

MEDIBOT : PEMBANTU PERUBATAN PINTAR MENGGUNAKAN MODEL BAHASA BESAR DAN TEKNIK PENJANAAN CAPAIAN TERIMBUH

WAN AHMAD ZUHAIR BIN WAN YUSOFF

LAILATUL QADRI BINTI ZAKARIA

*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi,
Selangor Darul Ehsan, Malaysia*

ABSTRAK

Penggunaan internet dalam kalangan rakyat Malaysia semakin meningkat, termasuk pencarian maklumat kesihatan yang mencatatkan 62.9% pada tahun 2023. Namun, maklumat yang diperoleh kerap kali tidak sahih, kurang relevan, dan sukar difahami, khususnya bagi pengguna Bahasa Melayu. Kajian ini bertujuan membangunkan Medibot, sebuah chatbot kesihatan pintar berasaskan sumber rasmi portal MyHealth oleh Kementerian Kesihatan Malaysia (KKM). Medibot menyediakan respons pantas, tepat, dan relevan dalam Bahasa Melayu, meningkatkan akses kepada maklumat kesihatan berkualiti tinggi, terutamanya untuk pengguna kurang fasih berbahasa Inggeris dan penduduk luar bandar. Pelaksanaan Medibot berpotensi mengurangkan jurang maklumat kesihatan, menyokong agenda transformasi digital kesihatan Malaysia menjelang 2030 serta Matlamat Pembangunan Mampan 3 mengenai kesejahteraan masyarakat.

Kata kunci: Chatbot, Model Bahasa Besar, Langchain, PCT

PENGENALAN

Dalam era teknologi maklumat yang pesat berkembang, masyarakat daripada pelbagai lapisan umur semakin cenderung untuk merujuk kepada internet bagi mendapatkan pelbagai jenis maklumat, termasuk maklumat berkaitan kesihatan. Penyebaran peranti mudah alih seperti telefon pintar serta kemudahan capaian internet yang meluas telah menjadikan penggunaan internet sebagai satu norma dalam kehidupan harian. Kecenderungan ini turut merangkumi usaha mendapatkan maklumat berkenaan simptom penyakit, rawatan, ubat-ubatan, dan petua kesihatan.

Kebiasaannya, pengguna akan menggunakan enjin carian seperti Google untuk mencari jawapan kepada persoalan kesihatan. Namun begitu, pencarian ini sering menghasilkan sejumlah besar maklumat yang tidak ditapis, tidak relevan dengan konteks tempatan, atau sukar difahami oleh pengguna awam. Banyak daripada hasil carian tersebut berpunca daripada laman web yang kandungannya tidak disahkan, seperti blog peribadi atau

forum tidak rasmi, yang boleh menyumbang kepada penyebaran maklumat tidak tepat. Kebergantungan terhadap maklumat sebegini berisiko mendorong pengguna membuat keputusan kesihatan sendiri yang tidak berasaskan sumber sahih.

Peningkatan akses kepada maklumat tidak semestinya membawa manfaat sekiranya maklumat yang diperoleh tidak sahih atau tidak disemak. Fenomena *infodemik* iaitu limpahan maklumat yang tidak tepat atau mengelirukan, menjadi isu kritikal terutamanya semasa krisis kesihatan seperti pandemik COVID-19. Penyebaran maklumat palsu berkaitan vaksin, rawatan alternatif atau teori konspirasi kesihatan telah menjadi cabaran besar kepada usaha Kementerian Kesihatan Malaysia (KKM) dalam memupuk kesedaran awam yang berasaskan bukti.

Bagi menangani cabaran ini, pendekatan berasaskan teknologi perlu diadaptasi bagi memastikan maklumat kesihatan yang sahih dapat disampaikan secara lebih tersusun, pantas dan mudah difahami. Salah satu pendekatan yang semakin mendapat perhatian ialah penggunaan *chatbot* iaitu sistem perbualan automatik yang mampu memberikan maklumat secara terus kepada pengguna. Teknologi *chatbot* kini digunakan secara meluas dalam pelbagai sektor seperti perkhidmatan pelanggan, pendidikan, perbankan dan semakin berkembang dalam bidang kesihatan.

Namun begitu, kebanyakan *chatbot* kesihatan sedia ada dibangunkan menggunakan bahasa Inggeris sebagai bahasa utama. Keadaan ini menyukarkan sebahagian pengguna di Malaysia yang lebih selesa berkomunikasi dalam Bahasa Melayu, khususnya dalam kalangan warga emas, penduduk luar bandar serta mereka yang kurang fasih dalam bahasa asing. Hal ini jelas menunjukkan wujudnya keperluan yang signifikan untuk membangunkan sebuah *chatbot* kesihatan yang menyokong Bahasa Melayu secara menyeluruh.

Pembangunan sistem *chatbot* kesihatan dalam Bahasa Melayu berpotensi untuk memberikan akses yang lebih mudah kepada maklumat perubatan yang sahih, berstruktur dan mesra pengguna. Dengan merujuk kepada sumber rasmi seperti portal MyHealth yang dikendalikan oleh KKM, *chatbot* ini boleh dijadikan sebagai rujukan harian yang dipercayai untuk menjawab pelbagai persoalan berkaitan kesihatan. Penyampaian maklumat dalam bentuk interaktif dan ringkas juga dapat meningkatkan keberkesanan komunikasi serta pengalaman pengguna.

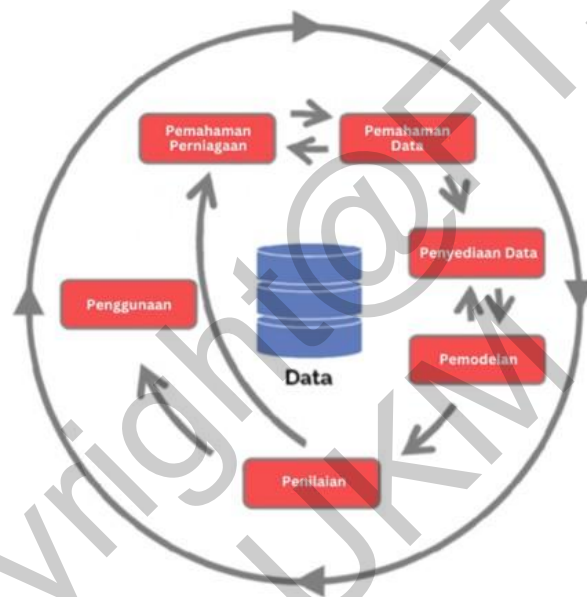
Inisiatif ini sejajar dengan aspirasi pendigitalan sistem penjagaan kesihatan negara menjelang tahun 2030, seperti yang digariskan oleh Kementerian Kesihatan Malaysia. Selain itu, ia turut menyokong Matlamat Pembangunan Mampan (SDG) ketiga, iaitu memastikan kehidupan yang sihat dan kesejahteraan untuk semua peringkat umur. Sistem ini bukan sahaja dapat memperluas akses kepada maklumat kesihatan yang berkualiti, malah membantu meningkatkan tahap literasi kesihatan digital dalam kalangan masyarakat.

