



**Pengembangan Model BERT dengan Metode *Abstraction Summarization* untuk Ringkasan Otomatis Tafsir Ayat Al-Qur'an**

**Proposal Skripsi**

diajukan sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Komputer

oleh  
Reiki Aziz Yoga Utama  
4611422055

**TEKNIK INFORMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI SEMARANG  
SEMARANG  
2024**

## **PERSETUJUAN PEMBIMBING**

Skripsi berjudul “Pengembangan Model BERT dengan Metode *Abstraction Summarization* untuk Ringkasan Otomatis Tafsir Ayat Al-Qur'an” yang disusun oleh :

nama	: Reiki Aziz Yoga Utama
NIM	: 4611422055
Prodi/Fakultas	: Teknik Informatika/Matematika dan Ilmu Pengetahuan Alam

Semarang, 06 September 2024  
Belum ditentukan

Belum ditentukan & NIP

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Al-Qur'an merupakan kitab suci umat Islam yang menjadi sumber hukum Islam yang pertama dan utama (Aji Fitra Jaya Institut Perguruan Tinggi Ilmu Al Qur, n.d.). Dalam praktiknya, pemahaman terhadap Al-Qur'an seringkali membutuhkan penjelasan atau tafsir yang mendalam agar pesan yang terkandung dapat dimengerti oleh berbagai kalangan. Tafsir Al-Qur'an adalah karya ilmiah yang menjelaskan makna ayat-ayat Al-Qur'an secara mendetail, baik dalam bahasa Arab maupun bahasa lainnya, termasuk bahasa Indonesia. Namun, panjangnya tafsir sering kali menjadi tantangan, terutama bagi mereka yang membutuhkan pemahaman cepat namun tetap akurat.

Seiring dengan perkembangan teknologi, berbagai pendekatan berbasis kecerdasan buatan (Artificial Intelligence) mulai diterapkan untuk mempermudah akses terhadap informasi, termasuk dalam bidang teks keagamaan. Salah satu teknologi yang berperan penting dalam pengolahan teks adalah **Natural Language Processing (NLP)**. Dalam konteks ini, **summarization** atau teknik peringkasan otomatis merupakan salah satu pendekatan NLP yang dapat membantu meringkas tafsir tanpa kehilangan esensi dari teks aslinya. Summarization menawarkan kemudahan bagi pengguna untuk memahami inti dari tafsir yang panjang dalam waktu yang lebih singkat.

Pada umumnya, terdapat dua metode utama dalam summarization, yaitu **extractive** dan **abstractive summarization**. **Extractive summarization** mengekstraksi kalimat-kalimat penting dari teks asli, sedangkan **abstractive summarization** mencoba meringkas informasi dengan menyusun kalimat baru yang lebih singkat dan koheren. Abstractive summarization memiliki keunggulan dalam menciptakan ringkasan yang lebih alami dan tidak sekadar mengutip langsung teks asli, namun memiliki tantangan lebih besar dalam memahami konteks yang kompleks (Alami Merrouni et al., 2023). Dalam konteks teks bilingual seperti tafsir Al-Qur'an dalam bahasa Arab dan Indonesia, abstractive summarization memberikan potensi yang lebih besar untuk menghasilkan ringkasan yang informatif dan tidak bergantung pada struktur kalimat asli, sehingga cocok untuk menangani kompleksitas dan perbedaan bahasa tersebut.

