

LINGUISTIK MELAYU DAN KECERDASAN BUATAN DALAM MASYARAKAT 5.0

*SALINAH BINTI JA'AFAR**

Universiti Malaya
b1salina@um.edu.my

MARDIAN SHAH OMAR

Universiti Malaya
mardianso@um.edu.my

MAIZURA OSMAN

Universiti Malaya
mai.zura@um.edu.my

**Pengarang Perantara*

ABSTRAK

Era Masyarakat 5.0 (*Society 5.0*), sedikit sebanyak telah mengubah lanskap dan dimensi kajian linguistik. Perkembangan zaman yang memanfaatkan teknologi dan masyarakat ini turut memperlihatkan perubahan dalam bidang kajian linguistik Melayu yang signifikan. Kajian ini meneliti permasalahan utama dalam bidang linguistik Melayu, khususnya berkaitan dengan integrasinya dalam era kecerdasan buatan (AI), termasuk cabaran dalam pemprosesan bahasa semula jadi (NLP), penterjemahan automatik, dan pembangunan korpus digital. Objektif kajian ini adalah untuk menganalisis kepentingan NLP dalam linguistik Melayu, mengenal pasti cabaran yang dihadapi, serta mencadangkan langkah-langkah untuk meningkatkan keupayaan AI dalam memproses dan memahami bahasa Melayu. Kaedah kajian yang digunakan adalah analisis data sekunder, analisis data pustaka, dan sorotan terhadap pandangan pengkaji terdahulu. Data yang dikumpulkan merangkumi kajian-kajian lepas, artikel ilmiah, dan perkembangan terkini dalam bidang NLP dan linguistik Melayu. Dapatan kajian menunjukkan bahawa walaupun terdapat kemajuan dalam teknologi AI, masih terdapat cabaran besar dalam penterjemahan automatik dan pembangunan korpus digital untuk bahasa Melayu. Implikasi kajian ini menekankan pentingnya usaha kolaboratif antara ahli linguistik, pakar komputer, dan pihak berkepentingan untuk memastikan linguistik Melayu terus relevan dan berkembang seiring dengan perubahan masa. Kajian ini mencadangkan agar penyelidikan masa depan memberi tumpuan kepada pembangunan model NLP yang lebih canggih, pengumpulan data linguistik yang lebih komprehensif, dan peningkatan kesedaran tentang kepentingan linguistik Melayu dalam membentuk Masyarakat 5.0 yang seimbang, adil, dan inklusif.

Kata kunci: Linguistik Melayu; kecerdasan buatan; Masyarakat 5.0; pemprosesan linguistik semula jadi; penterjemahan mesin.

MALAY LINGUISTICS AND ARTIFICIAL INTELLIGENCE IN SOCIETY 5.0

ABSTRACT

The advent of Society 5.0 has significantly transformed the landscape and dimensions of linguistic studies. The rapid advancement of technology and its integration into society have also brought about notable changes in the field of Malay linguistics. This study examines the primary issues in Malay linguistics, particularly its integration into the era of artificial intelligence (AI), including challenges in natural language processing (NLP), machine translation, and the development of digital corpora. The objectives of this study are to analyze the importance of NLP in Malay linguistics, identify the challenges faced, and propose measures to enhance AI's capability in processing and understanding the Malay language. The research methodology employed includes secondary data analysis, literature review, and an examination of perspectives from previous scholars. Data collected encompass past studies, scholarly articles, and recent developments in the fields of NLP and Malay linguistics. The findings reveal that despite advancements in AI technology, significant challenges remain in machine translation and the development of digital corpora for the Malay language. The implications of this study emphasize the importance of collaborative efforts among linguists, computer scientists, and stakeholders to ensure the continued relevance and evolution of Malay linguistics in tandem with societal changes. This study recommends that future research focus on the development of more advanced NLP models, the compilation of more comprehensive linguistic data, and raising awareness about the significance of Malay linguistics in shaping a balanced, equitable, and inclusive Society 5.0.

Keywords: Malay linguistics; artificial intelligence; Society 5.0; natural language processing; machine translation.

PENGENALAN

Masyarakat 5.0, atau *Society 5.0*, merupakan satu konsep yang diperkenalkan oleh Jepun sebagai langkah seterusnya dalam evolusi masyarakat manusia. Deguchi, A., et al. (2020) mengatakan bahawa konsep ini menggabungkan kemajuan teknologi seperti kecerdasan buatan (AI), *Internet of Things* (IoT), dan *big data* untuk mencipta satu masyarakat yang lebih seimbang antara kemajuan teknologi dan kualiti hidup manusia. Linguistik Melayu dan kecerdasan buatan memainkan peranan penting dalam memastikan bahawa masyarakat 5.0 bukan sahaja maju dari segi teknologi, tetapi juga memelihara identiti budaya dan linguistik tempatan. Era kecerdasan buatan (AI) telah membawa perubahan ketara dalam pelbagai aspek kehidupan, termasuk dalam bidang linguistik. Integrasi teknologi AI dengan kajian linguistik menawarkan pelbagai manfaat, termasuk pemerkasaan pemprosesan linguistik semula jadi (NLP), pembangunan korpus digital, dan penterjemahan automatik (Manning, C. D., & Schütze, H. (1999). Masyarakat 5.0 yang dicadangkan oleh Jepun menekankan keseimbangan

antara teknologi dan kemanusiaan, sekali gus menuntut peranan linguistik dalam memastikan keadilan linguistik serta keterangkuman sosial.

