

# PROPOSAL\_FATUR\_ALKHAIR\_1\_. pdf

*by ITentix Plagiarism*

---

**Submission date:** 11-Aug-2025 04:55AM (UTC-0400)

**Submission ID:** 2720196723

**File name:** PROPOSAL\_FATUR\_ALKHAIR\_1\_.pdf (363.55K)

**Word count:** 4728

**Character count:** 33027

**IMPLEMENTASI RETRIEVAL-AUGMENTED GENERATION (RAG)  
MENGGUNAKAN MODEL GPT UNTUK CHATBOT PENCARIAN  
INFORMASI REGULASI KESELAMATAN DAN KESEHATAN KERJA  
(K3)**

**PROPOSAL SKRIPSI**

Diajukan sebagai Salah Satu Syarat untuk mendapatkan  
Gelar Sarjana Komputer (S.Kom) Program Studi Informatika



**FATURROHMAN AL-KHAIR**

**105841107021**

**5  
PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH MAKASSAR**

**2025**

## KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

8

*Assalamu'alaikum Warahmatullahi Wabarakatuh*

Segala puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala atas limpahan rahmat, karunia, dan petunjuk-Nya sehingga Proposal Skripsi berjudul "Implementasi Retrieval-Augmented Generation (RAG) Menggunakan Model GPT untuk Chatbot Pencarian Informasi Regulasi Keselamatan dan Kesehatan Kerja (K3)" dapat diselesaikan dengan baik.

Proposal skripsi ini disusun sebagai salah satu syarat akademik yang harus dipenuhi dalam menyelesaikan program Sarjana di Fakultas Teknik Program Studi Informatika Universitas Muhammadiyah Makassar. Penulis menyadari bahwa karya ini masih jauh dari kesempurnaan, oleh karena itu masukan berupa kritik dan saran yang membangun dari berbagai pihak sangat penulis harapkan demi penyempurnaan proposal skripsi ini.

Dalam proses penyusunan proposal skripsi ini, penulis memperoleh berbagai bentuk bantuan, bimbingan, dan dukungan dari berbagai pihak, baik secara materiil maupun moril. Selama menempuh perkuliahan, penulis juga senantiasa mendapatkan dukungan yang luar biasa. Oleh sebab itu, penulis menyampaikan rasa hormat dan terima kasih yang setulus-tulusnya kepada:

1. Kedua orang tua kami tercinta, penulis mengucapkan terima kasih yang sebesar-besarnya atas segala limpahan kasih sayang, doa yang tiada putus, serta dukungan moral dan materiil yang sangat berarti bagi penulis.
2. Bapak **Ir Muh.Syafaat S.Kuba, S.T,M.T.,PM**, selaku Dekan Fakultas Teknik Universitas Muhammadiyah Makassar yang telah memberikan arahan dan fasilitas akademik.
3. Bapak **Muhyiddin A.M Hayat S.Kom.,M.T.**, selaku Ketua Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar yang telah memberikan arahan dan fasilitas akademik.

1. Ibu Titin Wahyuni, S.Pd.,M.T., selaku Dosen Pembimbing I yang dengan sabar memberikan bimbingan, saran dan koreksi yang konstruktif selama proses penyusunan proposal ini.
2. Bapak Fachrim Irhamna Rachman S.Kom.,M.T., selaku Dosen Pembimbing II yang telah meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, arahan, dan masukan yang sangat berharga dalam penyusunan proposal skripsi ini.
3. Seluruh Bapak/Ibu Dosen dan Tenaga Kependidikan Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Makassar yang telah memberikan ilmu pengetahuan, pengalaman, dan pelayanan akademik.
4. Saudara/saudari kami di Fakultas Teknik. Khususnya Angkatan 2021, dengan rasa persaudaraan yang tinggi banyak membantu, memberikan semangat, dukungan dan kerjasama yang baik dalam menjalani proses perkuliahan dan penyusunan proposal skripsi ini.

Akhir kata, penulis berharap proposal ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan teknologi, khususnya di bidang Keinformatikaan. Aamiin

*“Billahi Fii Sabili Haq Fastabiqul Khaerat”*

*Wassalamualaikum Warahmatullahi Wabarakatuh*

Makassar, Agustus 2025

**Penulis**

## DAFTAR ISI

<sup>27</sup>	<b>KATA PENGANTAR .....</b>	<b>2</b>
	<b>DAFTAR ISI.....</b>	<b>4</b>
	<b>DAFTAR TABEL.....</b>	<b>5</b>
	<b>DAFTAR GAMBAR .....</b>	<b>6</b>
<sup>1</sup>	<b>BAB I PENDAHULUAN .....</b>	<b>7</b>
	A. Latar Belakang .....	7
	B. Rumusan Masalah .....	9
	C. Tujuan Penelitian .....	9
	D. Manfaat Penelitian .....	9
	E. Ruang Lingkup Peneltian .....	10
	F. Sistematika Penulisan.....	11
	<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>12</b>
	A. Landasan Teori.....	12
	B. Penelitian Terkait .....	14
	C. Kerangka Pikir .....	16
	<b>BAB III METODE PENELITIAN .....</b>	<b>18</b>
	A. Tempat dan Waktu Penelitian .....	18
	B. Alat dan Bahan.....	18
	C. Perancangan Sistem .....	19
	D. Teknik Pengujian Sistem.....	25
	E. Teknik Analisis Data.....	25
	<b>DAFTAR PUSTAKA .....</b>	<b>27</b>

## **DAFTAR TABEL**

Table 1. Penelitian Terkait .....	14
Table 2. Waktu Penelitian .....	18

## DAFTAR GAMBAR

Gambar 1. Kerangka Berpikir .....	17
Gambar 2. Alur Penelitian.....	21
Gambar 3. Use case diagram.....	21
Gambar 4. Activity Diagram.....	23

**PENDAHULUAN****A. Latar Belakang**

13

Keselamatan dan Kesehatan Kerja (K3) merupakan aspek krusial yang menjamin perlindungan tenaga kerja dari risiko kecelakaan maupun penyakit akibat aktivitas kerja. Menurut Undang-Undang No. 1 Tahun 1970,<sup>10</sup> keselamatan kerja adalah seluruh upaya untuk menjaga keutuhan jasmani dan rohani pekerja di lingkungan kerja. Regulasi K3 telah banyak diterbitkan oleh pemerintah melalui peraturan menteri, standar operasional prosedur (SOP), hingga dokumen teknis yang berlaku di berbagai sektor industri. Regulasi tersebut bertujuan tidak hanya untuk menurunkan angka kecelakaan kerja, tetapi juga menciptakan lingkungan kerja yang produktif, sehat, dan berkelanjutan.

