

MODEL KECERDASAN BUATAN BAGI PENJANAAN PENGAJARAN SURAH AL-QURAN MELALUI PENDEKANAN TAFSIR INTERAKTIF BERASASKAN PEMROSESAN BAHASA SEMULAJADI DAN PEMBELAJARAN MESIN^(*)

(Artificial Intelligent (AI) Model for Generating Lessons Of Quranic Chapters Through An Interactive Tafsir Approach Based On Natural Language Processing (NLP) And Machine Learning (ML))

Fatimah Noni Muhamad¹, Mohd Aizul Yaakob²,
Mohammad Roshimi Abdullah³, Noorkartina Mohamad⁴,
Siti Khalilah Basarud-din⁵

ABSTRACT

The Qur'an is the primary source of guidance for Muslims; however, a deep understanding of its teachings remains limited to static authoritative works that require manual searching. This limitation creates a challenge for users who struggle to access the teachings of the verses in a systematic and interactive manner. This study proposes the development of an Artificial Intelligence (AI) model that applies Natural Language Processing (NLP) and Machine Learning approaches to automatically generate benefits and lessons from the chapters of the Qur'an. The model is

^(*) This article was submitted on: 20/07/2025 and accepted for publication on: 25/09/2025.

¹ Pensyarah Kanan (Dr.), Jabatan Teknologi Maklumat, Fakulti Perniagaan dan Sains Pengurusan, Universiti Islam Antarabangsa Tuanku Syed Sirajuddin (UniSIRAJ).
Email: fatimahnoni@unisiraj.edu.my

² Pensyarah Kanan (Dr.), Jabatan Usuluddin, Fakulti Pengajian Islam, Universiti Islam Antarabangsa Tuanku Syed Sirajuddin (UniSIRAJ).
Email: aizulyaakob@unisiraj.edu.my (Corresponding Author)

³ Pensyarah, Jabatan Bahasa Arab, Fakulti Pengajian Islam, Universiti Islam Antarabangsa Tuanku Syed Sirajuddin (UniSIRAJ).
Email: mohdroshimi@unisiraj.edu.my

⁴ Pensyarah Kanan (Dr.), Jabatan Perniagaan, Fakulti Perniagaan dan Sains Pengurusan, Universiti Islam Antarabangsa Tuanku Syed Sirajuddin (UniSIRAJ).
Email: kartina@unisiraj.edu.my

⁵ Pensyarah Kanan (Dr.), Jabatan Muamalat, Fakulti Muamalat dan Kewangan Islam, Universiti Islam Antarabangsa Tuanku Syed Sirajuddin (UniSIRAJ).
Email: khalilah@unisiraj.edu.my

designed to function by receiving the name of a surah as input from the user, extracting the relevant tafsir content, summarizing the key teachings, and delivering explanations that are concise, coherent, and easy to understand. The study's methodology consists of several main phases, including the selection of a tafsir dataset (based on Tafsir al-Azhar by HAMKA), followed by linguistic pre-processing, machine learning model development, implementation of the NLP pipeline, and a fact-verification procedure to ensure the accuracy and reliability of the generated output. The proposed approach not only facilitates greater access to tafsir knowledge but also empowers digital Islamic education through a system that is both responsive and adaptive.

Keywords: *Artificial Intelligent (AI); Natural Language Processing (NLP); Machine Learning (ML); Semi-Supervised Learning (SSL)*

ABSTRAK

Al-Quran menjadi sumber utama panduan hidup umat Islam, namun kefahaman mendalam terhadap pengajaran ayat masih terhad kepada karya-karya muftakar yang statik dan memerlukan carian manual. Keterbatasan ini menimbulkan masalah pengguna sukar mengakses pengajaran ayat secara sistematik dan interaktif. Kajian ini mencadangkan pembangunan satu model Kecerdasan Buatan (AI) yang menggunakan pendekatan Pemprosesan Bahasa Semulajadi (NLP) dan Pembelajaran Mesin (ML) untuk menjana secara automatik faedah dan pengajaran daripada surah-surah al-Quran. Model ini akan berfungsi dengan menerima input nama surah daripada pengguna, seterusnya mengekstrak kandungan tafsir yang berkaitan, merumuskan intipati pengajaran, dan menyampaikan penjelasan yang bersifat ringkas, padat dan mudah difahami. Metodologi kajian ini merangkumi beberapa fasa utama yang melibatkan pemilihan set data tafsir (berdasarkan Tafsir al-Azhar oleh HAMKA), diikuti pra-pemprosesan linguistik, pembangunan model pembelajaran mesin, pelaksanaan saluran pemprosesan NLP (NLP pipeline), serta prosedur pengesahan fakta untuk memastikan ketepatan dan kesahihan output yang Dijana. Pendekatan yang dicadangkan ini bukan sahaja dapat memudahkan akses kepada ilmu tafsir, bahkan mampu memperkasa pendidikan Islam digital melalui sistem yang bersifat responsif dan adaptif.

Kata Kunci: *Kecerdasan Buatan (Artificial Intelligent (AI); Pemprosesan Bahasa Semulajadi (Natural Language Processing (NLP); Pembelajaran Mesin (Machine Learning (ML); Pembelajaran Separuh Diselia (Semi-Supervised Learning (SSL)*

1. PENDAHULUAN

Al-Quran merupakan sumber utama dalam Islam yang mengandungi panduan menyeluruh untuk kehidupan manusia. Hal ini sebagaimana firman Allah SWT:

كِتَابٌ أَنْزَلْنَاهُ إِلَيْكَ مُبَارَكٌ لِيَدَّبَّرُوا آيَاتِهِ وَلِيَتَذَكَّرَ أُولُوا الْأَلْبَابِ

Terjemahan: “Ini ialah sebuah kitab yang Kami turunkan kepadamu penuh berkat supaya mereka memperhatikan ayat-ayatnya dan supaya mendapat pelajaran orang-orang yang mempunyai fikiran” (Surah Sad (38): 29)

Jelas menunjukkan bahawa tujuan utama al-Quran diturunkan adalah untuk difahami dan dihayati kandungannya. Namun begitu, memahami mesej yang tersirat di sebalik ayat-ayat al-Quran memerlukan rujukan kepada tafsir yang muktabar. Para sarjana tafsir seperti Abu Ja'far Muhammad Ibn Jarir al-Tabari (m. 310 H), Abu 'Abd Allah al-Qurtubi (m. 671 H), Abu al-Fida' Isma'il bin 'Umar bin Kathir (m. 774 H) dan ramai lagi telah menyumbang penulisan tafsir yang kaya dengan dimensi nilai, hukum-hakam dan linguistik. Seiring dengan kemajuan teknologi, akses bagi kandungan karya-karya seperti ini perlu diluaskan bukan hanya melalui platform fizikal, tetapi juga digital. Menurut Sati et al (2025), platform tafsir digital memberikan akses yang lebih cepat dan luas kepada umat Islam seluruh dunia. Dalam konteks ini, pelbagai teknologi moden turut menawarkan potensi signifikan dalam memperkukuh kefahaman terhadap tafsir al-Quran. Seperti teknologi *Big Data* membantu penyelidik dalam bidang tafsir mengenalpasti isu-isu utama masyarakat dan mengesan salah faham umum tentang al-Quran (Kawtharany, 2017). Demikian juga teknologi-teknologi lain seperti *Internet of Thing (IoT)*, *Augmented Reality (AR)* dan *Virtual Reality (VR)* yang berpotensi membantu pengguna mendalami al-Quran dan mengakseskan pengajaran yang terkandung di dalamnya. Malah, lebih ke hadapan, medium Kecerdasan Buatan (*Artificial Intelligence (AI)*) itu sendiri memberi peluang yang besar dalam kajian al-Quran, khususnya dalam menghasilkan tafsir yang mesra dan lebih mudah diakses (Purnawati, 2025; Najib & Kurjum 2025; Javed & Bibi, 2024).