Dengan kewujudan sistem seperti *Medibot*, jurang maklumat kesihatan antara

komuniti bandar dan luar bandar dapat dikurangkan. Di samping itu, *chatbot* ini juga berfungsi sebagai alternatif yang pantas, mudah diakses dan boleh dipercayai untuk mendapatkan maklumat awal sebelum mendapatkan nasihat profesional secara fizikal. Justeru, pembangunan *Medibot* bukan sekadar satu inovasi teknologi, malah merupakan langkah penting dalam usaha meningkatkan taraf kesihatan masyarakat secara menyeluruh.

METODOLOGI KAJIAN

Metodologi yang digunakan dalam pembangunan sistem *Medibot* ialah CRISP-DM (*Cross Industry Standard Process for Data Mining*). Pendekatan ini dipilih kerana ia menyediakan kerangka kerja yang sistematik dalam pembangunan aplikasi berasaskan data, terutamanya projek yang melibatkan pemodelan kecerdasan buatan dan pemprosesan bahasa semula jadi.



Rajah 1 Metodologi Pembangunan *Medibot*

CRISP-DM merangkumi enam fasa utama yang saling berkait dan iteratif, membolehkan pembangunan sistem dilakukan dengan lebih tersusun, berfokus, dan berpanduan objektif yang jelas. Pendekatan ini amat sesuai digunakan dalam pembangunan *Medibot*, iaitu sebuah *chatbot* kesihatan pintar berasaskan Model Bahasa Besar dan teknik Penjanaaan Capaian Terimbuh, kerana ia menyokong pengurusan data, pemodelan, serta penilaian sistem secara bersepadu..

Fasa Pemahaman Perniagaan

Fasa pertama adalah untuk mengenal pasti masalah utama pengguna, iaitu kesukaran dalam mendapatkan maklumat kesihatan yang sahih, ringkas dan mudah difahami dalam Bahasa Melayu. Oleh itu, objektif utama projek ini adalah untuk membangunkan sebuah *chatbot* yang mampu menyampaikan maklumat kesihatan rasmi dalam bentuk perbualan, yang boleh diakses 24 jam sehari, 7 hari seminggu, dan mesra pengguna. Objektif perniagaan ini diterjemahkan kepada beberapa objektif teknikal, antaranya penggunaan sumber data kesihatan yang sahih dan rasmi, penyampaian respons dalam Bahasa Melayu, serta

memastikan ketepatan dan kebolehpercayaan respons yang dijana oleh *chatbot*..

Fasa Pemahaman Data

Fasa kedua melibatkan pengumpulan dan eksplorasi awal terhadap data yang akan digunakan dalam pembangunan sistem. Sumber data utama diperoleh daripada portal rasmi *MyHealth* yang dikendalikan oleh Kementerian Kesihatan Malaysia. Dokumen seperti artikel, garis panduan, dan risalah kesihatan dalam format teks digunakan sebagai asas kandungan untuk sistem. Analisis awal dijalankan bagi mengenal pasti struktur data dan kesesuaian kandungan, serta isu-isu yang mungkin timbul, seperti penggunaan istilah perubatan yang kompleks atau panjangnya dokumen yang boleh memberi kesan terhadap keberkesanan sistem.

Fasa Pra-pemprosesan Data

Fasa pra-pemprosesan data melibatkan pemprosesan dan penyediaan data yang telah dikumpul untuk digunakan dalam model pemprosesan bahasa semula jadi. Proses ini merangkumi beberapa langkah penting, seperti pembersihan teks untuk menghapuskan karakter khas, simbol, dan teks yang tidak relevan. Seterusnya, teks akan melalui proses tokenisasi, iaitu pemecahan teks kepada unit-unit kecil seperti ayat atau frasa, diikuti dengan *lemmatization*, di mana perkataan ditukar kepada bentuk akar untuk memastikan konsistensi makna. Selain itu, perkataan yang tidak memberi nilai semantik, seperti "dan", "atau", dan "yang", akan dipadamkan melalui langkah penyingkiran *stopwords*. Teks yang telah dibersihkan kemudian akan dibahagikan kepada blok (*chunking*) agar saiznya tidak melebihi had *token* model bahasa. Setiap blok akan dijana kepada bentuk vektor menggunakan model *sentence-transformers/paraphrase-multilingual-mpnet-base-v2* dan disimpan dalam pangkalan data vektor *Pinecone* untuk tujuan pencarian semantik..

Fasa Pemodelan

Fasa pemodelan memberi fokus kepada pembangunan logik sistem *chatbot* yang bertujuan untuk menjawab soalan pengguna berdasarkan data yang relevan. Sistem ini menggunakan pendekatan Penjana Capaian Terimbuh, yang menggabungkan hasil carian dokumen dengan kebolehan pemahaman bahasa oleh model GPT-4o. Komponen teknikal utama dalam fasa ini adalah penggunaan GPT-4o sebagai model bahasa besar untuk menjana respons. Selain itu, *LangChain* digunakan sebagai rangka kerja untuk menyambungkan antara model, pangkalan data *embedding* (*Pinecone*), dan antara muka pengguna. Parameter seperti saiz *chunk*, nilai *top-K*, *temperature*, dan *max_tokens* akan ditentukan bagi memastikan respons yang relevan, ringkas, dan tidak terlalu teknikal. Proses sistem akan berlaku dengan langkah-langkah di mana soalan pengguna ditukar kepada bentuk *embedding*, sistem mencari dokumen yang relevan dalam *Pinecone*, dan maklumat yang dijumpai akan dihantar semula kepada GPT-4 untuk menjana jawapan.