Salah satu model yang dapat mendukung penerapan **abstractive summarization** adalah **BERT** (Bidirectional Encoder Representations from Transformers). BERT telah terbukti sangat efektif dalam berbagai tugas NLP, termasuk **abstractive summarization**, karena kemampuannya untuk memahami konteks dua arah secara mendalam dalam teks. Ini memungkinkan BERT untuk menghasilkan ringkasan yang lebih akurat dan koheren, terutama untuk teks panjang dan kompleks. Namun, penggunaan model BERT dalam konteks tafsir Al-Qur'an yang bilingual (Arab-Indonesia) masih belum banyak dieksplorasi. Padahal, potensinya sangat besar untuk meningkatkan aksesibilitas dan pemahaman terhadap tafsir dengan menghasilkan ringkasan otomatis yang lebih informatif dan mudah dipahami.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengoptimalkan penggunaan metode **abstractive summarization** berbasis model transformer **BERT** dalam meringkas tafsir Al-Qur'an dalam dua bahasa, yaitu Arab dan Indonesia. Dengan memanfaatkan teknologi ini, diharapkan hasil ringkasan dapat tetap mempertahankan makna dan esensi tafsir yang kompleks, namun disajikan dalam bentuk yang lebih ringkas dan mudah dipahami oleh pengguna. Penelitian ini juga diharapkan dapat mengisi kesenjangan penelitian sebelumnya yang belum sepenuhnya memanfaatkan potensi summarization berbasis transformer untuk teks religius bilingual, khususnya dalam konteks tafsir Al-Qur'an.

## 1.2 Batasan Masalah

Penelitian ini memiliki beberapa batasan untuk memastikan fokus yang jelas dan pencapaian tujuan yang diinginkan. Pertama, penelitian ini difokuskan hanya pada teks tafsir Al-Qur'an yang berbahasa Arab serta terjemahannya dalam bahasa Indonesia, sehingga tidak mencakup teks keagamaan lain atau teks tafsir dalam bahasa selain Arab dan Indonesia. Kedua, penelitian ini menggunakan metode **abstractive summarization** berbasis transformer **BERT**, tanpa menerapkan metode lain di luar pendekatan ini. Model yang digunakan terbatas pada BERT (Bidirectional Encoder Representations from Transformers) yang disesuaikan untuk menangani teks bilingual (Arab-Indonesia), dan tidak mengeksplorasi model lain seperti GPT, T5, atau model transformer lainnya.

Selain itu, evaluasi kinerja dilakukan menggunakan metrik standar seperti ROUGE, tanpa menggunakan metrik evaluasi lain yang lebih kompleks atau spesifik terhadap domain tertentu. Dataset yang digunakan juga dibatasi pada kumpulan tafsir yang telah tersedia secara publik dan terjemahannya yang resmi, sehingga tidak dilakukan pengumpulan atau pembuatan dataset baru dari sumber yang belum terverifikasi. Akhirnya, penelitian ini hanya mempertimbangkan konteks teks pada level kalimat dan paragraf dalam tafsir, tanpa memasukkan konteks historis atau sosial dari tafsir Al-Qur'an secara keseluruhan.

## 1.3 Rumusan Masalah

Berdasarkan latar belakang dan batasan yang telah dijelaskan, rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana menerapkan metode **Abstractive summarization** berbasis transformer BERT untuk menghasilkan ringkasan otomatis dari teks bilingual (Arab-Indonesia) pada tafsir Al-Qur'an?
2. Sejauh mana abstractive summarization mampu menangkap esensi dari tafsir Al-Qur'an dalam dua bahasa?
3. Bagaimana performa model abstractive summarization dalam meringkas tafsir Al-Qur'an bila dievaluasi menggunakan metrik seperti ROUGE?

## 1.4 Tujuan Penelitian

Tujuan penelitian ini dapat dirinci sebagai berikut:

1. Menerapkan metode *abstraction summarization* berbasis transformer BERT untuk menghasilkan ringkasan otomatis dari teks bilingual (Arab-Indonesia) pada tafsir Al-Qur'an.
2. Sejauh mana pendekatan **abstractive summarization** mampu menangkap esensi dari tafsir Al-Qur'an dalam dua bahasa?.
3. Bagaimana performa model **abstractive summarization** dalam meringkas tafsir Al-Qur'an bila dievaluasi menggunakan metrik seperti ROUGE?