Dalam era digital kini, terdapat peningkatan minat dalam kalangan masyarakat terhadap penggunaan teknologi untuk memudahkan pemahaman kandungan al-Quran, terutamanya melalui pendekatan tafsir yang lebih interaktif, mudah diakses dan sesuai dengan gaya pembelajaran moden (Sati et al, 2025; Purnama, 2025; Madadzadeh & Bahariniya, 2024). Ini sekaligus membuka ruang kepada pembangunan sistem pintar yang mampu membantu pengguna menjana faedah dan pengajaran daripada ayat-ayat al-Quran secara lebih sistematik. Dalam konteks ini, penggunaan teknologi moden seperti *AI* dilihat sebagai satu alat bantu dalam memahami dan mentafsir al-Quran menerusi struktur linguistik, hubungan semantik dan pola tematik (Madadzadeh & Bahariniya, 2024). Jelas Sati et al (2025), melalui teknologi *AI*, khususnya model *Machine Learning (ML)*, masyarakat awam mampu mengakses penjelasan al-Quran secara interaktif, sekalipun tanpa memahami bahasa Arab atau disiplin ilmu tafsir secara mendalam.

Walaupun pelbagai sumber platform tafsir al-Quran digital telah dibangunkan, kebanyakannya masih bersifat statik dan memerlukan pencarian manual oleh pengguna. menggariskan beberapa keterbatasan *AI* dalam tafsir al-Quran, termasuklah tidak mampu memahami konteks agama, sejarah dan budaya secara lebih mendalam (al-Janabi, 2024; Najib & Kurjum, 2025) serta terdedah kepada bias budaya dan gender atau ideologi tertentu (Najib & Kurjum, 2025; Mubarik & Sakina, 2024). Malah, sistem sedia ada juga tidak mengintegrasikan *AI* untuk menganalisis kandungan tafsir dengan tepat (Bashir et al, 2023) dan merumuskan faedah serta pengajaran utama secara automatik dan tersusun.

Artikel ini mengemukakan cadangan pembangunan satu Model Kecerdasan Buatan (*AI*) yang menggabungkan pendekatan Pemprosesan Bahasa Semulajadi (*Natural Language Processing (NLP)*) dan Pembelajaran Mesin (*Machine Learning (ML)*) untuk menjana secara automatik faedah serta pengajaran daripada surah-surah al-Quran. Model yang dicadangkan akan berfungsi dengan menerima input nama surah daripada pengguna, kemudian mengekstrak kandungan tafsir yang relevan, seterusnya merumuskan intipati pengajaran dalam bentuk yang ringkas, padat, dan mesra pengguna. Pendekatan ini diharapkan dapat mengatasi keterbatasan sistem tafsir digital sedia ada yang hanya menyediakan teks statik tanpa penyesuaian interaktif.

2. SOROTAN LITERATUR

2.1 Peranan Kecerdasan Buatan (AI) Dalam Pemahaman dan Tafsir Al-Quran

Kemajuan dalam bidang Kecerdasan Buatan (AI) telah mencetuskan gelombang transformasi dalam pelbagai bidang, termasuk pendidikan Islam dan pengajian al-Quran. Dalam konteks tafsir al-Quran, AI dapat digunakan untuk menganalisis serta membandingkan pelbagai tafsiran, di samping membantu umat Islam untuk memahami berbagai perspektif serta interpretasi (Fitriani et al, 2025). Jelas Putra (2024), AI dapat membantu identifikasi tema-tema penting dalam al-Quran serta menyediakan data komprehensif yang bermanfaat kepada pengkaji dalam bidang tafsir. Ali & Isnaini (2024) menjelaskan digitalisasi tafsir tidak hanya soal teknologi, tetapi mengubah cara masyarakat berinteraksi dan menerima wahyu Tuhan secara dinamik dalam kehidupan harian mereka. Nirwana (2024), turut menyimpulkan multimedia tafsir membawa nilai yang signifikan dalam memperluaskan cara umat Islam berinteraksi dengan al-Quran secara digital dan interaktif. Malah, Putra & Yusuf (2021) menjelaskan penggunaan teknik *ML* dan *NLP* dapat membantu dalam memahami ayat-ayat al-Quran dengan lebih objektif dan kontekstual. Dalam hal ini, Putra & Yusuf (2021) menegaskan masalah utama dalam tafsir adalah subjektiviti, iaitu pengaruh budaya, ideologi dan latar belakang ahli tafsir. Justeru, AI dicadangkan kerana “*alat neutral*” yang bebas dari bias manusia (Putra & Yusuf, 2021).

Selain itu, Basid et al (2024) turut mencadangkan integrasi AI sebagai fasilitator dalam interpretasi kitab suci al-Quran. Kajian mereka menekankan keupayaan AI untuk mengurangkan bias tafsir boleh meningkatkan kebolehpercayaan dalam penyampaian mesej al-Quran, terutamanya dalam konteks pendidikan dan dakwah digital. Seiring pandangan ini, Madadzadeh & Bahariniya (2024) turut menegaskan struktur dan corak dalam ayat-ayat al-Quran dapat dikenalpasti melalui algoritma tertentu, sekaligus menjadikan proses pemahaman lebih sistematik dan terarah.

Bagaimanapun, dari sudut yang lain, AI juga tidak terlepas dari bias manusia, kerana ia dilatih dengan data yang dikumpulkan oleh manusia dan kemungkinan boleh berlaku bias dan interpretasi ayat yang salah (Hm & Satra, 2024). Justeru, beliau mencadangkan agar algoritma AI perlu dikembangkan dengan mengambilkira konteks dan makna al-Quran yang benar, serta mempertimbangkan interpretasi dari ahli tafsir (Hm & Satra, 2024). Seiring pandangan ini, Tahir & Suswandi (2024) menekankan keperluan memastikan sistem AI yang dibangunkan perlu selari dengan prinsip akidah Islam dan model sistem tafsir al-Quran itu mestilah mengutamakan nilai keislaman dan sensitiviti

terhadap wahyu agar tidak tersasar daripada kerangka Syariah. Sehubungan itu, integrasi *AI* dalam konteks tafsir bukan semata-mata isu teknikal, tetapi turut melibatkan tanggungjawab etika dan epistemologi Islam.