Ilmu linguistik dan kecerdasan buatan bermula ketika pengkaji linguistik menyedari kepentingan aplikasi komputer dalam pengumpulan data analisis teks linguistik. Bermula dengan pengumpulan data dan pembinaan data korpus, bidang linguistik terus berkembang sejajar dengan perkembangan sains dan teknologi yang melahirkan bidang linguistik korpus. Linguistik korpus memerlukan pemahaman terhadap pelbagai isu linguistik dan teknologi yang berkaitan dengan ketersediaan kemudahan teknologi untuk penjaan, penyimpanan, pengurusan, anotasi, pemprosesan, analisis, dan penyebaran data serta maklumat bahasa. Selain itu, ia juga memerlukan kesedaran yang mendalam tentang kepentingan korpus bahasa dalam proses memajukan teknologi, bahasa, dan masyarakat (Dash, Niladri Sekhar (2024)). Pendekatan empirikal menggunakan data yang dapat diperhatikan sebagai sumber pengetahuan, dan ahli linguistik korpus juga mendasarkan penemuan mereka pada data korpus yang dapat diperhatikan. Pendekatan ini adalah asas kepada kaedah saintifik, yang membawa kepada pembentukan dan pengujian hipotesis. Hipotesis yang sah mesti boleh diuji dan oleh itu boleh disangkal, tetapi cara pengujiannya berbeza antara satu disiplin dengan disiplin yang lain. Eksperimen adalah kaedah biasa dalam sains semula jadi, dan ujian untuk kepentingan statistik digunakan secara meluas dalam sains sosial.

Dalam linguistik korpus, hipotesis cenderung bersifat tidak formal dan bahkan tersirat, tetapi pengujian dilakukan menggunakan sejumlah besar data autentik yang dihasilkan secara semula jadi. Dengan cara ini, linguistik korpus berada di hujung spektrum saintifik dalam bidang linguistik. Perkara ini berbeza dengan arus perdana teori dalam linguistik, contohnya tatalinguistik struktur frasa dipacu kepala (Müller et al., 2021), yang masih banyak berkaitan dengan kesempurnaan ayat (Abeillé & Borsley, 2021). Kesempurnaan sememangnya merupakan pertimbangan penting, dan memainkan peranan penting dalam pengaturcaraan komputer, kerana kod yang ditulis oleh pengaturcara diuji untuk kesempurnaan oleh pengkompil, dan kod tersebut akan gagal jika tidak sempurna.

Linguistik Melayu, sebagai linguistik kebangsaan dan lingua franca di Malaysia dan beberapa negara ASEAN, mempunyai peranan yang penting dalam membentuk identiti budaya dan sosial masyarakat. Linguistik Melayu, iaitu kajian saintifik tentang linguistik Melayu, bukan sahaja membantu dalam memahami struktur dan fungsi linguistik tersebut, tetapi juga dalam memastikan kelestariannya dalam era digital. Dalam Masyarakat 5.0, di mana interaksi manusia dan mesin semakin kerap, linguistik Melayu perlu diintegrasikan dengan sistem kecerdasan buatan untuk memastikan bahawa teknologi dapat berkomunikasi dengan efektif dalam linguistik tempatan. Integrasi kecerdasan buatan dan linguistik ini pembangunan sistem pemprosesan linguistik semula jadi (*Natural Language Processing, NLP*) yang mampu memahami, mentafsir, dan menghasilkan teks dalam linguistik Melayu dengan tepat.

Berdasarkan perbincangan di atas, maka adalah penting untuk meningkatkan usaha dalam pengumpulan data linguistik Melayu yang berkualiti dan pembangunan teknologi NLP yang sesuai. Usaha ini termasuklah kerjasama antara institusi penyelidikan, universiti, dan industri teknologi untuk memastikan linguistik Melayu terus relevan dan berdaya saing dalam era digital. Dengan memastikan integrasi linguistik Melayu dalam sistem kecerdasan buatan,

kita dapat memastikan bahawa teknologi yang dibangunkan adalah inklusif dan mencerminkan identiti budaya masyarakat, selaras dengan konsep Masyarakat 5.0.

LATAR BELAKANG KAJIAN

Ilmu linguistik amat penting kepada masyarakat dalam mendalami dan menguasai sesuatu Bahasa. Kecerdasan buatan pula adalah wadah penting dalam konteks masyarakat hari ini bagi memastikan data dan sumber rujukan penggunaan dapat diakses dengan mudah sesuai dengan perkembangan sains dan teknologi. Bidang yang berkait rapat dengan kecerdasan buatan dalam kajian linguistic ialah linguistik korpus. Linguistik korpus merupakan pendekatan dalam kajian bahasa yang memanfaatkan koleksi data bahasa (korpus) untuk menganalisis penggunaan sebenar bahasa dalam konteks semula jadi. Dalam konteks bahasa Melayu, kajian linguistik korpus telah berkembang dengan pesat dan memberikan sumbangan yang signifikan dalam pelbagai bidang.

Pembinaan korpus bahasa Melayu telah menjadi asas penting dalam kajian linguistik korpus. Azhar Jaludin (2019) telah membangunkan korpus rujukan bahasa Melayu berasaskan teks web dengan menggunakan prosedur telusur untuk mencari, mengekstraksi, mengumpul, dan memurnikan teks bahasa Melayu dari laman web. Korpus yang dihasilkan ini telah dinilai kualitinya dan didapati setara dengan korpus rujukan DBP-UKM, menjadikannya andal untuk digunakan dalam kajian linguistik. Penggunaan korpus sebagai sumber data dalam analisis tatabahasa bahasa Melayu telah menunjukkan potensi besar. Kajian oleh Hishamudin Isam dan Mashetoh Abd Mutalib (2019) mencadangkan penggunaan analisis korpus sebagai teknik alternatif dalam pengajaran dan pembelajaran tatabahasa. Dengan memanfaatkan pangkalan data korpus DBP dan program daftar kata AntConc, mereka menunjukkan bahawa teknik ini dapat memupuk minat dan penglibatan aktif pelajar dalam proses pembelajaran tatabahasa.

Linguistik korpus juga telah digunakan dalam analisis sentimen dan kajian kesihatan mental. Umaimah Kamarulzaman et al. (2024) menjalankan kajian mengenai pola bahasa kemurungan di media sosial dengan menggunakan analisis korpus linguistik. Mereka membangunkan korpus Bahasa Kemurungan Bahasa Melayu (KoBKMelayu) dan mendapati bahawa pola bahasa yang wujud dapat membantu dalam mengenal pasti simptom awal gejala kemurungan, serta meningkatkan pemahaman mengenai isu kesihatan mental.

Walaupun korpus merupakan sumber data yang bernilai, penggunaannya dalam kajian linguistik bahasa Melayu mempunyai kekuatan dan kelemahan. Hasmidar Hassan (2023) membincangkan bahawa walaupun data korpus membolehkan pelbagai data diperoleh dalam tempoh yang singkat, penggunaan korpus sebagai sumber data menuntut pendekatan, strategi, dan matlamat yang betul bagi memastikan dapatan yang diperoleh adalah empiris dan boleh dipertanggungjawabkan.