## 1.5 Manfaat Penelitian

### 1. Kontribusi Teoritis:

- Mengembangkan dan memperkaya metode **abstractive summarization** berbasis transformer BERT untuk teks bilingual (Arab-Indonesia), khususnya dalam konteks tafsir Al-Qur'an.
- Menyediakan wawasan baru dalam literatur pemrosesan bahasa alami (NLP) dan summarization dengan penerapan teknologi canggih pada teks keagamaan.

### 2. Manfaat Praktis:

- Mempermudah proses meringkas teks tafsir Al-Qur'an, membuat informasi lebih aksesibel dan mudah dipahami oleh pembaca bilingual.
- Menyediakan model summarization yang dapat digunakan oleh lembaga penelitian, pengembang aplikasi keagamaan, dan institusi pendidikan untuk menyederhanakan informasi dari tafsir dan meningkatkan kualitas bahan ajar serta sumber daya belajar.

### 3. Aplikasi Potensial:

- Model yang dikembangkan dapat diterapkan dalam aplikasi edukasi dan penelitian untuk memfasilitasi pemahaman yang lebih baik terhadap tafsir Al-Qur'an.
- Meningkatkan efisiensi dalam pengembangan sumber daya belajar dan materi ajar yang berkaitan dengan teks keagamaan.

## 1.6 Kebaruan Penelitian

Kebaruan penelitian ini terletak pada penerapan **abstractive summarization** berbasis transformer **BERT** dalam meringkas teks bilingual (Arab-Indonesia) pada tafsir Al-Qur'an. Dari segi konsep, penelitian ini memperkenalkan penggunaan pendekatan **abstractive** untuk meringkas teks keagamaan, khususnya tafsir Al-Qur'an, yang belum banyak dieksplorasi. Pendekatan ini menawarkan cara baru dalam menangani teks religius yang kompleks dan bilingual, yang bertujuan menghasilkan ringkasan yang lebih alami dan koheren. Dari segi metode, penelitian ini mengadopsi transformer **BERT** yang disesuaikan untuk menangani teks dalam dua bahasa secara bersamaan. Hal ini memungkinkan BERT untuk memahami

konteks bahasa Arab dan Indonesia secara efektif, sehingga dapat menghasilkan ringkasan yang lebih akurat dan informatif. Inovasi ini memperluas cakupan aplikasi model transformer dalam pemrosesan teks religius bilingual, memberikan metode baru dalam menghasilkan ringkasan otomatis yang efektif dan relevan.

## BAB II

### Kajian Pustaka

#### 2.1 Tinjauan Pustaka

Dalam penelitian mengenai **abstractive summarization** berbasis model **BERT** untuk teks bilingual (Arab-Indonesia) dalam konteks **tafsir Al-Qur'an**, berbagai kajian sebelumnya dapat dijadikan landasan penting. Secara umum, summarization terbagi dalam dua metode utama: **extractive** dan **abstractive**. Pada extractive summarization, kalimat-kalimat penting dipilih dari teks asli tanpa perubahan struktur, sedangkan abstractive summarization menghasilkan kalimat baru berdasarkan pemahaman semantik terhadap teks.

Penelitian oleh (Nallapati et al., n.d.) menggarisbawahi efektivitas metode extractive dalam menghasilkan ringkasan yang cepat, tetapi dengan keterbatasan dalam fleksibilitas dan akurasi semantik. Sebaliknya, penelitian oleh (Liu & Lapata, n.d.) menunjukkan bahwa abstractive summarization berbasis **BERT** memiliki kemampuan lebih baik dalam menghasilkan ringkasan yang lebih menyerupai interpretasi manusia, karena mampu memahami konteks dan makna dari teks yang lebih luas.

Selain itu, penelitian oleh (Alselwi & Taşçı, 2024) mengembangkan teknik summarization berbasis graf untuk teks Arab dengan menggunakan **word embedding** dan algoritma **PageRank**, yang dikenal sebagai **GEATS (Graph-based Extractive Arabic Text Summarization)**. Teknik ini memanfaatkan pendekatan berbasis graf untuk menangani hubungan morfologis yang kompleks dalam bahasa Arab. Hasil dari penelitian ini menunjukkan bahwa pendekatan GEATS memberikan hasil ringkasan yang lebih baik dibandingkan metode lainnya, dengan peningkatan lebih dari 7,5% pada nilai **F-measure**.