2.2 Kelemahan Sistem Digital Sedia Ada dan Keperluan Inovasi Interaktif

Dalam era digital, teknologi telah mempengaruhi cara mengakses, memahami dan mengamalkan al-Quran dalam kehidupan seharian (Alfarizi & Farabi, 2024). Meskipun begitu, Abdul Rouf (2025) dan Sati et al (2025) menjelaskan tafsir yang dihasilkan oleh *AI* sering menyimpang dari tafsir klasik dan tidak memenuhi piawaian disiplin *Ulum al-Quran*. Annur & M. Ridwan (2024) turut menjelaskan *AI* berpotensi bias kepada pandangan dominan dan mengabaikan minoriti. Malah, jelas Mukmin et al (2025), tafsir yang dihasilkan oleh *AI* bersifat sintetik, tidak menimbang kontekstual dan seringkali mengabaikan prinsip utama dalam ilmu tafsir. Sebagai contoh, ayat yang berbunyi: “*dhalika al-kitab la rayba fih*” (Surah al-Baqarah (2): 2) dijelaskan secara literal seperti “*ini adalah kitab yang tidak diragukan kebenarannya*”. Menurut Mukmin et al (2025), penjelasan yang diberikan ini tidak disertai dengan ulasan mendalam tentang konteks wahyu, perbezaan pandangan dalam kalangan sarjana serta kesannya terhadap akidah Islam.

Selain itu, dalam memberikan jawapan kepada persoalan tafsir, *AI* seringkali tidak menampilkan sanad, penilaian terhadap sumber atau kedalaman kontekstual yang biasa ditemui dalam karya klasik seperti Tafsir al-Tabari dan al-Qurtubi (Mukmin et al, 2025). Hal ini secara tidak langsung berpotensi menyesatkan umat Islam khususnya dalam kalangan awam apabila menerima mentah-mentah tanpa melakukan verifikasi dari ahlinya. Justeru, bagi memastikan kualiti ilmu terjaga, Nirwana (2024) mencadangkan agar penggunaan multimedia tafsir perlu diberi perhatian serius terhadap sumber yang berautoriti dan termasuklah kebijaksanaan masyarakat itu sendiri dalam menggunakannya.

Terdapat beberapa website digital seperti *Tarteel.ai* yang menawarkan integrasi antara al-Quran dan *AI*. *Tarteel.ai* dibangunkan sebahagian daripada aplikasi *Quran.com*. Secara umumnya, *Quran.com* menyediakan kandungan al-Quran termasuklah terjemahan, tafsir dan pelajaran-pelajaran tentang hukum hakam Islam (Alfarizi & Farabi, 2024). Manakala *Tarteel.ai* pula adalah sebuah aplikasi hafazan al-Quran berkuasa *AI* yang memberi tumpuan kepada pembelajaran bacaan dan hafazan. Aplikasi ini menawarkan beberapa fungsi pintar seperti pengesanan bacaan yang membolehkan pengguna mengetahui

kesalahan bacaan secara terus dan carian berasaskan suara. Selain itu, *Tarteel.ai* turut menyertakan ciri penjejakan kemajuan yang membantu pengguna menilai prestasi hafazan mereka dari semasa ke semasa. Gabungan ciri ini menjadikan *Tarteel.ai* sebagai alat yang praktikal untuk individu yang ingin menghafaz dan memperbaiki bacaan al-Quran secara lebih sistematik.

Rajah 1.0

Paparan Tarteel.ai, capaian dari <https://tarteel.ai>



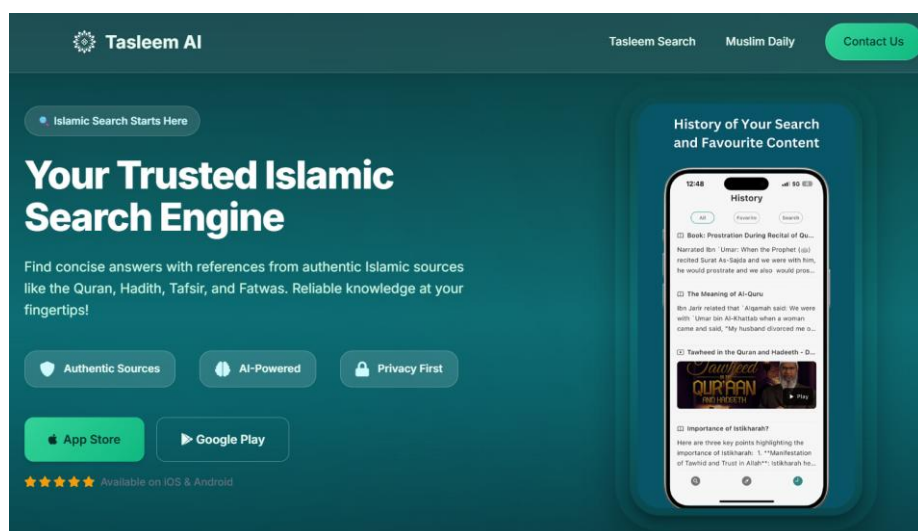
Kajian oleh Tarshany & Zeki (2024) menekankan bahawa aplikasi berasaskan teknologi seperti *Tarteel.ai* berperanan penting dalam memperkuat pembelajaran al-Quran melalui teknologi pengenalan suara dan *AI* yang menyokong hafazan serta bacaan. Sekalipun *Tarteel.ai* mempunyai kekuatan pada aspek bacaan dan hafazan, namun skopnya masih terhad kepada pemprosesan audio dan pengiktirafan pertuturan (*speech recognition*) sahaja. Ia tidak dibangunkan sebagai sistem tafsir, justeru tidak menyokong keupayaan untuk menjana ringkasan pengajaran atau menilai kandungan tafsir secara semantik. Kelemahan ini menjadikan *Tarteel.ai* kurang sesuai digunakan bagi tujuan mendalami makna ayat atau memperoleh pengajaran secara langsung,

kerana fokus utamanya adalah pada bacaan dan hafazan, bukan interpretasi kepada ayat al-Quran.

Selain itu, aplikasi lain yang menawarkan integrasi antara al-Quran dan *AI* adalah *Tasleem.ai*. Pada penelitian penulis, aplikasi ini merupakan sebuah aplikasi carian Islamik berasaskan *AI* yang membolehkan pengguna mendapatkan jawapan daripada al-Quran, hadis, tafsir, dan fatwa secara cepat dan mudah. Aplikasi ini menekankan ketepatan sumber dengan hanya menggunakan rujukan sahih, sekali gus memberi keyakinan kepada pengguna terhadap maklumat yang dipaparkan. Selain itu, ia menyediakan penjelasan yang ringkas dan padat serta jelas, sehingga menjadikan ilmu Islam itu lebih mudah diakses oleh masyarakat umum tanpa memerlukan bacaan panjang. Ciri privasi mesra pengguna serta bebas iklan menjadikan pengalaman penggunaan lebih fokus dan selesa. Walau bagaimanapun, penulis mendapati kelemahan utama *Tasleem.ai* ialah ia masih berfungsi secara asas sebagai enjin carian dan ringkasan, tanpa kemampuan untuk menjana tafsir interaktif atau menghasilkan penyesuaian pengajaran secara automatik mengikut konteks pengguna. Justeru, walaupun *Tasleem.ai* menunjukkan potensi besar dalam memperkasa akses ilmu Islam, terdapat keperluan untuk pengembangan ke arah sistem yang lebih interaktif, adaptif, dan berasaskan pembelajaran mesin yang mendalam.