Kajian dalam linguistik korpus ini juga diterajui oleh Zuraidah Mohd Don dan Gerry Knowles. Mereka banyak menulis berkaitan kajian linguistic korpus. Artikel mereka yang bertajuk *Processing Natural Malay Texts: A Data-Driven Approach*, membincangkan pembangunan sistem untuk pemprosesan automatik teks bahasa Melayu. Inti sistem ini adalah pangkalan data leksikal berhubung yang dinamakan MALEX, yang dibangunkan berdasarkan

analisis korpus teks semula jadi bahasa Melayu. Sistem ini digunakan untuk pelbagai aplikasi seperti penandaan tatabahasa, stemming, lemmatisasi, dan penjanaan representasi fonologi. Kajian ini menjadi rujukan penting dalam kajian linguistik berkomputer Bahasa Melayu. Dalam makalah *Language and Disciplinary Concepts in Corpus Linguistics: Investigating Corpus Data*, mereka menjelaskan istilah teknikal dalam linguistik korpus dan konsep yang berkaitan, serta prosedur seperti penandaan dan penguraian. Bahasa Inggeris dan Melayu digunakan sebagai ilustrasi, dengan contoh-contoh bahasa Melayu diterjemahkan ke dalam bahasa Inggeris untuk pemahaman pembaca yang tidak mengetahui bahasa Melayu. Makalah ini juga menelusuri sejarah penggunaan istilah dalam pengajaran bahasa tradisional dan kajian Latin di Eropah. Reka bentuk MALEX, yang merupakan alat kerja untuk linguistik korpus empiris bahasa Melayu, dibentangkan sebagai sumbangan kepada kemanusiaan digital.

Berdasarkan kajian-kajian lalu, artikel ini meneiti secara keseluruhan, dan mendapati bahawa kajian linguistik korpus dalam bahasa Melayu telah memberikan sumbangan yang signifikan dalam pelbagai bidang, termasuk pembinaan korpus, analisis tatabahasa, analisis sentimen, dan kajian kesihatan mental. Namun begitu, dalam konteks program linguistik Melayu yang sedia ada, kajian lanjut perlu dilaksanakan bagi memastikan pengisian kandungan program linguistik seimbang dan selaras dengan keperluan teknologi yang akan mengaplikasikan kecerdasan buatan dalam bidang linguistik, terutama linguistik Melayu dalam menghadapi kemunculan era masyarakat 5.0.

PERMASALAHAN KAJIAN

Kecerdasan buatan (AI) telah menjadi tulang belakang kepada Masyarakat 5.0. kecerdasan buatan digunakan dalam pelbagai bidang seperti kesihatan, pendidikan, pertanian, dan perindustrian untuk meningkatkan kecekapan dan kualiti hidup. Namun, untuk memastikan bahawa manfaat kecerdasan buatan dapat dinikmati oleh semua lapisan masyarakat, teknologi ini perlu disesuaikan dengan konteks tempatan, termasuk linguistik dan budaya. Dalam konteks Malaysia dan negara-negara yang menggunakan linguistik Melayu, kecerdasan buatan perlu dilengkapi dengan kemampuan untuk memahami dan berinteraksi dalam linguistik Melayu. penyelidikan mendalam dalam bidang linguistik Melayu untuk membangunkan korpus linguistik yang besar, model linguistik, dan algoritma yang dapat memproses linguistik Melayu dengan baik. Sebagai contoh, sistem kecerdasan buatan seperti chatbot, penterjemah mesin, dan pembantu maya perlu dilatih dengan data linguistik Melayu yang mencukupi untuk memastikan ketepatan dan keberkesanannya.

Linguistik Melayu, sebagai linguistik kebangsaan dan lingua franca di Malaysia serta beberapa negara ASEAN, memainkan peranan penting dalam membentuk identiti budaya dan sosial masyarakat. Linguistik Melayu, iaitu kajian saintifik tentang linguistik Melayu, bukan sahaja membantu memahami struktur dan fungsi linguistik tersebut, tetapi juga memastikan kelestariannya dalam era digital. Dalam konteks Masyarakat 5.0, yang menekankan integrasi teknologi pintar dan manusia, linguistik Melayu perlu diintegrasikan dengan sistem kecerdasan buatan untuk memastikan teknologi dapat berkomunikasi dengan efektif dalam linguistik

tempatan.

Satu daripada aspek penting dalam integrasi ini ialah pembangunan sistem pemprosesan linguistik semula jadi (*Natural Language Processing, NLP*) yang mampu memahami, mentafsir, dan menghasilkan teks dalam linguistik Melayu dengan tepat. NLP membolehkan mesin memproses dan memahami linguistik manusia, yang seterusnya meningkatkan interaksi antara manusia dan mesin. Namun, cabaran utama dalam pembangunan NLP untuk linguistik Melayu termasuk kekurangan data berkualiti dan jurang teknologi di kawasan terpencil. Di Indonesia, banyak usaha telah dilakukan dalam membangunkan model linguistik tempatan melalui inisiatif seperti Sahabat-AI, yang bertujuan untuk membolehkan pembangunan perkhidmatan dan aplikasi berasaskan kecerdasan buatan yang memahami linguistik dan konteks tempatan. Usaha sebegini menunjukkan kepentingan integrasi linguistik tempatan dalam teknologi kecerdasan buatan untuk memastikan kerelevanan dan keberkesanan dalam masyarakat setempat.

Kekurangan data sahih dan berkualiti juga menjadi satu daripada cabaran utama dalam pembangunan sistem NLP untuk linguistik Melayu. Ketiadaan korpus digital yang mencukupi menyukarkan pembangunan model kecerdasan buatan yang berkesan untuk memahami dan memproses linguistik Melayu. Jurang teknologi yang wujud di kawasan luar bandar dan pedalaman menghalang pengumpulan data linguistik yang pelbagai dan representatif. Keadaan ini menyukarkan pembangunan aplikasi kecerdasan buatan yang inklusif dan mampu memahami variasi dialek serta penggunaan linguistik yang berbeza. Di samping kurangnya pelaburan dan sokongan dalam penyelidikan serta pembangunan teknologi kecerdasan buatan terutama yang khusus untuk linguistik Melayu. Cabaran ini menyukarkan pembangunan sistem NLP yang lengkap, masalah kekurangan dana, infrastruktur, dan kerjasama antara institusi penyelidikan, universiti, dan industri teknologi juga menyumbang kepada masalah kekurangan data sahih dan berkualiti dalam bahasa Melayu.

