

Question Answering Al-Qur'an Menggunakan Generative Pre-Trained Transformer 3.5 Berbasis Chatbot Telegram

Elvino Dwi Saputra¹, Nazruddin Safaat Harahap^{2*}, Jasril³, Yusra⁴

Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

*e-mail *Corresponding Author*: nazruddin.safaat@uin-suska.ac.id

Abstract

The Al-Qur'an is a holy book that regulates everything related to life in this world and the afterlife. Searching for and understanding certain information in the Qur'an took a long time. Because it contains 30 juz, 114 surahs, and 6326 verses. However, with technological development, the search and understanding process can be faster by utilizing Artificial Intelligence (AI). Because AI can do what humans do with a faster and more accurate process. Combining AI with a Question Answering System (QAS) using a chatbot solves this problem. The searching and understanding process could be done quickly and accurately in two directions. Generative Pre-trained Transformer (GPT) is used as a model to understand natural human language. This model is considered accurate and fast, with the time needed approximately 1 minute to get an answer with an accuracy of 78.85%, answer relevance of 98.3%, and hallucination of 22.5%.

Keywords: *Al-Qur'an; Artificial Intelligence; Chatbot; Question Answering System; Generative Pre-trained Transformer*

Abstrak

Al-Qur'an merupakan kitab suci yang didalamnya mengatur segala hal terkait kehidupan di dunia dan akhirat. Dibutuhkan waktu yang begitu lama dalam proses pencarian dan pemahaman mengenai informasi tertentu dalam Al-Qur'an. Dikarenakan didalamnya terkandung 30 juz, 114 surah, dan 6326 ayat. Namun dengan adanya perkembangan teknologi proses pencarian dan pemahaman bisa lebih cepat dengan memanfaatkan *Artificial Intelligence* (AI). Ini dikarenakan AI dapat melakukan pekerjaan layaknya manusia dengan proses yang lebih cepat dan akurat. Perpaduan antara AI dengan *Question Answering System* (QAS) menggunakan *chatbot* menjadi solusi dari masalah tersebut. Proses pencarian dan pemahaman dapat dilakukan dengan cepat dan akurat serta dapat dilakukan dengan dua arah. *Generative Pre-trained Transformer* (GPT) digunakan sebagai model dalam proses pemahaman bahasa manusia secara alami. Penggunaan model ini dinilai akurat dan cepat dengan waktu yang dibutuhkan lebih kurang 1 menit untuk mendapatkan jawaban dengan akurasi sebesar 78,85%, *answer relevancy* sebesar 98,3% dan *hallucination* sebesar 22,5%.

Kata kunci: *Al-Qur'an; Artificial Intelligence; chatbot; Question Answering System; Generative Pre-trained Transformer*

1. Pendahuluan

Penyampaian dan penyaluran informasi berkembang secara pesat dari hari ke hari termasuk berkembangnya teknologi informasi [1]. Berbagai bidang telah banyak digunakan dalam menunjang kebutuhan juga ikut terpengaruh secara cepat dan akurat [2]. Al-Qur'an dalam bidang keagamaan juga tidak terlepas dari terpengaruhnya perkembangan teknologi dalam pencarian informasi Al-Qur'an. Dengan data yang begitu besar dan kompleks dalam Al-Qur'an, *Artificial Intelligence* (AI) menjadi teknologi yang dapat memecahkan masalah karena melakukan pembelajaran layaknya kepintaran manusia dalam kehidupan sehari-hari [3].

Al-Qur'an memiliki data yang begitu lengkap dan kompleks dengan 30 juz, 114 surah, 6326 ayat, dan lebih kurang 80 ribu kata Arab yang terkandung didalamnya [4]. Semuanya lengkap dibahas dalam Al-Qur'an dari manusia dilahirkan hingga meninggal, bahkan penciptaan langit dan bumi, surga, neraka dan akhirat yang merupakan sumber perkataan langsung dari Allah SWT. Tata cara hidup, tingkah laku, hukum, ibadah, iman tauhid, kisah terdahulu, dan sebagainya dibahas secara terperinci dalam Al-Qur'an.

Terjemahan Al-Qur'an membantu dalam proses pemahaman terhadap kandungan yang ada dalam Al-Qur'an. Walaupun demikian, proses pencarian informasi terhadap ayat atau surah tertentu yang membahas topik tertentu akan membutuhkan waktu yang lama. Ini dikarenakan pencarian dilakukan dengan membaca satu per satu secara manual. Namun ada beberapa Al-Qur'an yang memiliki point-point topik dari masing-masing ayat atau surah yang dikenal dengan *Indexing of Al-Qur'an* [5]. Ini dapat membantu menghemat waktu dalam mencari topik walaupun harus membaca satu per satu.

Question Answering System (QAS) menjadi salah satu teknologi AI yang dapat digunakan dalam pencarian informasi yang begitu besar dan dilakukan dengan komunikasi dua arah terhadap informasi yang ingin didadaptkan yaitu mengajukan pertanyaan dan mendapatkan jawabannya [6]. Dengan adanya sistem ini dapat mencari informasi terkait data Al-Qur'an yang begitu banyak menjadi lebih mudah dengan proses tanya jawab dalam bahasa alami manusia [7]. Dengan demikian, proses pencarian terhadap informasi Al-Qur'an jauh lebih cepat dan akurat dengan menggunakan QAS.

Chatbot menjadi teknologi yang sering dipadukan dengan QAS karena dinilai dapat melakukan proses tanya jawab secara otomatis dan menghasilkan *response* yang cepat [8]. *Chatbot* termasuk salah satu sistem AI yang dapat digunakan sebagai penunjang di berbagai sektor termasuk otomasi pelayanan tanya jawab [9]. Banyak hal-hal yang telah menerapkan AI baik itu dari segi tanya jawab, menggambar, dan sebagainya. Karena AI pada era sekarang sedang gencar-gencarnya dan dapat dikembangkan untuk tujuan apapun.

Penelitian yang menggabungkan antara Al-Qur'an dan *Question Answering* (QA) pernah dilakukan oleh Rana Malhas, Watheq Mansour, dan Tamer Elsayed pada tahun 2022. Penelitian yang berjudul Qur'an QA 2022: Overview of The First Shared Task on Question Answering over the Holy Qur'an ini menggunakan AraBERT dan AraELECTRA sebagai *transformer-based language* dan setidaknya 13 tim dan 21 lembaga ikut berpartisipasi didalamnya [10]. Model yang digunakan terbukti telah dapat menjawab pertanyaan yang beragam dari Al-Qur'an. Ini dikarenakan model termasuk pada *Large Language Model* (LLM) dapat memahamai bahasa alami dari manusia. Namun pada penelitian ini terdapat tantangan yang mana diharapkan nantinya sistem dapat memberikan jawaban berdasarkan surat tertentu atau bahkan dari seluruh Al-Qur'an.

Dengan demikian penilitian yang saya lakukan akan dapat memudahkan pengguna dalam pencarian informasi Al-Qur'an dengan menerapkan *Question Answering System* menggunakan GPT-3.5 yang dapat memahami bahasa alami dari manusia dan melakukan komunikasi dua arah. Sehingga dengan demikian tanya jawab akan lebih mendalam pemahamannya dikarenakan terdapat jawaban yang langsung bersumber kepada surah tertentu dan ayat tertentu didalam Al-Qur'an.

2. Tinjauan Pustaka

QAS bukanlah hal yang baru dan sudah dilakukan pada penelitian-penelitian sebelumnya. Salah satunya pada penilitian yang dilakukan oleh Fauzan Ishlakhuddin, Azhar Basir, dan Nurlaela pada tahun 2020. Penelitian yang berjudul Rancang Bangun Sistem Tanya Jawab Berbasis Aturan STMIK Muhammadiyah Paguyangan Brebes dengan Menggunakan Telegram *Chatbot* telah menghasil akurasi sebesar 98,95% [11]. Sayangnya pada penelitian ini, akurasi yang sangat tinggi dihasilkan karena pertanyaan berdasarkan aturan yang ada. Sehingga pertanyaan diluar dari aturan tersebut maka tidak akan dijawab oleh *chatbot* tersebut. Perbedaannya pada penelitian saya saat ini adalah chatbot dapat menjawab pertanyaan tanpa aturan dengan menggunakan bahasa alami dari manusia. Sehingga pertanyaan akan beragam dan jawabannya mengikuti dari pertanyaan dan sumber yang tersedia.

