



CMPT 413/713: Natural Language Processing

Project information

Spring 2026
2026-02-11

Project

- Project timeline
 - 2-18 - Initial abstract/title (not graded) for initial feedback
 - 2-23 - Proposal due
 - 3-19 - Project milestone due
 - 4-6 - Final project videos
 - 4-9 - Final project report due

Project Abstract and Proposal

- Initial title and abstract (~250 words)
 - Think about what you want to do for your project
 - Have an informative title (“CMPT 413 class project” is not a good title)
 - Describe the topic/problem your group will work on
 - Scope of the project
 - Data and compute resources you plan to use
- Proposal (1-2 pages)

<https://angelxuanchang.github.io/nlp-class/project.html>

Project Proposal

What will you work on and why is it interesting?

- What **task** are you addressing? What is the **input / output**? Why is it interesting?
- What specific aspects will your project be on?
 - Re-implement paper? Compare different methods? Analysis?
- What **data** do you plan to use? For training vs for evaluation?
 - Preliminary statistics for your data (number of sentences, tokens, etc)
- What is the specific method or methods you will use to address the task?
 - **What will you implement by yourself vs what existing code will you use?**
 - What **compute resources** do you plan to use?

Project Proposal

How will you evaluate, and coming up with a plan.

- How do you plan to evaluate?
 - Data splits?
 - What **metrics**?
 - What experiments will you run to **compare** different variations / different approaches?
- **Timeline and work breakdown**
 - What do you plan to have by the milestone? The end of the term?
 - Who will work on what?

What kind of approaches you want to take

- In this class, we started with fundamentals and simple ML models (Naive Bayes and Logistic regression)
- You should be experimenting with methods beyond that
 - Use pretrained models! Do fine-tuning!
 - Use LLMs with few-shot prompting
 - We will cover more about these topics after the break

Evaluating LLM capabilities

Adversarial attacks

Model distillation

Type of problem and resources

<https://angelxuanchang.github.io/nlp-class/project.html>

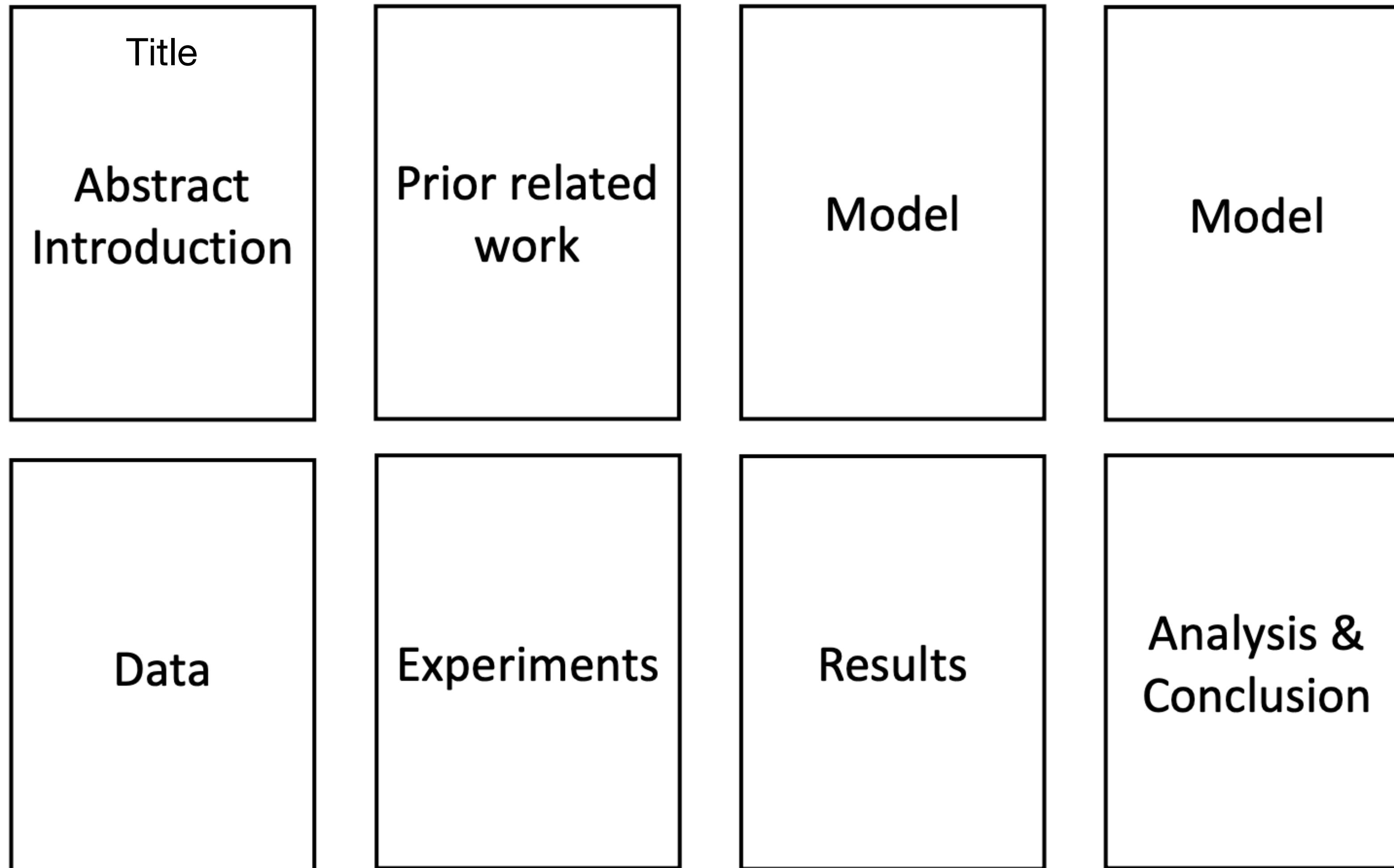
- Pick a problem based on what are you interested in?
- Types of problems
 - Building a cool NLP application
 - Re-implementing / re-reproducing a recent paper
 - Applying an existing neural model to a new task
 - Implementing a complex neural architecture
 - Proposing a new neural model or a new variation of an existing model
 - Proposing a new training, optimization, or evaluation scheme
 - Experimental and/or theoretical analysis of a NLP model
 - Probing the capabilities of current NLP models

Final project report

8 pages not including references

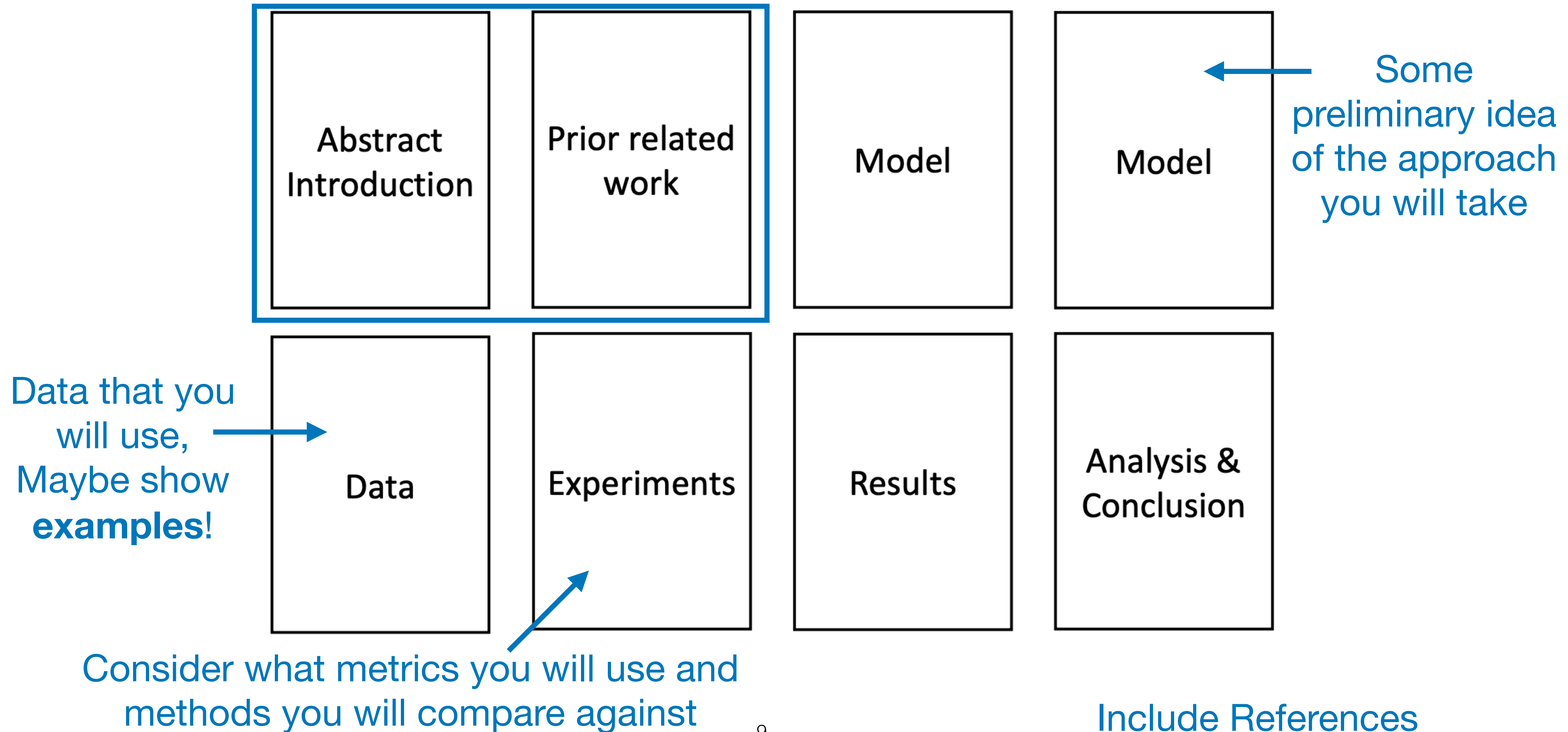
Model or what approach you are taking

Rough page layout:
your report can have
different number of
pages for each
section



What is expected to be included in the project proposal

We expect you to have around 1-2 pages (can be more)



Be **careful** if you try to use LLMs to help you with your abstract!