Program Linguistik Melayu seharusnya menjadi medan penyebaran ilmu linguistik yang perlu mengaplikasikan kemajuan sains dan teknologi dalam pengajaran dan pembelajaran. Walaupun ada program yang menawarkan kursus linguistik berkomputer seperti di Universiti Malaya, Universiti Utara Malaysia, dan Universiti Sains Malaysia, namun boleh dikatakan kandungan kursus yang ditawarkan dalam program-program tersebut lebih bersifat teoretikal. Lanjutan dari itu, program linguistik yang berlandaskan kecerdasan buatan perlu lebih menitikberatkan aspek program komputer agar penguasaan bahasa dan linguistik berserta kemahiran komputer menjadi seimbang.

Permasalahan utama dalam kajian **ini** merangkumi keterbatasan data korpus bahasa Melayu, kelemahan pemprosesan bahasa semula jadi (NLP), serta kekurangan model pembelajaran mesin yang khusus untuk bahasa ini. Selain itu, tiadanya standardisasi dalam penggunaan bahasa Melayu dalam AI menyebabkan ketidakkonsistenan, manakala cabaran etika dan sosio-budaya meningkatkan risiko bias algoritma. Sistem komputer juga menghadapi kesukaran dalam memahami makna kontekstual dan semantik bahasa Melayu, terutama dalam komunikasi tidak formal atau berbentuk metafora disebabkan beberapa faktor yang disebutkan di atas.

OBJEKTIF KAJIAN

Objektif utama kajian ini adalah melihat keperluan integrasi linguistik Melayu dalam sistem kecerdasan buatan (AI) dengan mengidentifikasi keperluan dan kepentingan integrasi linguistik Melayu dalam aplikasi kecerdasan buatan untuk memastikan teknologi tersebut relevan dan efektif dalam konteks tempatan. Kajian ini juga akan mengenal pasti halangan seperti kekurangan data berkualiti, jurang teknologi di kawasan terpencil, dan kurangnya pelaburan dalam penyelidikan yang menghambat pembangunan sistem NLP yang berkesan untuk linguistik Melayu. Di samping itu, kajian ini turut melihat dan menilai peranan pendidikan dalam meningkatkan kesedaran dan kemahiran berkaitan kecerdasan buatan dan linguistik Melayu, serta mengkaji bagaimana kurikulum yang sesuai dapat mempersiapkan generasi akan datang untuk memajukan teknologi AI dalam konteks linguistik Melayu.

METODOLOGI KAJIAN

Kajian ini dijalankan dengan menggunakan kaedah kajian kualitatif yang tertumpu kepada analisis konsep. Kaedah kajian konsep ini akan menggunakan kaedah analisis kualitatif yang berasaskan kajian terdahulu untuk memahami bagaimana linguistik Melayu dan kecerdasan buatan berinteraksi dalam konteks Masyarakat 5.0. Kajian ini meneliti penulisan artikel, buku, dan laporan yang berkaitan dengan linguistik Melayu, kecerdasan buatan, dan konsep Masyarakat 5.0. Pengumpulan ini akan merangkumi sumber akademik dan industri yang membincangkan trend semasa dan amalan terbaik dalam penggunaan bahasa Melayu dalam teknologi. Dari data sekunder ini, analisis tematik dibuat untuk mengenal pasti tema utama yang berkaitan dengan penggunaan bahasa Melayu dalam aplikasi kecerdasan buatan. Tema ini termasuk integrasi teknologi, cabaran linguistik, penerimaan masyarakat, dan implikasi kebudayaan.

Kajian ini akan turut membandingkan dapatan kajian dengan penemuan kajian lain yang relevan dengan melihat keperluan menggunakan kecerdasan buatan dan kesan terhadap budaya dan komunikasi. Berdasarkan analisis, implikasi penggunaan bahasa Melayu dalam konteks digital, termasuk cadangan untuk pengembangan teknologi yang lebih inklusif, termasuklah yang melibatkan peranan pendidikan dalam meningkatkan kefahaman dan penggunaan bahasa Melayu dalam dunia digital. Kaedah kajian konsep ini bertujuan untuk menyediakan wawasan yang mendalam tentang interaksi antara linguistik Melayu dan kecerdasan buatan dalam Masyarakat 5.0. Melalui analisis kualitatif yang berasaskan kajian terdahulu, diharapkan penemuan ini dapat memberikan panduan untuk pembangunan teknologi yang lebih berkesan dan relevan dalam konteks penggunaan bahasa Melayu.

ANALISIS DAN PERBINCANGAN

Integrasi linguistik Melayu dan kecerdasan buatan dapat memberikan impak yang besar dalam pendidikan. Integrasi kecerdasan buatan dalam pendidikan menawarkan peluang yang

signifikan untuk meningkatkan proses pengajaran dan pembelajaran, khususnya dalam konteks linguistik Melayu. Dengan memanfaatkan teknologi kecerdasan buatan, sistem pembelajaran adaptif dapat dibangun untuk memenuhi keperluan individu pelajar, menjadikan pengalaman pembelajaran lebih peribadi dan efektif. Platform seperti *Gradescope* menggunakan algoritma kecerdasan buatan untuk menilai tugas pelajar secara automatik, membolehkan guru memberikan maklum balas yang lebih cepat dan efisien.

Selain itu, kecerdasan buatan berpotensi untuk membangunkan alat bantu mengajar yang inovatif. Sistem pengecaman pertuturan dapat digunakan untuk latihan sebutan, membantu pelajar memperbaiki penguasaan fonetik mereka dalam linguistik Melayu. Perpustakaan digital yang dikuasakan oleh kecerdasan buatan juga boleh menyediakan bahan pembelajaran yang disesuaikan dengan tahap kefahaman dan minat pelajar, meningkatkan keterlibatan dan motivasi mereka. Namun, untuk merealisasikan potensi ini, beberapa cabaran perlu diatasi. Kekurangan data berkualiti dalam linguistik Melayu menjadi halangan utama dalam pembangunan sistem kecerdasan buatan yang berkesan.