Penelitian lainnya yang berhubungan denan pencarian informasi juga pernah dilakukan oleh Waad Alshammari dan Sarah Alhumoud pada tahun 2022 dengan judul TAQS: *An Arabic Question Similarity System Using Transfer Learning of BERT With BiLSTM*. Penelitian ini relevan dengan penelitian yang saya lakukan karena menggunakan LLM sebagai model. Perbedaannya terletak pada model yang digunakan. Penelitian ini menggunakan model BiLSTM dengan akurasi 94,45% serta AraBERTv2 dan AraBERTv0.2 dengan akurasi 93.10% dan 93.90% [12]. Ini membuktikan bahwa dengan menggunakan LLM terbukti dapat meningkatkan akurasi yang tinggi dengan beragam bahasa alami bahasa pertanyaan dan jawaban terhadap pencarian informasi yang berhubungan. Perbedaan penelitian Alshammari dengan penelitian saya adalah data yang digunakan merupakan 44.404 entri pasangan jawaban arab atau *tawasul* dengan model BERT dari *Google*. Sedangkan penelitian saya

menggunakan Al-Qur'an dengan model LLM yaitu GPT-3.5 sebagai proses pemahaman dalam tanya jawab.

Penelitian yang berjudul *Chat2VIS: Generating Data Visualizations via Natural Language Using ChatGPT, Codex and GPT-3 Large Language Models* pada tahun 2023 oleh Paula Maddigan dan Teo Susnjak telah menerapkan *Large Language Model* (LLM) sebagai model untuk mengubah bahasa alami yang bebas menjadi kode bahasa yang dapat dipahami. Pada penelitian ini digunakan model ChatPGT dan GPT-3 dalam mengembangkan model baru untuk memahami pertanyaan yang tidak jelas dan buruk seperti ambiguitas. Terbukti bahwa penelitian ini menjadi solusi yang lebih baik dari pada *Natural Lanugage Processing* (NLP) tradisional seperti aturan tata bahasa buatan atau model yang disesuaikan [13]. Pada penelitian ini hanya memvisualisasikan saja dan tidak adanya tanya jawab. Sehingga saya menggunakan data Al-Qur'an pada penelitian saya dan menggabungkan LLM menjadi model untuk QAS.

Telegram *chatbot* menjadi solusi untuk memadukan proses *question answering* yang dilakukan karena dinilai cepat dan responsif. Ini telah terbukti pada penelitian yang dilakukan oleh Tjut Awaliyah Zuraiyah, Dian Kartika Utami, dan Degi Herlambang yang berjudul *Impelementasi Chatbot Pada Pendaftaran Mahasiswa Baru Menggunakan Recurrent Neural Network* pada tahun 2019. Penelitian ini menghasilkan akurasi *chatbot* Telegram sebesar 88%, presisi sebesar 95% dan recall sebesar 92% [14]. Pada penelitian ini telah menggunakan *Recurrent Neural Network* (RNN) yang membuat pertanyaan dapat memahami bahasa alami manusia dibandingkan penelitian terkait pertama oleh Fauzan Ishlakhuddin yang mengikuti aturan yang ada dan tidak bisa keluar dari aturan menggunakan bahasa alami manusia.

Penelitian yang telah dilakukan memiliki beberapa perkembangan dan pembaruan terhadap penelitian-penelitian sebelumnya. Sistem yang dikembangkan pada penelitian ini berfokus pada *Question Answering* (QA) terhadap Al-Qur'an dengan menggabungkan GPT-3.5 sebagai modelnya. Dengan penggunaan model ini terhadap *Large Language Model* (LLM) akan dapat memahami bahasa alami dari manusia dalam memahami pertanyaan pengguna dan memberikan jawabannya berdasarkan data Al-Qur'an dan terjemahannya. Ini merupakan pembaruan dari penelitian sebelumnya yang hanya menggunakan *Question Answering* (QA) berbasis aturan saja. Sehingga saat adanya pertanyaan yang keluar dari aturan tidak akan mendapatkan jawaban dan pertanyaannya yang diajukan hanya baku dan kaku sesuai aturan.

Perkembangan dan pembaruan lainnya juga terletak pada pemahaman pertanyaan yang tidak jelas atau ambiguitas. Ini dikarenakan penggunaan GPT-3.5 dapat memahami pertanyaan yang ambigu dibandingkan dengan model NLP terdahulu. Sayangnya pada penelitian sebelumnya yang telah dibahas diatas, penggunaan GPT-3.5 dijadikan sebagai pemahaman bahasa alami saja terhadap data lalu divisualisasikan. Sedangkan dalam penelitian ini terjadi pembaruan dengan pemahaman bahasa alami manusia dengan GPT-3.5 dijadikan suatu sistem tanya jawab untuk membantu dalam proses pemahaman dan pencarian terkait informasi Al-Qur'an dan terjemahannya dengan cepat, efisien, dan akurat.

3. Metodologi

Langkah-langkah dalam penelitian *Question Answering System* (QAS) terhadap data Al-Qur'an ini dapat dilihat pada gambar 1.

3.1 Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data Al-Qur'an dan terjemahan dalam bahasa Indonesia yang diambil dari API (*Application Programming Interface*) milik quran.com yaitu <https://api-docs.quran.com/docs/category/quran.com-api>. Data tersebut lalu diolah dan dimasukkan kedalam database sendiri. Selanjutnya data diolah lagi menjadi JSON (*JavaScript Object Notation*) yang berisi informasi surah seperti nama surah, nomor surah, nama latin surah, arti surah, serta tempat dan urutan surah diturunkan.

Data juga meliputi seluruh ayat dalam surah berbahasa arab lengkap dengan arti bahasa Indonesiannya yang diterjemahkan oleh Al-Mukhtasar. Bukan hanya itu kata per kata dari ayat tersebut beserta arab dan artinya juga digunakan dalam penelitian ini dalam bentuk file JSON tersebut. Sehingga data yang digunakan sebanyak 114 file JSON atau 114 surah dalam Al-Qur'an.

JSON merupakan suatu format dari teks yang dapat merepresentasikan data dalam empat tipe seperti *string*, angka, *boolean*, dan *null* dengan menggunakan pengkodean karakter *unicode* [15]. Sehingga dengan data Arab yang dimiliki dapat lebih bagus dalam proses pembacaan data karena menggunakan *unicode* dari JSON itu sendiri. Dengan demikian hasil

yang proses data akan jauh lebih mudah nantinya pada saat tahapan selanjutnya dalam membaca tulisan arab tersebut.



Gambar 1. Metodologi Penelitian

3.2 Pengolahan Data

Pada penelitian ini pengolahan data dilakukan dengan menggunakan *framework* (kerangka kerja) *Langchain* dalam prosesnya. Penggunaan *Langchain* sebagai *framework* dari LLM dimaksudkan dengan tujuan agar memudahkan interaksi pemanfaatan sumber data dengan aplikasi lain dalam aplikasi yang dikembangkan [16]. Terdapat beberapa fungsi pada *Langchain* untuk memudahkan pengolahan data yang dapat dilihat secara lebih terperinci dalam gambar dibawah ini.