Using ChatGPT for abstract writing (Spring 2025)



You

Please write a one paragraph abstract for a class project that uses BERT for sentiment analysis for toxic comment classification



ChatGPT

Title: BERT-Enhanced Sentiment Analysis for Toxic Comment Classification: A Comprehensive Investigation

Abstract:

This class project delves into the realm of toxic comment classification using Bidirectional Encoder Representations from Transformers (BERT), an advanced natural language processing model. Toxic comments pose significant challenges in online discourse, necessitating robust sentiment analysis tools for content moderation. The project begins by elucidating the theoretical underpinnings of BERT and its capacity for capturing contextualized word embeddings. We focus on the specific application of toxic comment classification, addressing preprocessing steps, dataset selection, and domain-specific challenges associated with online language. Evaluation metrics such as accuracy, precision, recall, and F1 score are employed to assess the effectiveness of BERT in identifying and categorizing toxic comments. Additionally, the interpretability of the model is explored through attention mechanisms, shedding light on key linguistic indicators of toxicity. The project aims to provide insights into the application of BERT for enhancing content moderation efforts, ultimately contributing to a safer and more inclusive online environment.

Using ChatGPT for abstract writing



You

Please write a one paragraph abstract for a class project that uses BERT for sentiment analysis for toxic comment classification



ChatGPT

Title: BERT-Enhanced Sentiment Analysis for Toxic Comment Classification: A Comprehensive Investigation

Abstract:

This class project delves into the realm of toxic comment classification using Bidirectional Encoder Representations from Transformers (BERT), ~~an advanced natural language processing model~~. Toxic comments pose significant challenges in online discourse, necessitating robust sentiment analysis tools for content moderation. The project begins by elucidating the theoretical underpinnings of BERT and its capacity for capturing contextualized word embeddings. We focus on the specific application of toxic comment classification, addressing preprocessing steps, dataset selection, and domain-specific challenges associated with online language. Evaluation metrics such as accuracy, precision, recall, and F1 score are employed to assess the effectiveness of BERT in identifying and categorizing toxic comments. Additionally, the interpretability of the model is explored through attention mechanisms, shedding light on key linguistic indicators of toxicity. The project aims to provide insights into the application of BERT for enhancing content moderation efforts, ultimately contributing to a safer and more inclusive online environment.

Using ChatGPT for abstract writing



You

Please write a one paragraph abstract for a class project that uses BERT for sentiment analysis for toxic comment classification



ChatGPT

Title: BERT-Enhanced Sentiment Analysis for Toxic Comment Classification: A Comprehensive Investigation

Abstract:

This class project delves into the realm of toxic comment classification using Bidirectional Encoder Representations from Transformers (BERT), ~~an advanced natural language processing model~~. Toxic comments pose significant challenges in online discourse, necessitating robust sentiment analysis tools for content moderation. **The project begins by elucidating the theoretical underpinnings of BERT and its capacity for capturing contextualized word embeddings. We focus on the specific application of toxic comment classification, addressing preprocessing steps, dataset selection, and domain-specific challenges associated with online language.** Evaluation metrics such as accuracy, precision, recall, and F1 score are employed to assess the effectiveness of BERT in identifying and categorizing toxic comments. Additionally, the interpretability of BERT through attention mechanisms, shedding light on key linguistic indicators of toxicity. The project aims to provide insights into the application of BERT for enhancing content moderation efforts, ultimately contributing to a safer and more inclusive online environment.

Vague and not informative

Using ChatGPT for abstract writing



You

Please write a one paragraph abstract for a class project that uses BERT for sentiment analysis for toxic comment classification



ChatGPT

Title: BERT-Enhanced Sentiment Analysis for Toxic Comment Classification: A Comprehensive Investigation

Abstract:

This class project delves into the realm of toxic comment classification using Bidirectional Encoder Representations from Transformers (BERT), ~~an advanced natural language processing model~~. Toxic comments pose significant challenges in online discourse, necessitating robust sentiment analysis tools for content moderation. The project begins by elucidating the theoretical underpinnings of BERT and its capacity for capturing contextualized embeddings. We focus on the specific application of toxic comment classification, addressing preprocessing steps, dataset selection, and domain-specific challenges associated with online language. **Evaluation metrics such as accuracy, precision, recall, and F1 score are employed to assess the effectiveness of BERT in identifying and categorizing toxic comments.** Additionally, the interpretability of the model is explored through attention mechanisms, shedding light on key linguistic indicators of toxicity. The project aims to provide insights into the application of BERT for enhancing content moderation efforts, ultimately contributing to a safer and more inclusive online environment.

Okay

Using ChatGPT for abstract writing



You

Please write a one paragraph abstract for a class project that uses BERT for sentiment analysis for toxic comment classification



ChatGPT

Title: BERT-Enhanced Sentiment Analysis for Toxic Comment Classification: A Comprehensive Investigation

Abstract:

This class project delves into the realm of toxic comment classification using Bidirectional Encoder Representations from Transformers (BERT), ~~an advanced natural language processing model~~. Toxic comments pose significant challenges in online discourse, necessitating robust sentiment analysis tools for content moderation. The project begins by elucidating the theoretical underpinnings of BERT and its capacity for capturing contextualized word embeddings. We focus on the specific application of toxic comment classification. **How? What specifically will be done to explore the interpretability of the model?** Specific challenges associated with this task and F1 score are employed to assess the effectiveness of BERT in identifying and categorizing toxic comments.

Additionally, the interpretability of the model is explored through attention mechanisms, shedding light on key linguistic indicators of toxicity. The project aims to provide insights into the application of BERT for enhancing content moderation efforts, ultimately contributing to a safer and more inclusive online environment.

Using ChatGPT for abstract writing (Spring 2026)



You

Please write a one paragraph abstract for a class project that uses BERT for sentiment analysis for toxic comment classification



ChatGPT

Please give my project a title

Title: Context Matters: Toxic Comment Classification with BERT-Based Sentiment Analysis

Abstract: This class project applies BERT-based sentiment analysis to the task of toxic comment classification, aiming to automatically detect harmful language in online text. We fine-tune a pre-trained BERT model on a labeled dataset of user-generated comments to predict toxicity-related categories and compare its performance against simpler baselines such as logistic regression with TF-IDF features. The model is evaluated using standard metrics including accuracy, precision, recall, and F1-score, with additional attention to class imbalance and false positives that may affect moderation outcomes. Our results demonstrate that fine-tuned BERT captures contextual cues and subtle linguistic patterns more effectively than traditional approaches, leading to improved classification performance. This work highlights the potential of transformer-based language models for supporting scalable and reliable content moderation in real-world settings.

Overall better!

Using ChatGPT for abstract writing (Spring 2026)



You

Please write a one paragraph abstract for a class project that uses BERT for sentiment analysis for toxic comment classification



ChatGPT

Please give my project a title

Title: Context Matters: Toxic Comment Classification with BERT-Based Sentiment Analysis

Abstract: This class project **What specific dataset?** analysis to **What categories?** classification, aiming to automatically detect harmful language in online text. We fine-tune a pre-trained BERT model on a labeled dataset of user-generated comments to predict toxicity-related categories and compare its performance against simpler baselines such as logistic regression with TF-IDF features. The model is evaluated using standard metrics including accuracy, precision, recall, and F1-score, with additional attention to class imbalance and false positives. Our results demonstrate that fine-tuned BERT captures contextual nuances more effectively than traditional approaches, leading to improved classification performance. This highlights the potential of transformer-based language models for supporting scalable and reliable content moderation in real-world settings.

Maybe these baselines are too simple!
Now there are a lot of different BERT-based models.

How about having LLM suggest some topics to me?



You

Suggest a NLP class project to me and write an abstract

Using ChatGPT for project suggestion (Spring 2025)



ChatGPT

Title: Sentiment Analysis on Social Media Texts

Abstract:

In this Natural Language Processing (NLP) class project, we aim to develop a sentiment analysis model for social media texts. Social media platforms have become a ubiquitous source of information, opinion, and expression. Analyzing the sentiments expressed in these texts can provide valuable insights into public opinion, user satisfaction, and emerging trends.

The project will involve the following key steps:

- **Data Collection:** Gather a diverse dataset from popular social media platforms, containing text data with associated sentiment labels (positive, negative, neutral).
- **Preprocessing:** Clean and preprocess the data to handle challenges such as noise, abbreviations, and special characters. Tokenization, stemming, and lemmatization will be applied to prepare the text for analysis.
- **Feature Extraction:** Utilize techniques like Bag-of-Words, TF-IDF, or word embeddings to represent the textual data in a format suitable for machine learning models.
- **Model Development:** Experiment with different machine learning algorithms such as Naive Bayes, Support Vector Machines, or deep learning approaches like Recurrent Neural Networks (RNNs) and Transformers for sentiment classification.
- **Evaluation:** Assess the model's performance using metrics such as accuracy, precision, recall, and F1 score. Explore the model's ability to generalize to new, unseen data.
- **Visualization:** Create visualizations to illustrate sentiment trends over time or across different topics/categories, providing a comprehensive understanding of the data.
- **Deployment:** Develop a user-friendly interface or integration that allows users to input text and receive sentiment analysis results. This could be a web application, a chatbot, or a simple API.

This project offers hands-on experience in dealing with real-world text data, implementing machine learning models, and addressing the challenges of sentiment analysis in the context of social media. It also provides an opportunity to explore ethical considerations related to sentiment analysis and the responsible use of NLP technologies in social media analytics



You

Suggest a NLP class project to me and write an abstract

Title: Sentiment Analysis on Social Media Texts

Abstract:

In this Natural Language Processing (NLP) class project, we aim to develop a sentiment analysis model for social media texts. Social media platforms have become a ubiquitous source of information, opinion, and expression. Analyzing the sentiments expressed in these texts can provide valuable insights into public opinion, user satisfaction, and emerging trends.