Usaha pengumpulan data linguistik tempatan adalah penting untuk memastikan kecerdasan buatan dapat memahami dan memproses linguistik tersebut dengan tepat. Selain itu, jurang teknologi di kawasan terpencil dan kurangnya pelaburan dalam penyelidikan turut menghambat perkembangan ini. Pendidikan memainkan peranan penting dalam mengintegrasikan kecerdasan buatan dan linguistik Melayu. Dengan memasukkan elemen kecerdasan buatan dalam kurikulum, pelajar dapat didedahkan kepada teknologi terkini dan memahami bagaimana ia boleh digunakan dalam konteks linguistik dan budaya mereka. Ini bukan sahaja meningkatkan kesedaran dan kemahiran mereka dalam bidang ini, tetapi juga memastikan kelestarian linguistik Melayu dalam era digital.

Secara keseluruhan, integrasi kecerdasan buatan dalam pendidikan linguistik Melayu menawarkan peluang yang besar untuk meningkatkan kualiti pengajaran dan pembelajaran. Dengan mengatasi cabaran yang ada dan memanfaatkan teknologi dengan bijak, kita dapat memastikan bahawa linguistik Melayu terus relevan dan berkembang dalam dunia yang semakin digital. Dengan bantuan kecerdasan buatan, sistem pembelajaran adaptif dapat dibangun untuk memenuhi keperluan individu pelajar. Sistem ini boleh menggunakan linguistik Melayu sebagai medium pengajaran, menjadikan pembelajaran lebih inklusif dan efektif bagi pelajar yang lebih selesa dengan linguistik ibunda mereka.

Selain itu, kecerdasan buatan juga boleh digunakan untuk membangunkan alat bantu mengajar seperti penilaian automatik karangan, sistem pengecaman pertuturan untuk latihan sebutan, dan perpustakaan digital yang mengandungi bahan pembelajaran dalam linguistik Melayu. Penggunaan kecerdasan buatan bukan sahaja memudahkan proses pengajaran dan pembelajaran, tetapi juga membantu dalam memelihara dan mempromosikan penggunaan linguistik Melayu dalam konteks akademik.

KECERDASAN BUATAN DALAM BIDANG LINGUISTIK MELAYU

Era Revolusi Industri 4.0 dan kemunculan Masyarakat 5.0, kecerdasan buatan (AI) semakin

berperanan dalam kehidupan seharian manusia. AI yang berasaskan pemprosesan bahasa semula jadi (NLP) semakin berkembang pesat dalam pelbagai bidang, termasuk pendidikan, perkhidmatan pelanggan, kesihatan, dan pentadbiran. Namun, cabaran utama yang dihadapi dalam pengembangan AI untuk masyarakat berbilang bahasa seperti Malaysia adalah keupayaan sistem ini untuk memahami dan menghasilkan aplikasi berbahasa Melayu dengan tepat. Perubahan zaman yang menjadikan masyarakat pengguna bahasa beralih kepada teknologi digital menjadikan bidang linguistik Melayu sebagai sumber utama pengguna bahasa Melayu. Linguistik Melayu memainkan peranan penting dalam membangunkan AI yang lebih efektif, sekaligus menyumbang kepada kejayaan Masyarakat 5.0.

Integrasi kecerdasan buatan (AI) dalam bidang linguistik Melayu menawarkan potensi yang signifikan untuk memperkaya penggunaan dan pemahaman bahasa tersebut. Namun, terdapat beberapa cabaran utama yang perlu ditangani bagi memastikan integrasi ini berjalan dengan efektif dan beretika. Satu halangan utama dalam pembangunan model kecerdasan buatan untuk bahasa Melayu adalah kekurangan data linguistik yang berkualiti tinggi. Model kecerdasan buatan, khususnya dalam pemprosesan bahasa semula jadi (NLP), memerlukan sejumlah besar data untuk dilatih dengan berkesan. Ketiadaan korpus digital yang mencukupi menyukarkan pembangunan model kecerdasan buatan yang berkesan untuk memahami dan memproses bahasa Melayu. Jurang teknologi yang wujud di kawasan luar bandar dan pedalaman menghalang pengumpulan data bahasa yang pelbagai dan representatif. Keadaan ini menyukarkan pembangunan aplikasi kecerdasan buatan yang inklusif dan mampu memahami variasi dialek serta penggunaan bahasa yang berbeza.

Kajian awal dalam linguistik Melayu menumpukan perhati kepada komponen-komponen teras linguistik seperti aspek tatabahasa yang terdiri dari fonologi, morfologi, sintaksis, semantik, dan pragmatik. Kemunculan teknologi komputer menjadikan kajian dan penyelidikan linguistik Melayu semakin menarik. Kajian yang dibuat oleh Zuraidah Mohd Don dan Gerry Knowles (2004, 2006, 2008, 2021, 2022), dan banyak lagi kajian mereka yang lain lebih tertumpu kepada aspek tatabahasa dan kaedah yang boleh digunakan bagi memanfaatkan teknologi dalam kajian linguistik Melayu. Selain kajian bersama Knowles, Zuraidah juga banyak menghasilkan kajian bersama-sama pelajar ijazah tinggi seliaan beliau. Walaupun banyak kajian telah dihasilkan dari dahulu sehingga kini, namun masih belum ada satu pangkalan data berbahasa Melayu yang benar-benar dapat menandingi pangkalan data bahasa Inggeris yang lengkap dengan penandaan kelas kata, contoh penggunaan, analisis semantik, dan analisis wacana dalam bentuk digital, sesuai dengan kemunculan kecerdasan buatan.