Penelitian-penelitian tersebut menjadi fondasi penting untuk pengembangan model summarization berbasis **BERT** dalam konteks teks bilingual yang lebih spesifik seperti tafsir Al-Qur'an, di mana makna dan esensi dari teks keagamaan harus tetap terjaga dalam proses peringkasan.

#### 2.2 Landasn Teoritik

Landasan teoretik penelitian ini berfokus pada konsep-konsep summarization, Natural Language Processing (NLP), serta model transformer seperti BERT. Summarization sendiri adalah proses merangkum teks panjang menjadi bentuk yang lebih singkat tanpa kehilangan informasi penting yang terkandung di

dalamnya. Summarization terbagi dalam dua jenis utama, yaitu extractive dan abstractive. Extractive summarization mengambil langsung kalimat atau frasa penting dari teks asli tanpa perubahan struktur, sedangkan abstractive summarization menghasilkan ringkasan baru yang dibentuk berdasarkan pemahaman mendalam model terhadap makna teks (Nallapati et al., n.d.) Dalam konteks penelitian ini, abstractive summarization lebih diutamakan karena memungkinkan model untuk menghasilkan kalimat yang lebih fleksibel dan bermakna, khususnya ketika meringkas tafsir Al-Qur'an yang memiliki konteks religius dan nilai semantik yang kompleks..

NLP, yang merupakan bidang utama dalam penelitian ini, memainkan peran penting dalam pengolahan teks otomatis, termasuk summarization. NLP mencakup berbagai teknik yang memungkinkan komputer untuk memproses dan menganalisis bahasa manusia, baik dalam bentuk tulisan maupun lisan (Jurafsky & Martin, 2023). Di dalam NLP, model-model berbasis transformer, seperti BERT, semakin populer karena keunggulannya dalam memahami konteks yang luas dalam teks dan menghasilkan representasi semantik yang lebih kaya. BERT, atau Bidirectional Encoder Representations from Transformers, menggunakan pendekatan bidirectional, yaitu mempelajari hubungan kata-kata dari dua arah dalam satu kali pemrosesan. Hal ini memberikan BERT keunggulan signifikan dalam menangkap konteks makna dari teks yang kompleks, termasuk teks keagamaan seperti tafsir Al-Qur'an, yang sering kali memiliki struktur bahasa yang mendalam dan berlapis (Devlin et al., n.d.).

Dalam penelitian ini, BERT dipilih sebagai model dasar karena kemampuannya untuk menangani konteks dua arah yang sangat bermanfaat dalam abstractive summarization. Model BERT dapat dilatih untuk memahami hubungan semantik dalam teks panjang, yang menjadikannya cocok untuk aplikasi dalam meringkas tafsir Al-Qur'an (Devlin et al., n.d.). Model ini akan disesuaikan dengan domain teks religius dan bilingual yang diambil dari bahasa Arab dan Indonesia. Dengan demikian, BERT tidak hanya akan menangkap informasi dasar tetapi juga nuansa dari makna yang terkandung di dalam tafsir. Proses ini mencakup pemahaman mengenai hubungan antar kata dan frasa yang ada dalam kedua bahasa tersebut, sehingga hasil ringkasan tetap informatif, akurat, dan merepresentasikan esensi tafsir yang lebih luas.

Penggunaan teks bilingual dalam penelitian ini, yakni bahasa Arab dan Indonesia, menambahkan tantangan tersendiri karena perbedaan struktural dan gramatikal yang ada di antara kedua bahasa. Bahasa Arab memiliki struktur morfologi yang berbeda dari bahasa Indonesia, yang mengharuskan model BERT untuk beradaptasi dalam memahami makna antarbahasa. Transfer learning akan digunakan untuk melatih BERT pada domain teks keagamaan sehingga mampu menangani teks bilingual secara lebih efektif. Dengan proses adaptasi ini, model BERT yang digunakan diharapkan dapat mengidentifikasi serta menyampaikan esensi dari teks tafsir dalam bahasa yang lebih mudah dipahami tanpa mengorbankan makna.