Namun, meskipun dokumen regulasi K3 tersedia secara resmi, tingkat literasi dan pemahaman pekerja terhadap isi regulasi tersebut masih tergolong rendah. Studi menunjukkan bahwa sebagian besar pekerja tidak memahami secara menyeluruh isi prosedur keselamatan kerja, bahkan setelah mengikuti pelatihan (Hidayah & Kamali Zaman, 2022). Hambatan yang paling sering ditemukan meliputi keterbatasan akses terhadap informasi digital, penggunaan bahasa hukum yang sulit dipahami oleh orang awam, serta kurangnya media interaktif yang dapat menjembatani pemahaman regulasi dengan konteks kerja aktual (Ulzheimer et al., 2021). Di samping itu, pelatihan formal mengenai K3 memerlukan biaya yang cukup besar, membuatnya tidak dapat diakses secara merata oleh semua pihak.<sup>35</sup>

Seiring dengan berkembangnya teknologi kecerdasan buatan (*Artificial Intelligence/AI*), pendekatan baru dalam penyebaran informasi regulasi mulai mendapatkan perhatian, salah satunya melalui pengembangan chatbot berbasis bahasa alami (*Natural Language Processing*). Model *Generative Pre-trained Transformer* (GPT) merupakan salah satu teknologi NLP paling canggih saat ini, dengan kemampuan menjawab pertanyaan dan memberikan informasi

berbasis teks secara instan dan mudah dipahami (Yenduri et al., 2023). GPT <sup>17</sup> telah digunakan dalam berbagai sektor, termasuk pendidikan, kesehatan, dan layanan publik. Namun, GPT secara murni hanya mengandalkan data pelatihan umum, sehingga belum cukup kuat apabila dituntut untuk menjawab pertanyaan yang spesifik terhadap dokumen atau konteks tertentu seperti peraturan K3 (Beheshti et al., 2023).

Untuk mengatasi keterbatasan tersebut, metode *Retrieval-Augmented Generation* (RAG) diperkenalkan sebagai solusi yang menggabungkan pencarian dokumen relevan (*retrieval*) dengan kemampuan generatif model GPT. Dalam konteks ini, chatbot tidak hanya mengandalkan pengetahuan umum, tetapi juga menarik bagian teks dari dokumen K3 aktual sebagai basis jawabannya, sehingga menghasilkan respons yang faktual dan kontekstual. Studi sebelumnya telah membuktikan efektivitas RAG dalam sistem tanya jawab berbasis dokumen seperti fatwa, peraturan sekolah, dan modul pembelajaran. Namun, sejauh penelusuran penulis, belum ditemukan penelitian yang secara khusus mengimplementasikan RAG untuk chatbot pencarian <sup>43</sup> regulasi K3 di Indonesia—ini merupakan gap penting yang ingin dijawab dalam penelitian ini.

Penelitian ini bertujuan untuk mengembangkan sistem *chatbot* berbasis <sup>37</sup> *Generative Pre-trained Transformer* (GPT) yang diintegrasikan dengan pendekatan *Retrieval-Augmented Generation* (RAG) guna memfasilitasi pencarian dan pemahaman informasi regulasi Keselamatan dan Kesehatan Kerja (K3) secara otomatis. Sistem ini dirancang sebagai solusi alternatif yang efisien, interaktif, dan inklusif untuk mendukung peningkatan literasi K3 bagi pekerja, petugas K3, maupun pelaku industri. Melalui pemanfaatan teknologi kecerdasan buatan yang mampu menghadirkan informasi secara cepat, akurat, dan kontekstual, <sup>41</sup> penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap peningkatan keselamatan kerja serta memperluas akses terhadap regulasi K3 secara merata dan efektif. Selain itu, hasil penelitian ini berpotensi menjadi model penerapan teknologi AI pada sektor regulasi dan kepatuhan (*compliance*), yang dapat direplikasi pada berbagai domain

pengetahuan lainnya, sehingga memberikan dampak positif bagi pengelolaan informasi di berbagai bidang industri.

**21**  
**B. Rumusan Masalah**

Berdasarkan latar belakang diatas maka dapat dirumuskan masalah sebagai berikut:

1. Bagaimana mengintegrasikan berbagai dokumen regulasi K3 dalam satu sistem pencarian?
2. Apa solusi untuk membaca gambar atau tabel dalam pdf yang tidak bisa diekstraksi sebagai teks?

**20**  
**C. Tujuan Penelitian**

Berdasarkan rumusan masalah tersebut, adapaun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan sistem pencarian yang mampu mengintegrasikan berbagai dokumen regulasi K3 (Keselamatan dan Kesehatan Kerja) dalam satu platform terpadu hingga memudahkan pengguna dalam menemukan informasi yang relevan.
2. Merumuskan solusi teknis untuk mengekstraksi informasi dari elemen non-teks seperti gambar atau tabel dalam dokumen PDF yang tidak dapat dibaca secara langsung sebagai teks.

**D. Manfaat Penelitian**

Penelitian ini berjudulul “Implementasi Retrieval-Augmented Generation (RAG) menggunakan model GPT untuk chatbot pencarian <sup>16</sup> informasi regulasi Keselamatan dan Kesehatan Kerja (K3) yang diharapkan memberikan manfaat secara teoritis dan secara praktis, yaitu:

1. Secara Teoritis:

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pencarian informasi berbasis kecerdasan buatan yang bertujuan mempermudah akses terhadap regulasi Keselamatan dan Kesehatan Kerja secara cepat dan relevan.

<sup>30</sup>

## 2. Secara Praktis

### a. Manfaat bagi peneliti

Penelitian ini memberikan peneliti pengalaman mendalam mengenai implementasi model AI, serta penerapannya dalam dunia nyata. Selain itu, penelitian ini juga melatih peneliti dalam proses perancangan sistem, pengolahan data, serta evaluasi sistem cerdas berbasis teks.