Selain kekurangan data, terdapat keperluan mendesak untuk penyelidikan yang lebih mendalam dalam bidang pemprosesan linguistik Melayu. Banyak algoritma dan model NLP yang sedia ada dibangunkan berdasarkan struktur dan ciri-ciri bahasa Inggeris atau bahasa lain yang lebih dominan. Linguistik Melayu memerlukan adaptasi model-model ini untuk bahasa Melayu memerlukan pemahaman yang mendalam tentang struktur sintaksis, morfologi, dan semantik bahasa tersebut. Tanpa penyelidikan yang komprehensif, model kecerdasan buatan mungkin gagal menangani kepelbagaian dan kekayaan bahasa Melayu, yang boleh membawa kepada hasil yang tidak tepat atau mengelirukan. Ezuana Sukawai dan Nazlia Omar (2020)

menekankan bahawa kajian analisis sentimen dalam bahasa Melayu masih kurang. Mereka mencadangkan pembangunan korpus analisis sentimen menggunakan data Twitter dengan pendekatan separa selia, menggabungkan leksikon sentimen dan pembelajaran mesin. Kajian ini menunjukkan bahawa usaha untuk membangunkan korpus bahasa Melayu yang khusus masih berada pada peringkat awal dan memerlukan perhatian lanjut.

Pada Disember 2024, YTL AI Labs melancarkan Ilmu 0.1, sebuah Model Bahasa Besar yang dibangunkan khusus untuk rakyat Malaysia. Model ini menunjukkan prestasi luar biasa dalam memahami budaya dan konteks tempatan serta berjaya lulus peperiksaan nasional seperti PT3 dan SPM dalam bahasa Malaysia. Walaupun Model Bahasa Besar ini merupakan pencapaian penting, namun dalam pelancaran ini turut menekankan bahawa sebelum ini, model LLM dalam bahasa Melayu adalah terhad, menunjukkan kekurangan korpus yang luas dan berkualiti tinggi untuk pembangunan model AI (Muhamad, H, 2024).

Penggunaan kecerdasan buatan dalam masyarakat menimbulkan kebimbangan berkaitan etika dan privasi. Salah satu isu utama adalah potensi ketaksaan dan bias dalam kecerdasan buatan. Algoritma kecerdasan buatan dibina berdasarkan data latihan yang digunakan; jika data tersebut mengandungi bias, hasil yang dihasilkan oleh kecerdasan buatan juga mungkin mencerminkan bias tersebut, yang akhirnya boleh menyebabkan diskriminasi terhadap kumpulan tertentu. Kajian Vonia, V., Kristianto, H. ., & Parhusip, J.(2025) mengatakan bahawa di sisi teknikal, pengembangan sistem kecerdasan buatan perlu memastikan bahawa algoritma yang digunakan bebas dari bias yang dapat menyebabkan ketidakadilan. Perkara ini dapat dilakukan melalui pelatihan algoritma menggunakan dataset yang beragam, yang mencerminkan kondisi demografis dan teknis siswa. Selain itu, pengumpulan dan penyimpanan data linguistik menimbulkan persoalan mengenai privasi individu. Adalah penting untuk memastikan bahawa data yang dikumpul dilindungi dengan sewajarnya dan digunakan mengikut undang-undang privasi yang berkuat kuasa. Transparansi dalam pengembangan dan penerapan kecerdasan buatan juga perlu dipastikan untuk mengelakkan penyalahgunaan atau pencurian data peribadi. Dalam konteks pendidikan, penggunaan kecerdasan buatan menimbulkan kebimbangan mengenai privasi data pelajar. Teknologi kecerdasan buatan memerlukan sejumlah besar data untuk berfungsi dengan berkesan, termasuk maklumat peribadi pelajar. Kekurangan polisi dan akta undang-undang yang jelas mengenai hak milik data pelajar boleh menyebabkan pendedahan kepada penyalahgunaan maklumat peribadi mereka.

Untuk memastikan integrasi kecerdasan buatan dan linguistik Melayu yang berjaya, adalah penting untuk menangani cabaran-cabaran yang telah dibincangkan. Pelaburan dalam pengumpulan data berkualiti tinggi, penyelidikan mendalam dalam pemprosesan linguistik Melayu, dan penetapan garis panduan etika yang ketat adalah langkah-langkah kritikal yang perlu diambil. Dengan pendekatan yang komprehensif dan beretika, kecerdasan buatan berpotensi untuk memperkaya penggunaan bahasa Melayu dan menyumbang kepada perkembangan masyarakat yang lebih inklusif dan adil. Namun, cabaran ini juga membuka peluang untuk kerjasama antara ahli linguistik, saintis data, dan penggubal dasar untuk membangunkan penyelesaian inovatif yang dapat memanfaatkan kekuatan linguistik Melayu dan kecerdasan buatan. Dengan usaha yang bersepadu, Malaysia dan negara-negara serantau

dapat memainkan peranan utama dalam membentuk Masyarakat 5.0 yang inklusif dan berdaya saing.

Dalam era Masyarakat 5.0, linguistik Melayu dan kecerdasan buatan menjadi dua elemen yang saling melengkapi dalam usaha membina sebuah masyarakat yang seimbang antara kemajuan teknologi dan pemeliharaan nilai budaya. Linguistik Melayu memainkan peranan penting dalam mengekalkan identiti budaya dan warisan bahasa, manakala kecerdasan buatan membuka peluang baharu dalam pemprosesan bahasa, komunikasi, dan pembangunan teknologi pintar. Integrasi bijak antara kedua-dua bidang ini bukan sahaja menjamin kesinambungan bahasa Melayu dalam era digital, tetapi juga memperkukuh kedudukannya dalam landskap teknologi global.

Linguistik Melayu merangkumi pelbagai aspek, termasuk sintaksis, morfologi, semantik, dan pragmatik. Dalam konteks digital, peranan linguistik menjadi lebih luas dengan wujudnya cabang seperti linguistik korpus, pemprosesan bahasa semula jadi (NLP), dan analisis leksikal. Kajian linguistik Melayu membolehkan pembangunan model kecerdasan buatan yang memahami konteks, dialek, dan penggunaan bahasa dalam komunikasi harian. Keperluan mengintegrasikan kecerdasan buatan dan ilmu linguistik Melayu ini akan memberikan banyak manfaat kepada pelbagai pihak. Program Perkamusan Melayu Digital yang dijalankan oleh Dewan Bahasa dan Pustaka (DBP) juga telah membangunkan pelbagai pangkalan data leksikal yang digunakan untuk menyokong pembangunan AI seperti terjemahan mesin dan pengesanan kesalahan tatabahasa. Projek seperti WordNet Bahasa Melayu dan Malay Treebank juga membantu dalam meningkatkan keupayaan AI untuk memahami struktur dan semantik bahasa Melayu dengan lebih baik.