Fasa Penilaian

Fasa penilaian melibatkan dua kaedah utama untuk menilai keberkesanan sistem. Penilaian teknikal dijalankan menggunakan metrik *BERTScore*, yang mengukur tahap kesamaan semantik antara jawapan *chatbot* dengan jawapan rujukan. Nilai F1 *BERTScore* yang melebihi 0.85 dianggap sangat memuaskan. Selain itu, penilaian pengguna dilakukan melalui

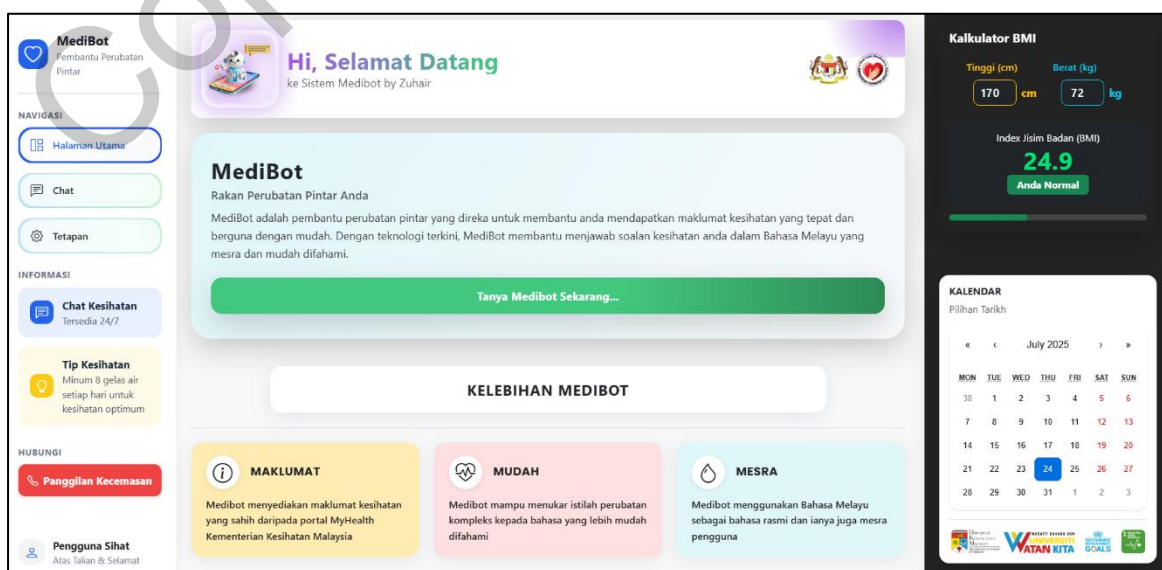
ujian penerimaan yang dilaksanakan menggunakan soal selidik *Google Forms*, yang diedarkan kepada 40 pengguna awal. Aspek yang dinilai termasuk kefahaman bahasa, kesesuaian kandungan, antara muka pengguna, serta tahap kepuasan keseluruhan. Maklum balas yang diterima daripada ujian ini digunakan untuk memperbaiki sistem sebelum ia dilancarkan secara rasmi.

Fasa Penggunaan

Fasa terakhir dalam metodologi ini adalah fasa penggunaan, di mana sistem *chatbot* dilancarkan kepada pengguna sebenar melalui platform web. Antara muka pengguna dibangunkan menggunakan *ReactJS* untuk *frontend*, manakala bahagian pelayan dan pemprosesan belakang menggunakan *Flask*. Pendekatan ini dipilih kerana ia memberikan lebih kawalan, fleksibiliti, dan kemudahan untuk integrasi dengan API dan sistem *embedding* berasaskan *Pinecone*. Ciri-ciri utama sistem termasuk kebolehcapaian 24/7 melalui pelayar web tanpa penyimpanan data peribadi pengguna, bagi mematuhi Akta Perlindungan Data Peribadi. Di samping itu, sistem juga menyediakan borang maklum balas yang membolehkan pengguna memberikan input untuk penambahbaikan berterusan. Sistem direka untuk menyokong kemas kini kandungan dan model secara dinamik, berdasarkan maklum balas pengguna dan perkembangan maklumat kesihatan semasa.

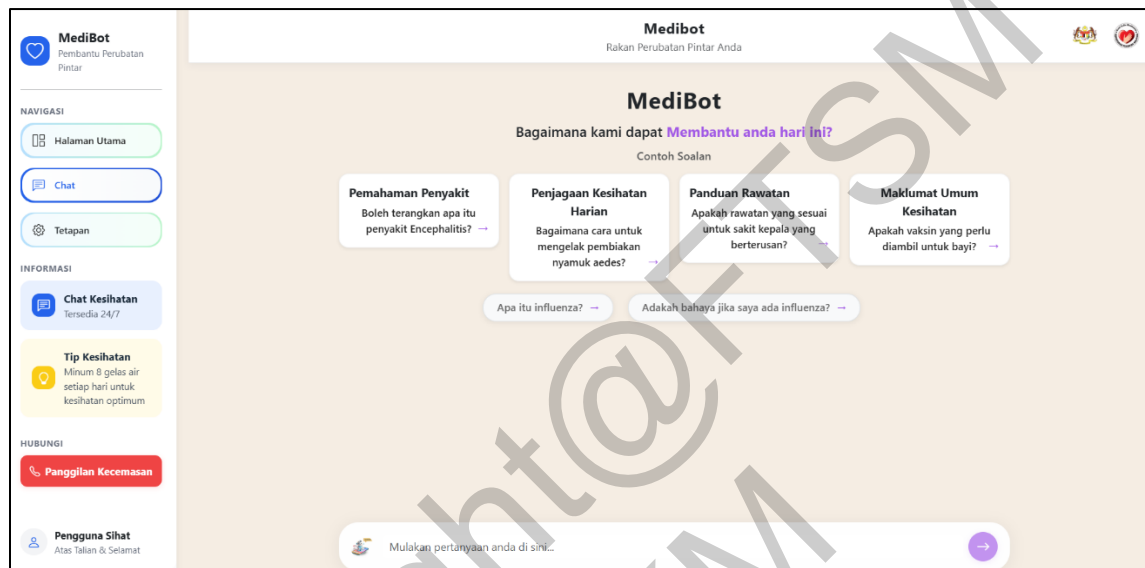
KEPUTUSAN DAN PERBINCANGAN

Projek pembangunan *Medibot*, sebuah *chatbot* kesihatan pintar menggunakan Model Bahasa Besar dan Penjanaan Capaian Terimbuhan, telah berjaya disiapkan dengan semua komponen berfungsi mengikut jangkaan. *Medibot* dibangunkan untuk memberi maklumat kesihatan yang sahih dan mudah difahami dalam Bahasa Melayu, terutama bagi golongan yang kurang mahir dalam bahasa Inggeris.