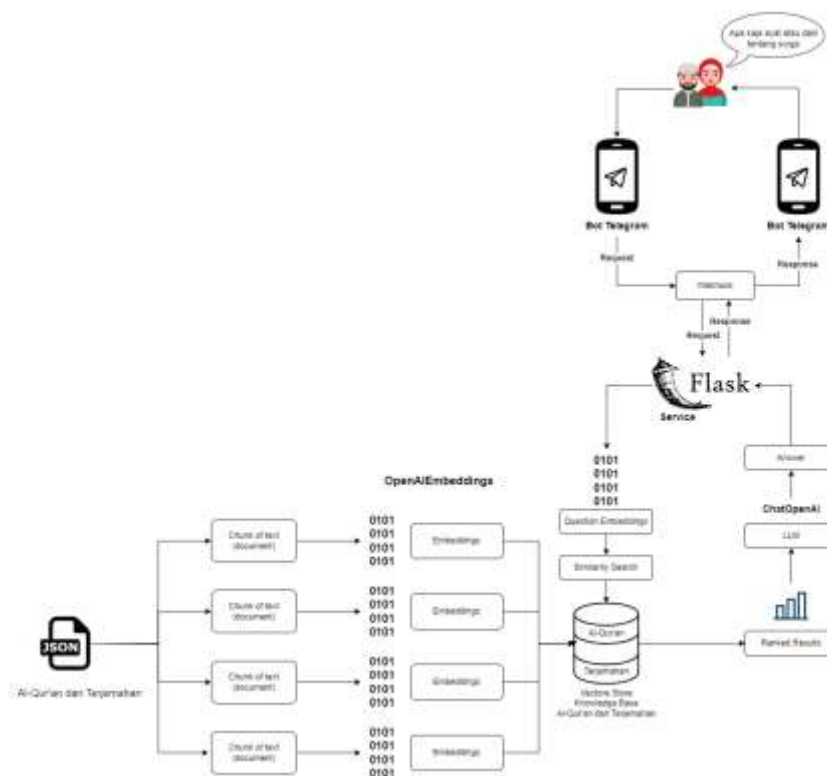


Gambar 2. Proses Pengolahan Data Menggunakan Langchain

- 1) *Document loaders*: memuat dokumen dengan berbagai tipe yang pada penelitian ini digunakan tipe file yaitu JSON.
- 2) *Document transformers*: mengubah dokumen yang telah dimuat menjadi beberapa bagian kecil (*chunk of text / split data*) agar dapat diproses oleh LLM.
- 3) *Text embeddings models*: proses pengubahan data tidak terstruktur hasil dari *chunk (transformers)* menjadi daftar angka (*list of floating-point numbers*) yang mewakili data tersebut. Dalam penelitian ini model yang digunakan untuk *embedding* adalah dari *OpenAI* yaitu *OpenAIEmbeddings*.
- 4) *Vector stores*: menyimpan hasil proses *embed* dan digunakan kembali pada saat proses pencarian data, juga dikenal sebagai *knowledge base* (basis pengetahuan).
- 5) *Retrivers*: menanyakan atau mencari data berdasarkan kemiripan data (*similarity search*) *embed* yang ada dalam *vector stores*.

3.3 Pengembangan Sistem

Proses pengembangan sistem pada penelitian *Question Answering* Al-Qu'ran dimulai setelah adanya pengumpulan dan pengolahan data yang telah dilakukan. Setelah data dikumpulkan dan diolah menggunakan *Langchain* sebagai *framework* barulah sistem akan dikembangkan. Dimana pengembangan dimulai dari pembuatan *Flask* sebagai penghubung hingga nantinya pembuatan *chatbot* Telegram sebagai *interface* (antarmuka) untuk pengguna bertanya dan mendapatkan jawaban. Adapun proses pengembangan sistem dapat dilihat dengan mengikuti alur atau diagram yang ada pada Gambar 3.



Gambar 3. Diagram Alir Pengembangan Sistem

3.3.1 Flask

Flask merupakan suatu *micro web framework* dalam bahasa pemrograman *Python* yang digunakan untuk memudahkan proses pembangunan aplikasi lebih cepat karena dapat menggunakan *library* dan kumpulan kode yang telah tersedia [17]. Pada tahapan ini *Flask* akan menjadi *service* atau penghubung antara data yang telah diolah menjadi *vector stores* dengan pertanyaan dari *chatbot* Telegram. *Flask* akan membuat API (*Application Programming Interface*) sebagai penghubung antara kedua hal tersebut. Karena API merupakan kode program yang berfungsi sebagai penghubung antara aplikasi *front-end* (dalam hal ini *chatbot*) dengan *back-end* (*Python*) [18].

Flask sebagai *service* akan menjalankan aplikasi *backend* dari *Python* yang digunakan saat adanya pertanyaan dari *chatbot*. Ketika pengguna bertanya di *chatbot* Telegram, maka nantinya API yang telah dibuat *Flask* dalam *method POST* akan dipanggil dalam *code chatbot*. Pertanyaan akan dikirimkan (*request*) melalui API ke *code Python* untuk diproses mendapatkan jawaban. Setelah jawaban didapatkan maka, *response* (hasil) akan diberikan oleh *Flask* dan diambil kembali oleh *chatbot* sebagai jawaban dari pertanyaan pengguna. Penggunaan *Flask* dalam penelitian ini dikarenakan *Flask* merupakan framework yang memiliki skalabilitas dan kompatibilitas yang tinggi dalam penggunaannya [19]. Proses *Flask* dalam penelitian ini secara detail dapat dilihat pada *Gambar 3*.

3.3.2 Chatbot Telegram

Chatbot Telegram pada penelitian ini dijadikan sebagai *interface* (antarmuka) yang menjadi penghubung antara sistem yang akan memberikan jawaban dengan pengguna yang bertanya. Penggunaan *chatbot* dikarenakan sistem mirip dengan manusia dalam melakukan tugasnya yaitu dengan percakapan secara otomatis kepada penggunanya dengan menggunakan *Artificial Intelligence (AI)* [20]. Dengan demikian proses bertanya akan maksimal dikarenakan adanya interaksi dua arah yang dilakukan secara cepat, responsif, dan akurat.

Pembuatan *chatbot* dilakukan dengan menggunakan aplikasi penyedia bot milik Telegram yang dikenal dengan nama *BotFather* dalam aplikasi Telegram. Dalam *BotFather* dilakukan proses pembuatan bot dan mendapatkan API Bot Telegram yang akan dihubungkan dengan sistem pembuat bot yang telah dibuat [21]. Penggunaan *chatbot* Telegram sangat bermanfaat karena dapat berfungsi secara gratis dan tanpa batas dengan pengimplementasiannya diberbagai bahasa pemrograman yang ada.

Setelah bot dibuat maka selanjutnya dibutuhkan penghubung atau API antara bot dengan server tempat sistem akan diletakkan yang bertugas menerima permintaan pengguna (*request*/perintah) dan mengembalikannya lagi ke bot (*response*), yaitu *Webhook* [22]. Penggunaan *Webhook* dalam pembuatan bot Telegram sangatlah penting. Ini dikarenakan dalam bot berjalan secara *realtime* sehingga harus ada penghubung antara bot dengan server. Apalagi bot Telegram hanya dapat menerima request dalam keamanan yang telah *secure* saja atau dengan kata lain hanya menerima *https* saja.

3.3.3 Ranked Results (Large Language Model)

Pemrosesan pertanyaan hingga menjadikannya sebuah jawaban melakukan berbagai tahapan. Dimulai dari yang awalnya pertanyaan diajukan oleh pengguna melalui *chatbot* Telegram lalu diteruskan melalui *webhook* dan *flask* sebagai *service* penghubung. Pertanyaan akan dilakukan proses *embedding* lagi dan dicarikan *similarity* atau kemiripan data pertanyaan dengan *knowledge base* yang ada sesuai dengan tahapan yang telah dijelaskan sebelumnya.

Proses pencarian kemiripan pertanyaan dengan basis pengetahuan yang dimiliki akan menghasilkan beberapa jawaban yang sesuai. Disinilah peran *Large Language Model* (LLM) akan bekerja untuk menghasilkan jawaban yang terbaik yang diberikan kepada pengguna sebagai penanya. LLM merupakan pengembangan lanjutan dari NLP (*Natural Language Processing*) yang lebih kompleks dalam memahami suatu bahasa. Hasil yang dihasilkan oleh LLM jauh lebih bagus dan berkualitas [23]. Ini dikarenakan banyak tugas-tugas NLP yang tidak bisa dilakukan namun dapat dilakukan oleh LLM. Sehingga LLM bukan hanya menaikkan akurasi melainkan juga menghasilkan penyelesaian yang lebih cepat dan efisien dalam pengembangannya baik dari segi biayanya.

Dalam penelitian ini LLM bekerja sebagai penghasil jawaban terbaik dari beberapa sumber yang telah didapatkan oleh *similarity search*. Dengan menggabungkan pertanyaan yang diajukan dan sumber-sumber jawaban yang telah didapatkan. LLM akan dengan mudah memahami bahasa alami dari pertanyaan lalu diolah dan dipadukan dengan berbagai jawaban yang telah didapatkan. Sehingga menghasilkan satu jawaban yang akurat dan bagus dengan bahasa alami yang mudah dipahami pengguna. Oleh karena itu, tahapan pencarian jawaban terbaik oleh LLM pada penelitian ini dinamakan *Ranked Results*. Dengan kata lain mengambil jawaban paling sesuai dengan pertanyaan dan sumber-sumber yang telah diberikan.

Pada tahapan ini model LLM yang digunakan adalah milik dari OpenAI yaitu GPT-3.5 (*Generative Pre-trained Transformer 3.5*). Model GPT-3.5 adalah model yang dirancang dengan data teks dalam jumlah besar untuk memproses bahasa alami seperti meringkas dan merespon teks pertanyaan [24]. Proses pembuatan model ini bukan hanya menggunakan data besar saja. Melainkan juga menggunakan *Learning from Human Feedback* (RLHF) yang mana menggabungkan antara model dengan kecerdasan dan keahlian manusia.