Rajah 2.0:

Paparan Tasleem.ai, capaian dari <https://tasleem.ai>



Sehubungan daripada itu, sebuah model lain ingin dicadangkan agar sistem tafsir al-Quran digital, khususnya pengajaran ayat secara teratur dan sisematik dapat diakses dengan lebih efisien. Kajian oleh Khidhir (2024) menunjukkan bahawa algoritma *parsing* dalam *NLP* boleh digunakan untuk mengenal pasti struktur ayat al-Quran, sekaligus boleh menjadi asas penting dalam pembinaan sistem tafsir adaptif. Justeru, pembangunan sistem tafsir digital perlu bergerak ke arah yang lebih bersifat dialogik dan dinamik, di mana pengguna boleh berinteraksi dengan kandungan dan menerima maklum balas kontekstual yang dijana secara automatik.

Berbeza dengan *Tarteel.ai* dan *Tasleem.ai*, model tafsir *AI* yang dicadangkan dalam kajian ini memberi tumpuan kepada penjaanaan pengajaran dan faedah surah berdasarkan teks tafsir berautoriti. Model ini menggunakan pendekatan *NLP* untuk melaksanakan proses seperti pengekstrakan entiti, analisis tema, analisis sentimen keagamaan, serta penjaanaan ringkasan ayat. Selain itu, ia juga berpotensi sebagai pengesahan fakta bagi menjaga integriti disiplin ilmu tafsir itu sendiri. Jika *Tarteel.ai* dan *Tasleem.ai* memanfaatkan *AI* untuk membantu hafazan melalui bacaan dan sebagai enjin carian Islamik, tetapi model yang dicadangkan ini pula memanfaatkan *AI* untuk membantu pemahaman dengan menyediakan tafsir yang ringkas, padat dan interaktif.

2.3 Integrasi NLP dan ML dalam AI untuk Penjaanaan Pengajaran Ayat Al-Quran

Keperluan terhadap sistem *AI* yang lebih canggih adalah diperlukan bagi menyokong proses penjaanaan pengajaran ayat al-Quran secara sistematik, efisien dan interaktif. Menurut Sridefi et al. (2024), pemprosesan teks suci seperti al-Quran memerlukan pendekatan khusus, dengan mengambil kira sensitiviti semantik dan struktur linguistik al-Quran yang kompleks. Sehubungan itu, Mahdi et al (2021) mencadangkan penggunaan enjin carian yang lebih dinamik bagi memudahkan capaian kepada kandungan tafsir yang relevan.

Kajian ini menawarkan sistem Pemprosesan Bahasa Semulajadi (*Natural Language Processing (NLP)*) dan Pembelajaran Mesin (*Machine Learning (ML)*) kerana ia dilihat berpotensi dalam pemprosesan teks keagamaan seperti al-Quran dan tafsir. Menurut Chowdhary (2020), *NLP* merupakan salah satu cabang utama Kecerdasan Buatan (*AI*) yang berfokus kepada interaksi antara komputer dan bahasa manusia. Seiring pandangan ini, Putra & Yusuf (2021) dan Le Glaz et al (2021) menjelaskan *NLP* adalah subdisiplin sains komputer yang menganalisis bahasa manusia untuk mengkategorikan data teks yang

kemudiannya boleh digunakan oleh algoritma *ML*. Manakala *ML* pula aplikasi *AI* yang menggunakan teknik statistik untuk membolehkan computer “belajar” daripada data bagi menghasilkan model automatik tanpa pengaturcaraan yang eksplisit (Putra & Yusuf, 2021; Le Glaz et al, 2021). Integrasi dua sistem ini membolehkan analisis data bahasa yang tidak berstruktur dilaksanakan secara besar-besaran, di mana *NLP* menyediakan data teks dan *ML* pula menggunakan data tersebut untuk membuat klasifikasi atau ramalan (Putra & Yusuf, 2021).

Dalam konteks kajian ini, sistem yang akan dibangunkan ini dijangka berupaya melaksanakan pelbagai proses secara automatik untuk membantu pengguna memahami pengajaran utama dalam ayat al-Quran. Proses-proses ini termasuklah pengekstrakan entiti (*Named Entity Recognition*), analisis tema, penilaian sentimen keagamaan dan penjana ringkasan ayat secara abstrak (*abstractive summarization*) dapat dilakukan secara automatik untuk membantu pengguna memahami mesej utama dalam ayat tertentu. Malah lebih utama, ia berpotensi menyediakan teknik penyemakan fakta (*fact-checking*) dan semakan rujukan silang (*cross-referencing*) dengan sumber tafsir asal.

Kerangka cadangan sebegini adalah selari dengan cadangan yang dikemukakan oleh Putra & Yusuf (2021) bahawa model *ML* sesuai dilaksanakan untuk tafsir al-Quran, di mana penyelidik membangunkan kerangka, memilih teknik analisis semantik dan mengaplikasikan *NLP* pada teks tafsir untuk menghasilkan tafsiran yang objektif, mendalam dan bebas daripada penyelewengan.

3. METODOLOGI

Kajian ini menggunakan pendekatan *Multi-Stage NLP framework* yang menggabungkan antara sistem *NLP* dan *ML* bagi membina sistem *AI* yang dapat menjana dan mengesahkan pengajaran surah al-Quran secara automatik daripada teks tafsir bahasa Melayu. Kaedah ini merangkumi beberapa fasa utama yang melibatkan beberapa proses iaitu Pemetaan dan Penyusunan Data Tafsir (*Tafsir Data Acquisition*), Pra-pemprosesan Linguistic (*Linguistic Preprocessing*), Pembangunan Model Pembelajaran Mesin (*Machine Learning Model Development*) dan Proses Pengesahan Fakta (*Fact Verification Process*).

Kajian ini menggunakan data tafsir al-Quran yang diperoleh daripada sumber muktabar iaitu Tafsir al-Azhar yang ditulis oleh Haji Abdul Malik bin Karim Amrullah (HAMKA) (m. 1401 H). Karya ini ditulis dalam bahasa Melayu yang jelas, moden dan mudah difahami oleh masyarakat. Keunikan ini

menjadikannya sesuai dijadikan set data kerana sistem *AI* yang dibina dalam kajian ini juga mensasarkan pengguna dalam kalangan mereka yang berbahasa Melayu. Banyak tafsir muktabar lain ditulis dalam bahasa Arab klasik (seperti Tafsir al-Tabari, Tafsir Ibn Kathir dan lain-lain), dan memerlukan kerja penterjemahan tambahan. Sehubungan itu, pemilihan Tafsir al-Azhar dilihat tepat kerana karya ini telah menyediakan asas linguistik yang lebih relevan, praktikal dan mesra pengguna tempatan.