The project will involve t

Too much work: find existing dataset

- **Data Collection:** Gather a diverse dataset from popular social media platforms, containing text data with associated sentiment labels (positive, negative, neutral).
- **Preprocessing:** Clean and preprocess the data to handle challenges such as noise, abbreviations, and special characters. Tokenization, stemming, and lemmatization will be applied to prepare the text for analysis.
- **Feature Extraction:** Utilize techniques like Bag-of-Words, TF-IDF, or word embeddings to represent the textual data in a format suitable for machine learning models.
- **Model Development:** Experiment with different machine learning algorithms such as Naive Bayes, Support Vector Machines, or deep learning approaches like Recurrent Neural Networks (RNNs) and Transformers for sentiment classification.
- **Evaluation:** Assess the model's performance using metrics such as accuracy, precision, recall, and F1 score. Explore the model's ability to generalize to new, unseen data.
- **Visualization:** Create visualizations to illustrate sentiment trends over time or across different topics/categories, providing a comprehensive understanding of the data.
- **Deployment:** Develop a user-friendly interface or integration that allows users to input text and receive sentiment analysis results. This could be a web application, a chatbot, or a simple API.

This project offers hands-on experience in dealing with real-world text data, implementing machine learning models, and addressing the challenges of sentiment analysis in the context of social media. It also provides an opportunity to explore ethical considerations related to sentiment analysis and the responsible use of NLP technologies in social media analytics



You

Suggest a NLP class project to me and write an abstract

Title: Sentiment Analysis on Social Media Texts

Abstract:

In this Natural Language Processing (NLP) class project, we aim to develop a sentiment analysis model for social media texts. Social media platforms have become a ubiquitous source of information, opinion, and expression. Analyzing the sentiments expressed in these texts can provide valuable insights into public opinion, user satisfaction, and emerging trends.

The project will involve the following steps:

Many of these steps are not necessary if you use existing datasets and modern deep learning techniques. You will need to tokenize.

- **Data Collection:** Gather social media texts from various platforms, ensuring a diverse range of sentiment (positive, negative, neutral).
- **Preprocessing:** Clean and preprocess the data to handle challenges such as noise, abbreviations, and special characters. **Tokenization**, stemming, and lemmatization will be applied to prepare the text for analysis.
- **Feature Extraction:** Utilize techniques like Bag-of-Words, TF-IDF, or word embeddings to represent the textual data in a format suitable for machine learning models.
- **Model Development:** Experiment with different machine learning algorithms such as Naive Bayes, Support Vector Machines, or deep learning approaches like Recurrent Neural Networks (RNNs) and Transformers for sentiment classification.
- **Evaluation:** Assess the model's performance using metrics such as accuracy, precision, recall, and F1 score. Explore the model's ability to generalize to new, unseen data.
- **Visualization:** Create visualizations to illustrate sentiment trends over time or across different topics/categories, providing a comprehensive understanding of the data.
- **Deployment:** Develop a user-friendly interface or integration that allows users to input text and receive sentiment analysis results. This could be a web application, a chatbot, or a simple API.

This project offers hands-on experience in dealing with real-world text data, implementing machine learning models, and addressing the challenges of sentiment analysis in the context of social media. It also provides an opportunity to explore ethical considerations related to sentiment analysis and the responsible use of NLP technologies in social media analytics.



You

Suggest a NLP class project to me and write an abstract

Title: Sentiment Analysis on Social Media Texts

Abstract:

In this Natural Language Processing (NLP) class project, we aim to develop a sentiment analysis model for social media texts. Social media platforms have become a ubiquitous source of information, opinion, and expression. Analyzing the sentiments expressed in these texts can provide valuable insights into public opinion, user satisfaction, and emerging trends.

The project will involve the following key steps:

- **Data Collection:** Gather a dataset of social media texts, ensuring a mix of positive, negative, and neutral sentiments.
- **Preprocessing:** Clean the data by removing noise, handling missing values, and applying tokenization, stemming, and lemmatization to prepare the text for analysis.
- **Feature Extraction:** Utilize techniques like Bag-of-Words, TF-IDF, or **word embeddings** to represent the textual data in a format suitable for machine learning models.
- **Model Development:** Experiment with different machine learning algorithms such as Naive Bayes, Support Vector Machines, or deep learning approaches like Recurrent Neural Networks (RNNs) and Transformers for sentiment classification.
- **Evaluation:** Assess the model's performance using metrics such as accuracy, precision, recall, and F1 score. Explore the model's ability to generalize to new, unseen data.
- **Visualization:** Create visualizations to illustrate sentiment trends over time or across different topics/categories, providing a comprehensive understanding of the data.
- **Deployment:** Develop a user-friendly interface or integration that allows users to input text and receive sentiment analysis results. This could be a web application, a chatbot, or a simple API.

This project offers hands-on experience in dealing with real-world text data, implementing machine learning models, and addressing the challenges of sentiment analysis in the context of social media. It also provides an opportunity to explore ethical considerations related to sentiment analysis and the responsible use of NLP technologies in social media analytics.



You

Suggest a NLP class project to me and write an abstract

Title: Sentiment Analysis on Social Media Texts

Abstract:

In this Natural Language Processing (NLP) class project, we aim to develop a sentiment analysis model for social media texts. Social media platforms have become a ubiquitous source of information, opinion, and expression. Analyzing the sentiments expressed in these texts can provide valuable insights into public opinion, user satisfaction, and emerging trends.

The project will involve the following key steps:

- **Data Collection:** Gather a diverse dataset from popular social media platforms, containing text data with associated sentiment labels (positive, negative, neutral).
- **Preprocessing:** Clean the data by removing noise, handling missing values, and performing tokenization, stemming, and lemmatization to prepare the text for machine learning.
- **Feature Extraction:** Extract relevant features from the preprocessed text, such as word embeddings, n-grams, and syntactic structures, making them suitable for machine learning.
- **Model Development:** Experiment with different machine learning algorithms such as Naive Bayes, Support Vector Machines, or **deep learning approaches like Recurrent Neural Networks (RNNs) and Transformers** for sentiment classification.
- **Evaluation:** Assess the model's performance using metrics such as accuracy, precision, recall, and F1 score. Explore the model's ability to generalize to new, unseen data.
- **Visualization:** Create visualizations to illustrate sentiment trends over time or across different topics/categories, providing a comprehensive understanding of the data.
- **Deployment:** Develop a user-friendly interface or integration that allows users to input text and receive sentiment analysis results. This could be a web application, a chatbot, or a simple API.

This project offers hands-on experience in dealing with real-world text data, implementing machine learning models, and addressing the challenges of sentiment analysis in the context of social media. It also provides an opportunity to explore ethical considerations related to sentiment analysis and the responsible use of NLP technologies in social media analytics.



You

Suggest a NLP class project to me and write an abstract

Title: Sentiment Analysis on Social Media Texts

Abstract:

In this Natural Language Processing (NLP) class project, we aim to develop a sentiment analysis model for social media texts. Social media platforms have become a ubiquitous source of information, opinion, and expression. Analyzing the sentiments expressed in these texts can provide valuable insights into public opinion, user satisfaction, and emerging trends.

The project will involve the following key steps:

- **Data Collection:** Gather a diverse dataset from popular social media platforms, containing text data with associated sentiment labels (positive, negative, neutral).
- **Preprocessing:** Clean and preprocess the data to handle challenges such as noise, abbreviations, and special characters. Tokenization, stemming, and lemmatization will be applied to prepare the text for analysis.
- **Feature Extraction:** Utilize techniques like Bag-of-Words, TF-IDF, or word embeddings to represent the textual data in a format suitable for machine learning.
- **Modeling:** This part is important! Pay attention to evaluation. Explore deep learning approaches like Recurrent Neural Networks (RNNs) and Transformers for sentiment classification.
- **Evaluation:** Assess the **model's performance using metrics** such as accuracy, precision, recall, and F1 score. Explore the model's ability to generalize to **new, unseen data**.
- **Visualization:** Create visualizations to illustrate sentiment trends over time or across different topics/categories, providing a comprehensive understanding of the data.
- **Deployment:** Develop a user-friendly interface or integration that allows users to input text and receive sentiment analysis results. This could be a web application, a chatbot, or a simple API.

This project offers hands-on experience in dealing with real-world text data, implementing machine learning models, and addressing the challenges of sentiment analysis in the context of social media. It also provides an opportunity to explore ethical considerations related to sentiment analysis and the responsible use of NLP technologies in social media analytics.



You

Suggest a NLP class project to me and write an abstract

Title: Sentiment Analysis on Social Media Texts

Abstract:

In this Natural Language Processing (NLP) class project, we aim to develop a sentiment analysis model for social media texts. Social media platforms have become a ubiquitous source of information, opinion, and expression. Analyzing the sentiments expressed in these texts can provide valuable insights into public opinion, user satisfaction, and emerging trends.

The project will involve the following key steps:

- **Data Collection:** Gather a diverse dataset from popular social media platforms, containing text data with associated sentiment labels (positive, negative, neutral).
- **Preprocessing:** Clean and preprocess the data to handle challenges such as noise, abbreviations, and special characters. Tokenization, stemming, and lemmatization will be applied to prepare the text for analysis.
- **Feature Extraction:** Utilize techniques like Bag-of-Words, TF-IDF, or word embeddings to represent the textual data in a format suitable for machine learning.
- **Model Development:** Implement and train machine learning models using the extracted features. This step involves deep learning.
- **Evaluation:** Assess the model's performance using appropriate metrics. This includes a breakdown of performance and the ability to generalize to new, unseen data.

Do try to provide qualitative analysis (give examples) and
breakdown of performance

- **Visualization:** Create visualizations to illustrate sentiment trends over time or across different topics/categories, providing a comprehensive understanding of the data.

- **Deployment:** Develop a user-friendly interface or integration that allows users to input text and receive sentiment analysis results. This could be a web application, a chatbot, or a simple API.

This project offers hands-on experience in dealing with real-world text data, implementing machine learning models, and addressing the challenges of sentiment analysis in the context of social media. It also provides an opportunity to explore ethical considerations related to sentiment analysis and the responsible use of NLP technologies in social media analytics.



You

Suggest a NLP class project to me and write an abstract

Title: Sentiment Analysis on Social Media Texts

Abstract:

In this Natural Language Processing (NLP) class project, we aim to develop a sentiment analysis model for social media texts. Social media platforms have become a ubiquitous source of information, opinion, and expression. Analyzing the sentiments expressed in these texts can provide valuable insights into public opinion, user satisfaction, and emerging trends.