### b. Manfaat bagi Sosial dan Nasional

Penelitian ini berpotensi mendukung pembangunan nasional melalui peningkatan pemahaman dan akses terhadap <sup>32</sup> regulasi keselamatan kerja, yang pada akhirnya dapat mengurangi kecelakaan kerja dan meningkatkan kesejahteraan tenaga kerja di Indonesia.

## E. Ruang Lingkup Penelitian

<sup>2</sup>

Berdasarkan rumusan masalah, adapun batasan pada penelitian ini sebagai berikut:

1. Ruang Lingkup Regulasi K3: Regulasi yang digunakan hanya mencakup <sup>39</sup> peraturan nasional <sup>39</sup> di Indonesia, seperti Undang-Undang, Peraturan Pemerintah, Peraturan Menteri Ketenagakerjaan, Keputusan Menteri, dan <sup>10</sup> SNI yang berkaitan dengan keselamatan dan kesehatan kerja (K3).
2. Bahasa Dokumen: Seluruh dokumen yang digunakan dalam proses retrieval dan pemrosesan berbahasa Indonesia; dokumen berbahasa asing tidak disertakan dalam penelitian ini.
3. Jenis dan Format Dokumen: Penelitian hanya memproses dokumen digital berformat PDF dan DOCX, dokumen hasil pemindaian (scan) tanpa teknologi OCR tidak digunakan.
4. Retrieval-Augmented Generation (RAG): Penelitian berfokus pada penerapan model RAG dengan GPT sebagai komponen generatif; pendekatan lain seperti rule-based, keyword matching, atau fine-tuning murni tidak dieksplorasi.

5. Sistem Chatbot: Chatbot dirancang sebagai prototipe berbasis teks; tidak mendukung antarmuka berbasis suara atau integrasi dengan platform komunikasi seperti WhatsApp atau Telegram.
6. Uji Coba Sistem: Pengujian dilakukan secara terbatas pada kelompok kecil, seperti mahasiswa, akademisi, dan praktisi K3; tidak diterapkan secara luas dalam skala industri atau organisasi.

## **F. Sistematika Penulisan**

Secara keseluruhan, struktur penulisan ini dibagi ke beberapa bab utama yang masing-masing memuat aspek-aspek penting sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini menjelaskan uraian mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, ruang lingkup dan sistematika penulisan.

### **BAB II TINJAUAN PUSTAKA**

Bab ini berisi landasan teoritis yang relevan dengan topik penelitian. Teori-teori dan hasil studi terdahulu yang mendukung pelaksanaan penelitian.

### **BAB III METODE PENELITIAN**

Bab ini membahas secara rinci metode yang digunakan dalam penelitian, mencakup pendekatan, tahap pelaksanaan, serta perangkat dan sumber daya yang dimanfaatkan dalam proses perancangan dan pembangunan sistem.

11  
**BAB II**

**TINJAUAN PUSTAKA**

**A. Landasan Teori**

**1. Keselamatan dan Kesehatan Kerja (K3)**

Keselamatan dan Kesehatan Kerja (K3) adalah seperangkat upaya sistematis yang dilakukan untuk menjamin kondisi kerja yang aman sehat, dan terhindar dari risiko kecelakaan maupun penyakit akibat kerja. Undang-Undang No. 1 Tahun 1970 mendefinisikan keselamatan kerja sebagai segala kegiatan untuk menjamin keutuhan dan kesempurnaan jasmani dan rohani tenaga kerja melalui perlindungan kerja di lingkungan kerja. (Waisapi, 2022) menjelaskan bahwasanya tujuan dari K3 adalah mencegah kecelakaan kerja, meningkatkan produktivitas, serta menciptakan suasana kerja yang aman dan nyaman.

Menurut (Djoko Subagijo) regulasi K3 mencakup berbagai ketentuan teknis dan hukum yang kompleks, sehingga pemahaman terhadapnya membutuhkan sistem pendukung berbasis teknologi yang mampu menjembatani antara dokumen regulatif dengan kebutuhan pengguna. Maka dari itu, pendekatan digital yang interaktif seperti chatbot menjadi penting untuk mempermudah akses informasi ini.

**2. Sistem Informasi**

Sistem informasi adalah kumpulan elemen yang saling berkaitan untuk mengumpulkan, mengelola, menyimpan, dan mendistribusikan informasi guna mendukung pengambilan keputusan dan pengendalian dalam organisasi. Sistem informasi terdiri dari input, proses, output, dan feedback, yang semuanya dikoordinasikan oleh perangkat keras, perangkat lunak, manusia, serta data (Vargas Saputra et al., 2025)

Dalam konteks penelitian ini, sistem informasi yang dibangun berfungsi untuk memfasilitasi interaksi antara pengguna dan data regulasi K3 melalui antarmuka chatbot yang cerdas dan adaptif.

### **3. Kecerdasan Buatan (Artificial Intelligence)**

Kecerdasan Buatan (AI) adalah bidang ilmu komputer yang berfokus pada pembuatan mesin yang dapat meniru kecerdasan manusia. AI memungkinkan komputer untuk melakukan tugas-tugas kompleks seperti mengenali suara, memahami bahasa alami, membuat prediksi, dan belajar dari data (Strohmeier, 2022).

AI menjadi fondasi dalam pengembangan sistem chatbot yang digunakan dalam penelitian ini. Salah satu bentuk AI yang digunakan adalah pemrosesan bahasa alami (Natural Language Processing/NLP), yang memungkinkan sistem memahami dan menjawab pertanyaan dalam bahasa manusia (Strohmeier, 2022)

### **4. Chatbot**

Chatbot adalah sistem komputer yang dirancang untuk berinteraksi dengan manusia melalui antarmuka percakapan berbasis teks atau suara. Chatbot dapat bersifat rule-based (berbasis aturan) atau AI-based (berbasis kecerdasan buatan). Chatbot AI mampu mempelajari konteks dan memberikan jawaban yang lebih fleksibel dan personal (Sharifi et al., 2021)

Dalam penelitian ini, chatbot dikembangkan untuk memudahkan pencarian dan pemahaman regulasi K3, menjadikannya sebagai sarana bantu edukatif dan informatif yang dapat diakses kapan saja dan oleh siapa saja.