Dengan mengacu pada konsep dan teori di atas, landasan teoretik penelitian ini merangkum berbagai elemen penting yang relevan untuk meringkas tafsir Al-Qur'an secara otomatis. Penggunaan BERT untuk abstractive summarization dalam teks bilingual menunjukkan pendekatan yang inovatif dan bertujuan untuk mengatasi tantangan dalam menyajikan ringkasan teks keagamaan yang kompleks. Pendekatan ini tidak hanya menyederhanakan teks tafsir, tetapi juga mempertahankan kedalaman makna yang dikandungnya, memungkinkan pengguna untuk memperoleh informasi inti dengan cara yang lebih cepat dan mudah diakses. Penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam pengembangan teknologi NLP, khususnya dalam aplikasi summarization teks-teks religius dan bilingual.

### 2.3 Kerangka Berpikir

Kerangka berpikir penelitian ini mengacu pada gagasan bahwa teks tafsir Al-Qur'an yang bilingual (Arab-Indonesia) memiliki struktur dan makna yang kompleks, sehingga memerlukan metode summarization yang mampu menangkap esensi dalam dua bahasa. Model **abstractive summarization** berbasis BERT dipilih karena mampu menangani konteks teks dua arah dan menggabungkan makna dari berbagai segmen teks. Model ini akan dilatih pada teks bilingual untuk memahami hubungan semantik antarbahasa, dengan output berupa ringkasan yang informatif dan akurat. Pendekatan ini diharapkan mampu menjawab tantangan dalam menyajikan ringkasan yang ringkas namun tidak kehilangan esensi tafsir.

### 2.4 Hipotesis Teoritis

Penelitian ini mengajukan hipotesis bahwa **abstractive summarization** berbasis **BERT** dapat menghasilkan ringkasan yang lebih koheren, akurat, dan informatif dibandingkan dengan metode summarization lainnya untuk teks tafsir Al-Qur'an dalam dua bahasa. Model yang dihasilkan diharapkan dapat menangkap esensi dari tafsir Al-Qur'an, dan performanya akan dievaluasi menggunakan metrik standar seperti ROUGE.

## BAB III

### Metode Penelitian

#### 3.1 Pendekatan, Jenis, dan Prosedur penelitian

Penelitian ini menggunakan pendekatan **kuantitatif** berbasis **eksperimen** untuk mengukur efektivitas model **hybrid summarization** berbasis **BERT** dalam menghasilkan ringkasan otomatis tafsir Al-Qur'an dalam dua bahasa, yaitu Arab dan Indonesia. Jenis penelitian ini adalah **penelitian komputasi NLP (Natural Language Processing)** yang bertujuan untuk merancang dan menguji model berbasis **transformer BERT** dalam menggunakan teknik **abstractive summarization**.



### 3.2 Lokasi dan Waktu Penelitian

### 3.3 Data dan Sumber Data

Sumber data Utama dalam penelitian ini adalah teks **tafsir Al-Qur'an**. Data diambil dari **korpus publik** tafsir yang sudah tersedia secara online dan bersifat open-access, misalnya dari situs-situs tafsir resmi atau platform digital tafsir yang terpercaya. Data yang diambil meliputi:

- a) **Teks Arab tafsir Al-Qur'an**
- b) **Terjemahan resmi dalam bahasa Indonesia**

Data ini diolah untuk tujuan summarization dengan proses tokenisasi, segmentasi kalimat, dan penyelarasan antara teks Arab dan Indonesia.