The project will involve the following key steps:

- **Data Collection:** Gather a diverse dataset from popular social media platforms, containing text data with associated sentiment labels (positive, negative, neutral).
- **Preprocessing:** Clean and preprocess the data to handle challenges such as noise, abbreviations, and special characters. Tokenization, stemming, and lemmatization will be applied to prepare the text for analysis.
- **Feature Extraction:** Utilize techniques like Bag-of-Words, TF-IDF, or word embeddings to represent the textual data in a format suitable for machine learning models.
- **Model Development:** Experiment with different machine learning algorithms such as Naive Bayes, Support Vector Machines, or deep learning approaches like Recurrent Neural Networks (RNNs) and Transformers for sentiment classification.
- **Evaluation:** Assess the model's performance using metrics such as accuracy, precision, recall, and F1 score. Explore the model's ability to generalize to new data.
- **Visualization:** Create visualizations to aid in the comprehensive understanding of the data.
- **Deployment:** Develop a user-friendly interface or integration that allows users to input text and receive sentiment analysis results. This could be a web application, a chatbot, or a simple API.

Deployment is not necessary for the class project

comprehensive

This project offers hands-on experience in dealing with real-world text data, implementing machine learning models, and addressing the challenges of sentiment analysis in the context of social media. It also provides an opportunity to explore ethical considerations related to sentiment analysis and the responsible use of NLP technologies in social media analytics.



You

Suggest a NLP class project to me and write an abstract



ChatGPT

Here's a solid NLP project idea that's *interesting, doable, and publishable-ish* without requiring a giant lab-scale dataset.

Project idea: Faithful Meeting Summarization with Evidence Linking

Core problem

Most meeting summarizers produce fluent summaries, but they often:

- invent decisions (“hallucinate”)
- blur who said what
- miss action items
- fail to cite where information came from

Your project

Build a meeting summarization system that outputs:

1. a structured summary (decisions, action items, open questions, key points)
2. **evidence links**: each summary bullet is linked to the exact supporting utterances in the transcript

Why it's a good NLP project

- Very relevant to real-world use (Zoom/Teams notes)
- Evaluatable beyond ROUGE (faithfulness + attribution)
- Works nicely with modern LLMs + retrieval + constraints
- You can make a strong “systems + evaluation” paper without inventing a new model

Dataset options

- AMI Meeting Corpus
- ICSI Meeting Corpus
- QMSum (query-based, but adaptable)

Evaluation ideas

- Evidence precision/recall: does the cited span actually support the claim?
- Hallucination rate: unsupported summary statements
- Action item extraction F1
- Human evaluation: usefulness + trustworthiness

Using ChatGPT for project suggestion (Spring 2026)

Better but...

Is it interesting to you?

Is it doable by you?

Did it suggest the same idea
to others?

Double check suggested
datasets and evaluation metrics

In the end, you need to
understand and execute
your project ideas!



You

Suggest a NLP class project to me and write an abstract



ChatGPT

Using ChatGPT for project suggestion (Spring 2026)

Abstract (you can use as a paper-style abstract)

Abstract —

Meeting summarization is a practical application of natural language processing, but existing systems frequently produce summaries that are difficult to trust due to hallucinated content, missing attribution, and incomplete capture of action items. We propose a meeting summarization framework that generates structured summaries consisting of decisions, action items, unresolved questions, and key discussion points, while explicitly linking each summary statement to supporting evidence spans in the transcript. Our approach combines segment-level retrieval with constrained generation to encourage factual grounding and speaker-aware attribution. We introduce an automatic evaluation protocol for faithfulness based on evidence precision and coverage, alongside standard summarization metrics and human judgments of usefulness. Experiments on publicly available meeting corpora demonstrate that evidence-linked summaries reduce unsupported statements and improve user trust compared to baseline summarizers, while maintaining competitive informativeness. This work highlights the importance of grounding and verifiability in summarization systems intended for real-world deployment.

If you want, I can also:

- turn this into a full project plan (milestones + experiments)
- suggest 2–3 alternative projects in different directions (NLP+vision, NLU+3D grounding, etc.)



You

Suggest a NLP class project to me and write an abstract



ChatGPT

Using ChatGPT for project suggestion (Spring 2026)

Project idea: “Fact-or-Flavor”: Faithful Summarization with Grounded Citations for Web & PDF Sources

One-line pitch: Build a summarization system that produces *useful* summaries **only from evidence it can quote**, and attaches sentence-level citations (like a mini literature review, but automated).

Why this is a good NLP project

Most summarizers are fluent but unreliable: they’ll invent details, blur sources, or mix facts across documents. This project tackles a real pain point in research, enterprise, and education: **trustworthy summaries**.

It’s also nicely scoped: you can start with a simple baseline and keep upgrading.

What you would build

A system that takes:

- a set of documents (web pages, PDFs, or papers)
- a query or task prompt (e.g., “Summarize the key findings about X”)

And outputs:

- a structured summary (bullets + short paragraph)
- **every sentence backed by citations** (document + span)
- a “coverage vs hallucination” score

Core NLP components (very doable)

- Document chunking + retrieval (BM25 or dense retrieval)
- Evidence selection
- Citation-aware generation
- Automatic evaluation of faithfulness (e.g., entailment / attribution metrics)



You

Suggest a NLP class project to me and write an abstract

Using ChatGPT for project suggestion (Spring 2026)



ChatGPT

Abstract (ready to submit)

Title: *Fact-or-Flavor: Citation-Grounded Summarization for Faithful Multi-Document NLP*

Abstract:

Neural summarization systems often produce fluent outputs that contain hallucinated or weakly supported claims, limiting their usefulness in high-stakes settings such as scientific review, policy analysis, and enterprise reporting. We present *Fact-or-Flavor*, a citation-grounded summarization framework that generates multi-document summaries constrained to evidence explicitly retrieved from source texts. Given a query and a collection of documents, the system performs retrieval and evidence selection over semantically coherent text chunks, then produces a structured summary where each sentence is accompanied by one or more citations pointing to the supporting source spans. To evaluate factual consistency, we measure attribution precision using span overlap and entailment-based scoring, and report a coverage metric quantifying how well the summary reflects the document set. Experiments on web and PDF corpora demonstrate that grounding generation in retrieved evidence substantially reduces unsupported claims while maintaining informativeness and readability. This work provides a practical step toward trustworthy summarization systems that are both useful and verifiable.

If you want, I can also:

- propose a 4-week project plan,
- suggest datasets + evaluation metrics,
- or tailor this to a specific domain (scientific papers, news, legal, medical, etc.).

Last year observations on project abstract

- Many of the abstract seems to be generated using LLMs
 - Good: Fluent English, with nice keywords
 - Bad: vague, over-scoped projects
- What you DON'T need to include / do:
 - Prerequisites - don't list the prerequisites of the class as prerequisites for your projects
 - Address everything - try to scope your project down
 - Collect datasets - try to use existing datasets

Examples of what not to have:

- “*we will be using Natural Language Processing*” to ...
- This entire class is on NLP - at this point, you should be able to be more specific about what model you plan to use

Examples of what not to have:

- *“The prerequisites for this project are knowledge of mathematics and statistics, machine learning fundamentals, deep learning, NLP, programming and frameworks. Most of the topics are something that at least one of the team members knows about or can easily refresh their knowledge.”*
- These are very broad prerequisites. Also, you are in this class, you should know this. It shouldn't be something that needs to be specified.

Examples of what not to have:

- Try to avoid dataset collection and data cleaning
- Be careful:
 - *“our healthcare data from WHO and FDA, stock data from Yahoo finance, and market indicators data from S&P500 and BioTech index”*
 - *“decent dataset made up of diverse texts that may include articles, books, and some online content. After that, the data should be cleaned to ensure better quality”*
 - How long will this take? Try to find existing datasets.
- Better: *“We will use a Kaggle dataset”*
 - Be precise: which dataset?
 - Also some Kaggle problems are not interesting...

Many possible NLP tasks and problems!

Some NLP Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis
- POS: Part-of-speech tagging
- Entity linking
- Constituency Parsing
- Dependency Parsing
- Intent detection and slot filling
- Machine translation
- Paraphrasing
- Summarization
- Semantic Parsing (text to code)
- Question Answering
- Multimodal tasks (e.g. image captioning)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

Stanford CS224 default project

– SQuAD

– RobustQA

Passage

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called “showers”.

Question / Answer

What causes precipitation to fall?

gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

graupel

Where do water droplets collide with ice crystals to form precipitation?

within a cloud

SQuAD: 100,000+ Questions for Machine Comprehension of Text

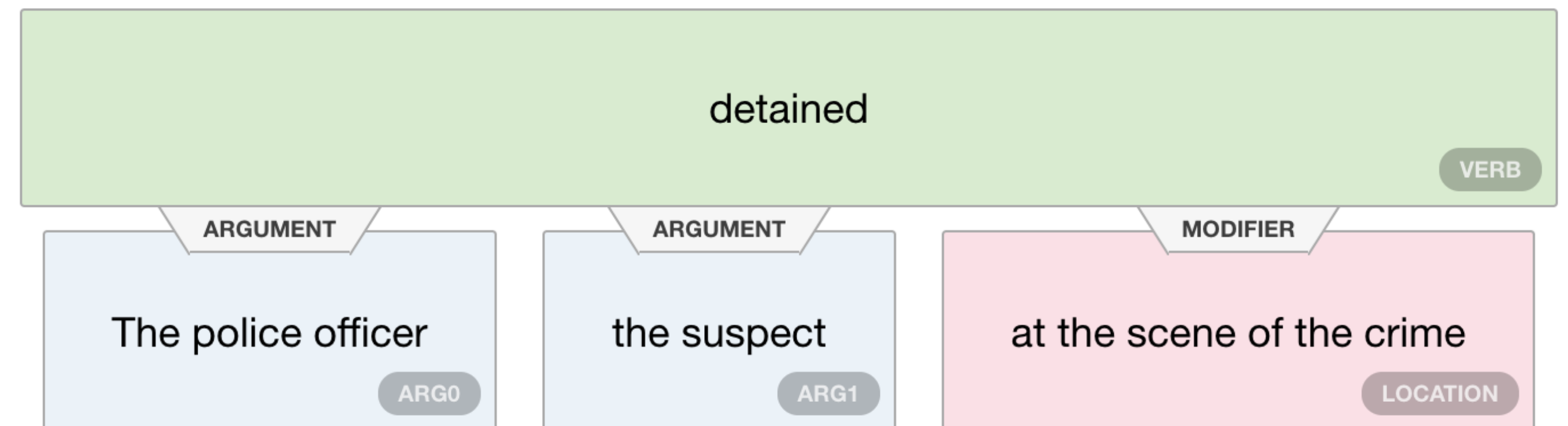
Rajpurkar et al, EMNLP 2016

Tasks	Premise	Hypothesis	contradiction entailment neutral
- SQuAD: question answering SNLI: natural language inference - SRL: semantic role labeling - Coref: coreference resolution - NER: named entity recognition - SST-5: sentiment analysis	A man inspects the uniform of a figure in some East Asian country. The man is sleeping. An older and younger man smiling. Two men are smiling and laughing at the cats playing on the floor. A soccer game with multiple males playing. Some men are playing a sport.		