### **5. Generative Pre-trained Transformer (GPT)**

GPT adalah model bahasa berbasis deep learning yang dilatih pada jumlah besar data teks untuk memahami dan menghasilkan bahasa alami. GPT bekerja berdasarkan arsitektur transformer yang memungkinkan pemrosesan urutan teks dalam konteks yang lebih luas (Brown et al., 2020)

Menurut (Brown et al., 2020) model GPT memiliki kemampuan zero-shot, one-shot, dan few-shot learning, sehingga sangat efektif dalam menjawab pertanyaan berbasis teks tanpa pelatihan ulang. GPT digunakan

dalam penelitian ini sebagai model inti dalam memberikan jawaban dari chatbot secara natural dan sesuai konteks.

17

#### 6. Retrieval-Augmented Generation (RAG)

RAG adalah pendekatan dalam pengembangan sistem tanya-jawab (question answering system) yang menggabungkan proses pencarian informasi (retrieval) dari dokumen eksternal dengan kemampuan generatif model bahasa seperti GPT. Dalam arsitektur RAG, dua komponen utama digunakan: retriever untuk mencari dokumen relevan, dan generator untuk menyusun jawaban berdasarkan dokumen tersebut (Lewis et al., 2021a)

4

#### 7. Unified Modeling Language (UML)

UML adalah bahasa pemodelan standar yang digunakan untuk merancang dan mendokumentasikan sistem perangkat lunak. UML menyediakan berbagai jenis diagram seperti use case, class, sequence, dan activity diagram yang membantu dalam perancangan sistem secara visual dan struktural (Booch et al., 2005(M Teguh, 2018)).

Dalam penelitian ini, UML digunakan untuk memodelkan alur kerja sistem chatbot, interaksi antar komponen, serta hubungan antara pengguna, sistem, dan basis data regulasi K3.

1

### B. Penelitian Terkait

Peneliti memperoleh banyak inspirasi dan referensi untuk penyusunan proposal skripsi ini dari penelitian sebelumnya, terkait dengan latar belakang masalah pada proposal ini. Penelitian sebelumnya yang terkait antara lain:

1  
Table 1. Penelitian Terkait

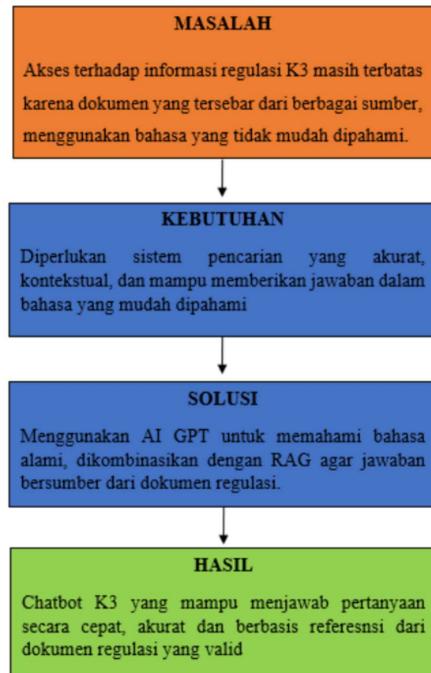
No	Nama/Tahun	Judul	Metode	Hasil
1	(Ahadi et al., 2025)	Retrieval-Augmented Generation in a Web-Based Question Answering	RAG + GPT-3, web-based QA	Mengurangi halusinasi LLM, meningkatkan akurasi jawaban berbasis

		System for Fiqh Books <sup>23</sup>	dokumen referensi.
2	(Gayathri Devi. et al., 2024)	MineBot: Chatbot to Respond to Text Queries Pertaining to Acts, Rules, and Regulations in Mining	LangChain + GPT-3.5-turbo + RAG, integrasi Wikipedia & Google API
3	(Lewis et al., 2021b)	Retrieval-Augmented Generation for Knowledge-Intensive NLP Task	RAG yang menggabungkan <i>Dense Passage Retriever</i> dengan model generatif BART Peningkatan akurasi dan relevansi jawaban pada tugas <i>knowledge-intensive question answering</i> .
4	(Silvanie & Subekti, 2022) <sup>19</sup>	Applikasi Chatbot Untuk Faq Akademik Di Ibi-K57 Dengan Lstm Dan Penyematan Kata	Long Short-Term Memory (LSTM) + Word Embedding (CBOW) + Flask (server) + Kotlin (client) Akurasi tertinggi 99,20% dengan optimizer Adam/RMSprop; chatbot mampu menjawab pertanyaan FAQ akademik berbasis Android

5	(Ranjit et al., 2023)	<p><b>9</b></p> <p>Retrieval-Augmented Chest X-ray Report Generation using a Pretrained Contrastive Language-Image Model</p>	<p>Retrieval-Augmented Generation (RAG) + CLIP (Contrastive Pretraining)</p> <p>Language–Image Pretraining)</p> <p>untuk retrieval + GPT-2</p>	<p>dengan koneksi server</p> <p>meningkatkan kualitas laporan X-ray dengan memanfaatkan retrieval data historis yang relevan, menghasilkan deskripsi lebih akurat dan kaya informasi dibanding model generatif murni</p>
---	--------------------------	--	--	--

### 1 C. Kerangka Pikir

Kerangka pikir pada penelitian ini menunjukkan bagaimana membangun sebuah sistem *chatbot* yang mampu membantu pekerja, petugas K3, dan pelaku industri dalam mengakses serta memahami regulasi Keselamatan dan Kesehatan Kerja (K3) dengan lebih mudah. Alasan yang melatarbelakangi pengembangan sistem ini adalah masih rendahnya pemahaman terhadap regulasi K3 di kalangan masyarakat pekerja, yang disebabkan oleh keterbatasan sumber informasi yang relevan, bahasa regulasi yang sulit dipahami, serta minimnya media interaktif yang dapat menyajikan informasi tersebut secara cepat, akurat, dan kontekstual.