Dalam konteks kajian berasaskan *NLP*, struktur set data yang kemas dan konsisten adalah asas penting untuk memastikan kejayaan fasa pemodelan. Struktur set data terdiri daripada Pembinaan Teks Bertanda (*Tagged Text*). Set data tafsir disusun dalam bentuk teks bertanda (*Annotated Text*), iaitu setiap bahagian tafsir dikaitkan dengan *metadata* utama seperti nama surah (untuk mengenal pasti surah mana tafsir itu merujuk), nombor ayat (supaya sistem dapat menghubungkan tafsir dengan ayat tertentu dalam al-Quran) dan sumber rujukan asal (memelihara integriti akademik dengan menunjukkan tafsir asal). Struktur ini membolehkan set data mempunyai susun atur hierarki yang jelas (Surah → Ayat → Tafsir), sekali gus memudahkan pemprosesan dan analisis oleh algoritma *NLP*.

Sebelum pembangunan model, set data perlu melalui proses Pra-Pemprosesan Linguistik (*Linguistic Preprocessing*) yang merangkumi Pembersihan Teks (*Text Cleaning*), Penyeragaman Format *Unicode* (*Unicode Normalization*), Pemisahan Ayat (*Sentence Segmentation*) dan Penyingkiran Elemen Bukan Linguistik (*Removal of Non-Linguistic Elements*). Keseluruhan proses ini memastikan set data tafsir yang digunakan adalah bersih, konsisten, dan berstruktur, sesuai untuk dilatih dalam model *NLP*. Dengan struktur metadata yang jelas, teks yang seragam, dan bebas daripada gangguan linguistik, model *AI* dapat mengekstrak tema, entiti, dan pengajaran surah dengan lebih tepat serta menjana output yang mesra pengguna.

Menurut Van Engelen & Hoos (2020), *ML* memainkan peranan penting dalam menganalisis data berskala besar, namun keberkesanannya sering dipengaruhi oleh keterbatasan data berlabel. Bagi mengatasi isu ini, Pendekatan Pembelajaran Separuh Diselia (*Semi-Supervised Learning (SSL)*) diperkenalkan dengan menggabungkan sebahagian kecil data berlabel bersama sejumlah besar data tidak berlabel dalam proses latihan model. Kajiannya menunjukkan bahawa *SSL* mampu meningkatkan prestasi model pembelajaran mesin dengan kos anotasi yang lebih rendah, menjadikannya sangat relevan dalam domain yang kekurangan data berlabel, seperti teks keagamaan atau tafsir al-Quran.

Pendekatan ini bukan sahaja meningkatkan ketepatan model, tetapi juga membolehkan sistem *AI* beradaptasi dalam dunia sebenar di mana data berlabel sukar diperoleh.

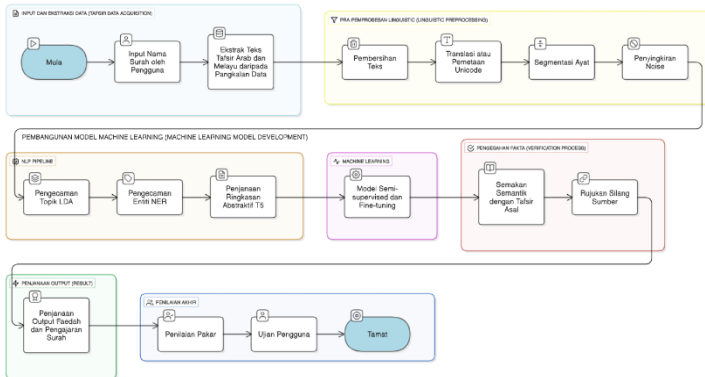
Maka bagi kajian ini, tujuan *ML* dan *SSL* digunakan untuk memaksimumkan keberkesanan model dalam keadaan set data yang berlabel (*labelled data*) adalah terhad. Sebahagian ayat tafsir dipasangkan secara manual dengan anotasi (*annotation*) pengajaran oleh pakar untuk dijadikan label bagi model, manakala sebahagian besar data tidak berlabel digunakan untuk pelatihan lanjutan. Anotasi (*annotation*) merujuk kepada proses pelabelan data dengan maklumat bermakna yang boleh digunakan untuk melatih model. Model asas yang digunakan ialah *Transformer-Based Architecture* (seperti *BERT* dan *T5*), yang kemudiannya menjalani proses *Fine-Tuning* agar lebih serasi dengan gaya bahasa keagamaan dan struktur tafsir. Proses Penalaan (*Fine-Tuning*) ini melibatkan penggunaan teknik *Backpropagation* ke atas lapisan akhir model menggunakan set data tafsir yang telah dipra-anotasi (*Pre-annotated*), iaitu merujuk kepada data yang telah dilabel terlebih dahulu sebelum dimasukkan ke dalam proses latihan model, sama ada oleh manusia atau melalui sistem separa automatik.

Pipeline NLP direka bentuk secara modular yang merangkumi tiga komponen utama: Pengesanan Topik Ayat (*Topic Modeling*), Pengesanan Entiti Bernama (*Named Entity Recognition - NER*), dan Penjanaan Ringkasan Abstraktif. Untuk pengesanan topik, algoritma *LDA (Latent Dirichlet Allocation)* digunakan bagi mengenal pasti tema utama setiap ayat tafsir. Komponen *NER* pula dibina menggunakan model yang dilatih untuk mengenal pasti entiti khusus dalam konteks Islam seperti nama nabi, tempat, dan konsep syariah. Bagi penjanaan ringkasan, model abstraktif yang disesuaikan (*Fine-Tuned T5*) digunakan untuk menghasilkan rumusan padat berkenaan pengajaran utama berdasarkan konteks ayat dan tema yang dikenal pasti.

Satu elemen penting dalam metodologi ini ialah proses Pengesahan Fakta (*Fact-Checking*) untuk memastikan ketepatan dan kesahihan output yang dijana. Komponen ini menggunakan pendekatan pengesahan semula terhadap sumber asal tafsir melalui kaedah semakan semantik. Model transformer digunakan untuk membandingkan kesepadanan semantik antara ringkasan yang dijana dengan kandungan asal tafsir. Sekiranya wujud ketidaksepadanan, model akan menandakan output sebagai perlu disemak semula. Di samping itu, pendekatan *Knowledge-Based Verification* turut diguna pakai dengan menyemak rujukan silang (*Cross-Referencing*) terhadap pangkalan data tafsir.