Tasks

- SQuAD: question answering
 - SNLI: natural language inference
 - SRL: semantic role labeling
 - Coref: coreference resolution
 - NER: named entity recognition
 - SST-5: sentiment analysis
-
- Different types of frames
 - FrameNet vs PropBank
 - Different semantics roles depending on the verb

The police officer detained the suspect at the scene of the crime

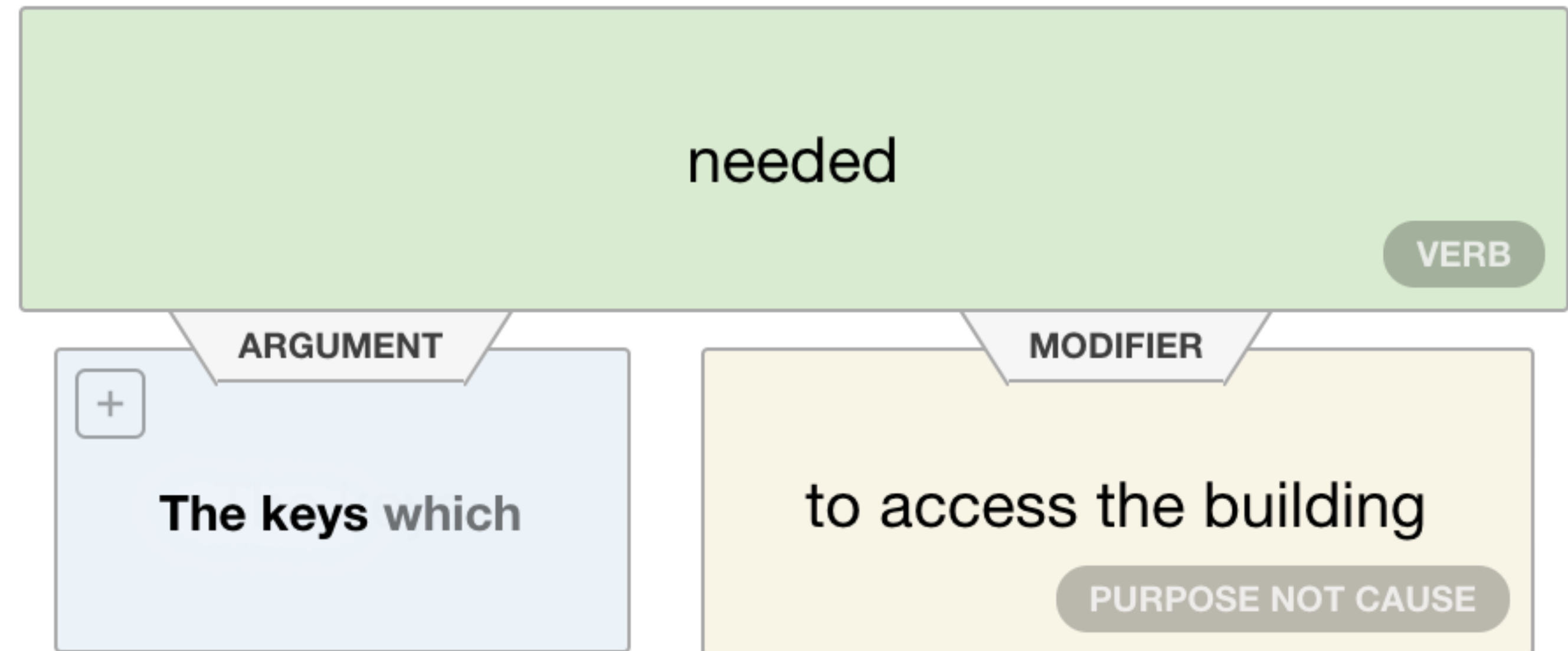


(figure credit: <https://demo.allennlp.org/semantic-role-labeling/>)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

The **keys**, which were **needed** to access the building, were locked in the car.



(figure credit: <https://demo.allennlp.org/semantic-role-labeling/>)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- **SRL: semantic role labeling**
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

The keys, which were needed to **access** the building, were locked in the car.

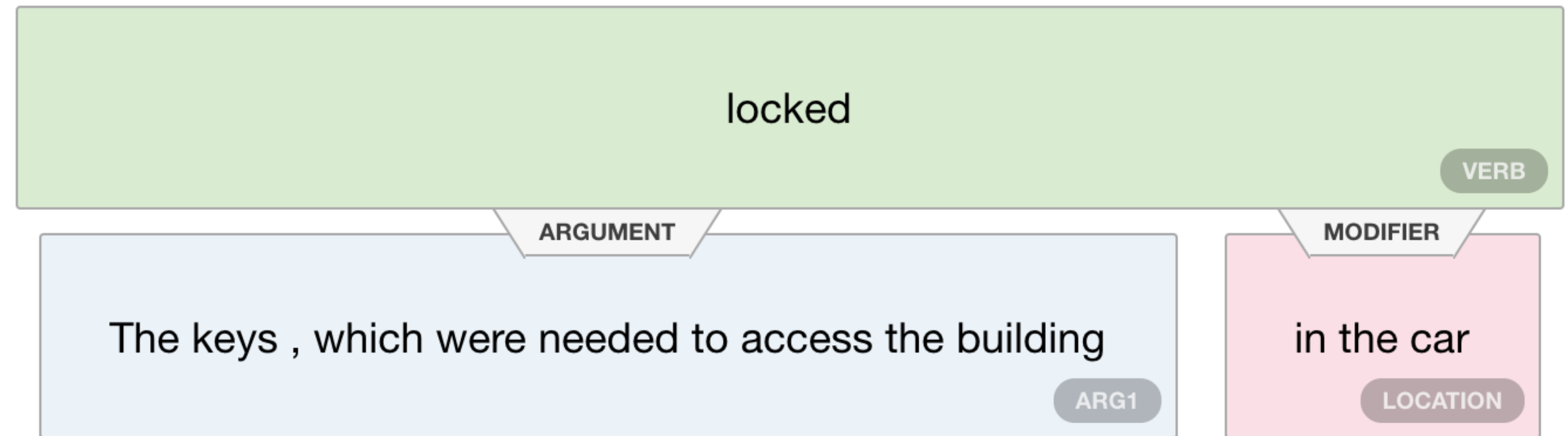


(figure credit: <https://demo.allennlp.org/semantic-role-labeling/>)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- **SRL: semantic role labeling**
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

The keys, which were needed to access the building, were **locked** in the car.



(figure credit: <https://demo.allennlp.org/semantic-role-labeling/>)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

“I voted for Nader because he was most aligned with my values,” she said.

The diagram shows three curved arrows representing coreference relations: one from 'I' to 'Nader', one from 'he' to 'Nader', and one from 'my' to 'she'.

(figure credit: <https://nlp.stanford.edu/projects/coref.shtml>)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- **NER: named entity recognition**
- SST-5: sentiment analysis

Michael Jordan is a professor at Berkeley .

PER LOC

This shirt was bought at Grandpa Joe 's in downtown Deep Learning .

ORG LOC

Did Uriah honestly think he could beat The Legend of Zelda in under three hours ?

PER MISC

(figure credit: <https://demo.allennlp.org/named-entity-recognition>)

A large annotated corpus for learning natural language inference

Sang and Meudler, CoNLL Shared Task 2002

Tasks

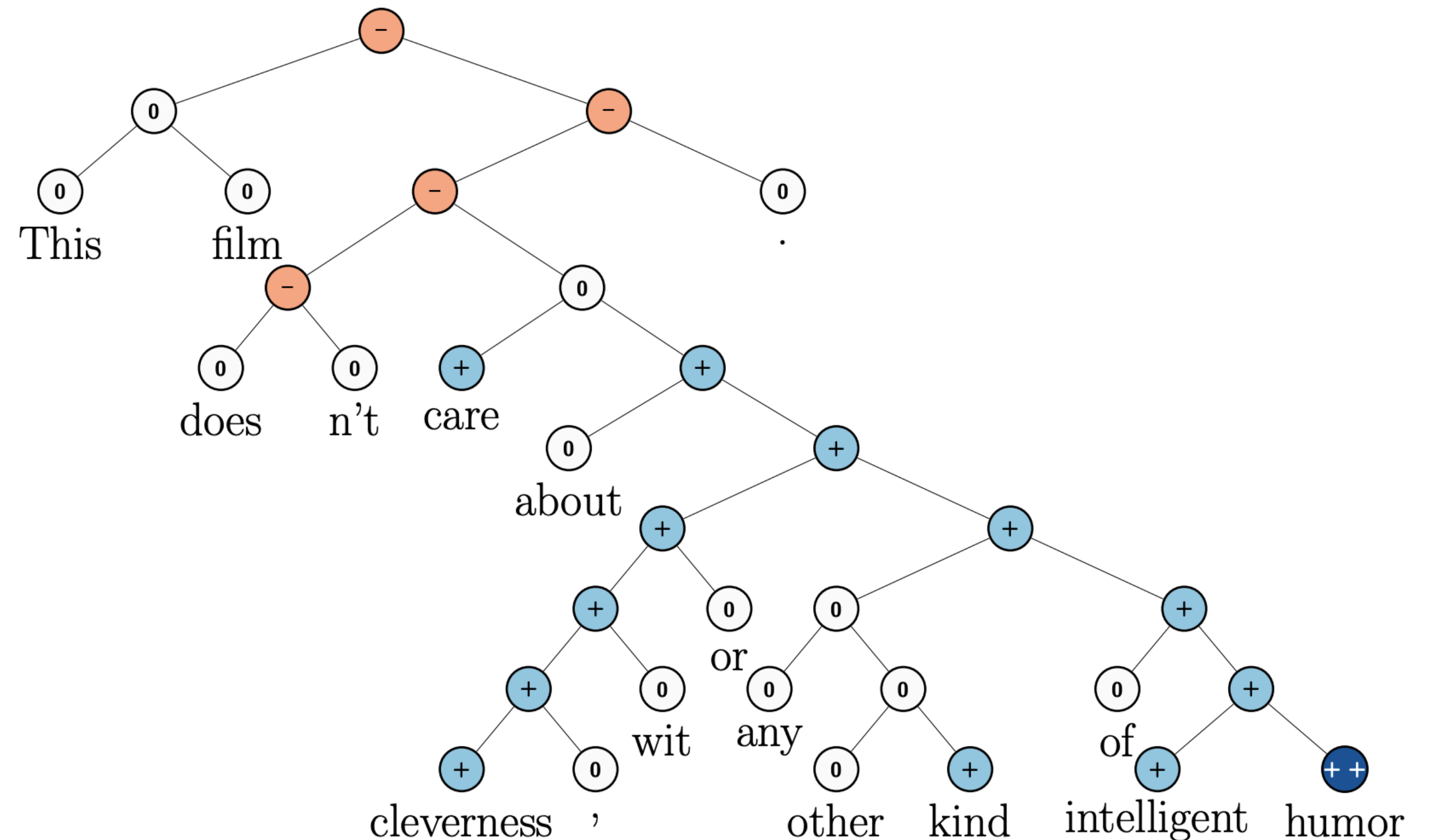
- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

