. PC functionality (short paragraph)

Step 1: know your audience

For whom u r writing

What terms can I use or not use

Jargon vs explanations -> first acronym use like San jose state university(SJSU), then u can use SJSU

Tailoring arguments / content

The granny test

Internal communications

Managing and delegating

Research vs. commercial products

Step 2 : planning

Understand the assignment

Expectations: Communicate early and often

Timeframe : know the due date(s)

Time management is the key

Deliverables and feedback

Ask questions

Respond promptly

Offer suggestions, edits

Be receptive and open-minded

Step 3: write

Ready, aim, fire not ready, fire, aim

Set time aside for writing

Minimize distractions

Remember:

Argument/purpose

Readability

Accessibility

Structure

Clarity

Everything else is extra

Step 4: editing and revision

Revise ad resubmit

Peer review and edits

Read aloud

Have another person read the writing to you

Edit, edit , edit

Sound practices

Find a style that works for you

Be understandable first

Shorter sentences > longer sentences

Structure is imperative

Logical flow: ideas build upon themselves

Accessibility:

Know your audience

Terms/language vary

Be direct, not evasive

Communicate early nd often

Practise and ask for help

“Explain the challenges with acquiring the necessary data and training AI models while respecting intellectual property and customer privacy”

Review and edit the last sample paragraph

Artificial Intelligence (AI) has been a hot topic of discussion in the current world. It’s not a very simple field to work in, so one must be an expert to learn the intricate details of AI to understand its functionality and build applications on it.

AI or Artificial Intelligence as its name suggests is the field of study in computer science that deals with collecting data and training the machine to build models and predict and find solutions to the problem statement. The main challenge in building the model to solve the problem is to collect the data. That’s because the data you use to train to model is the basis on which the model would be trained, and the accuracy and perfection of the machine depend on the trained data that is fed into the system.

There are multiple challenges in acquiring the necessary data to train the AI model. Consider an example of an AI to provide suggestions for shopping. The issue is that data to train this model must be very personal. The AI needs to know what you wear, what you searched for previously and what is your lifestyle. The AI must train using your browsing history, your phone records, your photos, etc. So, the customer might feel like it is invading their privacy while getting the data.

As seen in the above example, customers’ privacy and the search for data in their private properties like phones or laptops are required to build the model in the example. So, the developer of the AI needs to be quite careful as to how much is too much. The limit of the crawling for the information should be restricted as per the customer’s need. For example: the customer might have provided some restrictions on the search history, so the crawler or the AI shouldn't be collecting data from that space.

Let ‘s look into another instance where maintaining intellectual property is a challenge in building an AI model. Everyone knows ChatGPT as one of the AI tools that helps you provide information as per your search query. Now, do you ever wonder where is the training data coming from or how accurate is the search result? Now here comes the question of privacy and copyright. Should ChatGPT crawl anything on the internet to provide the information to the user? Again, the limit of the training data and the boundary of the search space should be well decided and built by the developer of the AI tool in order to avoid getting into the private data of the customers and other companies.

There are many such examples of when the developer must be quite understandable and careful as to what should be the boundary of the search space for training the model. Hence, the topic for the discussion is quite important to be considered and it’s a real challenge while building an AI model.

omsairam

Lec-03

Free writing : processing words can help to organize

Feedbacks and revision notes

* Explaining everything
* Structure and sub headings -> accessible (can be read by anyone and everyone, explain everything).
* First sentence gives a brief summary of what u are writing
* Readable document, provide context and background. Emphasize the uniqueness if your research
* Research abstract should contain all ideas in a summary view
* Purpose should be clear, what is it about , what is the conclusion
* Revision . comments from peer, advisor , profs,

omsairam

Lec -04

Structured, logical flow, simple sentence structure,

Keep same tense, voice same in same sentence

Avoid is where, is when, and reason

Keep it straight forward and keep short and effective sentence

Start of sentences: Therefore, consequently, unfortunately…. Use start of sentences as empathetic

Running logs:

Strategy

Urs , then , vs now the same document

What to use from others writing

Make the logs, edits and comprehensions

Lessons from peers

Your feedbacks to peers

Self assessments

Running logs for assignment 1:

V1: Artificial Intelligence (AI) has been a hot topic of discussion in the current world. It’s not a very simple field to work in, so one must be an expert to learn the intricate details of AI to understand its functionality and build applications on it.

AI or Artificial Intelligence as its name suggests is the field of study in computer science that deals with collecting data and training the machine to build models and predict and find solutions to the problem statement. The main challenge in building the model to solve the problem is to collect the data. That’s because the data you use to train to model is the basis on which the model would be trained, and the accuracy and perfection of the machine depend on the trained data that is fed into the system.

There are multiple challenges in acquiring the necessary data to train the AI model. Consider an example of an AI to provide suggestions for shopping. The issue is that data to train this model must be very personal. The AI needs to know what you wear, what you searched for previously and what is your lifestyle. The AI must train using your browsing history, your phone records, your photos, etc. So, the customer might feel like it is invading their privacy while getting the data.