1  
Gambar 1. Kerangka Berpikir

## **BAB III**

### **METODE PENELITIAN**

#### **A. Tempat dan Waktu Penelitian**

##### **1. Tempat Penelitian**

Tempat lokasi penelitian yang dipilih peneliti adalah di sebuah kantor Keselamatan dan Kesehatan Kerja (K3) di Kampus Universitas Muhammadiyah Makassar.

##### **2. Waktu Penelitian**

Pelaksanaan penelitian ini berlangsung pada bulan Juni hingga Agustus 2025, dimulai dari tahap pengumpulan data hingga penyusunan laporan akhir.

Table 2. Waktu Penelitian

NO	KEGIATAN	BULAN							
		JUNI		JULI			AGUSTUS		
		3	4	1	2	3	4	1	2
1	Perumusan topik penelitian								
2	Perencanaan penelitian								
3	Pelaksanaan penelitian								
	a. Pembuatan program								
	b. Penelitian uji program								
4	Penulisan laporan								

#### **B. Alat dan Bahan**

Alat penelitian berupa laptop yang akan digunakan peneliti untuk merancang sistem chatbot untuk pencarian informasi regulasi Keselamatan dan Kesehatan Kerja (K3) berbasis RAG. Adapun perangkat keras dan perangkat lunak yang digunakan adalah sebagai berikut:

1. Perangkat keras (Pengembang)
  - a. Processor Intel Core-I5 (2 core, 8 thread)

- b. Besar Memory Ram 24GB
  - c. Kapasitas SSD 119GB
2. Perangkat Lunak
- a. Windows Home
  - b. Text editor Visual Studio Code
  - c. Python sebagai bahasa programming
  - d. FastAPI dan Uvicorn sebagai framerowk dan server backend
  - e. HTML, CSS dan Javascript sebagai User Interface (frontend)
  - f. MongoDB dan ChromaDB sebagai sistem manajemen basis data (Vector dan dokumentasi)
  - g. Google Chrome sebagai browser.

<sup>36</sup> Data yang digunakan dalam penelitian ini berupa dokumen-dokumen regulasi K3, yang diperoleh dari berbagai sumber resmi. Dokumen tersebut diunggah ke dalam sistem dan diproses untuk menjadi sumber pengetahuan yang digunakan dalam menjawab pertanyaan pengguna secara kontekstual melalui chatbot.

### <sup>1</sup> C. Perancangan Sistem

Perancangan sistem sangat penting dalam pembangunan suatu sistem karena menguraikan bagaimana suatu sistem dibangun dari tahap perencanaan hingga tahap pembuatan fungsi-fungsi yang diperlukan untuk pengoperasian sistem. Tujuan dari perancangan sistem adalah untuk menentukan apakah sistem yang akan dikembangkan akan menghasilkan hasil yang diinginkan.

Alur jalannya penelitian pembuatan *chatbot* informasi regulasi Keselamatan dan Kesehatan Kerja (K3) seperti pada gambar 2 terbagi menjadi 7 (tujuh) blok penelitian, antara lain:

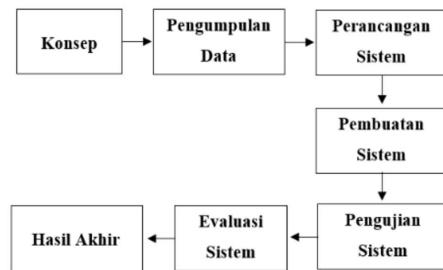
1. Konsep, tahapan ini dimulai dengan merumuskan ide dasar dari sistem chatbot yang akan dikembangkan. Pada tahap ini ditentukan permasalahan utama, batasan sistem, dan dirancang kebutuhan fungsional. Konsep chatbot diarahkan pada integrasi dan teknologi Artificial Intelligence (GPT) dan pencarian semantik (RAG) untuk

meningkatkan aksesibilitas regulasi Keselamatan dan Kesehatan Kerja (K3) dari dokumen.

2. Pengumpulan data, setelah konsep sistem telah ditetapkan, tahapan selanjutnya adalah pengumpulan data berupa dokumen regulasi K3 dalam format PDF. Selain itu, dilakukan studi literatur terhadap teknologi pendukung seperti GPT, LangChain, OpenAI Embedding, dan ChromaDB yang akan digunakan dalam proses pengolahan data.
3. Perancangan Sistem, Pada blok ini dilakukan perancangan teknis sistem, yang meliputi alur kerja sistem (workflow), perancangan modul utama, desain basis data, dan pemilihan algoritma yang akan digunakan. Komponen yang dirancang meliputi proses ekstraksi teks dari PDF, pembagian teks menjadi chunk, pembuatan embedding vektor menggunakan API OpenAI, penyimpanan embedding di MongoDB, hingga pengembangan sistem retrieval dan generation menggunakan LangChain dan GPT.
4. Pembuatan sistem, Tahap ini merupakan proses implementasi dari rancangan sistem sebelumnya. Pembuatan sistem dilakukan menggunakan bahasa pemrograman Python, dengan integrasi beberapa library seperti PyMuPDF, LangChain, ChromaDB, OpenAI, dan MongoDB. Pada tahap ini dibangun pipeline RAG yang mencakup ekstraksi teks, embedding, penyimpanan data, pencarian vektor, dan generasi jawaban oleh model GPT berdasarkan dokumen sumber.
5. Pengujian sistem, Setelah sistem selesai diimplementasikan, dilakukan pengujian terhadap berbagai jenis pertanyaan, baik pertanyaan langsung maupun pertanyaan lanjutan (multi-turn conversation). Pengujian mencakup keakuratan jawaban, relevansi konteks, serta responsivitas chatbot dalam menjawab berdasarkan isi dokumen regulasi K3. Hasil pengujian dianalisis secara deskriptif untuk menilai performa sistem.
6. Evaluasi dan penyimpulan, Setelah sistem selesai diimplementasikan, dilakukan pengujian terhadap berbagai jenis pertanyaan, baik pertanyaan langsung maupun pertanyaan lanjutan (multi-turn

conversation). Pengujian mencakup keakuratan jawaban, relevansi konteks, serta responsivitas chatbot dalam menjawab berdasarkan isi dokumen regulasi K3.

