**Part B:**

**Design Document**

**1. Introduction**

This document outlines the approach to fine-tuning a DistilBERT model for a question-answering task using GloVe embeddings. The goal is to evaluate the model's performance with and without GloVe embeddings and propose enhancements to generalize the model.

**2. Approach**

**2.1 Data Preparation**

1. **Data Collection**:
   * Fetch the introduction section from the Wikipedia page on Statistics.
   * Define relevant questions to evaluate the model.
2. **Data Preprocessing**:
   * Tokenize the context and questions using the DistilBERT tokenizer.
   * Prepare the dataset for training and evaluation.

**2.2 Model Setup**

1. **Baseline Model**:
   * Use the pre-trained DistilBERT model for question answering.
   * Fine-tune the model on the prepared dataset.
2. **Enhanced Model with GloVe Embeddings**:
   * Load GloVe embeddings and create an embedding matrix.
   * Modify the DistilBERT model to include GloVe embeddings.
   * Concatenate BERT and GloVe embeddings and pass them through a linear layer to get the start and end logits.

**2.3 Evaluation**

1. **Evaluation Metrics**:
   * Use standard metrics like Exact Match (EM) and F1 score to evaluate the model's performance.
2. **Comparison**:
   * Compare the performance of the baseline model and the enhanced model with GloVe embeddings.

**3. Detailed Steps and Workflows**

**3.1 Data Preparation**

1. **Fetch Wikipedia Content**:
   * Use requests and BeautifulSoup to fetch and parse the introduction section from the Wikipedia page on Statistics.
2. **Define Questions**:
   * Define a set of questions relevant to the context.

**3.2 Model Setup**

1. **Load Tokenizer and Model**:
   * Load the DistilBERT tokenizer and model using the transformers library.
2. **Load GloVe Embeddings**:
   * Load GloVe embeddings from a file and create an embeddings index.
3. **Modify the Model**:
   * Extend the DistilBERT model to include GloVe embeddings.
   * Initialize the embedding layer with GloVe embeddings.
   * Concatenate BERT and GloVe embeddings in the forward pass.

**3.3 Evaluation**

1. **Answer Questions**:
   * Use the fine-tuned model to answer the defined questions.
   * Evaluate the answers using standard metrics.
2. **Compare Performance**:
   * Compare the performance of the baseline model and the enhanced model with GloVe embeddings.

**4. Impact of Word Embedding Techniques**

1. **BERT Embeddings**:
   * Contextual embeddings that capture the meaning of words in different contexts.
   * Fine-tuning on specific tasks can lead to high performance.
2. **GloVe Embeddings**:
   * Pre-trained static embeddings that capture semantic relationships between words.
   * Combining GloVe with BERT can enhance the model by providing additional semantic information.
3. **Comparison**:
   * BERT embeddings are more powerful for capturing context-specific meanings.
   * GloVe embeddings can provide complementary semantic information.
   * Combining both can potentially improve model performance.

**5. Enhancements and Alternative Strategies**

1. **Use of Other Pre-trained Models**:
   * Experiment with other pre-trained models like RoBERTa, ALBERT, or T5.
2. **Data Augmentation**:
   * Use data augmentation techniques to create more training examples.
3. **Ensemble Methods**:
   * Combine predictions from multiple models to improve performance.
4. **Hyperparameter Tuning**:
   * Perform hyperparameter tuning to find the optimal settings for training.
5. **Transfer Learning**:
   * Use transfer learning to leverage knowledge from related tasks.

**6. Review of Contemporary Methodologies**

1. **BERT and Its Variants**:
   * BERT, RoBERTa, ALBERT, and other transformer-based models have shown state-of-the-art performance in various NLP tasks.
2. **Static vs. Contextual Embeddings**:
   * Static embeddings like Word2Vec and GloVe vs. contextual embeddings like BERT and ELMo.
3. **Hybrid Approaches**:
   * Combining static and contextual embeddings to leverage the strengths of both.
4. **Recent Advances**:
   * Explore recent advances in NLP, such as GPT-3, T5, and other large-scale pre-trained models.

**7. References**

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
2. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1532-1543).
3. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692.
4. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. arXiv preprint arXiv:1909.11942.
5. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. OpenAI Blog, 1(8), 9.