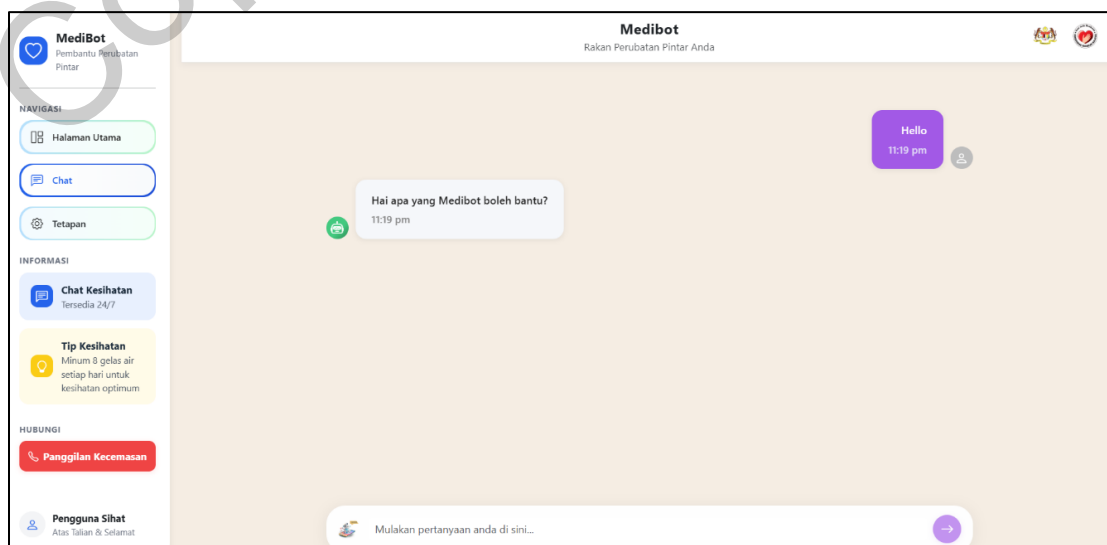
Rajah 1 Antara Muka Halaman Utama

Rajah 1 menunjukkan antara muka halaman utama *Medibot* yang menyambut pengguna dengan ucapan selamat datang. Halaman ini dilengkapi dengan butang hijau bertulis 'Tanya Medibot Sekarang' yang membolehkan pengguna mengakses *chatbot* secara pantas. Antara muka ini turut menyediakan paparan ringkas mengenai kelebihan *Medibot* iaitu maklumat yang sahih, mudah digunakan dan mesra pengguna. Panel sebelah kanan pula menampilkan kalkulator BMI untuk kemudahan pengguna memantau tahap kesihatan diri, serta kalendar interaktif bagi memudahkan perancangan aktiviti harian.



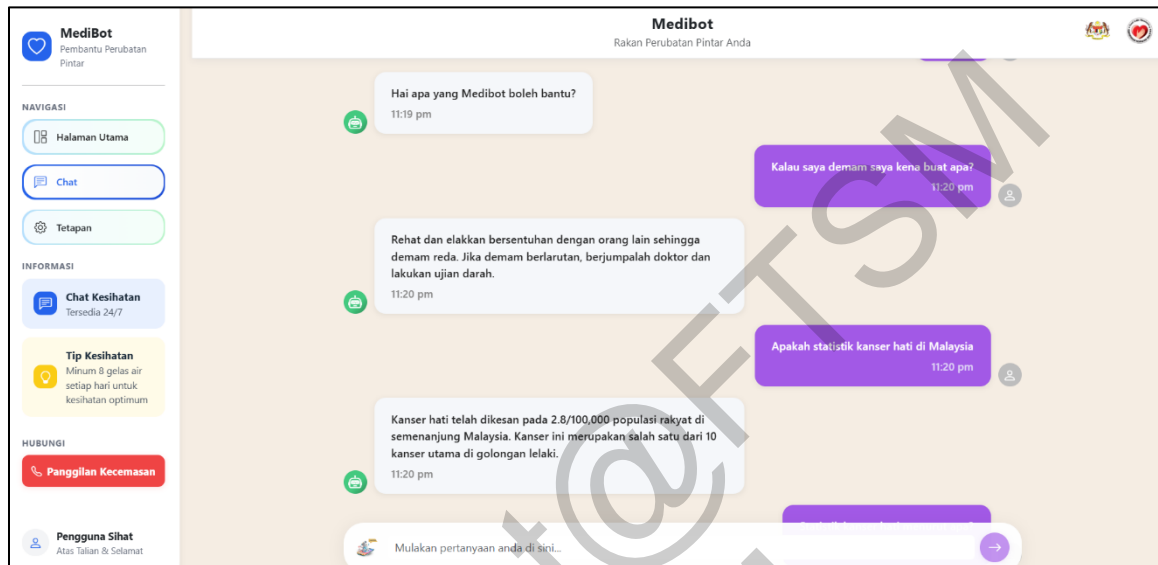
Rajah 2 Antara Muka Chat

Rajah 2 memaparkan antara muka fungsi Chat apabila pengguna menekan pilihan 'Chat' di *sidebar* atau butang 'Tanya Medibot Sekarang'. Skrin ini memaparkan beberapa soalan contoh seperti "Apa itu influenza?" atau "Bagaimana cara untuk mengelak pembiakan nyamuk aedes?" bagi membantu pengguna memulakan interaksi dengan *chatbot*.