5 classes

- Very Positive
- Positive
- Neutral
- Negative
- Very Negative

Stanford Sentiment Treebank

This file doesn't care about cleverness, wit or any other kind of intelligent humor.



Recursive deep models for semantic compositionality over a sentiment treebank

Socher et al, EMNLP 2013

Many NLP applications!

- Question answering
- Text generation
- Information extraction
- Dialogue and chatbots
- ...

Adapted from slides from Danqi Chen and Karthik Narasimhan
(with some content from slides from Chris Manning and Anoop Sarkar)

Many NLP applications!

- **Question answering**
- Text generation
- Information extraction
- Dialogue and chatbots
- ...

Adapted from slides from Danqi Chen and Karthik Narasimhan
(with some content from slides from Chris Manning and Anoop Sarkar)

QA Taxonomy

- Context (and available information sources)
 - A passage, a document, a large collection of documents, all web documents
 - Knowledge base
 - Semi-structured tables
 - Images
- Question type
 - Factoid vs non-factoid
 - Open-domain vs closed-domain
 - Simple vs compositional
- Answer type
 - A short span of text
 - A paragraph
 - Yes/No
 - A database entry
 - A list

Textual Question Answering

Also called “Reading Comprehension”

The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, *Il milione* (or, *The Million*, known in English as the *Travels of Marco Polo*), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge **through contact with Persian traders** since many of the places he named were in Persian.

How did some suspect that Polo learned about China instead of by actually visiting it?

Answer: **through contact with Persian traders**

Textual Question Answering

James the Turtle was always getting in trouble.

Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.

After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

1) What is the name of the trouble making turtle?

- A) Fries
- B) Pudding
- C) James
- D) Jane

2) What did James pull off of the shelves in the grocery store?

- A) pudding
- B) fries
- C) food
- D) splinters

Conversational Question Answering

The Virginia governor's race, billed as the marquee battle of an otherwise anticlimactic 2013 election cycle, is shaping up to be a foregone conclusion. Democrat Terry McAuliffe, the longtime political fixer and moneyman, hasn't trailed in a poll since May. Barring a political miracle, Republican Ken Cuccinelli will be delivering a concession speech on Tuesday evening in Richmond. In recent ...

Q: What are the candidates **running** for?

A: Governor

Q: **Where?**

A: Virginia

Q: Who is the democratic candidate?

A: Terry McAuliffe

Q: Who is **his** opponent?

A: Ken Cuccinelli

Q: What party does **he** belong to?

A: Republican

Q: Which of **them** is winning?

Long-form Question Answering

? Question | How do jellyfish function without a brain and a nervous system?

✓ Abstractive

Jellyfish don't have brains. Jellyfish have a nervous system. They know how to swim, they don't even know what to do with their bodies. They have a "nervous system", which is basically a bunch of nerve cells that connect to the body. The nervous system is made up of neurons, neurons, nerves, etc. Each neuron doesn't have a specific set of "neurons", it just has a different set of nerves. It has a set of sensory neurons, which are all connected to each other. It just knows that it does react to the stimulus, and it knows that the stimulus is there. They have a bunch of nervous systems that act like a filter to get information back.

✓ Extractive

But none of these hypotheses make much sense for a jellyfish because they don't have brains at all. They just have a nerve net — a loose ring of neurons that runs around the rim of their pulsating bells. They have an unusual nervous system, 451 because jellyfish are not bilaterally symmetrical — that is, they don't have a left side and a right side. Jellyfish don't have brains, but their nervous systems detect smells, light and other stimuli, and they coordinate their physical responses.

Abstractive:
Answer made up of
novel words and
sentences composed
through paraphrasing

Extractive:
Select excerpts
(extracts) and
concatenate them
to form the answer.

Open-domain Question Answering

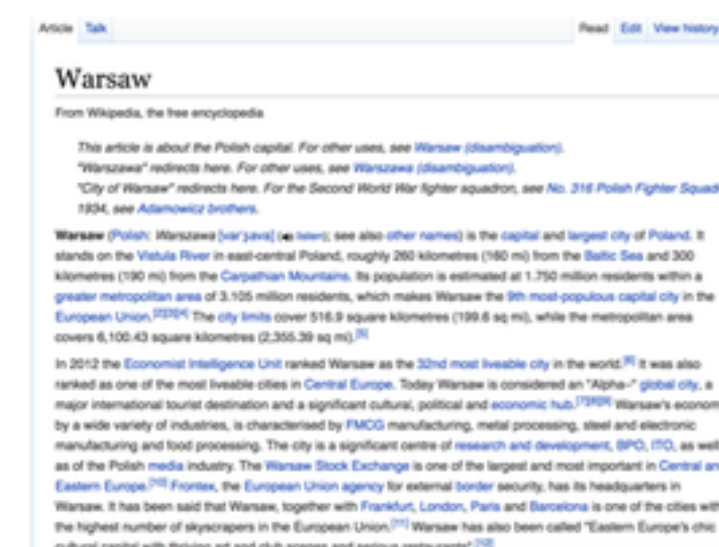
DrQA

- Factored into two parts:
 - Find documents that might contain an answer (handled with traditional **information retrieval**)
 - Finding an answer in a paragraph or a document (**reading comprehension**)

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

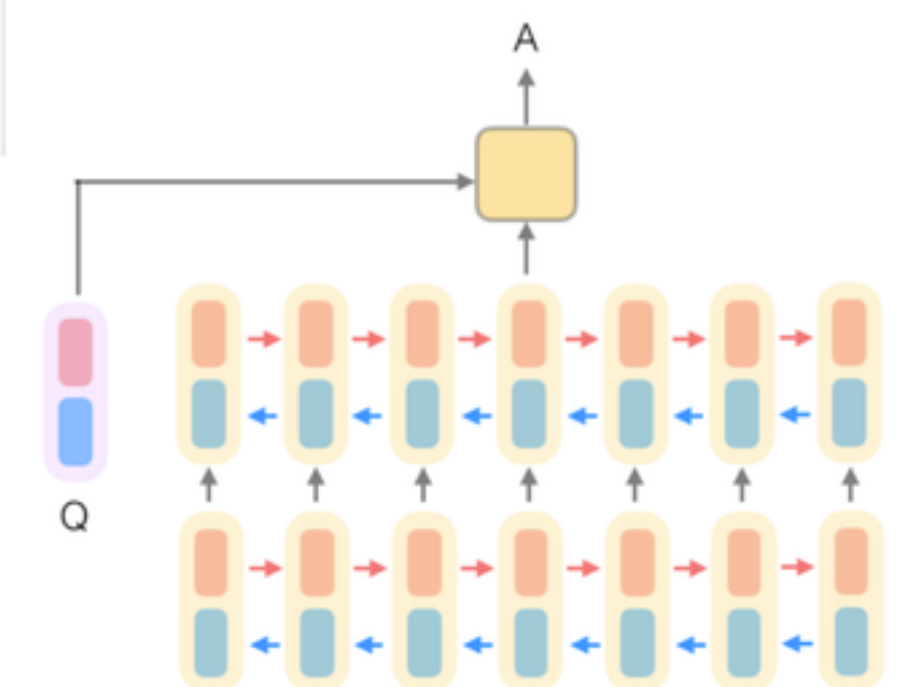


Document Retriever



Document Reader

833,500



```
>>> process('What is the answer to life, the universe, and everything?')
```