### 3.4 Teknik Keabsahan Data Preprocessing

Untuk memastikan keabsahan data selama tahap preprocessing, beberapa teknik diterapkan guna mempersiapkan teks untuk pemrosesan lanjutan dan analisis yang akurat. Salah satu tekniknya adalah *tokenisasi*, yaitu membagi teks menjadi unit-unit kecil yang disebut token untuk memudahkan pemrosesan oleh model. Dengan tokenisasi, kata-kata dalam teks dapat diidentifikasi satu per satu atau dalam kombinasi tertentu, yang memungkinkan model untuk memahami struktur kalimat dengan lebih baik.

Tahap selanjutnya adalah *stopword removal*, yang menghapus kata-kata yang umumnya dianggap tidak berkontribusi penting pada pemaknaan, seperti kata sambung dan kata fungsi. Penghapusan kata-kata ini membantu meringkas teks dan fokus pada kata-kata utama yang relevan bagi ringkasan yang dihasilkan. Selain itu, *normalization* dilakukan untuk menormalkan teks dengan menghilangkan karakter khusus, tanda baca, atau simbol yang tidak diperlukan. Hal ini penting untuk memastikan bahwa teks yang dimasukkan ke dalam model memiliki format yang bersih dan seragam.

Untuk menjaga konsistensi antara teks Arab dan terjemahannya dalam bahasa Indonesia, digunakan teknik *alignment* atau penyelarasan. Penyelarasan ini bertujuan memastikan kedua bahasa berada dalam satu konteks yang sama, sehingga model dapat mengenali hubungan antara segmen-segmen yang sama dalam kedua bahasa. Terakhir, *cross-validation* digunakan sebagai metode pengujian dan validasi dengan membagi dataset ke dalam beberapa subset. Dengan cross-validation, model dapat dilatih dan diuji pada data yang beragam, yang meningkatkan kehandalan model dalam menghasilkan ringkasan yang akurat pada teks bilingual.

### 3.5 Teknik Analisis Data

Teknik analisis data yang digunakan adalah evaluasi berbasis metrik **ROUGE** (Recall-Oriented Understudy for Gisting Evaluation) dalam ringkasan yang

disajikan secara otomatis, matrik ROUGE ini merupakan standar untuk mengukur kualitas ringkasan teks otomatis yang pertama kali disajikan oleh (Hovy et al., n.d.). Metrik ini membandingkan ringkasan yang dihasil oleh model dengan satu set ringkasan yang disediakan oleh manusia. Evaluasi ini mencakup beberapa aspek :

- a) **ROUGE-N**: mengukur kesamaan antara ringkasan otomatis dan ringkasan referensi berdasarkan n-grams. Nilainya dihitung menggunakan rasio recall dan precision dari n-grams yang sama :

$$Recall(ROUGE - N) = \frac{\text{jumlah kata yang tumpang tindih pada referensi dan model}}{\text{jumlah total kata dalam referensi}}$$

$$Precision(ROUGE - N) = \frac{\text{jumlah kata yang tumpang tindih pada referensi dan model}}{\text{jumlah total kata dalam ringkasan otomatis}}$$

Rumus F-Measure ROUGE-N merupakan kombinasi dari recall dan precision

$$F - Measure(ROUGE - N) = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall}$$

- b) **ROUGE-L**: Mengukur Longest Common Subsequence (LCS) antara ringkasan otomatis dan referensi. LCS mengidentifikasi urutan kata terpanjang yang muncul secara berurutan dalam kedua ringkasan.

Rumus LCS dapat didefinisikan sebagai berikut: Misalkan  $X = \{x_1, x_2, \dots, x_m\}$  dan  $Y = \{y_1, y_2, \dots, y_m\}$ , maka panjang dari Longest Common Subsequence (LCS) antara X dan Y dinyatakan sebagai  $LCS(X, Y)$ .

$$Recall(ROUGE - L) = \frac{LCS(X, Y)}{\text{Panjang ringkasan referensi}}$$

$$Precision(ROUGE - L) = \frac{LCS(X, Y)}{\text{Panjang ringkasan otomatis}}$$

$$F - Measure(ROUGE - L) = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall}$$