AI, terutamanya dalam bidang pemprosesan bahasa semula jadi (NLP), telah membuka peluang besar dalam kajian dan penggunaan linguistik Melayu. Beberapa aplikasi penting yang telah muncul termasuk seperti terjemahan mesin, terjemahan dalam talian seperti *Google Translate* dan *Yandex* telah menyertakan bahasa Melayu dalam sistem mereka, namun masih terdapat banyak cabaran dalam mengekalkan ketepatan makna dan konteks budaya. Model kecerdasan buatan seperti Ilmu 0.1 yang dibangunkan oleh YTL AI Labs menunjukkan peningkatan dalam pemahaman teks Melayu secara lebih kontekstual. AI boleh digunakan dalam pembangunan alat pembetulan tatabahasa seperti *Cekat-Tatabahasa* dan *Grammarly* untuk Bahasa Melayu bagi membantu pengguna menulis dengan lebih tepat.

Kecerdasan buatan yang dilatih menggunakan linguistik Melayu boleh membantu membangunkan chatbot berbahasa Melayu yang lebih pintar untuk sektor perkhidmatan pelanggan, pendidikan, dan e-dagang. kecerdasan buatan digunakan dalam mengenal pasti emosi dan pandangan masyarakat melalui analisis sentimen di platform media sosial. Ini membantu kerajaan dan syarikat memahami keperluan dan aspirasi rakyat.

Bagi memastikan usaha untuk mengintegrasikan program linguistik Melayu dengan kecerdasan buatan ini pembangunan Korpus Melayu yang Lebih Luas diperlukan. Institusi seperti DBP dan universiti perlu mempercepatkan usaha membina korpus bahasa Melayu yang komprehensif. Universiti dan penyelidik perlu memperluaskan kajian dalam pemprosesan bahasa Melayu untuk memastikan AI dapat memahami bahasa ini dengan lebih baik. Dasar yang lebih ketat perlu dilaksanakan bagi memastikan AI tidak menyalahgunakan data pengguna

atau menghasilkan bias yang tidak diinginkan. Seterusnya ahli akademik, syarikat teknologi, dan kerajaan perlu bekerjasama dalam memperkasa kecerdasan buatan yang berorientasikan linguistik Melayu.

KESIMPULAN

Kesimpulannya, integrasi kecerdasan buatan (AI) dengan linguistik Melayu merupakan langkah strategik yang signifikan dalam memastikan bahasa Melayu terus relevan dan berdaya saing dalam era digital. Melalui pembangunan sistem pemprosesan bahasa semula jadi (NLP) yang canggih, AI berpotensi untuk memahami, mentafsir, dan menghasilkan teks dalam bahasa Melayu dengan lebih tepat dan efisien. Namun, usaha ini memerlukan komitmen yang tinggi dalam penyelidikan dan pembangunan, khususnya dalam mengumpul dan menganalisis data linguistik Melayu yang berkualiti tinggi. Di samping itu, aspek etika dan privasi perlu diberi perhatian serius bagi memastikan penggunaan kecerdasan buatan tidak menjejaskan hak dan kebebasan individu.

Implikasi daripada integrasi ini adalah luas, termasuk peningkatan kecekapan dalam aplikasi seperti penterjemahan automatik, pembelajaran adaptif, dan analisis sentimen dalam bahasa Melayu. Namun, tanpa data yang mencukupi dan berkualiti, model kecerdasan buatan mungkin gagal memahami nuansa dan konteks budaya bahasa Melayu, yang boleh menjejaskan keberkesanan aplikasi tersebut. Oleh itu, adalah penting untuk memastikan penglibatan pelbagai pihak dalam usaha pengumpulan data dan pembangunan model kecerdasan buatan yang sensitif terhadap budaya tempatan. Bagi mengatasi cabaran yang dikenalpasti, kajian lanjut disarankan untuk memberi tumpuan kepada beberapa aspek kritikal. Pertama, pembangunan korpus digital bahasa Melayu yang komprehensif dan representatif perlu dipergiatkan. Kedua, penyelidikan mendalam mengenai algoritma NLP yang sesuai dengan struktur dan ciri unik bahasa Melayu adalah penting untuk meningkatkan ketepatan dan keberkesanan model kecerdasan buatan. Ketiga, kajian mengenai aspek etika dan privasi dalam penggunaan kecerdasan buatan perlu diperluas untuk memastikan pematuhan kepada standard etika dan perlindungan data. Akhirnya, kerjasama antara institusi penyelidikan, industri, dan pihak berkuasa perlu diperkukuhkan untuk menyokong pembangunan dan penerapan AI dalam konteks linguistik Melayu.

Dengan pendekatan yang holistik dan beretika, integrasi kecerdasan buatan dan linguistik Melayu berpotensi untuk memperkukuh identiti budaya dan sosial masyarakat, selaras dengan aspirasi Masyarakat 5.0 yang menekankan keseimbangan antara kemajuan teknologi dan kesejahteraan insan.

RUJUKAN

Abeillé, A., & Borsley, R. D. 2021. Basic properties and elements (S. Müller, Abeillé, R. D. Borsley, & J.-P. Koenig, Eds.; pp. 3-45). *Language Science Press*.