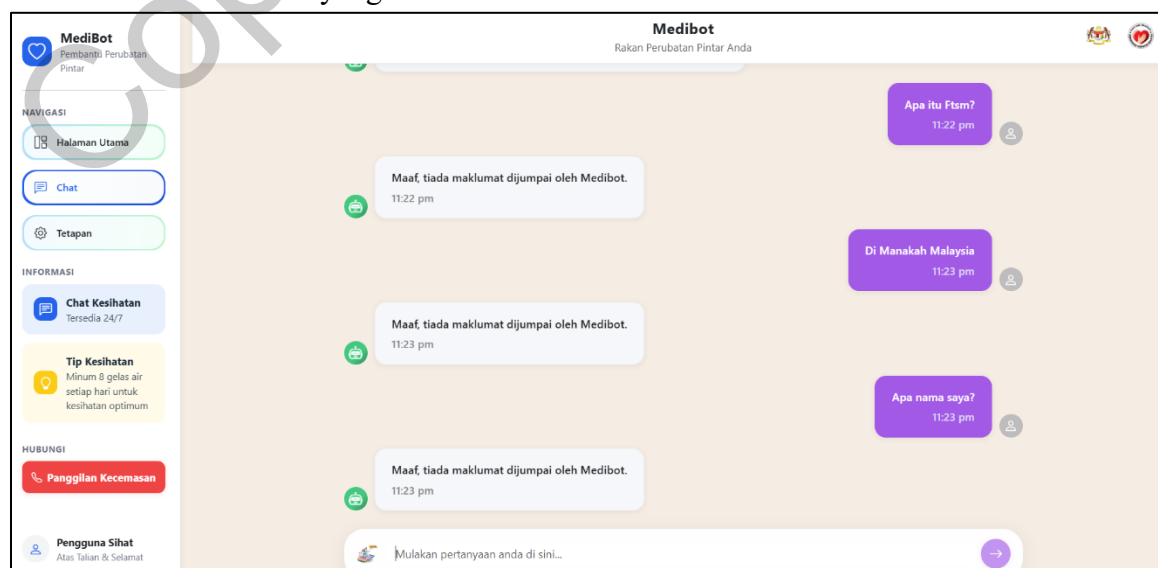
Rajah 3 Antara Muka Medibot

Dalam Rajah 3, apabila pengguna memberi ucapan salam atau ucapan umum seperti "Hello", *Medibot* memberikan respons automatik yang mesra seperti "Hai apa yang *Medibot* boleh bantu?" untuk memulakan perbualan dan membantu pengguna merasa lebih selesa serta mengurangkan jurang komunikasi antara pengguna dengan *chatbot*.



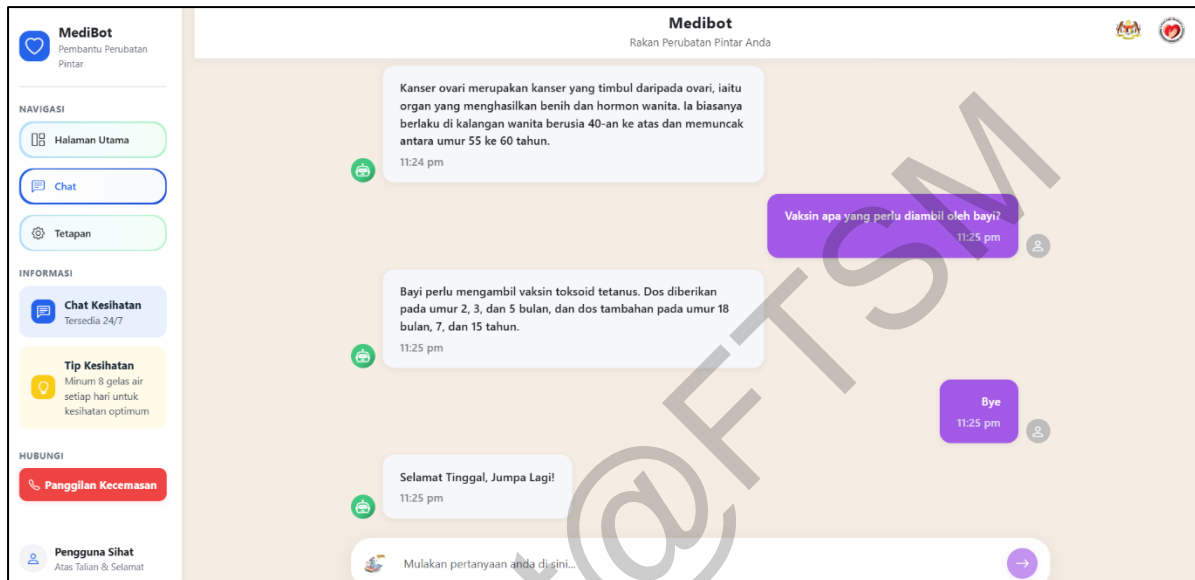
Rajah 4 Antara Muka Medibot 2

Rajah 4 menunjukkan bahawa *Medibot* dapat memberikan jawapan yang tepat dan relevan kepada pertanyaan pengguna berdasarkan dataset yang telah disediakan. Contohnya, apabila pengguna bertanya mengenai langkah yang perlu diambil ketika demam atau statistik berkaitan kanser hati, *Medibot* memberikan jawapan dengan fakta jelas dan mudah difahami. Ini menunjukkan keberkesanan *Medibot* dalam menyalurkan maklumat yang tepat berdasarkan sumber data yang sah.



Rajah 5 Antara Muka Medibot 3

Rajah 5 memperlihatkan keadaan apabila pengguna bertanya soalan yang tiada dalam dataset *Medibot*. Dalam situasi ini, *Medibot* akan menjawab dengan sopan iaitu "Maaf, tiada maklumat dijumpai oleh Medibot." Hal ini memastikan pengguna sedar mengenai had pengetahuan sistem dan tidak menerima maklumat yang mengelirukan atau salah, sekali gus memelihara integriti dan kredibiliti *Medibot*.



Rajah 6 Antara Muka Medibot 4

Akhir sekali, Rajah 6 menunjukkan interaksi pengguna yang mengakhiri sesi perbualan dengan ucapan selamat tinggal seperti "Bye" atau "Selamat tinggal". *Medibot* memberikan respons yang mesra seperti "Selamat Tinggal, Jumpa Lagi!" sebagai tanda penghargaan terhadap interaksi pengguna, sekali gus memberikan pengalaman yang positif, mesra pengguna, dan menggalakkan pengguna untuk kembali menggunakan *Medibot* pada masa akan datang.

Pengujian Kebolehgunaan

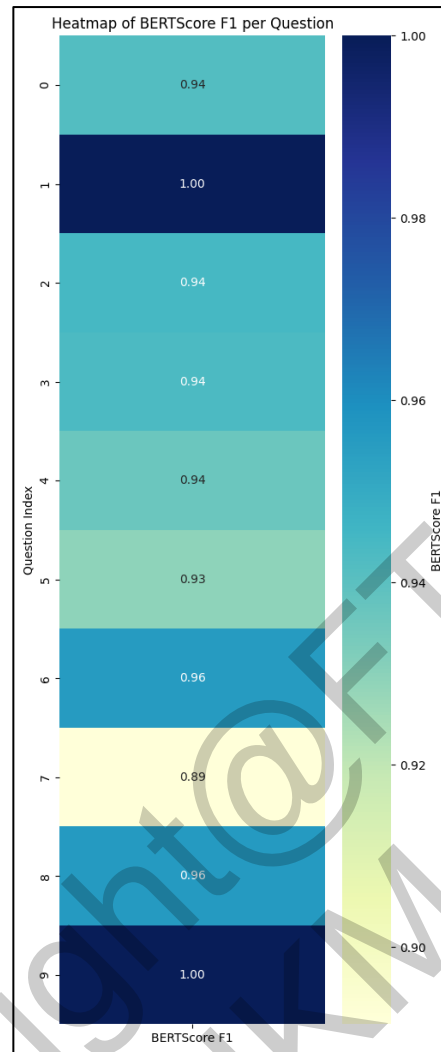
Pengujian kebolehgunaan yang dilaksanakan bertujuan untuk menilai tahap kefahaman pengguna terhadap jawapan *Medibot*, ketepatan maklumat yang diberikan oleh sistem, tahap kepuasan pengguna terhadap kelajuan respons, serta relevansi jawapan dalam konteks kesihatan. Pengujian ini dilaksanakan melalui kaedah Ujian Penerimaan Pengguna (UAT) yang melibatkan seramai 40 responden daripada pelbagai latar belakang umur dan jantina bagi memastikan keputusan pengujian ini mencerminkan keadaan sebenar penggunaan sistem *Medibot*.