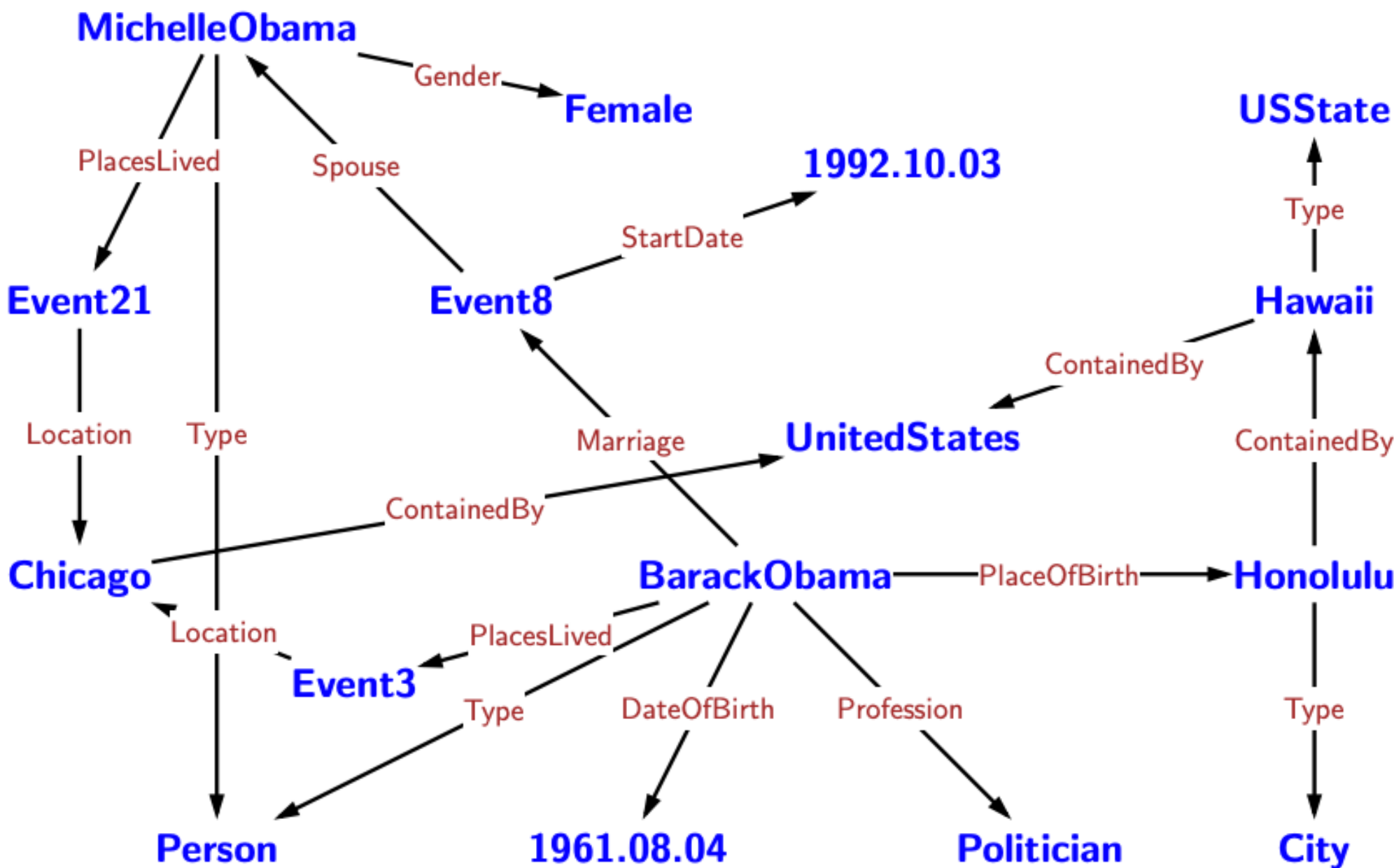
Top Predictions:

Rank	Answer	Doc	Answer Score	Doc Score
1	42	Phrases from The Hitchhiker's Guide to the Galaxy	47242	141.26

Knowledge Base Question Answering

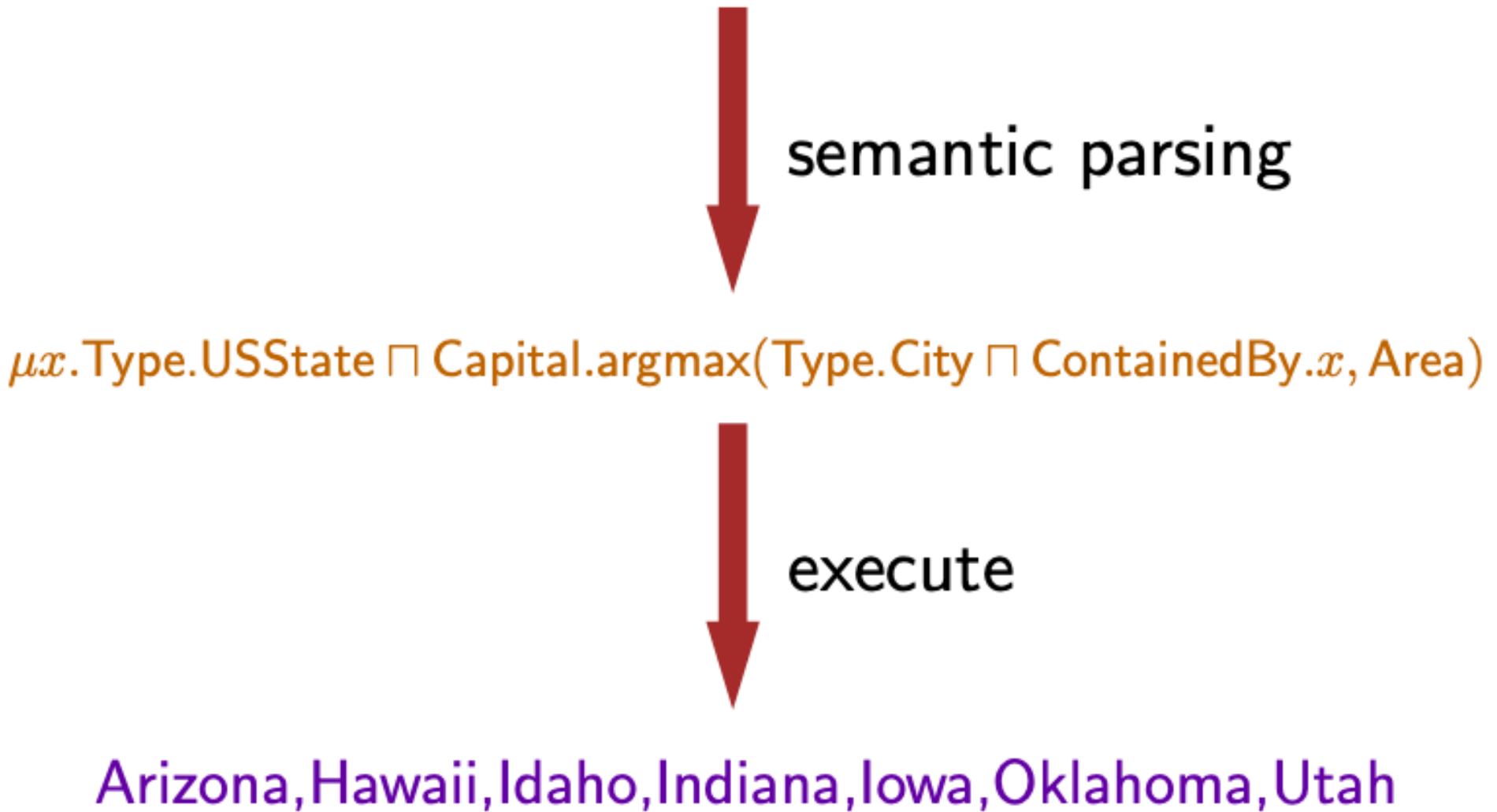


100M entities (nodes) 1B assertions (edges)



Structured knowledge representation

Which states' capitals are also their largest cities by area?



QA via semantic parsing

Table-based Question Answering

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

x = Greece held its last
Summer Olympics in
which year?

y = 2004

Visual Question Answering



What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?

Why do we care about this problem?

- Useful for many practical applications
- Reading comprehension is an important testbed for evaluating how well computer systems understand human language
 - Wendy Lehnert 1977: “Since questions can be devised to query any aspect of text comprehension, the ability to answer questions is the strongest possible demonstration of understanding.”
- Many other NLP tasks can be reduced to a reading comprehension problem:

Information extraction

(Barack Obama, educated_at, ?)

Question: Where did Barack Obama graduate from?

Passage: Obama was born in Honolulu, Hawaii. After graduating from Columbia University in 1983, he worked as a community organizer in Chicago.

(Levy et al. 2017)

Semantic role labeling

UCD **finished** the 2006 championship as Dublin champions ,
by **beating** St Vincents in the final .

finished

Who finished something? - UCD
What did someone finish? - the 2006 championship
What did someone finish something as? - Dublin champions
How did someone finish something? - by beating St Vincents in the final

beating

Who beat someone? - UCD
When did someone beat someone? - in the final
Who did someone beat? - St Vincents

(He et al. 2015)

Stanford Question Answering Dataset (SQuAD)

Passage

Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi's Stadium in the San Francisco Bay Area at Santa Clara, California.

Question: Which NFL team won Super Bowl 50?

Answer: Denver Broncos

Question: What does AFC stand for?

Answer: American Football Conference

Question: What year was Super Bowl 50?

Answer: 2016

SQuAD 2.0:

Have classifier/threshold to decide whether to take the most likely prediction as answer

- (passage, question, answer) triples
- Passage is from Wikipedia (~100-500 words), question is crowd-sourced
- Answer must be a span of text in the passage (aka. “**extractive question answering**”)
- SQuAD 1.1: 100k answerable questions, SQuAD 2.0: another 50k **unanswerable** questions

Stanford Question Answering Dataset (SQuAD)

Private schools, also known as independent schools, non-governmental, or nonstate schools, are not administered by local, state or national governments; thus, they retain the right to select their students and are funded in whole or in part by charging their students tuition, rather than relying on mandatory taxation through public (government) funding; at some private schools students may be able to get a scholarship, which makes the cost cheaper, depending on a talent the student may have (e.g. sport scholarship, art scholarship, academic scholarship), financial need, or tax credit scholarships that might be available.

3 gold answers are collected for each question

Along with non-governmental and nonstate schools, what is another name for private schools?

Gold answers: ① independent ② independent schools ③ independent schools

Along with sport and art, what is a type of talent scholarship?

Gold answers: ① academic ② academic ③ academic

Rather than taxation, what are private schools largely funded by?

Gold answers: ① tuition ② charging their students tuition ③ tuition

Stanford Question Answering Dataset (SQuAD)

SQuAD 1.1 evaluation:

- Two metrics: exact match (EM) and F1
 - Exact match: 1/0 accuracy on whether you match one of the three answers
 - F1: take each gold answer and system output as bag of words, compute precision, recall and harmonic mean. Take the max of the three scores.
- Final exact match and F1 are average of instance exact and F1 scores
- Estimated human performance: EM = 82.3, F1 = 91.2

Example

Q: What did Tesla do in December 1878?