3.4 Pengujian Sistem

Pengujian sistem dilakukan dengan dua cara. Pertama yaitu melakukan validasi terhadap jawaban chatbot dengan cara mendatangkan mahasiswa Ilmu Tafsir dan Al-Qur'an UIN SUSKA RIAU sebanyak 7 orang. Nantinya responden akan menilai jawaban yang diberikan oleh chatbot atas dasar pemahaman mereka dengan mengisikan kuisioner untuk mendapatkan akurasi dari jawaban chatbot atas *question answering* yang telah dibuat. Pengujian kedua dilakukan untuk mengevaluasi dari *Large Language Model* (LLM) dengan menggunakan *framework open source* yaitu *DeepEval* [25]. Nantinya akan didapatkan sebuah persentase terhadap *Hallucination* dan *Answer Relevancy* dari jawaban yang dihasilkan *chatbot*. Dengan kata lain mengukur seberapa kuat sumber yang dihasilkan dari jawaban berdasarkan fakta data dan seberapa relevan jawaban dengan pertanyaan yang diberikan pengguna.

4. Hasil dan Pembahasan

4.1. Hasil Pengumpulan dan Pengolahan Data

Pada penelitian ini menggunakan data JSON yang terdiri dari 114 file JSON yang mewakili masing-masing surah dengan struktur data JSON seperti gambar 4.

Data ini merupakan informasi surah, ayat, arti bahasa Indonesia, hingga kata per kata dari masing-masing ayat. Data inilah yang nantinya akan di *embed* menjadi satu file berbentuk *vector stores* dengan gabungan dari semua surah yang bisa dilihat pada gambar 5.


```

1 {
2   "chapter": {
3     "Nama Surah : Al-Kaather, Nomor Surah : 108, Nama Latin Surah : الفاتحة, Arti Surah : Nikmat Berlipat,
4     Tempat Surah Diturunkan : makkah, Urutuan Surah Diturunkan : 15",
5     "Bunyi Surah Al-Kaather Ayat 1 : بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ, Arti Surah Al-Kaather Ayat 1 : Sebenarnya sand telah
6     memberimu wahai Rasul! kebajikan yang banyak, di antaranya sungai Al-Kaather di Surga.",
7     "Kata Ke 1 Ayat 1 Surah Al-Kaather : 0", Arti Kata Ke 1 Ayat 1 Surah Al-Kaather : sesungguhnya",
8     "Kata Ke 2 Ayat 1 Surah Al-Kaather : اللّٰه, Arti Kata Ke 2 Ayat 1 Surah Al-Kaather : Aku telah
9     memberikan ",
10    "Kata Ke 3 Ayat 1 Surah Al-Kaather : الف, Arti Kata Ke 3 Ayat 1 Surah Al-Kaather : nikmat yang
11    banyak",
12    "Bunyi Surah Al-Kaather Ayat 2 : اِنَّ اَوْلٰى اَعْيُنٍ رَّا رُبُّكَ, Arti Surah Al-Kaather Ayat 2 : Maka tunjukan rasa
13    syukur kepada Allah atas nikmat ini dengan melaksanakan salat dan menyembelih (qurban) untuk Allah semata,
14    berbeda dengan ibadah yang dilakukan oleh orang-orang musyrik berupa pendekatan diri dengan menyembelih untuk
15    patung-patung mereka.",
16    "Kata Ke 1 Ayat 2 Surah Al-Kaather : اِنَّ, Arti kata Ke 1 Ayat 2 Surah Al-Kaather : maka dirikanlah
17    sholat",
18    "Kata Ke 2 Ayat 2 Surah Al-Kaather : اَوْلٰى, Arti Kata Ke 2 Ayat 2 Surah Al-Kaather : karena Tuhanmu",
19    "Kata Ke 3 Ayat 2 Surah Al-Kaather : اَعْيُنٍ, Arti Kata Ke 3 Ayat 2 Surah Al-Kaather : dan bersembahkanlah",
20    "Bunyi Surah Al-Kaather Ayat 3 : اِنَّ رُبُّكَ هُوَ الَّذِیْ یَرْزُقُ مَنْ یَّشَآءُ بِغَیْرِ حِسَابٍ, Arti Surah Al-Kaather Ayat 3 : Sebenarnya orang
21    yang benci kepadamu adalah orang yang terputus dari setiap kebajikan, yang terlupakan, yang apabila namanya
22    disebut maka yang teringat darinya hanyalah kebucukkannya.",
23    "Kata Ke 1 Ayat 3 Surah Al-Kaather : اِنَّ, Arti Kata Ke 1 Ayat 3 Surah Al-Kaather : sesungguhnya",
24    "Kata Ke 2 Ayat 3 Surah Al-Kaather : رُبُّكَ, Arti Kata Ke 2 Ayat 3 Surah Al-Kaather : orang yang
25    membencimu",
26    "Kata Ke 3 Ayat 3 Surah Al-Kaather : هُوَ, Arti Kata Ke 3 Ayat 3 Surah Al-Kaather : dia",
27    "Kata Ke 4 Ayat 3 Surah Al-Kaather : یَرْزُقُ, Arti Kata Ke 4 Ayat 3 Surah Al-Kaather : terputus"
28  }
29 }

```

Gambar 4. Data JSON

Gambar 5. Data Hasil Embedding

Data hasil *embedding* dapat dilihat pada gambar diatas dimana hasilnya tidak bisa terbaca karena data tersebut telah diubah sesuai dengan fungsi *embedding* yang telah dijelaskan sebelumnya. Namun data tersebut bisa diubah kembali menjadi data sebelumnya yang jika divisualisasikan hasilnya seperti table dibawah ini.

Tabel 1. Visualisasi Data *Embedding*

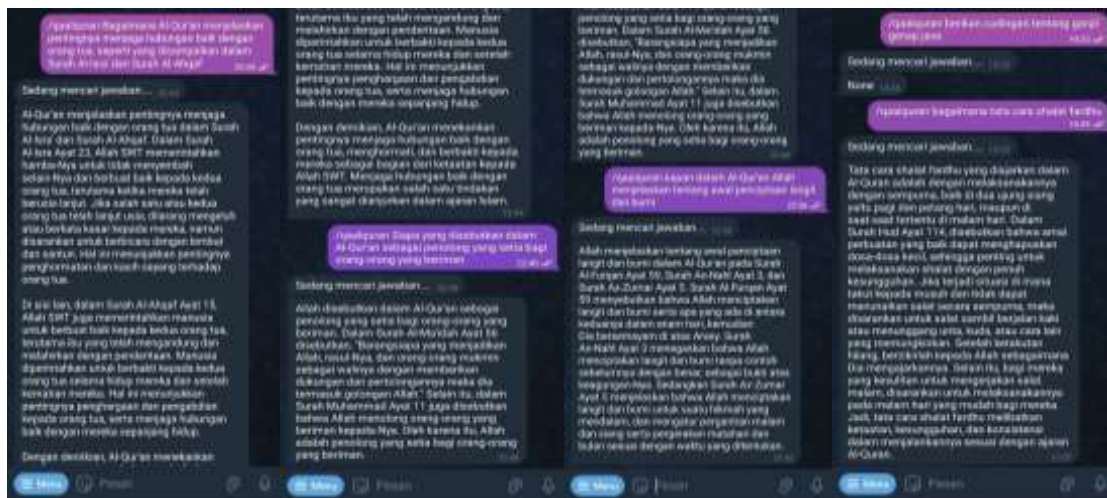
ID	Page Content	Metadata		
		Source	Seq Num	Surah
33362ddf-6ab2-4b1d-8ac4-caf79547146f	Nama Surah: Al-Fatihah, Nomor Surah : 1, Nama Latin Surah : الفاتحة, Arti Surah : Pembukaan, Tempat Surah Diturunkan : makkah, Urutuan Surah Diturunkan : 5 Bunyi Surah Al-Fatihah Ayat 1 : بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ, Arti Surah Al-Fatihah Ayat 1 : Dengan menyebut nama Allah aku mulai membaca Al-Qur`ān ini. Aku memohon pertolongan Allah -	1.json	1	Al-Fatihah
014a071d-33b0-4364-b618-431aefc8d25d	Subhānahu wa Ta`ālā- dan berharap keberkahan dengan menyebut nama-Nya. Lafal Basmalah mengandung tiga nama Allah yang sangat baik, yaitu: 1. Allāh, Zat yang berhak disembah, dan nama ini merupakan nama khusus hanya berlaku bagi Allah -Ta`ālā-, sehingga siapa pun tidak	1.json	2	Al-Fatihah

ID	Page Content	Metadata		
		Source	Seq Num	Surah
...	diperkenankan menyandang nama tersebut selain Dia -Subhānahu-. 2. Ar-Rahmān, pemilik rahmat yang luas, Dia Zat yang maha pengasih. 3. Ar-Rahīm, pemilik rahmat yang menyeluruh. Dia merahmati siapa pun yang dikehendaki-Nya dari makhluk-Nya, dan di antaranya hamba-hamba-Nya yang beriman.
a79bc344-669d-4466-b73e-676c5036c3d3	Kata Ke 3 Ayat 6 Surah An-Nas : وَالنَّاسِ , Arti Kata Ke 3 Ayat 6 Surah An-Nas : dan manusia	114.json	27	An-Nas

Data tersebut memiliki ID yang setiap potongan data tersebut memiliki ID yang berbeda-beda. Lalu terdapat juga *page content* atau sumber datanya, dan *metadata* yang berisi *source* atau nama *file*, *seq num* atau potongan ke berapa dari data surah, dan terakhir nama surahnya. Data inilah yang digunakan sebagai basis pengetahuan atau *knowledge base* sebagai rujukan untuk mencari jawaban dari pengguna.