Setiap responden diberikan peluang untuk mengajukan soalan berkaitan kesihatan secara bebas tanpa sebarang campur tangan pihak pembangun, bagi menilai sejauh mana *Medibot* dapat memberikan jawapan yang memuaskan berdasarkan dataset yang tersedia dalam sistem. Setelah selesai interaksi, responden diminta mengisi borang soal selidik secara dalam talian untuk menilai tahap kefahaman mereka terhadap jawapan *Medibot*, ketepatan informasi, kelajuan respons, serta relevansi jawapan terhadap soalan yang diajukan.

Jadual 1 Ringkasan Keputusan Ujian Penerimaan Pengguna (UAT)

Aspek Penilaian	Skala 1	Skala 2	Skala 3	Skala 4	Skala 5	Purata
Kefahaman terhadap jawapan <i>Medibot</i>	0 (0%)	0 (0%)	0 (0%)	2 (5%)	38 (95%)	4.95
Kepuasan terhadap ketepatan jawapan <i>Medibot</i>	0 (0%)	0 (0%)	0 (0%)	10 (25%)	30 (75%)	4.75
Kepuasan terhadap kelajuan respons <i>Medibot</i>	0 (0%)	0 (0%)	0 (0%)	7 (17.5%)	33 (82.5%)	4.83
Relevansi jawapan dengan konteks kesihatan	Tidak: 0 (0%)	Ya: 40 (100%)	-	-	-	100% relevan

Berdasarkan jadual di atas, didapati responden secara umumnya memberikan maklum balas yang sangat positif terhadap *Medibot*. Tahap kefahaman pengguna mencatatkan purata tertinggi iaitu 4.95, menunjukkan bahawa jawapan *Medibot* mudah difahami. Aspek ketepatan maklumat juga dinilai tinggi (purata 4.75), namun terdapat ruang kecil untuk peningkatan dalam memberikan jawapan kepada soalan yang lebih rumit. Selain itu, responden berpuas hati dengan kelajuan respons *Medibot* dengan purata skor 4.83. Secara keseluruhan, semua responden (100%) bersetuju bahawa jawapan *Medibot* adalah relevan dan bersesuaian dengan konteks kesihatan yang diajukan.



Rajah 7 Heatmap skor F1 *BERTScore*

Tambahan pula, pengujian ketepatan jawapan *Medibot* turut dijalankan menggunakan metrik *BERTScore* yang membandingkan jawapan *Medibot* dengan jawapan rujukan ideal yang disediakan oleh pakar kesihatan. Rajah 7 menunjukkan *heatmap* skor F1 *BERTScore* bagi setiap soalan dalam dataset ujian, di mana kebanyakan skor tertumpu antara 0.93 hingga 1.00, menandakan prestasi semantik yang tinggi dan konsisten. Secara khususnya, skor F1 purata yang diperoleh adalah sebanyak 0.9499, menunjukkan sistem *Medibot* mampu menghasilkan jawapan dengan ketepatan semantik yang sangat baik. Visualisasi ini membantu mengenal pasti soalan-soalan tertentu yang memerlukan penambahbaikan lanjut.

Kesimpulannya, pengujian kebolehgunaan yang dijalankan menunjukkan bahawa *Medibot* adalah sistem yang sangat efektif, boleh dipercayai, serta berpotensi tinggi untuk dijadikan alat rujukan digital dalam bidang kesihatan. Keputusan ujian ini bukan sahaja menunjukkan *Medibot* memenuhi objektif pembangunan, tetapi turut menyediakan maklumat penting untuk penambahbaikan berterusan sistem ini pada masa hadapan.

Cadangan Penambahbaikan

Berdasarkan maklum balas yang diperoleh sepanjang proses pengujian sistem *Medibot*, beberapa cadangan penambahbaikan telah dikenal pasti untuk mempertingkatkan keberkesanan dan keupayaan sistem ini pada masa hadapan. Cadangan ini berpandukan analisis hasil pengujian serta maklum balas langsung daripada responden yang terlibat, menjadikannya relevan dengan keperluan sebenar pengguna.

Salah satu cadangan utama adalah penambahan fungsi audio dalam sistem *Medibot*. Ciri ini dilihat berpotensi besar dalam menjadikan interaksi pengguna dengan sistem lebih fleksibel dan interaktif. Dengan adanya fungsi audio, pengguna boleh mengemukakan pertanyaan secara lisan dan menerima jawapan dalam bentuk audio secara langsung. Pendekatan ini sangat sesuai untuk golongan pengguna yang menghadapi kesukaran membaca atau menaip, seperti warga emas, individu kurang upaya penglihatan, atau mereka yang lebih selesa berinteraksi secara lisan. Selain meningkatkan aksesibiliti sistem, fungsi audio ini dijangka akan memberikan pengalaman penggunaan yang lebih semula jadi, meningkatkan kepuasan pengguna dan menjadikan *Medibot* lebih mesra pengguna secara keseluruhan.

Cadangan seterusnya melibatkan integrasi teknologi Realiti Terimbuh (*Augmented Reality*, AR) dan Realiti Maya (*Virtual Reality*, VR) ke dalam *Medibot*. Penggunaan teknologi ini akan memberikan pengalaman interaktif dan visual yang lebih mendalam kepada pengguna. Misalnya, melalui AR, pengguna boleh melihat simulasi visual simptom penyakit, demonstrasi penggunaan alat perubatan, atau prosedur rawatan dengan jelas dan interaktif menggunakan kamera telefon pintar atau peranti lain yang menyokong AR. Teknologi VR pula boleh digunakan untuk mewujudkan simulasi maya seperti persekitaran bilik rawatan, atau latihan menghadapi senario kecemasan kesihatan yang mendidik pengguna secara efektif. Integrasi ini dijangka dapat meningkatkan tahap pemahaman pengguna terhadap maklumat kesihatan melalui pengalaman visual dan kinestetik yang lebih realistik.

