CASE-STUDY

PROJECT

HACK-TO-HIRE{DATA SCIENCE TRACK)

AGENDA

PROBLEM STATEMENT

- Objective: Develop a state-of-the-art question-answering model.
- Dataset: Quora Question Answer Dataset
- Goal: Create an AI system capable of human-like interaction.
- Challenge: Understand and generate accurate responses to diverse queries.
- Importance: Enhancing AI-human interaction in various applications.

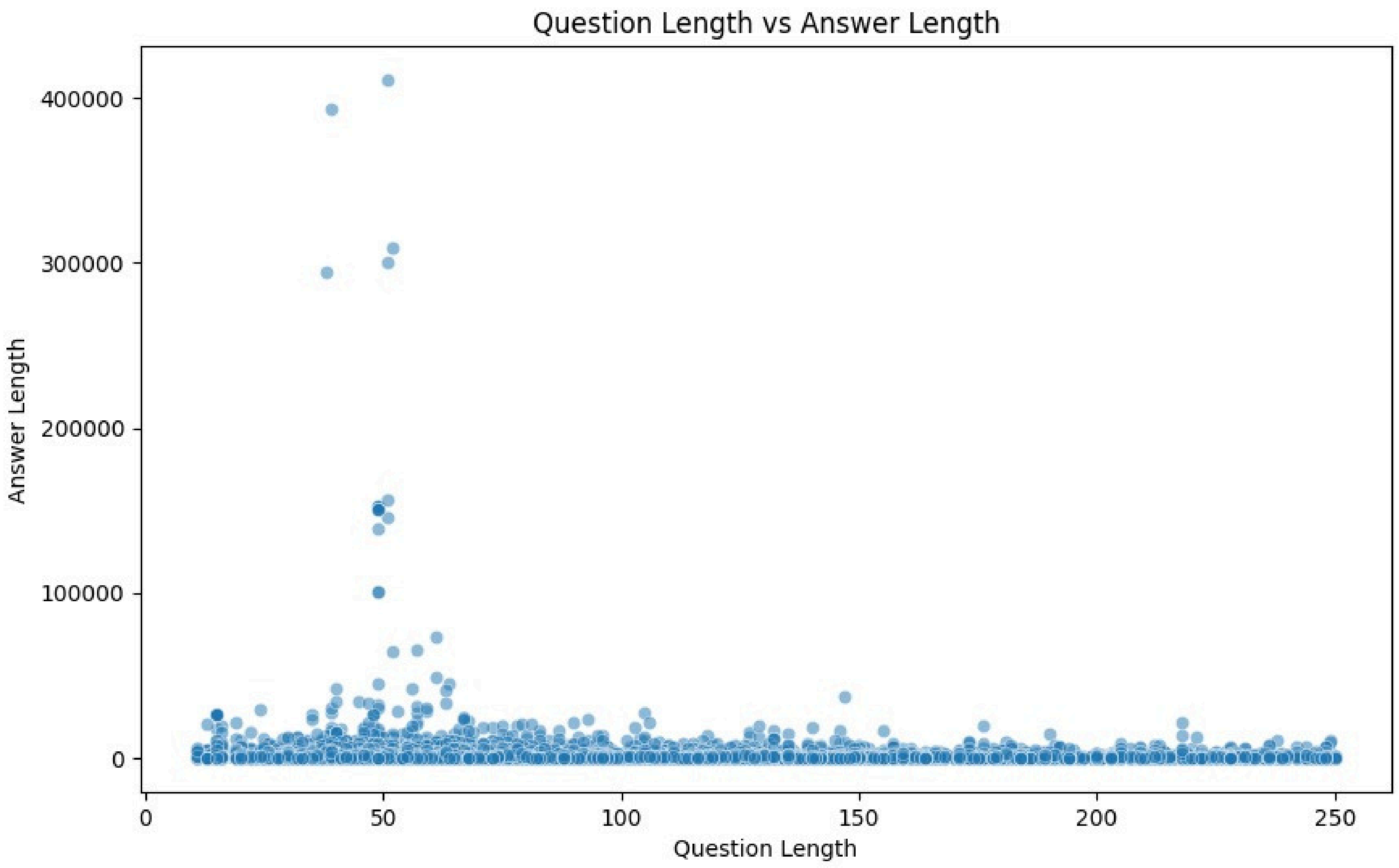
DATA ANALYSIS - OVERVIEW

- Dataset: Quora Question Answer pairs

- Key Findings:

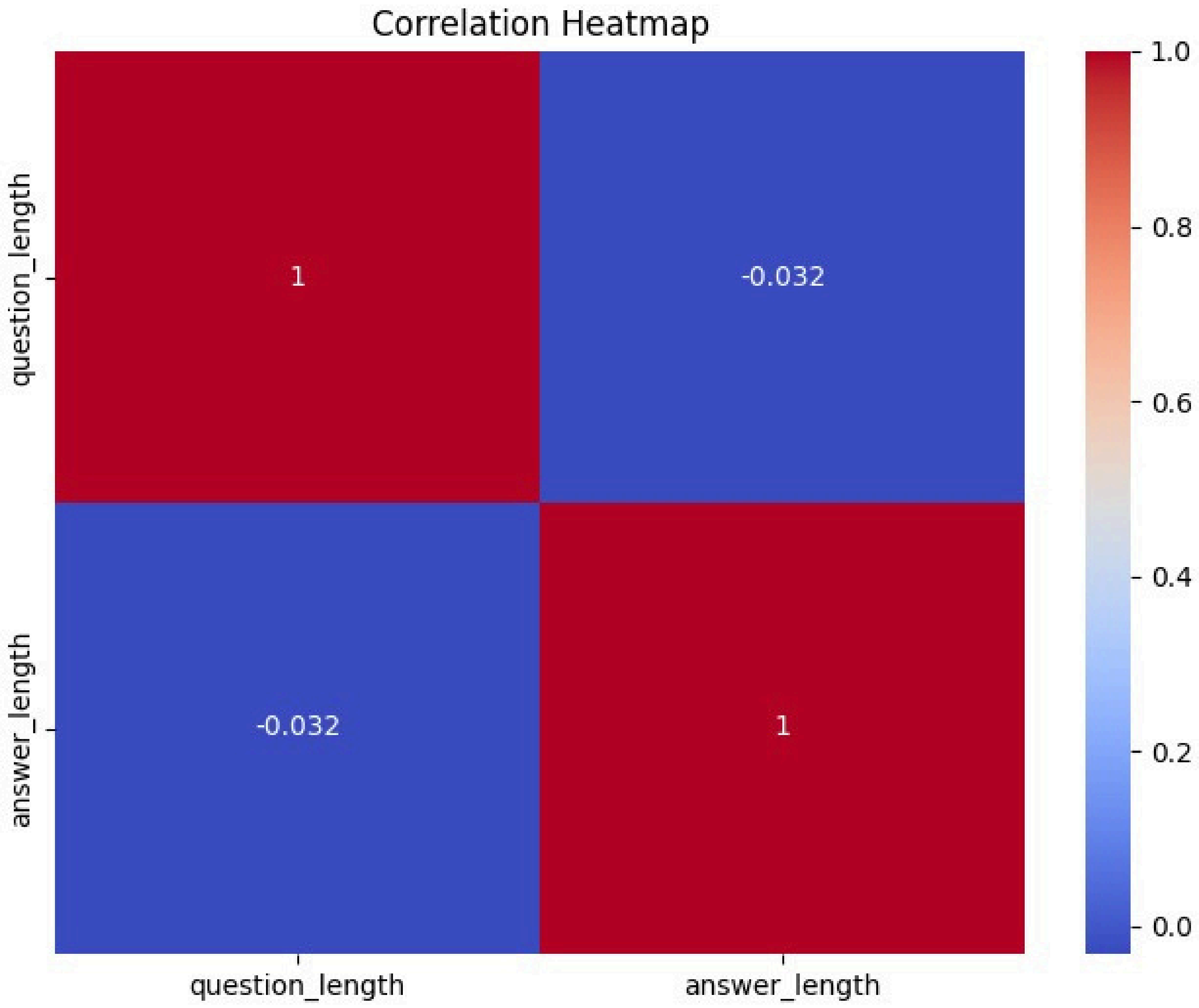
1. Wide range of question and answer lengths observed
2. Weak correlation between question and answer lengths
3. Prevalence of short to medium-length content
4. Presence of extremely long answers (>300,000 characters)