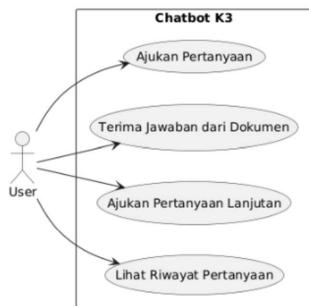
7. Hasil akhir, setelah melalui proses pengujian dan perbaikan,<sup>26</sup> chatbot sudah dapat diimplementasikan. Jalannya penelitian dapat diliat pada gambar 2 di bawah ini:



Gambar 2. Alur Penelitian

Alur Adapun perancangan web *chatbot* regulasi Keselamatan dan Kesehatan Kerja digambarkan sebagai berikut:

1. *Use case diagram*



Gambar 3. Use case diagram

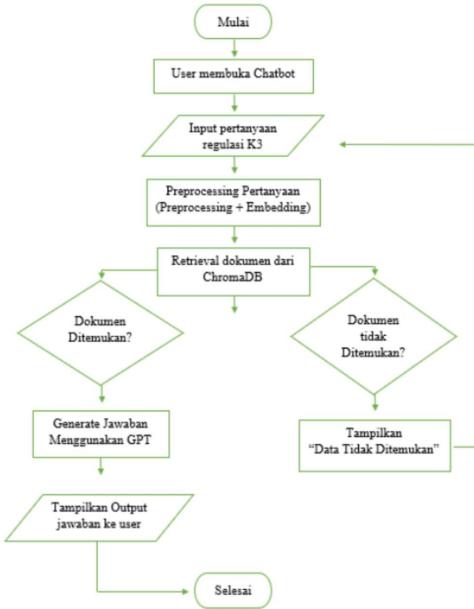
Pada tahap perancangan use case diagram, diagram ini berfungsi untuk menggambarkan interaksi antara pengguna (user) dengan sistem

chatbot regulasi Keselamatan dan Kesehatan Kerja (K3). Diagram ini menunjukkan proses awal hingga akhir penggunaan sistem, serta fitur-fitur utama yang tersedia bagi pengguna. Dalam sistem ini hanya terdapat satu jenis akun yaitu akun pengguna (user), karena sistem bersifat publik dan ditujukan untuk akses informasi K3 secara mudah.

- a. Ajukan pertanyaan, pada tahap awal, pengguna dapat langsung mengakses halaman antarnuha chatbot dan mengajukan pertanyaan seputar regulasi K3. Pertanyaan ini dapat berupa definisi, prosedur, atau ketentuan hukum terkait keselamatan kerja, seperti “Apa itu APD?”, atau “Apa isi pasal tentang pemadam kebakaran?”. Sistem kemudian menerima input ini dan memprosesnya melalui tahap *retrieval* dan *generation* menggunakan model RAG.
- b. Terima jawaban dari Dokumen, setelah sistem menerima pertanyaan, modul retrieval akan mencari bagian dokumen PDF yang relevan menggunakan pencarian berbasis vektor. Kemudian hasil pencarian tersebut akan dijadikan konteks oleh model GPT untuk menghasilkan jawaban yang informatif dan mudah dipahami. Jawaban ini ditampilkan secara instan kepada pengguna di tampilan chatbot.
- c. Ajukan pertanyaan lanjutan, Jika pengguna merasa jawabannya belum cukup lengkap, mereka dapat mengajukan pertanyaan lanjutan tanpa perlu mengulang konteks. Fitur ini memungkinkan percakapan berkelanjutan (*multi-turn conversation*), seperti “lanjutkan” atau “bagaimana dengan pengecualian pasal tersebut?”, dan sistem tetap mempertahankan konteks pertanyaan sebelumnya.
- d. Lihat riwayat pertanyaan, Pengguna juga dapat melihat kembali riwayat percakapan mereka dalam sesi sebelumnya. Fitur ini berguna untuk merevisi, menyimpan, atau menyalin informasi yang sudah diterima tanpa harus bertanya ulang. Riwayat disimpan secara lokal selama sesi berlangsung, atau secara opsional dapat diarsipkan.

## 1 2. Activity diagram

Activity Diagram adalah representasi grafis dari seluruh tahapan alur kerja yang mengandung aktivitas, pilihan tindakan, perulangan dan hasil dari aktivitas tersebut. Diagram ini dapat digunakan untuk menjelaskan proses dan alur kerja operasional secara langkah demi langkah dari komponen suatu sistem. Adapun activity diagram dari sistem ini adalah sebagai berikut:



Gambar 4. Activity Diagram

a. User membuka chatbot

Tahapan pertama adalah ketika pengguna (user) mengakses antarmuka chatbot yang sudah dibangun berbasis web. Pada tahap ini, pengguna siap untuk melakukan interaksi awal dengan sistem.

- b. User menginput pertanyaan  
Pengguna kemudian mengetikkan pertanyaan seputar regulasi K3 yang ingin diketahui, misalnya “Apa isi pasal tentang APD?” atau “Bagaimana prosedur evakuasi darurat?”.
- c. Preprocessing pertanyaan (embedding)  
Setelah Pertanyaan dari pengguna kemudian diproses melalui tahap preprocessing. Pada tahap ini, pertanyaan diubah menjadi representasi vektor (embedding) agar bisa dibandingkan secara semantik dengan data yang ada di basis data ChromaDB.
- d. Retrieval dokumen dari ChromaDB  
Setelah embedding pertanyaan dihasilkan, sistem melakukan pencarian ke dalam basis data vektor (ChromaDB) untuk menemukan dokumen yang memiliki kemiripan semantik tinggi. Proses ini dikenal dengan retrieval..
- e. Percabangan: Apakah dokumen ditemukan?  
Sistem akan mengecek apakah ada dokumen atau bagian dokumen yang relevan dengan pertanyaan pengguna. Jika **dokumen ditemukan**, maka proses akan dilanjutkan ke tahap generasi jawaban. Jika **tidak ditemukan**, sistem akan menampilkan informasi bahwa “Data Tidak Ditemukan”.
- f. Generative jawaban menggunakan GPT  
Apabila dokumen berhasil ditemukan, potongan teks yang relevan akan dikirimkan ke model GPT untuk menghasilkan jawaban yang utuh dan mudah dipahami. Model GPT akan melakukan proses reasoning dengan konteks yang diberikan.
- g. Tampilkan output jawaban ke user  
Jawaban yang telah dihasilkan oleh GPT akan ditampilkan langsung di antarmuka pengguna. Sistem juga mempertahankan konteks untuk memungkinkan pertanyaan lanjutan (multi-turn conversation).Selesai

Setelah jawaban diterima oleh pengguna, proses sesi tanya-jawab dapat dihentikan, atau pengguna dapat memilih untuk melanjutkan percakapan dengan pertanyaan berikutnya.