Selain itu, penambahan ciri multimedia dalam bentuk video pendek turut dicadangkan oleh beberapa responden. Video ringkas dan padat mengenai topik-topik kesihatan yang kompleks dapat memudahkan pengguna memahami maklumat yang sukar disampaikan melalui teks sahaja. Penggunaan elemen visual ini dijangka meningkatkan minat pengguna terhadap kandungan yang disediakan oleh *Medibot* dan membantu pengguna yang cenderung kepada gaya pembelajaran visual. Ciri ini juga dapat menjadikan *Medibot* bukan sahaja sebagai sumber maklumat tetapi juga sebagai medium pembelajaran interaktif yang lebih menyeronokkan dan efektif.

Tambahan pula, aspek keupayaan *Medibot* dalam memahami variasi bahasa turut memerlukan penambahbaikan. Sistem *Medibot* perlu dipertingkatkan dengan menggunakan teknologi Pemprosesan Bahasa Semula Jadi (*Natural Language Processing*, NLP) yang lebih canggih bagi membolehkan ia memahami dialek, slanga tempatan, serta gaya pertuturan tidak formal yang kerap digunakan oleh pengguna tempatan. Dengan keupayaan ini, sistem *Medibot* akan lebih tepat dalam menafsir pertanyaan pengguna, meningkatkan ketepatan respons, serta menjadikan sistem lebih mesra dan relevan dengan budaya tempatan.

Akhir sekali, aspek keselamatan dan privasi pengguna perlu diberi perhatian serius dalam pembangunan sistem *Medibot* di masa hadapan. Sebagai sistem yang mengendalikan data sensitif berkaitan kesihatan pengguna, langkah keselamatan yang ketat seperti penggunaan teknologi penyulitan data yang kukuh haruslah diimplementasikan. Ini bertujuan untuk melindungi privasi pengguna dan memastikan data peribadi pengguna tidak disalahguna atau didedahkan kepada pihak ketiga tanpa kebenaran. Pendekatan ini juga penting bagi memastikan sistem mematuhi peraturan dan undang-undang perlindungan data peribadi.

Secara keseluruhannya, cadangan-cadangan ini mencerminkan usaha berterusan dalam meningkatkan kemampuan, inklusiviti, dan keselamatan sistem *Medibot*. Pelaksanaan penambahbaikan ini dijangka akan menjadikan *Medibot* sebagai platform penyampaian maklumat kesihatan yang lebih maju, efektif, dan relevan dengan keperluan semasa pengguna.

KESIMPULAN

Secara keseluruhannya, sistem *Medibot* telah berjaya dibangunkan berdasarkan objektif dan keperluan kajian yang ditetapkan. Semua fungsi dan ciri yang dirancang berjaya diimplementasikan dengan efektif. Kajian ini telah berjaya mencapai matlamat asal, iaitu menyediakan satu platform maklumat kesihatan digital yang tepat, sahih, inklusif dan mudah digunakan oleh pelbagai lapisan masyarakat. Walaupun terdapat beberapa cabaran sepanjang proses pembangunan, segala kekangan ini telah dikenal pasti dan dijadikan asas untuk cadangan penambahbaikan pada masa hadapan. Diharapkan sistem *Medibot* ini dapat menjadi titik permulaan bagi kajian lanjutan dalam bidang teknologi komunikasi kesihatan digital pada masa hadapan, seterusnya memberi manfaat yang lebih luas kepada masyarakat.

Kekuatan Sistem

Antara kekuatan utama sistem *Medibot* adalah keupayaannya menyediakan maklumat kesihatan yang sahih, tepat, dan boleh dipercayai, bersumberkan data rasmi daripada portal *MyHealth* oleh Kementerian Kesihatan Malaysia. Kekuatan lain yang signifikan termasuklah keupayaannya memberikan respons sepenuhnya dalam Bahasa Melayu, sekali gus meningkatkan kefahaman pengguna yang lebih selesa dengan bahasa ibunda. Di samping itu, sistem ini mampu menyediakan personalisasi maklumat berdasarkan profil pengguna serta memanfaatkan teknologi canggih Model Bahasa Besar (MBB) dengan Penjana Capaian Terimbuh (PCT). Antara muka yang mesra pengguna serta mudah diakses secara atas talian juga merupakan kelebihan utama, membolehkan pengguna daripada pelbagai peringkat umur menggunakan sistem tanpa kesukaran.

Kelemahan Sistem

Walau bagaimanapun, sistem *Medibot* masih mempunyai beberapa kelemahan yang perlu ditambah baik. Antaranya ialah keterbatasan kefahaman sistem terhadap variasi dialek, slanga tempatan, dan gaya bahasa tidak formal yang digunakan dalam komunikasi harian masyarakat Malaysia. Selain itu, isu kebolegunaan pada peranti dengan spesifikasi rendah atau

sambungan internet yang kurang stabil turut menjadi kekangan. Kebergantungan kepada sumber data tunggal iaitu portal *MyHealth* juga menimbulkan risiko sekiranya berlaku gangguan capaian atau maklumat yang tidak dikemaskini. Tambahan pula, sistem ini berdepan cabaran teknikal dari segi skalabiliti apabila berlaku peningkatan pengguna secara mendadak. Ketiadaan elemen multimedia interaktif serta isu keselamatan dan privasi data pengguna turut dikenal pasti sebagai kelemahan yang perlu diatasi bagi menjamin keberkesanan sistem dalam jangka masa panjang.

PENGHARGAAN

Dengan rasa penuh rendah hati, penulis ingin merakamkan setinggi-tinggi penghargaan dan ucapan terima kasih kepada penyelia projek, Dr. Lailatul Qadri Binti Zakaria, atas bimbingan, dorongan serta nasihat bernilai yang telah diberikan sepanjang tempoh pembangunan projek *Medibot* ini. Segala tunjuk ajar yang telah dikongsikan oleh beliau memainkan peranan yang amat besar dalam memastikan kejayaan projek ini.