4.2. Hasil Proses Bertanya Chatbot Telegram

Setelah data dirproses dan dihasilkan barulah pengguna bisa bertanya melalui *chatbot* Telegram dan akan diteruskan kepada *webhook* sebagai *service* Telegram Bot dan *Flask* sebagai *service Python*. Bot akan melakukan *request* terhadap pertanyaan yang dikirimkan dalam bentuk API yang telah dibuatkan dan mengembalikan jawaban atau *response*. Jawaban tersebut akan dikirimkan kembali kepada pengguna. Adapun hasil mekanisme bertanya melalui *chatbot* Telegram dapat dilihat pada gambar 6.



Gambar 6. Hasil Jawaban Chatbot Telegram

Pada dasarnya *Large Language Model* (LLM) dengan GPT-3.5 sebagai model dapat memahami bahasa alami dari manusia sebagai pengguna bot dan memberikan hasil dalam bentuk bahasa alami yang mudah dimengerti pula. Sebenarnya proses mendapatkan jawaban diambil dari API dengan *response* terdiri dari pertanyaan, jawaban, dan sumber dari jawaban itu sendiri. Sumber jawaban atau dalam visualisasi data pada *Tabel 1* disebut juga dengan *page content*. LLM yang akan menggabungkan antara pertanyaan dengan sumber (*page content*) hingga mendapatkan hasil jawaban yang sesuai dengan pertanyaan pengguna. Sumber dari jawaban pertanyaan pengguna dapat dilihat pada gambar 7.

Penggunaan *Large Language Model* dengan GPT-3.5 membantu dalam memahami bahasa alami manusia. Terbukti dengan pertanyaan yang diajukan menggunakan bahasa yang sering digunakan pada sehari-hari. Chatbot tetap dapat menjawab pertanyaan yang diberikan selagi pertanyaan masih berada di dalam konteks yaitu Al-Qur'an dan terjemahannya dalam Bahasa Indonesia. Bahkan sumbernya diberikan sesuai dengan konteks pertanyaan yang

diberikan seperti yang dapat dilihat pada Gambar 7. Berbeda dengan penelitian sebelumnya dari Hammad Afzal dan Tayyeba Mukhtar dalam penelitian berjudul *Semantically Enhanced Concept Search of the Holy Quran: Qur'anic English WordNet* pada tahun 2019.

```

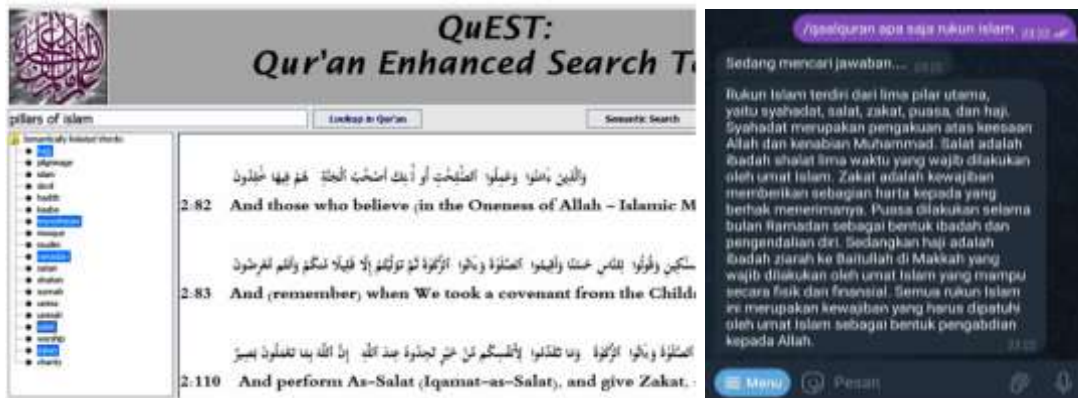
"content": "Bunyi Surah Al-Ahqaf Ayat 15 : وَرَوَّيْنَا لِلْإِنسَانِ إِذْ أَنشَأْنَاهُ بِحَبْلٍ وَإِنَّا لَهُ نَزُّوا رُوحَانَهُ فَهُوَ مُرَوَّيْنَا وَنَحْنُ لَهُ رُوحَانَةٌ وَرَوَّيْنَا لِلْإِنسَانِ إِذْ أَنشَأْنَاهُ مِنْ نَجْوٍ وَعَاقِبَةً أُولَىٰ عُنُقٍ وَإِنَّا كَاشِفُو الْعُقَدِ وَإِنَّا لَآئِمَّةٌ بِنَازِلَةٍ وَكَاشِفَةٌ لِلْهَالِكِ وَإِنَّا لَمُنشِقَةٌ لِلشَّامِ وَالْأَسَدِ وَإِنَّا لَمُنشِقَةٌ لِلشَّامِ وَالْأَسَدِ وَإِنَّا لَمُنشِقَةٌ لِلشَّامِ وَالْأَسَدِ",
"content_source": "Al-Ahqaf",
"file": "46.json",
"potongan": 238,

"content": "Bunyi Surah Al-Qasas Ayat 26 : فَذَلِكَ بِمَا نَكَلْنَاكَ لِآبَائِكَ إِذْ أَخَذْنَا مِنَ النَّبِيِّينَ مِيثَاقَهُمْ لَعَنَّاهُمْ أَن يَقولُوا إِنَّا أَنشَأْنَا قُرُونًا فَتَطَاوَأَتْ أَعْيُنُهُمْ الْغُورُ الْآبِئِينَ",
"content_source": "Al-Qasas",
"file": "28.json",
"potongan": 422,

"content": "Bunyi Surah Al-Isra Ayat 23 : وَإِنِّي لَأَنذَرْتُكَ إِن يَأْتِيَنَّكَ إِذَا تَأْتَىٰ وَجْهَكَ رَبُّكَ فَأَعْبُدْهُ وَاتَّقِ اللَّهَ وَمَا يُشْرِكُ بِهِ إِلهًا فَذَلِكَ عَصِيَ الْكافِرِينَ",
"content_source": "Al-Isra",
"file": "17.json",
"potongan": 327,

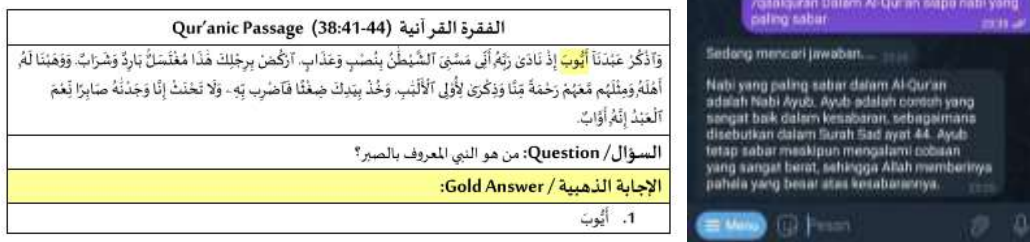
```