Rajah 3.0

Rajah aliran proses bagi metodologi kajian, yang merangkumi langkah-langkah utama daripada input surah oleh pengguna sehingga penjanaaan output berbentuk faedah dan pengajaran surah melalui pemprosesan NLP dan model Machine Learning.



Akhir sekali, sistem yang dibangunkan diuji melalui satu set senario input melibatkan nama surah, di mana pengguna akan menerima hasil ringkasan pengajaran dan faedah bagi surah tersebut. Hasil output dibandingkan dengan penilaian pakar tafsir untuk menilai ketepatan, kesesuaian konteks, dan kefahaman pengguna akhir. Protokol penilaian menggunakan metrik seperti *ROUGE* untuk ringkasan, serta *F1-score* untuk komponen pengesanan entiti dan topik. Metodologi yang dirancang ini bertujuan untuk menghasilkan satu model *AI* yang bukan sahaja berasaskan sains data yang kukuh, malah selari dengan disiplin keilmuan Islam yang sah dan dipercayai.

4. PERBINCANGAN

Berdasarkan kepada pembangunan model ini, terdapat empat fasa yang dicadangkan iaitu seperti berikut:

Fasa 1: Pemetaan Dan Penyusunan Data dari Kitab Tafsir Muktabar (Tafsir Data Acquisition)

Berdasarkan proses pemilihan data tafsir (*tafsir data acquisition*) yang bersumber daripada rujukan muktabar dalam bahasa Melayu seperti Tafsir al-Azhar yang ditulis oleh HAMKA. Tafsir ini dipilih berdasarkan ketulenan, kesahihan dan penerimaan umum dalam kalangan masyarakat Islam di Nusantara, bukan sahaja di Indonesia, bahkan merentasi hampir seluruh rantau Asia Tenggara khususnya

di Malaysia, Singapura, Brunei, Thailand (Yunus & Yusof, 2021). Jelas Bahar et al (2019), Tafsir al-Azhar adalah tafsir yang mendapat perhatian di Nusantara kerana olahan bahasanya yang mudah difahami terutamanya di negara yang menggunakan bahasa Melayu. Jelas Rouf dan Yusoff (2013) pula, HAMKA telah memaparkan penafsiran yang terperinci, penuh dengan deskripsi, serta membawa contoh dari pengalamannya sendiri dengan masyarakat sekeliling. Hal ini dikuatkan lagi dengan kajian Alviyah (2016) yang menjelaskan Tafsir al-Azhar memperlihatkan corak *adabi ijtima'i* yang menonjol, disertai keindahan bahasa Melayu yang diolah selaras dengan konteks sosial masyarakat pada zamannya. Penggunaan teknik bahasa yang pelbagai dalam penyampaian tafsirnya menampilkan gaya tutur yang dekat dengan bahasa pertuturan harian, menjadikannya relatif mudah difahami oleh khalayak pembaca. Justeru, tafsir ini bersesuaian untuk diangkat dalam kajian ini di mana sasaran penggunaannya adalah dalam kalangan masyarakat nusantara.

Hasil jangkaan utama dari proses ini adalah pembentukan set data tafsir yang bersih, tersusun dan boleh dilatih (*trainable*) untuk tujuan pemodelan *AI*. Data tafsir ini dijangka mengandungi struktur yang konsisten dari segi format *metadata* (nama surah, nombor ayat, sumber tafsir), serta kandungan teks yang telah disaring daripada gangguan linguistic seperti pengulangan, huraian berlebihan dan elemen luar konteks. Struktur data ini akan membolehkan proses pembelajaran mesin berlaku secara lebih terarah dan menghasilkan output yang lebih tepat dan berkualiti tinggi apabila diaplikasikan dalam konteks penjaan pengajaran ayat.

Fasa 2: Pra-Pemprosesan Linguistik (Linguistic Preprocessing)

Fasa pra pemprosesan linguistik (*Linguistic Preprocessing*) merupakan antara prasyarat teknikal penting sebelum data tafsir dimasukkan ke dalam sistem *NLP*. Menurut Pekar (2024), pra pemprosesan linguistik merujuk kepada langkah awal pemprosesan linguistik yang digunakan untuk menyediakan representasi kata (*word representation*) sebelum ia dimasukkan ke dalam model agihan (*distributional model*). Jelas Pekar (2024) lagi, tujuan proses ini adalah untuk mengekstrak, manapis dan mengubah data linguistik supaya lebih signifikan untuk klasifikasi semantik. Hasil jangkaan daripada fasa Pra-Pemprosesan Linguistik ini, data dijangka mencapai tahap keseragaman format dan kebersihan linguistik yang membolehkan sistem *NLP* mengekstrak makna semantik dan sintaktik dengan lebih berkesan. Dengan pelaksanaan teknik seperti *Unicode Normalization*, *Sentence Segmentation* dan *Noise Filtering*. Sistem akan dapat membezakan antara kandungan penting dan elemen tidak berkaitan dalam teks

tafsir. Ini akan memberi kesan langsung kepada prestasi model *NLP* dalam mengenal pasti entiti bernama, topik utama ayat, serta menghasilkan ringkasan yang relevan. Proses ini dijangka mengurangkan risiko kesilapan konteks dan meningkatkan kualiti jawapan yang dihasilkan secara automatik (Madadzadeh dan Bahariniya, 2024; Basid et al., 2024).

Kajian Chowdhary (2020) turut menyatakan normalisasi teks, tokenisasi dan pembersihan data merupakan asas penting untuk memastikan keberkesanan *NLP* dalam memahami makna semantik. Tambahan pula, Touati-Hamad, Laouar, & Bendib (2020) menjelaskan pra-pemprosesan teks al-Quran perlu dilakukan dengan teliti kerana ia melibatkan elemen linguistik Arab yang kompleks dan sarat dengan makna kontekstual. Ini bermakna, gabungan teknik seperti *Unicode Normalization*, *Sentence Segmentation*, dan *Noise Filtering* bukan sahaja menambah tahap keseragaman teks, tetapi juga meningkatkan keupayaan model untuk mengekstrak makna tafsir dengan lebih tepat.

Fasa 3: Pembangunan Model ML

Fasa pembangunan model *ML* ialah proses sistematik untuk membina, melatih, menilai dan menguruskan model pembelajaran mesin (Lorenzoni, 2021). Proses ini penting bagi menjamin kualiti, keberkesanan, kebolehselenggaraan, dan keberhasilan aplikasi pembelajaran mesin pada skala yang besar (Lorenzoni, 2021). Dalam konteks kajian ini, pendekatan yang digunakan ialah *ML* berasaskan seni bina *Transformer (Machine Learning Model Development)* seperti *BERT* atau *T5* yang menjalani proses *Fine-Tuning* ke atas set data tafsir. Penggunaan senibina berasaskan *Transformer* seperti *BERT* atau *T5*, telah terbukti berkesan dalam tugas *NLP* kerana keupayaannya memahami konteks linguistik yang kompleks (Devlin et al., 2019; Raffel et al., 2020).