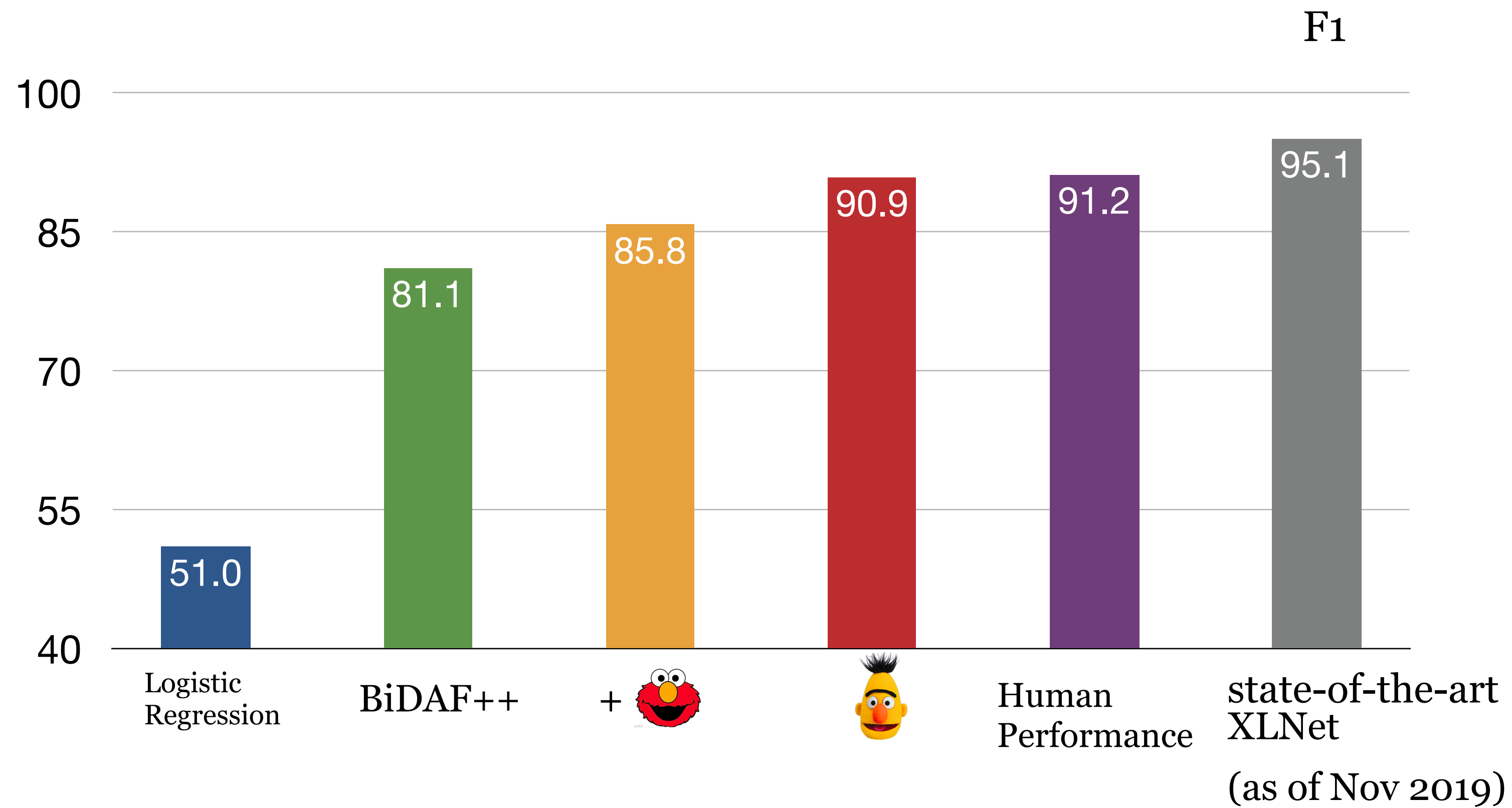
A: {left Graz, left Graz, left Graz and severed all relations with his family}

Prediction: {left Graz and served}

Exact match: $\max\{0, 0, 0\} = 0$

F1: $\max\{0.67, 0.67, 0.61\} = 0.67$

Experiments on SQuAD v1.1



*: single model only

Many NLP applications!

- Question answering
- **Text generation**
- Information extraction
- Dialogue and chatbots
- ...

Adapted from slides from Danqi Chen and Karthik Narasimhan
(with some content from slides from Chris Manning and Anoop Sarkar)

Many tasks and applications for natural language generation (NLG)

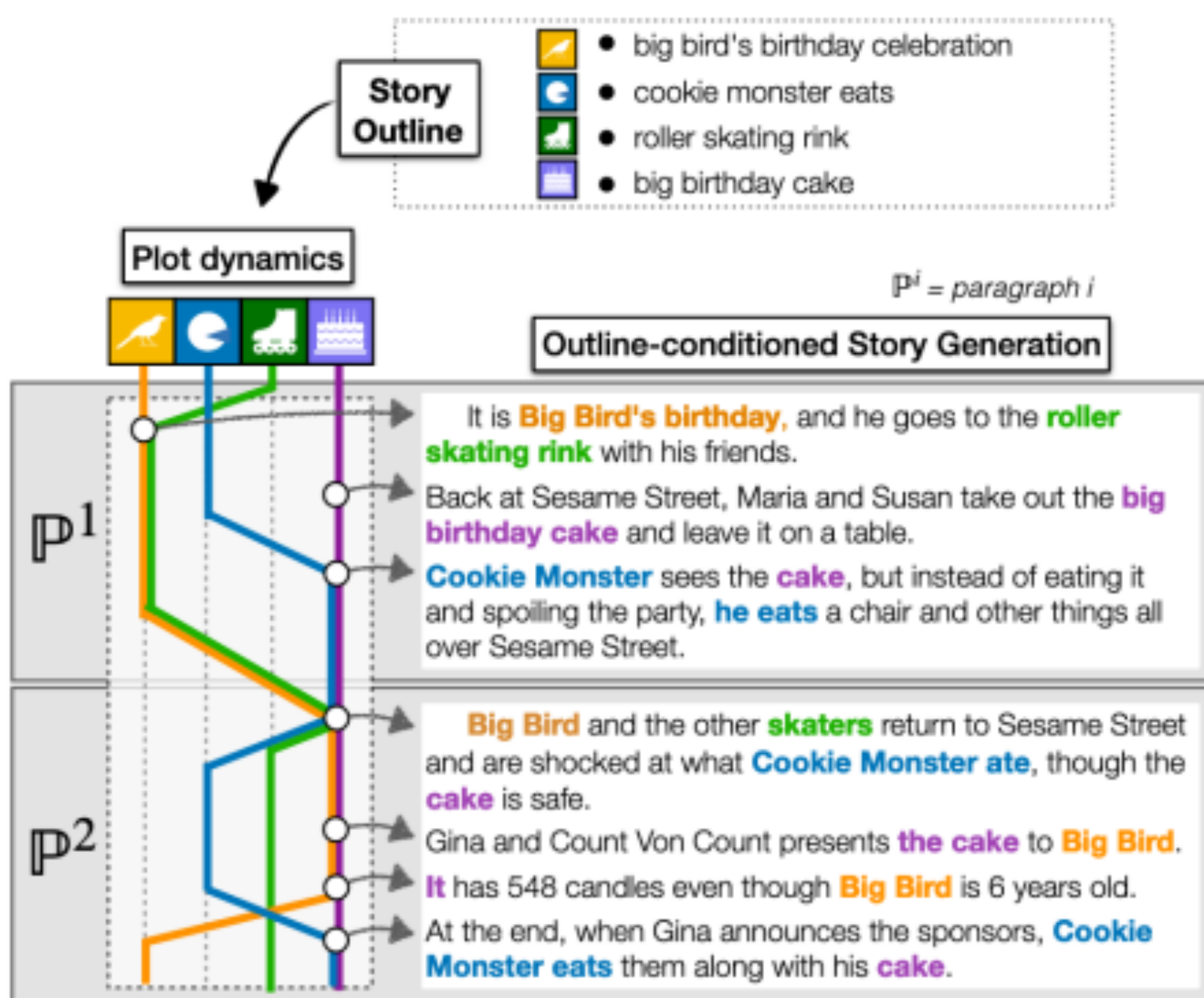
Task/Application	Input	Output
Machine Translation	French	English
Summarization	Document	Short Summary
Dialogue	Utterance Dialog history	Response
Image Captioning	Image	Caption
Story Generation	Prompt	Story

Many tasks and applications for natural language generation (NLG)

Task/Application	Input	Output
Free form question answering	Question + context	Answer
Lyric generation	Music + topic	Lyrics
Table summarization	Table	Description
Code summarization	Code	Description
Code generation	Prompt	Code

Examples of NLG

Creative stories



(Rashkin et al., EMNLP 2020)

Data-to-text

Table Title: Robert Craig (American football)
Section Title: National Football League statistics
Table Description: None

YEAR	TEAM	RUSHING					RECEIVING				
		ATT	YDS	AVG	LNG	TD	NO.	YDS	AVG	LNG	TD
1983	SF	176	725	4.1	71	8	48	427	8.9	23	4
1984	SF	155	649	4.2	28	4	71	675	9.5	64	3
1985	SF	214	1050	4.9	62	9	92	1016	11	73	6
1986	SF	204	830	4.1	25	7	81	624	7.7	48	0
1987	SF	215	815	3.8	25	3	66	492	7.5	35	1
1988	SF	310	1502	4.8	46	9	76	534	7.0	22	1
1989	SF	271	1054	3.9	27	6	49	473	9.7	44	1
1990	SF	141	439	3.1	26	1	25	201	8.0	31	0
1991	RAI	162	590	3.6	15	1	17	136	8.0	20	0
1992	MIN	105	416	4.0	21	4	22	164	7.5	22	0
1993	MIN	38	119	3.1	11	1	19	169	8.9	31	1
Totals	-	1991	8189	4.1	71	56	566	4911	8.7	73	17

Craig finished his eleven NFL seasons with 8,189 rushing yards and 566 receptions for 4,911 receiving yards.

(Parikh et al., EMNLP 2020)

Visual description



Two children are sitting at a table in a restaurant. The children are one little girl and one little boy. The little girl is eating a pink frosted donut with white icing lines on top of it. The girl has blonde hair and is wearing a green jacket with a black long sleeve shirt underneath. The little boy is wearing a black zip up jacket and is holding his finger to his lip but is not eating. A metal napkin dispenser is in between them at the table. The wall next to them is white brick. Two adults are on the other side of the short white brick wall. The room has white circular lights on the ceiling and a large window in the front of the restaurant. It is daylight outside.

(Krause et al. CVPR 2017)