Gambar 7. Sumber Jawaban



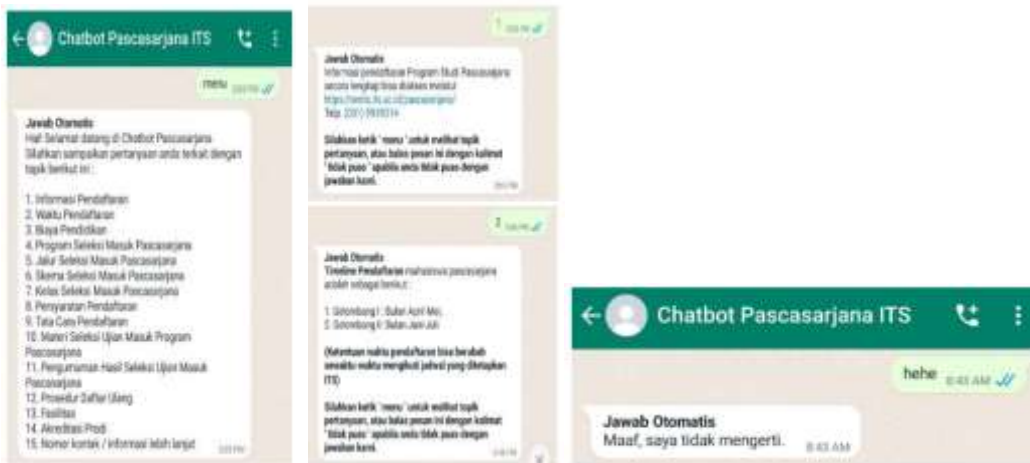
Gambar 8. Perbandingan Dengan Penelitian Sebelumnya (i)

Dalam penelitiannya mengatakan bahwa pencarian semantic Al-Qur'an dengan konsep abstrak biasanya menggunakan NLP dan IR (*Information Retrival*) sering kali gagal dalam pertanyaan abstrak [26]. Sehingga ia menggunakan *Quranic English WordNet* (QEWN) untuk mencari semantik Al-Qur'an dalam bahasa Inggris berdasarkan daya lingustik terjemahannya. Berbeda dengan LLM, cukup dengan model GPT-3.5 ini saja dapat memahami bahasa alami manusia. Seperti yang bisa dilihat pada Gambar 8 diatas, terbukti bahwa dengan GPT-3.5 dapat memahai bahasa alami dari manusia bahkan menjawab dengan bahasa alami manusia pula. Jika dilihat pada penelitian sebelumnya hasil yang dihasilkan saat bertanya rukun Islam masih terdapat beberapa jawaban yang tidak terkait dengan pertanyaan. Dibandingkan dengan penelitian yang saya lakukan tentu lebih baik karena langsung mendapatkan jawaban sesuai pertanyaan. Tentu ini akan menambah pemahaman terhadap Al-Qur'an karena langsung memberikan jawaban yang sesuai.



Gambar 9. Perbandingan Dengan Penelitian Sebelumnya (ii)

Penelitian serupa tentang penggabungan Al-Qur'an dan *Question Answering* (QA) juga dilakukan oleh Rana Malhas dkk seperti mana yang telah dijelaskan pada pendahuluan. Dimana berdasarkan Gambar 9 diatas terdapat perbedaan jawaban yang diberikan dengan pertanyaan yang sama yaitu "Siapa nabi yang paling sabar dalam Al-Qur'an". Penelitian yang dilakukan Rana Malhas dkk hanya menjawab "Ayub". Sedangkan pada penelitian ini didapatkan jawaban yang sama juga yaitu "Nabi Ayub" namun dengan penjelasan dan juga sumber ayatnya. Ini menjadi kelebihan dari penelitian ini dan juga sumber jawaban juga menjadi kelemahan pada penelitian sebelumnya seperti yang telah dijelaskan sebelumnya.



Gambar 10. Perbandingan Dengan Penelitian Sebelumnya (iii)

Penelitian lain terkait dengan *chatbot* pernah dilakukan oleh Mursidah [27]. Dimana pada penelitian ini *chatbot* hanya dapat menjalankan perintah berdasarkan aturan yang ada. Jika keluar dari aturan yang ada maka *chatbot* tidak akan menjawabnya. Jawaban yang diberikan pun telah ada sebelumnya disediakan berdasarkan pilihan aturan yang akan dipilih. Tentu pertanyaan tidak bisa layaknya bertanya pada manusia umumnya. Dibandingkan penelitian yang saya lakukan tentu jauh lebih baik karena dapat langsung memberikan jawaban tanpa adanya aturan yang harus digunakan. Penelitian serupa juga pernah dilakukan oleh Bryan Wijaya dan Hapnes Toba [28] yang penggunaan tanya jawab berbasis skenario. Jauh lebih baik dari pada Marsidah karena sudah bisa tanpa adanya aturan yang kaku. Namun penelitian ini juga masih kurang. Karena jika diluar dari skenario maka jawaban tidak akan muncul. Tentu diperlukan skenario yang begitu banyak dalam pembuatannya. Berbeda dengan penelitian saat ini yang saya buat menggunakan GPT-3.5 yang dapat mengerti bahasa alami manusia dari pertanyaan dan basis pengetahuan yang telah diolah dari dokumen yang ada. Tanpa perlu membuat skenario yang begitu banyak.

4.3. Hasil Pengujian

Hasil pengujian dilakukan dengan proses validasi jawaban dengan melibatkan mahasiswa jurusan Ilmu Al-Qur'an dan Tafsir dari UIN SUSKA RIAU sebanyak 7 orang. Pengujian dilakukan untuk mengukur seberapa persentase keberhasilan dan kemiripan jawaban antara *chatbot* yang dihasilkan dengan jawaban dari responden. Pertanyaan akan dilakukan sebanyak 10 pertanyaan berbeda ke *chatbot* dan juga responden. Lalu jawaban *chatbot* akan dinilai dengan jawaban dari responden. Para responden akan memberikan

rentang nilai jawaban chatbot sesuai dengan jawaban yang mereka miliki berdasarkan pengetahuan mereka.

Terdapat 5 pilihan yang akan dipilih oleh responden diantaranya Sangat Tidak Setuju (STS), Tidak Setuju (TS), Cukup (C), Setuju (S), dan Sangat Setuju (SS). Adapun pertanyaan dan nilai dari jawaban *chatbot* dapat dilihat pada tabel dibawah ini.

Tabel 2. Hasil Pengujian Validasi Jawaban

No	Pertanyaan	Jawaban				
		STS	TS	C	S	SS
1.	Apa hukum poligami				1	6
2.	Apa ciri-ciri orang beriman				3	4
3.	Bagaimana gambaran keindahan surga			1	3	3
4.	Apa pandangan islam mengenai pacaran				4	3
5.	Apa bukti Allah Maha Kaya			2	3	1
6.	Apa arti kata sijjin	1	4		1	1
7.	Cerita tentang apa surah al kahf	2	1	1	2	1
8.	Apa itu sapi betina			2	4	1
9.	Apa makna surga ada istana yang dibawahnya air mengalir			2	4	1
10.	Apa hukum riba				3	4
Total		3	6	8	28	25

Perhitungan persentase validasi jawaban dapat dilakukan dengan pembagian total skor dengan nilai X yang dapat dirumuskan sebagai berikut:

$$\text{Sangat Setuju} = 25 \times 5 = 125$$

$$\text{Setuju} = 28 \times 4 = 112$$

$$\text{Cukup} = 8 \times 3 = 24$$

$$\text{Tidak Setuju} = 6 \times 2 = 12$$

$$\text{Sangat Tidak Setuju} = 3 \times 1 = 3$$

$$\text{Total Skor} = 276$$

$$X = \text{Skor tertinggi} \times \text{jumlah pertanyaan} \times \text{jumlah responden}$$

$$= 5 \times 10 \times 7$$

$$= 350$$

$$\text{Persentase akurasi} = \text{Total Skor} / X \times 100\%$$

$$= 276 / 350 \times 100\%$$

$$= 78,85\%$$

Berdasarkan perhitungan persentase dari responden, maka dapat disimpulkan bahwa *Question Answering* Al-Qur'an menggunakan GPT-3 sebagai model dan Telegram sebagai basis antarmuka adalah Sangat Baik dengan persentase sebesar 78,85%. Responden berasal dari mahasiswa jurusan Ilmu Al-Qur'an dan Tafsir yang langsung menilai ketetapan jawaban dengan pemahaman yang dimilikinya. Responden juga memberikan saran bahwa sebenarnya jawaban dari sistem yang telah dibuat tidaklah salah dan buruk. Melainkan pertanyaan yang digunakan pada kuisoner kurang tepat. Ini dikarenakan data yang digunakan hanya Al-Qur'an dan terjemahan sehingga tidak bisa bertanya yang lebih rinci hanya umum saja karena Al-Qur'an membutuhkan tafsir untuk jawaban yang lebih spesifik dan terperinci.