Penghargaan istimewa turut diucapkan kepada semua pihak yang terlibat secara langsung atau tidak langsung dalam menjayakan projek ini, khususnya para responden yang sudi meluangkan masa menyertai sesi pengujian serta memberi maklum balas konstruktif bagi penambahbaikan sistem. Setiap sumbangan, pandangan, dan kerjasama yang telah dihulurkan amat dihargai, kerana tanpa bantuan dan kerjasama semua pihak tersebut, projek ini tidak akan dapat disempurnakan dengan lancar. Semoga segala jasa dan pengorbanan yang diberikan mendapat balasan dan ganjaran terbaik.

RUJUKAN

- Ahmad, A., Soon, L. K., & Ong, H. F. (2023). Malaysian English News Decoded: A Linguistic Resource for Named Entity and Relation Extraction. *arXiv preprint arXiv*. Retrieved from <https://arxiv.org/abs/2402.14521>.
- Alan, A. Y., Karaarslan, E., & Aydin, O. (2024). A RAG-based Question Answering System Proposal for Understanding Islam: MufassirQAS LLM. *International Journal of Scientific Research in Engineering and Management (IJSREM)*.
- Ananya, G., & Vanishree, K. (2024). RAG-based Chatbot using LLMs. *International Journal of Scientific Research in Engineering and Management (IJSREM)*.
- Agarwal, S., Laradji, I. H., Charlin, L., & Pal, C. (2024). LitLLM: A Toolkit for Scientific Literature Review. *arXiv*.
- Cheah, H., Song, H., Baek, G., Shin, M., Jung, C., & Cha, M. (2024). The Potential of Chatbots for Emotional Support and Promoting Mental Well-being in Different Cultures: Mixed Methods Study. *Journal of Medical Internet Research*. Retrieved from <https://www.jmir.org/2023/1/e51712/>.

- Dagang News. (2023). Penggunaan internet dalam kalangan rakyat Malaysia kekal tinggi pada 98% pada 2023: DOSM. Retrieved from https://www.dagangnews.com/article/penggunaan-internet-dalam-kalangan-rakyat-malaysia-kekal-tinggi-pada-98-pada-2023-dosm-36689?utm_source=chatgpt.com.
- Das, D., Rath, R. L., Singh, T., Mishra, S., Malik, V., Sobti, R., & Brahma, B. (2024). LLM-Based Custom Chatbot Using LangChain. Dalam *Innovative Computing and Communications* (hlm. 257–267). Springer.
- Dean, M., McTear, M. F., Bond, R. R., & Mulvenna, M. D. (2023). ChatPapers: An AI Chatbot for Interacting with Academic Research. *IEEE Xplore*.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). 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* (pp. 4171–4186). Retrieved from <https://arxiv.org/abs/1810.04805>.
- Ding, Y., Jia, M., Miao, Q., & Cao, Y. (2022). A Novel Time–Frequency Transformer Based on Self–Attention Mechanism and Its Application in Fault Diagnosis of Rolling Bearings. *Mechanical Systems and Signal Processing*, 177, 109124. Retrieved from <https://arxiv.org/pdf/2104.09079>.
- Ghojogh, B., & Ghodsi, A. (2020). Attention Mechanism, Transformers, BERT, and GPT: Tutorial and Survey. *OSF Preprints*. Retrieved from https://www.researchgate.net/publication/347623569_Attention_Mechanism_Transformers_BERT_and_GPT_Tutorial_and_Survey.
- Hashim, R. (2022, April 21). Budaya membaca rakyat Malaysia rendah. *Sinar Harian*. Retrieved from https://www.sinarharian.com.my/article/199053/berita/nasional/budaya-membaca-rakyat-malaysia-rendah#google_vignette.
- Ismail, S. (2023, September 5). KKM sasar pendigitalan penuh menjelang 2030. *Berita Harian*. Retrieved from <https://www.bharian.com.my/berita/nasional/2023/09/1148836/kkm-sasar-pendigitalan-penuh-menjelang-2030>.
- IslwisKhadija, M. A., Aziz, A., & Nurharjadmo, W. (2023). PDF-Based Chatbot Development Using LLAMA2 and LangChain: Training and Deployment for Document Interaction. Dalam *2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA)* (hlm. 394–399). IEEE.
- Kulkarni, M., Tangarajan, P., Kim, K., & Trivedi, A. (2024). Reinforcement Learning for Optimizing RAG for Domain Chatbots. *arXiv preprint arXiv:2401.06800*.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., & Riedel, S. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *arXiv preprint arXiv*. Retrieved from <https://arxiv.org/abs/2005.11401>.
- Luangaphirom, T., Jocknoi, L., Wunchum, C., Chokerungreang, K., & Siriborvornratanakul, T. (2024). ThaiNutriChat: Development of a Thai Large Language Model-Based Chatbot for Health Food Services. *Multimedia Systems*, 30, artikel 298.

- Neupane, S., Hossain, E., Keith, J., Tripathi, H., Ghiasi, F., & Mittal, S. (2024). From Questions to Insightful Answers: Building an Informed Chatbot for University Resources. *arXiv*.
- Nurnab Azli. (2023). Examining Factors Influencing Attitude and Adoption of Chatbots Among Consumers. Retrieved from https://etd.uum.edu.my/10791/2/s830498_01.pdf.
- Radhakrishnan, K. R., Kumar, R. S., & Krishnan, S. S. (2024). Multifaceted Chatbot: A Retrieval Augmented Generation Approach for Intelligent Website Query Handling. Dalam *Proceedings of 4th International Conference on Artificial Intelligence and Smart Energy* (hlm. 185–197). Springer.
- Sinar Harian. (2022). Berita palsu antara punca masyarakat Melayu tolak vaksin. Retrieved from https://www.sinarharian.com.my/article/190182/KHAS/Covid-19/Berita-palsu-antara-punca-masyarakat-Melayu-tolak-vaksin?utm_source=chatgpt.com#google_vignette.
- Tan, L. K., Soon, L. K., & Su, T. T. (2023). Emotion-Aware Chatbot with Cultural Adaptation for Mitigating Work-Related Stress. In *Proceedings of the Asian HCI Symposium*. Retrieved from <https://dl.acm.org/doi/abs/10.1145/3604571.3604578>.
- Zhong, G., & Yu, H. (2021). A Review on the Attention Mechanism of Deep Learning and Its Application in NLP. *Neurocomputing*, 452, 48–66. Retrieved from <https://www.sciencedirect.com/science/article/pii/S092523122100477X>.

Wan Ahmad Zuhair Bin Wan Yusoff (A200260)
 Dr. Lailatul Qadri Binti Zakaria
 Fakulti Teknologi & Sains Maklumat
 Universiti Kebangsaan Malaysia