- Anwar, M. R., & Ahyarudin, H. A., .2023. AI-Powered Arabic Language Education in the Era of Society 5.0. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, 5(1), 50-57. Retrieved from <http://aptikom-journal.id/index.php/itsdi/article/view/607> (PDF) *AI-Powered Arabic Language Education in the Era of Society 5.0*. Available from: https://www.researchgate.net/publication/375081184_AI-Powered_Arabic_Language_Education_in_the_Era_of_Society_50
- Azhar Jaludin. 2019. Reka bentuk dan pembinaan korpus Rujukan Bahasa Melayu berasaskan teks we. UKM, Bangi. <https://ptsldigital.ukm.my/jspui/handle/123456789/497831>
- Dash & Niladri Sekhar. 2024. "Corpus Linguistics and Language Technology". In: Sin-Wai, Chan; Kin-Wah, Mak; and Sze-Ming, Leung (Eds.) *The Routledge Encyclopedia of Technology and the Humanities*. London: Routledge. ISBN: 9781032049427. Chapter 12. Pp. 219-245. https://www.researchgate.net/publication/380073601_Corpus_Linguistics_and_Language_Technology
- Deguchi, A., et al. 2020. *Society 5.0: A Japanese Concept for a Super-smart Society*. Springer.
- Ezuana Sukawai & Nazlia Omar. 2020. Pembangunan korpus bagi analisis sentimen dalam Bahasa Melayu secara separa selia. *Asia-Pacific Journal of Information Technology and Multimedia*, 9 (1). pp. 94-109. ISSN 2289-2192 <https://doi.org/10.11113/lspi.v9.19469>
- Isam, H. & Abd Mutalib, M. 2019. Pemanfaatan Analisis Korpus sebagai Teknik Alternatif Pengajaran dan Pembelajaran Tatabahasa. *International Journal of Language Education and Applied Linguistics (IJLEAL)* 2019, 9 (1), 13-31.
- Jumadi, D., Mesman, N. S., Rashidi, N. H., & Ganing, P. 2024. Potensi dan Cabaran Kecerdasan buatan (AI) dalam Bidang Linguistik Melayu. *PENDETA*, 15(2), 108–118. <https://doi.org/10.37134/pendeta.vol15.2.8.2024>
- Khullar, V., Sharma, V., Angurala, M., & Chhabra, N. (Eds.). 2024. *Artificial Intelligence and Society 5.0: Issues, Opportunities, and Challenges* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781003397052>:
- Manning, C. D., & Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.
- Maria Jose Sa et al. 2021. Digital Literacy in Digital Society 5.0: Some Challenges. (2021). *Academic Journal of Interdisciplinary Studies*, 10(2).
- Mohamad Kamil, I. M., & Mohamad, M. 2020. Cabaran Pemeraksanaan Bahasa Melayu Dalam Usaha Mencapai Negara Bangsa Di Malaysia. *Asian People Journal (APJ)*, 3(2), 181-191.
- Mohd Don, Z.& Knowles, G. & J, Yong. 2006. *Word Class in Malay*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Mohd Don, Z.& Knowles, G. 2008. How words can be misleading: a study of syllable timing and "stress" in Malay. *The Linguistics Journal*, 3 (2), 66.
- Mohd Don, Z.& Knowles, G. 2021. Synthetic phonics, the digital humanities, and the teaching of beginning learners of English: Fonik Sintetik, Kemanusiaan Digital dan Pengajaran Bahasa Inggeris di Peringkat. *EDUCATUM Journal of Social Sciences*, 7 (2), 1-9.
- Mohd Don, Z.& Knowles, G. 2022. The Potential Contribution of General and Specialized Corpora to Research on Malay and Malaysian English. 2022. *LSP International Journal*,

- 9 (2), 85-96.
- Muhamad H (2024). YTL AI Labs lancar Ilmu 0.1, atasi model Bahasa Malaysia lain. Dalam https://www.astroawani.com/berita-malaysia/ytl-ai-labs-lancar-ilmu-01-atasi-model-bahasa-malaysia-lain-500365?utm_source=chatgpt.com
- Müller, S., Abeillé, A., Borsley, R. D., & Koenig, J.-P. (Eds.). 2021. *Head-Driven Phrase Structure Grammar: The handbook*. Language Science Press.
- Niam, Thian Khian; Shukri Nordin, Zaimuariffudin. Pengaruh Penggunaan Kecerdasan Buatan (AI) dalam Pengajaran dan Pembelajaran Terbeza Guru Bahasa Melayu di Sarawak, Malaysia. *Jurnal Dunia Pendidikan*, [S.l.], v. 6, n. 1, p. 617-625, apr. 2024. ISSN 2682-826X. Available at: <<https://myjms.mohe.gov.my/index.php/jdpd/article/view/26047>>
- Paige Pantulo (2023), Bagaimana Ai digunakan dalam Pendidikan & 10 Cara Anda Juga Boleh. Retrieved February 12, 2025 from https://www.classpoint.io/blog/ms/bagaimana-ai-digunakan-dalam-pendidikan-10-cara-anda-juga-boleh?utm_source=chatgpt.com
- Saleha Omar, Juhaida Abu Bakar, Maslinda Mohd. Nadzir, Nor Hazlyna Harun, & Nooraini Yusoff. 2021. Text simplification for Malay corpus: A review. International Conference Umaimah Kamarulzaman. Rozaimah Rashidin, Nurul Afnieza Md Zain & Zuraidah Jantan. 2023. Pola Bahasa Kemurungan di Media Sosial: Analisis Korpus Linguistik. *Jurnal Melayu*, 22 (2), 84-101.
- Vonia, V., Kristianto, H. ., & Parhusip, J. 2025. Etika dan Privasi dalam Penggunaan AI untuk Pengawasan Ujian Daring: Studi Kasus dan Perspektif Regulasi. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 2(1b), 1833–1839. <https://doi.org/10.32672/mister.v2i1b.2760>
- Zainab Omar, Mahzan Arshad, & Norliza Jamaludin. 2020. Pengajaran dan pembelajaran berasaskan tugasan meningkatkan motivasi dan keyakinan bertutur bahasa Melayu dalam kalangan murid sekolah rendah. *PENDETA*, 11(1), 122–132. <https://doi.org/10.37134/pendeta.vol11.1.9.2020>

Biodata Penulis:

Salinah binti Ja'afar (Ph.D) merupakan profesor madya di Akademi Pengajian Melayu, Universiti Malaya. Bidang kepakaran beliau ialah Linguistik Am, Semantik, Leksikologi, Geolinguistik, dan Penterjemahan.

Mardian Shah Omar (Ph.D) merupakan pensyarah kanan di Akademi Pengajian Melayu, Universiti Malaya. Bidang kepakaran beliau ialah Linguistik Am, Fonologi, Pembelajaran Bahasa Kedua, dan Pengucapan Awam.

Maizura Osman (Ph.D) merupakan pensyarah kanan di Akademi Pengajian Melayu, Universiti Malaya. Bidang kepakaran beliau ialah Linguistik Am, Pragmatik, dan Geolinguistik.