- c) **ROUGE-W**: mengukur kesamaan berdasarkan weighted Longest Common Subsequence (WLCS) dengan memberikan bobot lebih pada kesamaan berurutan yang lebih panjang. Pemberian bobot ini membantu dalam mengevaluasi kesesuaian urutan kata yang lebih panjang.

$$WLCS_{\alpha} = \frac{LCS(X, Y)}{\text{panjang dari ringkasan referensi}^{\alpha} \times \text{panjang dari ringkasan otomatis}^{(1-\alpha)}}$$

Dimana  $\alpha$  adalah parameter yang mengontrol bobot dari panjang kecocokan berurutan.

Analisis kualitatif dalam penelitian ini bertujuan untuk mengevaluasi sejauh mana ringkasan yang dihasilkan mampu mempertahankan makna dan esensi tafsir Al-Qur'an dalam bahasa Arab dan Indonesia. Dengan mengevaluasi kualitas semantik dan kesesuaian konteks, penelitian ini memastikan bahwa ringkasan yang dihasilkan tidak hanya singkat, tetapi juga tetap menyampaikan pemahaman mendalam dari tafsir tersebut. Hal ini sangat penting dalam konteks teks keagamaan yang kompleks, di mana setiap kata dan frasa memiliki makna yang signifikan dan perlu dijaga dengan hati-hati dalam proses peringkasan.

Selain itu, interpretasi hasil evaluasi dilakukan dengan membandingkan kinerja model hybrid summarization berbasis BERT yang dikembangkan dalam penelitian ini dengan pendekatan summarization lainnya. Model ini dievaluasi dengan mengukur kualitas ringkasan dalam aspek seperti ketepatan informasi, konsistensi semantik, dan kemampuan model dalam menangkap konteks teks bilingual. Pendekatan berbasis BERT memungkinkan penggunaan konteks yang lebih luas dalam analisis teks dua bahasa sehingga menghasilkan ringkasan yang lebih relevan dan representatif dibandingkan pendekatan lainnya yang umumnya bersifat monolingual atau kurang kontekstual.

Lebih lanjut, penelitian ini juga mempertimbangkan pentingnya kesesuaian semantik dari ringkasan yang dihasilkan, terutama dalam konteks teks keagamaan yang memerlukan akurasi lebih tinggi dalam menjaga makna yang diinginkan. Teks keagamaan, seperti tafsir Al-Qur'an, memiliki sensitivitas tertentu yang tidak hanya memerlukan pemahaman literal tetapi juga interpretasi kontekstual dan emosional. Oleh karena itu, pendekatan yang digunakan dalam penelitian ini berupaya memastikan bahwa ringkasan yang dihasilkan tidak hanya informatif tetapi juga mencerminkan nuansa keagamaan yang tersirat dalam teks asli.

## DAFTAR PUSTAKA

- Aji Fitra Jaya Institut Perguruan Tinggi Ilmu Al Qur, S. (n.d.). *AL-485 ¶\$IIDAN HADIS SEBAGAI SUMBER HUKUM ISLAM*.
- Alami Merrouni, Z., Frikh, B., & Ouhbi, B. (2023). EXABSUM: a new text summarization approach for generating extractive and abstractive summaries. *Journal of Big Data*, 10(1). <https://doi.org/10.1186/s40537-023-00836-y>
- Alsawi, G., & Taşcı, T. (2024). Extractive Arabic Text Summarization Using PageRank and Word Embedding. *Arabian Journal for Science and Engineering*. <https://doi.org/10.1007/s13369-024-08890-1>
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (n.d.). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://github.com/tensorflow/tensor2tensor>

Hovy, E., Lin, C.-Y., & Zhou, L. (n.d.). *Evaluating DUC 2005 using Basic Elements*.

Liu, Y., & Lapata, M. (n.d.). *Text Summarization with Pretrained Encoders*.  
<https://github.com/>

Nallapati, R., Zhou, B., dos Santos, C., & Xiang, B. (n.d.). *Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond*.