Hasil jangkaan ialah keupayaan model untuk menjana output pengajaran dan faedah utama bagi setiap surah secara automatik. teknik *SSL* membolehkan model belajar daripada gabungan data berlabel (yang dianotasi pakar) dan tidak berlabel secara serentak, sekaligus menjimatkan kos anotasi manual sepenuhnya. Model cadangan dijangka berupaya menghasilkan output yang bersifat padat, tersusun mengikut tema, dan sesuai digunakan oleh pelajar, pendidik, serta masyarakat awam untuk memahami mesej utama surah dengan lebih cepat dan mudah. Dalam kajian ini, pendekatan *SSL* dipilih kerana ia mampu memanfaatkan data berlabel yang terhad secara bersama dengan data tidak berlabel, sekaligus menjimatkan kos anotasi manual serta meningkatkan prestasi model (Van Engelen & Hoos, 2020). Melalui proses *Fine-Tuning*, model

dijangka dapat menghasilkan ringkasan tafsir yang lebih padat, tepat, dan sesuai untuk menyampaikan pengajaran utama surah kepada pelajar, pendidik, serta masyarakat awam dengan lebih efektif (Putra & Yusuf, 2021).

Fasa 4: Proses Pengesahan Fakta (Fact Verification Process)

Fasa terakhir adalah Proses Pengesahan Fakta (*Fact Verification Process*). Ini adalah proses untuk menentukan sama ada sesuatu kenyataan itu benar, salah, atau tidak dapat dipastikan dengan merujuk kepada sumber dan bukti (Guo et al, 2022). Proses ini begitu penting kerana ia bertindak antaranya sebagai pembanteras maklumat palsu yang telah dikumpulkan sebelum ini (Guo et al, 2022). Kaedah ini penting kerana ia mencegah penyebaran maklumat palsu atau tafsiran yang menyimpang (Lazarski, Al-Khassaweneh, & Howard, 2021). Dalam konteks kajian ini, sistem pengesahan akan menggunakan semakan semantik dan semakan silang terhadap teks tafsir asal bagi memastikan kesahihan output. Tegas Thorne et al (2018), verifikasi fakta perlu berasaskan perbandingan teks dan bukti yang kukuh. Sistem juga dijangka dapat menapis output yang tidak sah atau tidak konsisten. Ini akan meningkatkan kebolehpercayaan sistem dalam konteks penyampaian ilmu agama, selaras dengan prinsip penyelidikan Islam yang menitikberatkan kesahihan sumber. Output yang telah disahkan ini akan menjadi asas kepada paparan akhir kepada pengguna dalam bentuk visualisasi ringkasan pengajaran.

Hasil akhir dari cadangan ini ialah satu visualisasi pengajaran surah secara interaktif, yang memaparkan ringkasan faedah, tema utama, dan nilai-nilai utama surah kepada pengguna. Ini menjadikan sistem bukan sahaja sah dan tepat, tetapi juga mudah digunakan dalam pendidikan digital yang semakin berkembang dalam kalangan masyarakat Muslim global, seperti yang disarankan oleh Sridefi et al. (2024) dan Touati-Hamad et al. (2020).

5. KESIMPULAN

Kajian ini mencadangkan pembangunan satu model *AI* yang dijangka berupaya mentafsir surah al-Quran secara interaktif dengan menonjolkan faedah dan pengajaran utama yang relevan. Melalui pendekatan *NLP* dan *Machine Learning*. Kerangka model yang dicadangkan ini dijangka dapat menyumbang kepada pembelajaran tafsir yang lebih moden, mudah dicapai dan bersifat adaptif terhadap keperluan pengguna. Keupayaan sistem model ini menjana pengajaran secara automatik berdasarkan sumber tafsir muktabar diharapkan dapat

memperkasa proses pembelajaran sendiri, mempercepat pemahaman kandungan al-Quran, serta memudahkan integrasi dalam sistem pembelajaran Islam digital.

Dengan menggabungkan teknik-teknik tersebut, model yang dicadangkan ini berpotensi menjadi satu inovasi penting dalam dunia penyelidikan teks keagamaan. Kajian pembangunan sistem ini dijangka dapat memberi impak besar kepada pelbagai pengguna termasuk pelajar, pendakwah dan masyarakat umum yang ingin memahami pengajaran al-Quran secara lebih berkesan. Kajian ini turut membuka ruang kepada pemeraksanaan sistem tafsir digital dengan ciri-ciri interaktif, boleh suai dan berasaskan data yang sahih.

Selain itu, sistem ini juga berpotensi diperluas untuk merangkumi penjelasan dalam konteks hukum (fiqh), akidah, dan isu semasa berasaskan teks tafsir dan hadis sahih. Penyelidik turut mencadangkan penghasilan Antara Muka Pengguna (*User Interface*) yang lebih interaktif, integrasi *Chatbot* atau Pembantu Maya, serta mekanisme maklum balas pengguna untuk tujuan pembaikan berterusan. Dengan ini, sistem yang dicadangkan bukan sahaja menyokong penyebaran ilmu al-Quran, malah mampu menjadi platform pembelajaran tafsir Islam yang bersifat pintar dan masa hadapan.

RUJUKAN

- Abdul Rouf. (2025). Digitization of Tafsir and its Challenge to the Authority of Interpretation: A Review of Ulumul Qur'an. *QOSIM: Jurnal Pendidikan Sosial & Humaniora*, 3(2), 605–610.
- Alfarizi, A. T., & Farabi, R. (2024). Quran in the Digital Age. *al-Bunyan: Interdisciplinary Journal of Qur'an and Hadith Studies*, 2(2), 116-126.
- Ali, R., & Isnaini, S. N. (2024). Digitising interpretation: Transforming Tafsir Al-Mishbah in the context of the living Quran. *Jurnal Studi Ilmu-ilmu Al-Qur'an dan Hadis*, 25(1), 1-23.
- Alviyah, A. (2016). Metode Penafsiran Buya Hamka Dalam Tafsir Al-Azhar. *Jurnal Ilmiah Ilmu Ushuluddin*, 15(1), 25-35.
- Annur Wahid, & M. Ridwan Hasbi. (2024). Menyoal Akurasi ChatGPT Dalam Pengujian Validitas Penafsiran QS. An-Nur Ayat 3. *MULTIPLE: Journal of Global and Multidisciplinary*, 2(7), 2183–2195.
- Al-Janabi, M. K. A. (2024). Artificial Intelligence in Quranic Exegesis: A Critical Analytical Study of ChatGPT Technology. *QURANICA-International Journal of Quranic Research*, 16(2), 112-144.