Dikarenakan adanya saran yang diberikan oleh responden. Penelitian ini juga diuji terhadap *Large Language Model* (LLM) dengan menggunakan *DeepEval* dan pertanyaan yang lebih beragam dan tidak kaku seperti sebelumnya. Pertanyaan juga didapatkan dengan beberapa saran pertanyaan dari responden sebelumnya. GPT-4 digunakan dalam proses pengujian *DeepEval*. Nantinya akan didapatkan sebuah persentase dari pengujian *DeepEval* terhadap *Hallucination* dan *Answer Relevancy* dari *question answering* yang telah dibuat. *Hallucination* dalam LLM artinya seberapa besar model menghasilkan informasi yang benar secara faktual berdasarkan dasar dan sumber yang kuat. Sehingga untuk mengujinya diperlukan perbandingan antara *input* (pertanyaan), *actual output* (jawaban yang diberikan LLM), *context* (dokumen yang diinputkan sebagai proses LLM). Nilai *hallucination* sendiri dimulai dari 0.0 yang artinya baik atau tidak terjadi *hallucination* hingga 1.0 yang artinya sebaliknya yaitu terjadi *hallucination*. Sedangkan *answer relevancy* merupakan seberapa besar relevan yang terjadi antara jawaban dan pertanyaan. Sehingga pengukurannya cukup dengan menggunakan pertanyaan (*input*) dan jawaban yang dihasilkan LLM (*actual output*). Dengan rentang nilai yang dihasilkan 1.0 untuk baik dan 0.0 untuk sebaliknya. Adapun dari 10 pertanyaan didapatkan nilai *hallucination* dan *answer relevancy* sebagai berikut.

Tabel 3. Hasil Pengujian *Hallucination* dan *Answer Relevancy*

No	Pertanyaan	Evaluasi	
		Hallucination	Answer Relevancy
1.	Mengapa Allah melarang riba	0.0	1.0
2.	Kapan dalam Al-Qur'an Allah menjelaskan tentang awal penciptaan langit dan bumi	0.0	1.0
3.	Dimana lokasi penting yang diutamakan bagi umat Islam dalam menunaikan ibadah haji	0.0	1.0
4.	Apa pandangan islam mengenai pacaran	0.0	1.0
5.	Siapa yang disebutkan dalam Al-Qur'an sebagai penolong yang setia bagi orang-orang yang beriman	0.25	1.0
6.	Kenapa Allah memperingatkan umat manusia tentang hari di mana tidak berguna harta dan anak-anak dalam surah Ali-'imran	0.75	1.0
7.	Cerita tentang apa surah al kahf	0.25	1.0
8.	Siapa nabi yang ditelan ikan paus	0.75	1.0
9.	Apa makna surga ada istana yang dibawahnya air mengalir	0.0	0.83
10.	Bagaimana Al-Qur'an menjelaskan pentingnya menjaga hubungan baik dengan orang tua, seperti yang disampaikan dalam Surah Al-Isra' dan Surah Al-Ahqaf	0.25	1.0
Total		2.25	9.83
Rata-Rata		0.225	0.983

Berdasarkan hasil yang didapatkan maka persentase dari *hallucination* yang dihasilkan terhadap *Question Answering System* Al-Qur'an dengan menggunakan LLM adalah 22,5% yang artinya baik atau tidak terjadi *hallucination*. Karena semakin tinggi nilainya maka makin tinggi juga *hallucination* yang terjadi pada LLM yang dihasilkan. Terjadinya *hallucination* bukan dikarenakan jawaban yang diberikan salah melainkan sumber faktual yang dijadikan parameter pengukuran masih ada yang belum tepat dan kadang tidak berhubungan dengan jawaban yang diberikan. Sehingga terjadilah nilai *hallucination* yang besar dari suatu jawaban yang diberikan. Sedangkan persentase *answer relevancy* yang didapatkan adalah sebesar 98.3% yang artinya antara jawaban dan pertanyaan memiliki relevansi yang tinggi. Sedangkan sisanya yaitu 1,7% tidak relevan karena terkadang jawaban yang diberikan masih belum tepat namun hanya kecil dan tidak mempengaruhi jawaban secara besar. Ini dikarenakan dengan menggunakan GPT-3.5 jawaban yang diberikan terkadang selalu berubah-ubah cara menjawabnya namun tetap berhubungan dengan jawaban, sehingga terkadang ada jawaban yang relevansinya belum maksimal.

Hasil yang diperoleh pada penelitian ini bukan hanya terkait seberapa akurat atau relevan jawaban dengan pertanyaan. Namun penelitian ini juga membantu mencari informasi terkait Al-Qur'an yang begitu lama menjadi lebih cepat. Seperti yang dapat dilihat pada hasil yang diperoleh pada *Gambar 6*. Waktu yang diperoleh untuk mendapatkan jawaban terkait suatu informasi tertentu yang ditanyakan hanya membutuhkan waktu kurang dari 1 menit. Ini dibuktikan dengan waktu yang ditampilkan pada *chatbot* Telegram tersebut. Pengguna bertanya di jam 22:44 dan bot menjawab di jam 22:44 pula. Ini artinya hanya butuh 1 menit lebih kurang untuk menjawab pertanyaan tersebut. Dengan demikian, dari hasil yang diperoleh menyatakan bahwa penelitian ini bukan hanya mendapatkan jawaban relevan tetapi juga cepat.

Penelitian ini juga relevan dan pembaruan terhadap penelitian yang telah dilakukan oleh Hammad Afzal dan Tayyeba Mukthar [26] seperti yang telah dijelaskan sebelumnya. Dimana penelitian ini mendapatkan jawaban yang langsung berhubungan dengan pertanyaan yang diajukan. Beda dengan penelitian oleh Afzal [26] yang jawabannya masih terkandung hal yang tidak terkait. Penelitian ini juga perkembangan dari penelitian Rana Malhas dkk [10] yang telah dijelaskan pada hasil diatas. Dimana jawaban yang diberikan dari penelitian Rana Malhas hanya sekedar jawaban singkat tanpa adanya sumber pendukung yang layaknya pada penelitian yang telah saya lakukan. Tentu dengan adanya sumber pendukung dari ayat tersebut akan menambah pemahaman pengguna terkait pertanyaan yang diajukan. Namun penelitian yang saya lakukan ini bertolak belakang dengan penelitian dari Mursidah [27]. Dimana seperti yang telah dijelaskan penelitian Mursidah menggunakan *question answering* berbasis aturan yang jawaban dan pertanyaan telah disediakan berdasarkan aturan yang ada. Sedangkan pada penelitian yang saya lakukan jawaban akan terus berubah-ubah menggunakan bahasa alami manusia sejalan dengan pertanyaan yang diajukan. Tentu ini lebih bagus dibandingkan dengan penelitian yang telah dilakukan oleh Mursidah [27] tersebut.

5. Simpulan

Pada penelitian ini telah dilakukan pembuatan sistem tanya jawab (*Question Answering System*) berdasarkan data dari Al-Qur'an dan terjemahan Indonesia. Proses pengolahan data hingga tanya jawab menggunakan metode *Large Language Model* (LLM) dan *Generative Pre-trained Transformer* (GPT) 3.5 sebagai modelnya dengan menggunakan bahasa *Python* sebagai bahasa pemrograman. OpenAI dan *Langchain* digunakan sebagai *platform* untuk proses pengolahan data dan juga proses pencarian jawaban dari pengguna. Dari 114 surah dalam bentuk *file json* dijadikan dataset yang diolah untuk dapat ditanya pengguna dengan memberikan hasil akurasi sebesar 78,85%, *hallucination* sebesar 22,5% dan *answer relevancy* sebesar 98,3%. Hasil akurasi didapatkan dengan cara membandingkan jawaban *chatbot* dengan jawaban dari responden yang menilai jawaban *chatbot* tersebut.

Hallucination dan *answer relevancy* diukur dengan menggunakan *framework* dari *DeepEval* untuk evaluasi LLM. Dengan demikian dapat disimpulkan bahwa GPT-3.5 dan *Langchain* dapat membantu dalam proses tanya jawab pengguna. Karena model LLM yaitu GPT-3.5 dapat memahami bahasa manusia secara alami dan memberikan akurasi yang tinggi dalam jawabannya. Sehingga dapat membantu lebih mudah dalam memahami Al-Qur'an secara cepat, efisien, dan akurat. Namun adanya *hallucination* yang bisa dikatakan menuju tinggi dan akurasi yang belum dikatakan sempurna. Dikarenakan penggunaan model pada penelitian ini masih menggunakan *similarity search* atau kemiripan pertanyaan dengan jawaban. Belum menggunakan *semantic search* atau pencarian semantik (makna yang sama) terhadap pertanyaan dan jawaban. Nantinya diharapkan dalam penelitian selanjutnya dapat menggunakan penggabungan antara *similarity search* dengan *semantic search*. Sehingga berdasarkan hasil yang didapatkan dari penelitian ini dapat disimpulkan bahwa bukan hanya akurat dan relevan. Melainkan juga jawaban yang dihasilkan lebih kuat dan beragam baik itu secara kemiripan pertanyaan dan jawaban serta makna dari keduanya. Serta waktu yang dibutuhkan untuk mendapatkan informasi atau jawaban sangatlah cepat yaitu lebih kurang 1 menit seperti yang dapat dilihat langsung dari hasil yang diberikan oleh *chatbot* Telegram.