#### **D. Teknik Pengujian Sistem**

Teknik Pengujian Sistem dilakukan dengan metode black box testing. Uji coba yang dilakukan bersifat mandiri dan diuji coba langsung dengan memperhatikan rancangan yang dibuat.

Pengujian blackbox (blackbox testing) adalah salah satu metode pengujian perangkat lunak yang berfokus pada sisi fungsionalitas, khususnya pada input dan output (apakah sudah sesuai dengan apa yang diharapkan atau belum).

Black box testing adalah pengujian yang dilakukan untuk mengamati hasil input dan output dari perangkat lunak tanpa mengetahui struktur kode dari perangkat lunak. Pengujian ini dilakukan di akhir pembuatan perangkat lunak untuk mengetahui apakah perangkat lunak dapat berfungsi dengan baik.

#### **E. Teknik Analisis Data**

Setelah semua data yang diperlukan dalam pengembangan sistem chatbot regulasi Keselamatan dan Kesehatan Kerja (K3) berhasil dikumpulkan, proses analisis data menjadi salah satu langkah penting dalam menjamin kualitas dan akurasi sistem yang dikembangkan. Analisis data berperan dalam mengevaluasi kinerja sistem serta memvalidasi apakah sistem mampu memenuhi kebutuhan pengguna secara optimal. Ketepatan dalam pemilihan metode analisis sangat menentukan kualitas kesimpulan yang dihasilkan, sehingga dapat berdampak besar pada kebermanfaatan dan penerapan hasil penelitian ini dalam dunia kerja nyata, khususnya pada bidang keselamatan kerja.

Dalam penelitian ini, pendekatan analisis data yang digunakan adalah analisis kualitatif deskriptif. Hal ini disebabkan oleh sifat data yang dianalisis berasal dari hasil pengujian sistem berupa interaksi teks (pertanyaan dan jawaban) serta observasi terhadap performa chatbot dalam menjawab pertanyaan berbasis dokumen regulasi. Analisis dilakukan dengan cara

membandingkan antara hasil jawaban chatbot dengan isi dokumen K3 asli untuk menilai akurasi, kelengkapan jawaban, dan koherensi konteks, terutama dalam pertanyaan bertingkat (multi-turn).

Proses pengujian dilakukan melalui serangkaian skenario pertanyaan yang meliputi: pertanyaan definisi, prosedural, hingga pertanyaan lanjutan kontekstual seperti “lanjutkan” atau “jelaskan lebih dalam.” Log dari sesi tanya-jawab tersebut disimpan dalam bentuk file json sebagai bagian dari dokumentasi, yang kemudian dianalisis secara manual untuk menilai apakah sistem mampu memberikan jawaban relevan berdasarkan isi dokumen sumber dan mengikuti konteks percakapan sebelumnya.

Selain itu, dilakukan juga evaluasi fungsional sistem dengan cara menguji keandalan proses ekstraksi teks, pembuatan embedding, penyimpanan data ke database, serta kemampuan sistem dalam menangani error seperti file PDF kosong atau pertanyaan tanpa hasil pencarian. Hasil evaluasi ini digunakan untuk mengidentifikasi kelemahan sistem dan memberikan dasar dalam menyusun rekomendasi pengembangan selanjutnya.

Dengan demikian, teknik analisis data pada penelitian ini difokuskan untuk mengevaluasi sistem secara menyeluruh dari segi fungsi, konteks pemahaman, serta keakuratan isi jawaban, sehingga dapat dipastikan bahwa sistem yang dikembangkan mampu menjadi solusi efektif dalam membantu pekerja atau petugas K3 dalam mengakses regulasi yang relevan secara efisien.

## **DAFTAR PUSTAKA**

- Ahadi, R., Safaat Harahap, N., Fikry, M., & Kurnia, F. (2025). Retrieval-Augmented Generation in a Web-Based Question Answering System for Fiqh Books. *Journal of Artificial Intelligence and Software Engineering*, 5(2), 626–635. <https://doi.org/10.30811/jaise.v5i2.7005>
- Beheshti, A., Yang, J., Sheng, Q. Z., Benatallah, B., Casati, F., Dustdar, S., Nezhad, H. R. M., Zhang, X., & Xue, S. (2023). *ProcessGPT: Transforming Business Process Management with Generative Artificial Intelligence*. <http://arxiv.org/abs/2306.01771>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language Models are Few-Shot Learners*. <http://arxiv.org/abs/2005.14165>
- Hidayah, P., & Kamali Zaman, M. (2022). Science Midwifery Implementation of Occupational Safety and Health (K3) Inspection as a Work Accident Prevention Effort in Palm Oil Factory, Kampar Regency, Riau Province A R T I C L E I N F O ABSTRACT. In *Science Midwifery* (Vol. 10, Issue 3). Online. [www.midwifery.iocspublisher.org](http://www.midwifery.iocspublisher.org)
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2021a). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. <http://arxiv.org/abs/2005.11401>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2021b). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. <http://arxiv.org/abs/2005.11401>