- Bahar N. Z. N., Zabidi, A. F. M., & Kadir, N. A. (2019). Interaksi Sosial Dalam Tafsir Al-Azhar: Tinjaun Awal Terhadap Surah Madaniyah. *Journal Al-Turath; Journal of Al-Quran and Al-Sunnah*, 4 (2). pp. 68-74.
- Bashir, M. H., Azmi, A. M., Nawaz, H., Zaghouani, W., Diab, M., Al-Fuqaha, A., & Qadir, J. (2023). Arabic Natural Language Processing for Qur'anic Research: A Systematic Review. *Artificial Intelligence Review*, 56(7), 6801-6854.
- Basid, A., Qomariyah, L., Sunandar, M. F. A. F., & Asasuddin, M. (2024). Teks Kitab Suci dan Mesin: Menakar Kerja Mesin Kecerdasan Buatan dalam Memfasilitasi Pemahaman Al-Qur'an. *Ta'wiluna: Jurnal Ilmu Al-Qur'an, Tafsir dan Pemikiran Islam*, 5(3), 641-653.
- Bengueddach, A. (2025). Multi-Dimensional NLP Analysis of the Quran: Structure, Semantics and Emotion across Qirā'āt. *QURANICA-International Journal of Quranic Research*, 17(1), 32-52.
- Chowdhary, K. (2020). Natural Language Processing. *Fundamentals of Artificial Intelligence*, 603-649.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-Training Of Deep Bidirectional Transformers For Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long And Short Papers)* (pp. 4171-4186).
- Fitriani, A., Rahmawati, S., Afrawasih, N., & Fariq, W. M. (2025). Peran AI dalam Pendidikan Islam: Perspektif Al-Quran. *Jurnal Intelek Dan Cendekiawan Nusantara*, 2(3), 3228-3239.
- Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics*, 10, 178-206.
- Hm, M. Y., & Satra, M. (2024). Kajian Tafsir Al-Quran di Era Digital: Literasi dan Pengaruh Teknologi. *Journal of Quranic Studies*, 41(1), 55-72.
- Javed, F. H., & Bibi, N. (2024). Leveraging AI for Qur'anic Exegesis: A New Pedagogical Approach. *Al-Amir Research Journal for Islamic Studies*, 5(04), 41-51.
- Kawtharany, S. (2017). Big Data: New Development Opportunities in Islamic Studies. *International Multidisciplinary Journal of Pure Life (IMJPL)*, 4(11), 39-54.
- Khidhir, A. M. (2024). An AI model for Parsing the Text of Holy Quran Sentences. *Mesopotamian Journal of Quran Studies*, 2024, 16-23.

- Lazarski, E., Al-Khassaweneh, M., & Howard, C. (2021). Using NLP for Fact Checking: A Survey. *Designs*, 5(3), 42.
- Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., ... & Lemey, C. (2021). Machine Learning And Natural Language Processing In Mental Health: Systematic Review. *Journal of Medical Internet Research*, 23(5), e15708.
- Lorenzoni, G., Alencar, P., Nascimento, N., & Cowan, D. (2021). Machine Learning Model Development From a Software Engineering Perspective: A Systematic Literature Review. *arXiv preprint arXiv:2102.07574*.
- Madadzadeh, F., & Bahariniya, S. (2024). The Role of Artificial Intelligence in Understanding and Interpreting the Quran. *Journal of Community Health Research*, 2(2), 310-311.
- Mahdi, M. N., Ahmad, A. R., Natiq, H., Subhi, M. A., & Qassim, Q. S. (2021). Comprehensive Review and Future Research Directions on Dynamic Faceted Search. *Applied Sciences*, 11(17), 8113.
- Mubarik, S., & Sakinah, E. H. (2024). AI and the Discourse of Qur'anic Interpretation: An Analysis of Quran-Gpt Response to Gender Verses From The Semiotics Perspective of Roland Barthes. *Innovatio: Journal for Religious Innovations Studies*, 24(2), 189-203.
- Mukmin, M. F., Saputra, N. A., Untoro, J., Ali, M. (2025). Analisis Kritis Terhadap Tafsir Generatif AI: Chatgpt dan Penafsiran Ayat-Ayat Al-Qur'an. *Jurnal Islam Pesisir dan Kajian Keislaman*, 1(1), 39-59.
- Najib, A., & Kurjum, M. (2025). The Role of Natural Language Processing In AI-Based Interpretation Analysis: Opportunities and Limitations. *International Journal of Environmental Sciences*, 1206-1216.
- Nirwana, A. N. (2024). Multimedia Tafsir: Exploring the Meaning of the Quran in the Digital Era. *Available at SSRN 4785707*.
- Pekar, V. (2004). Linguistic Preprocessing for Distributional Classification of Words. In *Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries* (pp. 15-21).
- Purnamawati, N. W. (2025). Artificial Intelligence in Qur'anic Studies: Opportunities and Ethical Challenges of AI-Based Tafsir. *Journal of Qur'anic and Hadith Studies*, 1(1), 37-48.
- Putra, D. I. A., & Yusuf, M. (2021, March). Proposing Machine Learning of Tafsir al-Quran: In Search of Objectivity With Semantic Analysis and Natural Language Processing. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1098, No. 2, p. 022101). IOP Publishing.
- Putra, D. J. (2024). Revolusi Digital dalam Studi Al-Qur'an: Menggali Wawasan Baru dengan Artificial Intelligence (AI). *Manarul Qur'an: Jurnal Ilmiah Studi Islam*, 24(2), 68-92.

- Rouf A. Yusoff, M. Y. (2013). Tafsir al-Azhar dan Tasawuf Menurut HAMKA. *Jurnal Usuluddin*, (38). 1-30.
- Sati, A., Halim, A., Nasution, A. H., & Ridwan, M. (2025). The Digital Transformation of Tafsir and Its Implications for Islamic Legal Derivation in the Contemporary Era. *MILRev: Metro Islamic Law Review*, 4(1), 389-415.
- Sridefi, S., Hafiz, M., Permana, R. A., & Motivaria, M. A. (2024). Al-Qur'an Transformation in the Digital Era: Analysis of Literature Concerning Online Tafsir And The Application. *International Journal of Research*, 2(2), 387-396.
- Tahir, T. T., & Suswandi, S. (2024). Konsep Al-Qur'an Dalam Menghadapi Peluang Dan Tantangan Artificial Intelligence (AI). *TAFASIR: Journal of Quranic Studies*, 2(2), 79-95.
- Tarshany, Y., & Zeki, A. M. (2024). ICT studies for Islamic Sciences and the Holy Qur'an: Series 1. Kuala Lumpur: International Computing Institute for Quran and Islamic Sciences.
- Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). Fever: A Large-Scale Dataset for Fact Extraction and Verification. *arXiv preprint arXiv:1803.05355*.
- Touati-Hamad, Z., Laouar, M. R., & Bendib, I. (2020, June). Quran Content Representation in NLP. In Proceedings of the 10th International Conference on Information Systems and Technologies (pp. 1-6).
- Van Engelen, J. E., & Hoos, H. H. (2020). A Survey on Semi-Supervised Learning. *Machine Learning*, 109(2), 373-440.
- Yunus, M. M. M. & Yusof, M. Y. Z. M. (2021). Perkembangan Kajian Berkaitan Tafsir al-Azhar (2012-2020) [Development of Studies Related to Tafsir al-Azhar (2012-2020)]. *BITARA International Journal of Civilizational Studies and Human Sciences*, 4(3), 95-114.