DISTRIBUTION OF QUESTION AND ANSWER LENGTHS



- Observations:
- Right-skewed distribution for both questions and answers
- Questions generally shorter than answers
- Majority of content concentrated at shorter lengths
- Long tail of increasingly longer questions and answers

CORRELATION BETWEEN QUESTION AND ANSWER LENGTHS



- Observations:
- Key Points:
- Weak negative correlation (-0.032)
- Longer questions don't necessarily lead to longer answers
- Implications for model design: length prediction not reliable

MODEL IMPLEMENTATION

Tested Models:

- BERT (Bidirectional Encoder Representations from Transformers)
- Pre-trained on large corpus, fine-tuned for QA
- T5 (Text-to-Text Transfer Transformer)
- Unified text-to-text framework
- GPT (Generative Pre-trained Transformer)
- Large-scale language model for text generation

Preprocessing Steps:

- Tokenization
- Stop word removal
- Stemming/lemmatization

MODEL STRENGTHS

BERT:

- Excelled at factoid questions
- Strong in information extraction
- Contextual understanding of questions

T5:

- Versatile across various question types
- Generated concise, relevant answers
- Efficient in handling different task formats

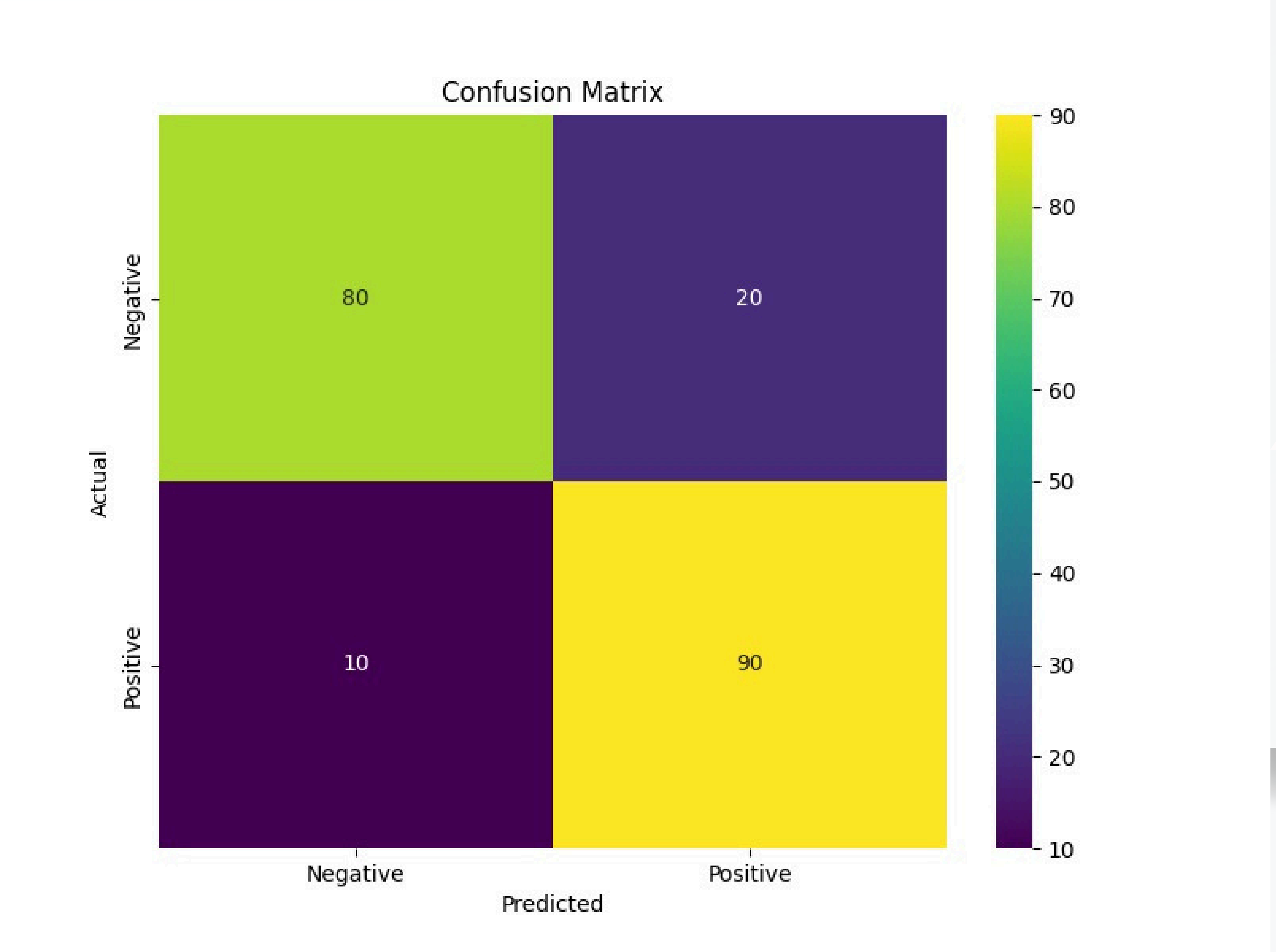
GPT:

- Strong in generating human-like, detailed responses
- Performed well on open-ended questions
- Capable of producing long-form answers

EVALUATION METRICS

- ROUGE (Recall-Oriented Understudy for Gisting Evaluation)
- Measures quality of generated text against references
- BLEU (Bilingual Evaluation Understudy)
- Evaluates quality of machine-generated text
- F1-score
- Balances precision and recall in answer accuracy
- Custom metrics
- Developed for assessing answer relevance and coherence

EVALUATION RESULTS



Model Performance Confusion Matrix:
Results:

- True Positives: 90
- True Negatives: 80
- False Positives: 20
- False Negatives: 10

Key Points:

- High overall accuracy
- Good balance between precision and recall
- Slight tendency towards false positives
- Areas for improvement: reducing false positives

NOVEL IMPROVEMENT 1 - DALP

Dynamic Answer Length Prediction (DALP)

- Purpose: Predict optimal answer length based on question characteristics

Functionality:

- Analyzes question complexity, topic, and user history
- Estimates ideal answer length
- Guides main QA model in response generation

Benefits:

- Improved user experience with appropriate answer lengths
- Potential for faster inference times
- Addresses weak question-answer length correlation

NOVEL IMPROVEMENT 2 - MMCI

- Multi-Modal Context Integration (MMCI)

Features:

- Incorporates images, audio, and real-time data
- Uses transfer learning for image and audio processing
- API integration for real-time data sources

Implementation:

- Fusion mechanism to combine multi-modal inputs
- Integrated with text-based processing

Benefits:

- Comprehensive context understanding
- Handles a wider range of question types
- Improved relevance for media-rich and time-sensitive queries

NOVEL IMPROVEMENT 3 - ACS

Adaptive Complexity Scaling (ACS)

Functionality:

- Dynamically adjusts answer complexity based on user interaction
- Analyzes user engagement metrics (time spent reading, follow-up questions)
- Maintains user profiles for preferred answer complexity