- Gayathri Devi., K, Siranjeevi., S, Rupeshwar., & T, Oviya. (2024). Minebot: Chatbot to Respond to Text Queries Pertaining to Various Acts, Rules, And Regulations Applicable to Mining Industries. *International Journal of Research Publication and Reviews*, 5(3), 4317–4325.  
<https://doi.org/10.55248/gengpi.5.0324.07107>
- M Teguh, P. (2018). *Unified Modeling Language (UML) Model Untuk Pengembangan Sistem Informasi Akademik Berbasis Web*.
- Ranjit, M., Ganapathy, G., Manuel, R., & Ganu, T. (2023). *Retrieval Augmented Chest X-Ray Report Generation using OpenAI GPT models*.  
<http://arxiv.org/abs/2305.03660>
- Sharifi, A., Zibaei, A., & Rezaei, M. (2021). A deep learning based hazardous materials (HAZMAT) sign detection robot with restricted computational resources. *Machine Learning with Applications*, 6, 100104.  
<https://doi.org/10.1016/j.mlwa.2021.100104>
- Silvanie, A., & Subekti, R. (2022). APLIKASI CHATBOT UNTUK FAQ AKADEMIK DI IBI-K57 DENGAN LSTM DAN PENYEMATAN KATA. *Jurnal Informatika Dan Komputer) Akreditasi KEMENRISTEKDIKTI*, 5(1).  
<https://doi.org/10.33387/jiko>
- Strohmeier, Stefan. (2022). *Handbook of research on artificial intelligence in human resource management*. Edward Elgar Publishing.
- Ulzheimer, L., Kanzinger, A., Ziegler, A., Martin, B., Zender, J., Römhild, A., & Leyhe, C. (2021). Barriers in times of digital teaching and learning – a german case study: Challenges and recommendations for action. *Journal of Interactive Media in Education*, 2021(1). <https://doi.org/10.5334/jime.638>
- Vargas Saputra, T., Harlina, T., Studi, P., Informatika, M., Pgri Banyuwangi, S., & Korespondensi, P. (2025). *Penerapan Sistem Informasi Perpustakaan Berbasis Web Pada STIKOM PGRI Banyuwangi*. 6(1).  
<https://doi.org/10.31284/j.kernel.2025.v6i1.7804>

Yenduri, G., M, R., G, C. S., Y, S., Srivastava, G., Maddikunta, P. K. R., G, D. R., Jhaveri, R. H., B, P., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2023). *Generative Pre-trained Transformer: A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions*. <http://arxiv.org/abs/2305.10435>



## PRIMARY SOURCES

1	dilibadadmin.unismuh.ac.id Internet Source	15%
2	123dok.com Internet Source	1 %
3	Submitted to Universitas Budi Luhur Student Paper	1 %
4	I Komang Sugiarktha, Dw Ayu Agung Indra Swari, Dedy Panji Agustino, Tria Hikmah Fratiwi. "Pengembangan Aplikasi Mobile E- Pasraman Hindu untuk Optimalisasi Pengelolaan Data Akademik", remik, 2025 Publication	1 %
5	Submitted to Universitas Muhammadiyah Makassar Student Paper	1 %
6	repository.uinjkt.ac.id Internet Source	1 %
7	Submitted to Keimyung University Student Paper	1 %
8	Submitted to Universitas Muslim Indonesia Student Paper	1 %
9	Agrawal, Tejas Prakash. "RadGraph: Integrating Fine-Grained and Global Features Using GCNs and LLMs", State University of New York at Buffalo, 2024 Publication	1 %

10	docplayer.info Internet Source	<1 %
11	perpustakaan.pancabudi.ac.id Internet Source	<1 %
12	Submitted to Universitas Maritim Raja Ali Haji Student Paper	<1 %
13	Submitted to Universitas Sebelas Maret Student Paper	<1 %
14	id.123dok.com Internet Source	<1 %
15	Submitted to Universitas Musamus Merauke Student Paper	<1 %
16	docobook.com Internet Source	<1 %
17	penerbit.gunadarma.ac.id Internet Source	<1 %
18	repository.iainpalopo.ac.id Internet Source	<1 %
19	ejournal.unkhair.ac.id Internet Source	<1 %
20	repository.iainbengkulu.ac.id Internet Source	<1 %
21	repository.uinfasbengkulu.ac.id Internet Source	<1 %
22	repository.ukwms.ac.id Internet Source	<1 %
23	www.ijrpr.com Internet Source	<1 %

Submitted to STT PLN

24

&lt;1 %

25

**Submitted to Universitas Pendidikan Ganesha**

&lt;1 %

Student Paper

26

**ejournal.unib.ac.id**

&lt;1 %

Internet Source

27

**pdfcoffee.com**

&lt;1 %

Internet Source

28

**pkm.tunasbangsa.ac.id**

&lt;1 %

Internet Source

29

**repository.upnvj.ac.id**

&lt;1 %

Internet Source

30

**digilib.uinkhas.ac.id**

&lt;1 %

Internet Source

31

**Submitted to iGroup**

&lt;1 %

Student Paper

32

**kappppppp.blogspot.com**

&lt;1 %

Internet Source

33

**pure.psu.edu**

&lt;1 %

Internet Source

34

**www.kajianpustaka.com**

&lt;1 %

Internet Source

35

**www.pinterpolitik.com**

&lt;1 %

Internet Source

36

**www.scilit.net**

&lt;1 %

Internet Source

37

**Muhammad Yusuf, Asyrofi Anam,  
Muhammad Rifki Raihan. "SISTEM PAKAR  
PENENTUAN KOMPOSISI SKINCARE  
BERDASARKAN MASALAH KULIT WAJAH**

&lt;1 %

MENGGUNAKAN NATURAL LANGUAGE  
PROCESSING (NLP)", Jurnal Informatika dan  
Rekayasa Elektronik, 2025

Publication

38	Submitted to Universitas Brawijaya Student Paper	<1 %
39	digilib.uns.ac.id Internet Source	<1 %
40	jer.or.id Internet Source	<1 %
41	journal.ipm2kpe.or.id Internet Source	<1 %
42	penerbitbukudeepublish.com Internet Source	<1 %
43	repository.unmuhjember.ac.id Internet Source	<1 %
44	www.scribd.com Internet Source	<1 %
45	www.springerprofessional.de Internet Source	<1 %
46	Yeni Risyani, Susi Japit, Conrad Bombongan, Tanda Selamat, Yuliana Yuliana. "Pelatihan Pembuatan Video dengan AI untuk Pemuda- Pemudi GPSI Medan Utara", Jurnal Minfo Polgan, 2024 Publication	<1 %
47	tintahmerah.wordpress.com Internet Source	<1 %
48	www.maxmanroe.com Internet Source	<1 %

---

Exclude quotes      On

Exclude bibliography      On

Exclude matches      Off