As seen in the above example, customers’ privacy and the search for data in their private properties like phones or laptops are required to build the model in the example. So, the developer of the AI needs to be quite careful as to how much is too much. The limit of the crawling for the information should be restricted as per the customer’s need. For example: the customer might have provided some restrictions on the search history, so the crawler or the AI shouldn't be collecting data from that space.

Let ‘s look into another instance where maintaining intellectual property is a challenge in building an AI model. Everyone knows ChatGPT as one of the AI tools that helps you provide information as per your search query. Now, do you ever wonder where is the training data coming from or how accurate is the search result? Now here comes the question of privacy and copyright. Should ChatGPT crawl anything on the internet to provide the information to the user? Again, the limit of the training data and the boundary of the search space should be well decided and built by the developer of the AI tool in order to avoid getting into the private data of the customers and other companies.

There are many such examples of when the developer must be quite understandable and careful as to what should be the boundary of the search space for training the model. Hence, the topic for the discussion is quite important to be considered and it’s a real challenge while building an AI model.

Feedback by peers : to improve: sentence structure to be improved ,

Well done on : shift of paragraphs, providing brief description and examples

Self feedback : same as above

V2:

Artificial Intelligence (AI) continues to be a prominent and complex subject in the contemporary world. Mastering the intricacies of AI is essential for effectively comprehending its operations and creating applications within this field.

AI, an abbreviation for Artificial Intelligence, embodies a branch of computer science that revolves around data collection and machine training to formulate predictive models and devise solutions for a variety of challenges. The principal hurdle encountered during model development lies in data aggregation. The quality of data employed for model training is the cornerstone upon which the model's effectiveness hinges. The accuracy and proficiency of the machine are inherently tied to the quality of the training data it is exposed to.

Obtaining the requisite data to train AI models is met with multiple challenges. Consider the case of an AI tasked with offering shopping recommendations. The crux of the issue lies in the personal nature of the required data. For the AI to operate optimally, it necessitates access to personal preferences, browsing history, lifestyle choices, and even images. This scenario raises concerns about encroaching on individuals' privacy when extracting such data.

As evidenced by the illustration, respecting customers' privacy while extracting data from their private domains, such as smartphones or laptops, is a key consideration during model development. Striking the right balance is paramount; developers must exercise discretion in their data collection efforts, ensuring that the extent of information gathering aligns with users' comfort levels. For instance, if a customer has imposed limitations on search history access, the AI or data crawler must adhere to these restrictions.

Another facet of AI development that presents a challenge pertains to safeguarding intellectual property. ChatGPT, a widely known AI tool renowned for its query-based information provision, raises pertinent questions. Users may ponder over the origin of training data and the accuracy of search results. The ethical dimension of privacy and copyright comes to the forefront. Should ChatGPT indiscriminately scour the internet to supply information? This query underscores the need to delineate the boundaries of training data acquisition and search parameters. Developers of AI tools must diligently define these parameters to steer clear of intruding into users' private information or infringing on copyrighted material.

The instances outlined above underscore the essentiality of discernment and caution on the part of developers. Determining the limits of search space for model training is pivotal. Consequently, the focal point of this discourse gains substantial significance, encapsulating the intricate challenges inherent in AI model construction.

5- 7 sentences about a topic of your research

What the topic is

What the background/context of the research is

What you aim to discover and argue in your research

What the broader impact of your research will be

Any ethical implications of your research

In today’s world, everyone uses social media. For any content on the internet, we might want to measure the sentiment a post or a topic creates among the people. Hence, as part of my research project, I would like to work on a field of Natural Language Processing (NLP).

NLP comprises machine learning to get the sentiment analysis depending on the dataset we feed into the training model. The trained model will provide statistics about a post on the social media. The statistics will include: the feedback on the post based on geographic location, influence of celebrities, trending topics, etc.

These statistics will be used to create a sentiment diagnostic of a particular topic in that period. This in turn can be used in different sectors like technology, politics, trading, and more to define their strategy for the forthcoming future. For instance, depletion of non-renewable resources has led to shifting to renewable sources of energy. Using the NLP model, we can get the diagnostics that new cars will start using electricity rather than diesel. Hence, people can buy stocks in companies that sell electric cars.

Unfortunately, there are some challenges related to training the NLP model. As crawling the data on social media to train the NLP might affect the privacy of the people, determining the limits of search space for model training is pivotal.