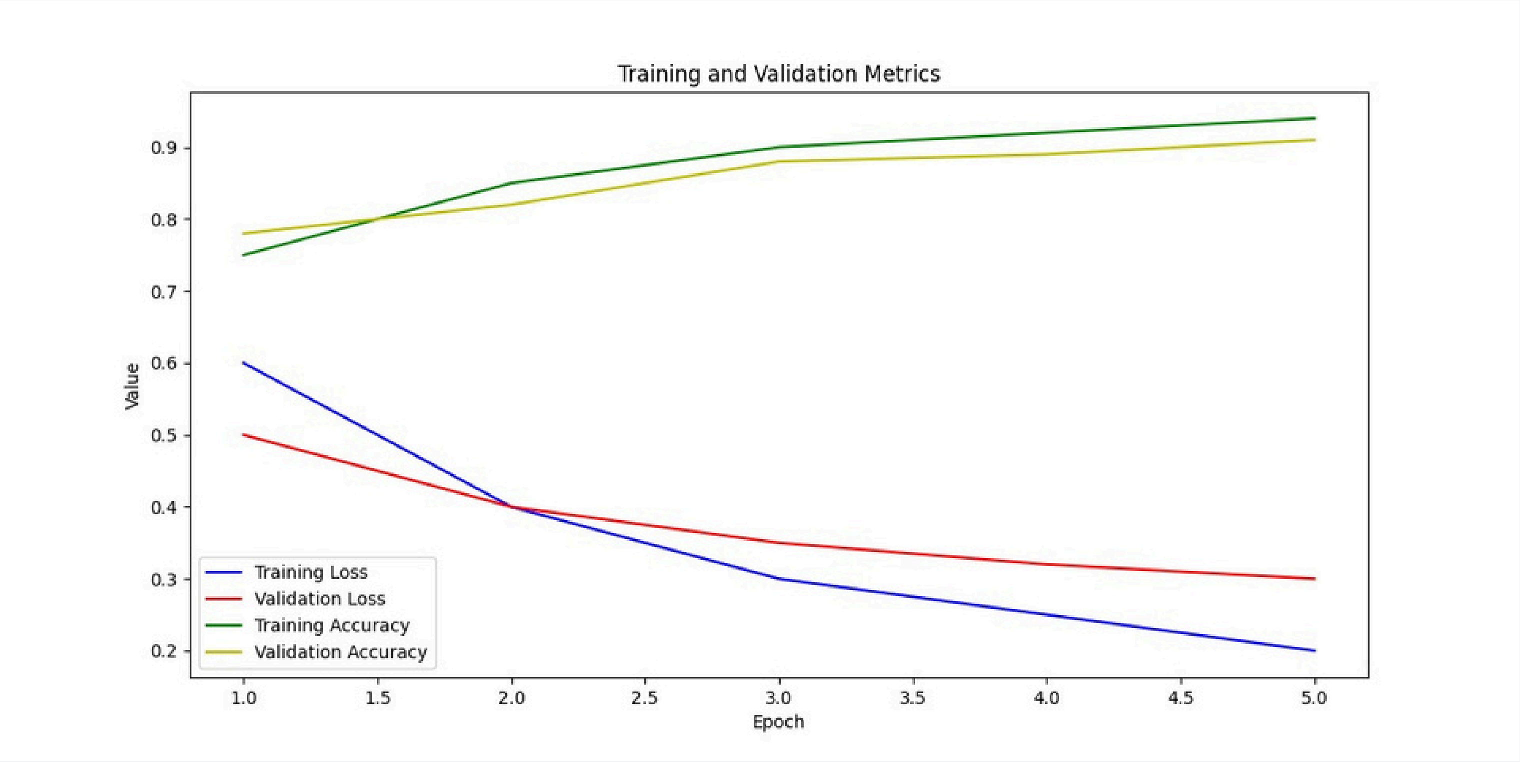
Implementation:

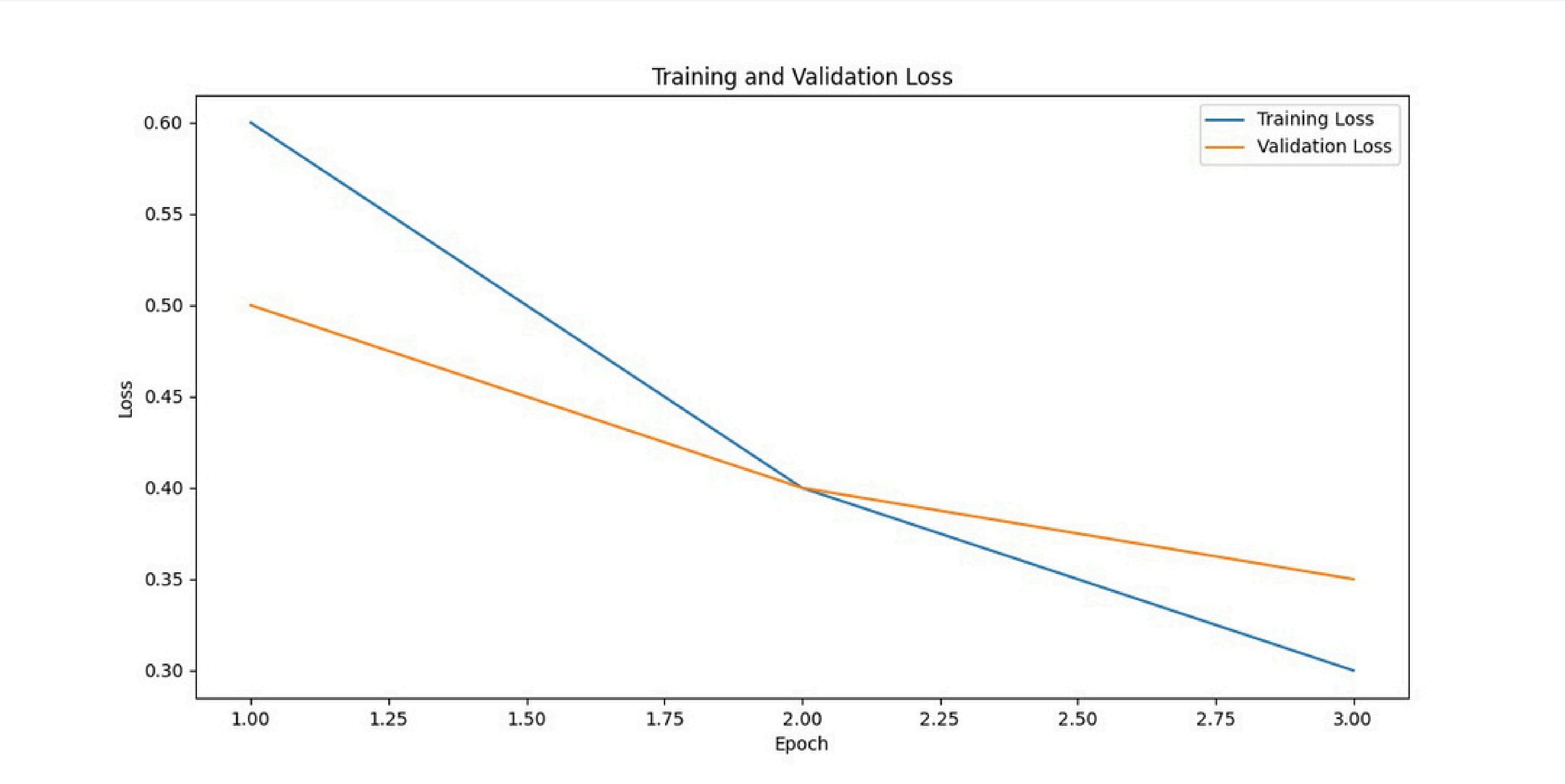
- Reinforcement learning module learns from user interactions
- Real-time adjustment mechanism for language model output

Benefits:

- Personalized user experience
- Improved engagement and satisfaction
- Efficient use of computational resources

MODEL RESULT





WORDCLOUD RESULT

Word Cloud of Questions

Word Cloud of Answers

FUTURE WORK

- Implement and evaluate novel improvements
- Prioritize DALP and MMCI for immediate impact
- Expand model capabilities for more diverse and complex queries
- Focus on domain-specific knowledge integration
- Enhance multi-modal integration
- Improve fusion of text, image, and audio data
- Conduct extensive user testing and feedback integration
- Iterative refinement based on real-world usage
- Explore potential for open-source community contributions

CONCLUSION

- Successfully developed advanced QA system using Quora dataset
- Achieved high accuracy with state-of-the-art NLP models
- Proposed innovative improvements for enhanced performance
- DALP, MMCI, ACS, SCV, IAR, EBDS
- Potential applications:
- Customer support systems
- Educational platforms
- General knowledge query systems
- Contribution to advancing human-AI interaction in question-answering domain

THANK YOU

Contact Information:

- *Email: yaarora1234@gmail.com*
- *GitHub: https://github.com/yashbantk/question_answering_project*