

- **1.1. Introduction**

- Project: Quantized Text Summarizer and Quiz Generator.
- Objective: Create a compact model for efficient CPU performance using **quantization** to reduce memory and inference time.
- Benefits: Extracts information, generates quizzes, enhances learning access.

- **1.2. Background**

- Current tools: Rely on large language models (LLMs), inaccessible due to resource needs.
- Need: Precise, query-driven summarization and interactive tools.
- Solution: Quantized model reducing resources while retaining effectiveness.

- **1.3. Problem Statement**

- Focus: Quantizing a large language model (LLM) for efficiency and accessibility.
- Objective: Reduce resource needs while maintaining performance.
- Applications: Summarization, quiz generation, query-specific summaries.

- **1.4. Stakeholders & Interests**

- Students/Researchers: Need accurate summaries and quizzes.
- Educational Institutions: Need tools for personalized learning and content digestion.
- Teachers/Educators: Need simplified quiz generation for efficient assessment.

- **1.5. Objectives**

- **Model Quantization:** Reduce memory and computational demands for standard CPUs, exploring Pruning, Trained Quantization, and Huffman Coding.
- **Efficient Text Summarization:** Develop a lightweight model for concise summaries.
- **Query-Based Summarization:** Implement summaries tailored to user queries.
- **Automated Quiz Generation:** Design quizzes assessing understanding of key concepts.

- **1.6. Scope**

- Goal: Lightweight tool for summarizing text and generating quizzes optimized for standard CPUs using **model quantization**.
- Functionality: Query-based summarization, user-friendly interface.
- Limitations: Text summarization quality and coverage, basic quiz formats, primarily English content.

- **1.7. Assumptions**

- System used on standard CPUs.
- Pre-trained models suitable for various texts.
- Users expect basic quiz formats.
- Users upload well-structured, English documents.

- **1.8. Risks**

- User hesitancy to adopt AI-generated content.
- System may not generalize to all text genres.

- Processing large documents may cause slower performance.
- Non-intuitive interface may lower user engagement.

• 1.9. Success Criteria

- **User Management:** Secure registration, login, and account management.
- **Core Functionality:** Effective text summarization and quiz generation.
- **Data Management:** Efficient storage/retrieval of user data, quizzes, summaries, results, and reports.
- **Reporting & Visualization:** Detailed performance reports with analytics and visuals.
- **Search & Filter:** Easy searching and filtering through summaries, quizzes, and reports.
- **Performance:**
 - Quiz generation: ≤ 10 seconds.
 - Summary generation: ≤ 20 seconds.
 - Report display: ≤ 5 seconds.
 - Page load: ≤ 3 seconds.
- **Scalability:** Handles ≥ 100 concurrent users.
- **Data Privacy:** User data encrypted in transit and at rest.
- **Access Control:** Role-Based Access Control (RBAC).
- **User Interface:** Intuitive, user-friendly design.
- **Cross-Platform Compatibility:** Accessible on web browsers across devices.
- **Error Handling:** Graceful error detection and informative messages.
- **Deployment Readiness:** Functional, secure, and optimized for production.

• 1.10. Tools, Libraries & Technologies

- **Tools:**
 - Google Colab (3.10): Hosted Jupyter Notebook for resource access.
 - VS Code (1.86): Code editor for development/debugging.
 - Git (2.43.0): Version control system.
- **Libraries:**
 - Tensorflow (2.17.0): Machine learning library.
 - React Js (18): UI library.
 - Node Js (20.18.0): Server environment.
 - FastApi (0.0): API framework.
- **Technology:**
 - Python (3.10): AI/ML language.
 - JavaScript (ES6): Web development language.

• 1.11. Work Division

- Muhammad Haider Sultan (BCSF21M530): Developer + Researcher.
- Ausaf Gill (BCSF21M546): Developer + Software Tester.
- Ahmed Arif (BCSF21M535): Documentation.
- Kamran Faisal (BCSF21M523): Backend Developer + Documentation.

• 1.12. Conclusion

- Project aims to provide text processing and interactive learning via **model quantization**.
- Addresses computational costs/accessibility issues of LLMs.
- Provides query-driven summaries and quizzes for improved learning.
- Paves the way for inclusive AI-driven educational tools.

- Well-positioned for functional and impactful solution.