Omsairam

Lec-05

Between 2 people, among three or more

Be definite, explain terms, acronym

Outline :

1. Brainstorm
2. Evernote for planning: like comment code
3. Calendar

Research assignment:

1. Topic selection : ethical considerations included
2. Proposal : draft, final copy
3. Literature review: peer-reviewed scholarship
4. Oral ppt: 3 minutes, argumentative, informative

Topic:

Rath\_1\_RESEARCH\_TOPIC

2 sentences: what is it and why are you researching it

Research questions/hypothesis (2-3 sentences)

Broader significance (3 -4 sentences)

Why important

Applications and utility of research

Potential ethical consideration of project

.docx or pdf version submitted to canvas portal

Can submit as many times as necessary util deadline

My research topic is performing sentiment analysis on social media website like twitter using natural language processing (NLP). I want to analysis the sentiment of the people on the huge data on the internet in these twitter posts.

This research will be based on answering questions like what is trending on the internet, whether a post is given a positive, negative, or neutral feedback. Hypothesis will be that the machine that will be trained to check for words to create the sentiment matrix, will not be understanding sarcasm.

The sentiment analysis will be used to create a sentiment diagnostic of a particular topic in that period. This in turn can be used in different sectors like technology, politics, trading, and more to define their strategy for the forthcoming future. For instance, depletion of non-renewable resources has led to shifting to renewable sources of energy. Using the NLP model, we can get the diagnostics that new cars will start using electricity rather than diesel. Hence, people can buy stocks in companies that sell electric cars.

Unfortunately, there are some challenges related to training the NLP model. As crawling the data on social media to train the NLP might affect the privacy of the people, determining the limits of search space for model training is pivotal.

Omsairam

1. Cite sources

Exercise 1:

The issue at hand revolves around Zoom's updates to its terms and conditions, allowing access to user videos, chats, and audio data. This move sparked a significant backlash from companies relying on Zoom's services. Subsequently, Zoom pledged not to utilize user data for AI model training without consent, though doubts about the sincerity of this promise emerged. As a result, numerous companies have either halted or are considering discontinuing their use of Zoom to safeguard user privacy.

Zoom's terms and conditions, as cited, broadly encompassed "Customer Input," granting the company extensive rights to employ this data for model training. This elicited strong opposition across the Zoom user base as concerns about privacy violations grew.

Critics expressed widespread dissatisfaction with Zoom's seemingly nonchalant modification of its terms and conditions, which appeared to infringe upon user privacy. A commentator on Ycombinator even quipped, "I, for one, do not welcome our dystopian overlords."

In response to the uproar, Zoom revised its terms to clarify that user data, including audio, video, or chat, would not be used for AI model training without explicit user consent. However, Sean Hogle, a legal expert specializing in business and intellectual property, pointed out that this amendment applied solely to "Customer Content" and did not cover "Service Generated Data." This has raised concerns, as it implies that Zoom's commitment to safeguarding user data remains incomplete and lacks comprehensive protection.

In line with this sentiment, Jan Schakowsky (D-Ill.) said, "Zoom has a poor track record of protecting consumers' data and living up to its promises -- as their consent order and 2021 settlement prove".

Omsairam

How to search for project

1. covid 19 OR coronavirus OR pandemic) AND ("supply chain")
2. ACM digital library
3. IEEE xplore
4. Engineering village
5. Scopus
6. Google scholar (library proxy)

Research topic: explore the development and deployment of AI-driven defenses against cybercrime

Concepts : AI-driven defenses : AI, machine learning

Cybercrime : security

Go to ACM and put below query

("artificial intelligence" OR "machine learning" OR "AI" OR "ML") AND (cybercrime OR "information security")

<https://dl.acm.org/search/advanced>

use only ACM and IEEE research articles.

Omsairam

Problem statement

Title: Enhancing Sentiment Analysis through Natural Language Processing (NLP)

* General Area of Research:
* The research revolves around the field of Natural Language Processing (NLP) applied to sentiment analysis. It seeks to develop and improve algorithms and models that can accurately determine sentiment in textual data, enabling deeper understanding of public opinion and emotional responses.
* Problem Description and Justification:
* Sentiment analysis plays a critical role in various domains like marketing, customer service, political analysis, and product development. However, traditional sentiment analysis techniques often struggle to capture nuances, sarcasm, context, and cultural variations present in human language. By employing NLP techniques, we aim to overcome these limitations and enhance sentiment analysis accuracy and efficiency, making it a more valuable tool for decision-making and understanding public sentiment.
* Proposed Solution:
* The proposed solution involves leveraging advanced NLP techniques, such as deep learning models (e.g., recurrent neural networks, transformers), linguistic analysis, and sentiment lexicons, to build a sophisticated sentiment analysis model. This model will be trained on labeled datasets to learn and interpret the complexities of human language, thus improving sentiment classification.
* Evaluation of the Solution:
* The solution's effectiveness will be evaluated based on various metrics, including accuracy, precision, recall, F1 score, and overall model performance. Additionally, user feedback and real-world applications in different domains will be considered to assess the model's practicality and relevance. The goal is to develop a robust and adaptable sentiment analysis tool that can accurately analyze sentiments across diverse textual data.

Omsairam

Sample project proposal draft ?

Omsairam

Where is the 20 points in ?peer review for which one ?