Daftar Referensi

- [1] D. Apriliani, S. F. Handayani, T. N. Anugrahaeni, A. Miftahudin, L. Nurarifiah, and I. T. Saputra, "Aplikasi Question Answer Sebagai Media Pembelajaran Interaktif Untuk Mata Pelajaran Akuntansi," *JMM (Jurnal Masy. Mandiri)*, vol. 7, no. 2, pp. 2003–2011, 2023, doi: 10.31764/jmm.v7i2.13867.
- [2] M. Sidik, B. Gunawan, and D. Anggraini, "Pembuatan Aplikasi Chatbot Kolektor dengan Metode Extreme Programming dan Strategi Forward Chaining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, pp. 293–302, 2021, doi: 10.25126/jtiik.2021824298.
- [3] T. N. Fitria, "Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay," *ELT Forum J. English Lang. Teach.*, vol. 12, no. 1, pp. 44–58, 2023, doi: 10.15294/elt.v12i1.64069.
- [4] N. I. Purwita, M. A. Bijaksana, K. M. Lhaksana, and M. Z. Naf'an, "Typo handling in searching of quran verse based on phonetic similarities," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 6, no. 2, pp. 130–140, 2020, doi: 10.26594/register.v6i2.2065.
- [5] I. Humaini, T. Yusnitasari, L. Wulandari, D. Ikasari, and H. Dutt, "Information Retrieval of Indonesian Translated version of Al Quran and Hadith Bukhori Muslim," *2018 Int. Conf. Sustain. Energy, Electron. Comput. Syst. SEEMS 2018*, pp. 1–5, 2019, doi: 10.1109/SEEMS.2018.8687330.
- [6] Y. H. Chen, E. J. L. Lu, and Y. Y. Lin, "Efficient SPARQL Queries Generator for Question Answering Systems," *IEEE Access*, vol. 10, no. September, pp. 99850–99860, 2022, doi: 10.1109/ACCESS.2022.3206794.
- [7] R. A. Yunmar and I. W. W. Wisesa, "Pengembangan Mobile-Based Question Answering System Mobile-Based Question Answering System Development With Ontology Based Knowledge," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 4, pp. 693–700, 2020, doi: 10.25126/jtiik.202072255.
- [8] F. Ishlakhuddin, Y. M. Santosa, and N. B. Nugraha, "Document Generation untuk Chatbot Berbasis Aturan dengan Pendekatan Template Method," *J. Inform. J. Pengemb. IT*, vol. 7, no. 3, pp. 194–198, 2022, doi: 10.30591/jpit.v7i3.5098.
- [9] L. Anindyati, "Analisis dan Perancangan Aplikasi Chatbot Menggunakan Framework Rasa dan Sistem Informasi Pemeliharaan Aplikasi (Studi Kasus: Chatbot Penerimaan Mahasiswa Baru Politeknik Astra)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp.

- 291–300, 2023, doi: 10.25126/jtiik.20231026409.
- [10] R. Malhas, W. Mansour, and T. Elsayed, "Qur' an QA 2022 : Overview of The First Shared Task on Question Answering over the Holy Qur' an," no. June, pp. 79–87, 2022.
- [11] F. Ishlakhuddin, A. Basir, and N. Nurlaela, "Rancang Bangun Sistem Tanya-jawab Berbasis Aturan STMIK Muhammadiyah Paguyangan Brebes dengan Menggunakan Telegram Chatbot," *J. Inform. J. Pengemb. IT*, vol. 5, no. 3, pp. 100–105, 2020, doi: 10.30591/jpit.v5i3.2900.
- [12] W. Alshammari and S. Alhumoud, "TAQS: An Arabic Question Similarity System Using Transfer Learning of BERT with BiLSTM," *IEEE Access*, vol. 10, no. September, pp. 91509–91523, 2022, doi: 10.1109/ACCESS.2022.3198955.
- [13] P. Maddigan and T. Susnjak, "Chat2VIS: Generating Data Visualizations via Natural Language Using ChatGPT, Codex and GPT-3 Large Language Models," *IEEE Access*, vol. 11, no. May, pp. 45181–45193, 2023, doi: 10.1109/ACCESS.2023.3274199.
- [14] T. A. Zuraiyah *et al.*, "Impelementasi Chatbot Pada Pendaftaran Mahasiswa Baru Menggunakan Recurrent Neural Network," *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 2, pp. 91–101, 2019.
- [15] G. Langdale and D. Lemire, "Parsing gigabytes of JSON per second," *VLDB J.*, vol. 28, no. 6, pp. 941–960, 2019, doi: 10.1007/s00778-019-00578-5.
- [16] O. Topsakal and T. C. Akinci, "Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast," *Int. Conf. Appl. Eng. Nat. Sci.*, vol. 1, no. 1, pp. 1050–1056, 2023, doi: 10.59287/icaens.1127.
- [17] Candra Wijayanto and Yeremia Alfa Susetyo, "Implementasi Flask Framework Pada Pembangunan Aplikasi Sistem Informasi Helpdesk (SIH)," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 07, no. 03, pp. 858–868, 2022, [Online]. Available: <https://jurnal.stkipppgritulungagung.ac.id/index.php/jupi/article/view/3161/1328>.
- [18] B. P. Putra and Y. A. Susetyo, "Implementasi Api Master Store Menggunakan Flask, Rest Dan Orm Di Pt Xyz," *Sistemasi*, vol. 9, no. 3, pp. 543–556, 2020, doi: 10.32520/stmsi.v9i3.899.
- [19] Z. Liang, Z. Liang, Y. Zheng, B. Liang, and L. Zheng, "Data analysis and visualization platform design for batteries using flask-based python web service," *World Electr. Veh. J.*, vol. 12, no. 4, pp. 187–198, 2021, doi: 10.3390/wevj12040187.
- [20] M. Sarosa, M. Kusumawardani, A. Suyono, and Z. Sari, "Implementasi Chatbot Pembelajaran Bahasa Inggris menggunakan Media Sosial," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, pp. 317–322, 2020, doi: 10.26418/jp.v6i3.43191.
- [21] C. Ukamaka Bertrand, O. Uzoamaka Ekwealor, and C. Juliet Onyema, "Artificial Intelligence Chatbot Advisory System," *Int. J. Intell. Inf. Syst.*, vol. 12, no. March, pp. 1–9, 2023, doi: 10.11648/j.ijis.20231201.11.
- [22] A. F. K. Sibero and A. Murdani, "Sistem Informasi Desa Menggunakan Telegram Bot Sebagai Antar Muka," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 1, pp. 206–211, 2023, doi: 10.35870/jimik.v4i1.161.
- [23] X. Yang *et al.*, "A large language model for electronic health records," *npj Digit. Med.*, vol. 5, no. 1, pp. 1–9, 2022, doi: 10.1038/s41746-022-00742-2.
- [24] I. A. Bernstein *et al.*, "Comparison of Ophthalmologist and Large Language Model Chatbot Responses to Online Patient Eye Care Questions," *JAMA Netw. open*, vol. 6, no. 8, p. e2330320, 2023, doi: 10.1001/jamanetworkopen.2023.30320.
- [25] DeepEval, "DeepEval - The open-source LLM evaluation framework," *DeepEval*, 2024. <https://docs.confident-ai.com/> (accessed Mar. 15, 2024).
- [26] H. Afzal and T. Mukhtar, "Semantically Enhanced Concept Search of the Holy Quran: Qur'anic English WordNet," *Arab. J. Sci. Eng.*, vol. 44, no. 4, pp. 3953–3966, 2019, doi: 10.1007/s13369-018-03709-2.
- [27] E. Mursidah, L. Ambarwati, and F. A. Karima, "Implementasi Chatbot Layanan Informasi Pendaftaran Mahasiswa Baru Program Pascasarjana Departemen Teknik Informatika ITS," *J. Ilm. NERO*, vol. 7, no. 1, pp. 43–52, 2022.
- [28] H. Toba and B. Wijaya, "Implementasi Sistem Tanya Jawab Berbasis Skenario untuk Mendukung Proses Akademik dengan IBM Watson Assistant," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 2, pp. 154–166, 2020, doi: 10.26418/jp.v6i2.40715.