Below are some ideas for potential NLP-based projects that might help you and your team brainstorm on what you would want to work on for the term project:

Text Classification

- Spam filtering
  - Classify emails as spam/not spam
    - Example dataset: Trec spam dataset, Spambase etc.
- De-anonymization
  - Classify the text of an e-mail message to decide who sent it?
    - Datasets include 150,000 Enron emails among others
- Autonomous Tagging of StackOverflow Questions
  - Make a multi-label classification system that automatically assigns tags for questions posted on a forum such as StackOverflow or Quora.
    - Dataset: StackLite or 10% sample
- Keyword/Concept identification
  - Identify keywords from millions of questions
    - Dataset: StackOverflow question samples by Facebook
- Topic identification
  - Multi-label classification of printed media articles to topics
    - Dataset: Greek Media monitoring multi-label classification

Text Summarization

- Automatically create a summary with the major points of the original document?
  - Abstractive (write your own summary) and Extractive (select pieces of text from original) are two popular approaches
    - Dataset: CNN and DailyMail News Pieces by Google DeepMind

Sentiment Analysis

- Generic Sentiment Analysis
  - Tweets sorted by geography and timestamp.
    - Dataset: Tweets sentiment tagged by humans
  - Movie reviews
    - Datasets such as Movie Review Data and others
  - Amazon product reviews
- Aspect-based sentiment analysis

Speech Recognition/Conversational AI

- Social Chat/Conversational Bots
    - Can you build a bot which talks to you just like people talk on social networking sites?
        - Reference: Chat-bot architecture
        - Dataset: Reddit Dataset
- Copy-cat Bot
    - Generate plausible new text which looks like some other text
    - Obama Speeches? For instance, you can create a bot which writes some new speeches in Obama's style
    - Trump Bot? Or a Twitter bot which mimics @realDonaldTrump

Language Understanding

- Automated essay grading
    - The purpose of this project is to implement and train machine learning algorithms to automatically assess and grade essay responses.
        - Dataset: Essays with human graded scores
- Sentence to Sentence semantic similarity
    - Can you identify question pairs that have the same intent or meaning?
        - Dataset: Quora question pairs with similar questions marked
- Fight online abuse
    - Can you confidently and accurately tell whether a particular comment is abusive?
        - Dataset: Toxic comments on Kaggle
- Open Domain question answering
    - Can you build a bot which answers questions according to the student's age or her curriculum?
    - Facebook's FAIR is built in a similar way for Wikipedia.
        - Dataset: NCERT books for K-12/school students in India, NarrativeQA by Google DeepMind and SQuAD by Stanford

Other ideas include:

- **Recommender systems (based on user's social media profiles)**
- **Machine Translation**

- Speech Recognition

References:

https://github.com/NirantK/awesome-project-ideas#text

Machine Learning Mastery Blog

Word Embeddings vs. Words