Association of Researcher of Skills and Vocational Training, Malaysia

ANP-JSSH

ISSN 2773-482X eISSN 2785-8863
DOI: https://doi.org/10.53797/anp.jssh.v7i1.7.2026

Penggunaan AI Chatbot sebagai Alat Sokongan Pembelajaran Diagnostik Kerosakan Enjin dalam Kalangan Pelajar Teknologi Automotif Kolej Vokasional: Satu Kajian Tindakan

(The Use of an AI Chatbot as a Learning Support Tool for Engine Fault Diagnosis Among Automotive Technology Students at a Vocational College: An Action Research Study)

Abd Hamid, Mohd Sazali[1*], Marjuni, Azrulhisham[1], Nasruddin, Muhammad Amri[1], Abdullah, Muhamad Solleh[1], Shahnon, Abdullah Hakim[1] & Zuraidi, Nur Eisha Fatini[1]

[1]Program Teknologi Automotif, Kolej Vokasional Sungai Buloh, Jalan Kuala Selangor, U20 Shah Alam, 47000 Sungai Buloh, Selangor, MALAYSIA

*Corresponding author email: mohdsazaliabdhamid@kvsbkpm.edu.my

Received: 16 May 2026; Revised: 18 May 2026; Accepted: 25 June 2026; Available Online: 30 June 2026

Abstrak: Kajian tindakan ini meneliti penggunaan AI Chatbot sebagai alat sokongan pembelajaran diagnostik kerosakan enjin dalam kalangan 30 orang pelajar Sijil Vokasional Malaysia (SVM) Teknologi Automotif Semester 1 di sebuah kolej vokasional di Selangor. Kajian dilaksanakan mengikut model kajian tindakan Kemmis dan McTaggart menggunakan reka bentuk satu kumpulan ujian pra dan ujian pos dalam tempoh kajian tujuh minggu. AI Chatbot tidak digunakan sebagai sumber diagnosis tunggal kerana setiap respons wajib disahkan melalui manual servis, pemeriksaan fizikal dan bimbingan guru mengikut protokol pengesahan berlapis. Skor min meningkat daripada 39.10 (SP = 3.42) kepada 82.70 (SP = 5.91), iaitu perbezaan 43.60 mata, 95% SK [41.37, 45.83], $t(29) = 39.93$, $p < .001$, dengan peningkatan ternormal 0.716. Kesemua 30 peserta merekodkan peningkatan, tetapi hanya 19 orang mencapai kategori N-Gain tinggi. Saiz kesan Cohen dz sebanyak 7.29 tidak ditafsirkan sebagai kesan intervensi kerana sisihan piawai ujian pra yang sempit dan ketiadaan kumpulan kawalan menjadikan nilai tersebut mencerminkan kehomogenan sampel dan pembelajaran kurikulum biasa. Soal selidik 20 item merekodkan min keseluruhan 4.31 (SP = 0.32) dengan alfa Cronbach .897, 95% SK [.835, .943], namun 96% daripada 600 respons berada pada skala 4 atau 5 dan tiada satu pun respons di bawah 3, iaitu pola yang konsisten dengan kecenderungan jawapan yang diingini secara sosial memandangkan pengkaji juga guru kelas. Dapatan dilaporkan sebagai perubahan pencapaian dan penerimaan selepas intervensi, bukan bukti kausal. Sumbangan utama kajian ialah protokol pengesahan berlapis bagi pengajaran diagnostik TVET.

Kata Kunci: AI Chatbot, Diagnostik Kerosakan Enjin, Teknologi Automotif, Kajian Tindakan, TVET

Abstract: *This action research study examined the use of an AI chatbot as a learning support tool for engine fault diagnosis among 30 first-semester Malaysian Vocational Certificate (SVM) Automotive Technology students at a vocational college in Selangor. It followed the Kemmis and McTaggart action research model using a one-group pre-test and post-test design across a seven-week study period. The chatbot was not a sole diagnostic source: every response had to be verified against the service manual, physical inspection and teacher confirmation under a layered verification protocol. The mean score rose from 39.10 (SD = 3.42) to 82.70 (SD = 5.91), a difference of 43.60 points, 95% CI [41.37, 45.83], $t(29) = 39.93$, $p < .001$, with a normalized gain of 0.716. All 30 participants improved, but only 19 reached the high N-Gain category. The Cohen's dz of 7.29 is not interpreted as an intervention effect: the narrow pre-test standard deviation and the absence of a control group mean it reflects sample homogeneity and ordinary curriculum learning. A 20-item questionnaire recorded an overall mean of 4.31 (SD = 0.32) with a Cronbach's alpha of .897, 95% CI [.835, .943]; however, 96% of the 600 responses fell on scale points 4 or 5 and none below 3, a pattern consistent with socially desirable responding given that the researcher was also the class teacher. Findings are reported as changes in achievement and acceptance following the intervention, not as causal evidence. The study's contribution is a layered verification protocol for TVET diagnostic teaching.*

Keywords: *AI Chatbot, Engine Fault Diagnosis, Automotive Technology, Action Research, TVET*

**Corresponding author: mohdsazaliabdhamid@kvsbkpm.edu.my*

1. Pengenalan

Pendidikan dan Latihan Teknikal dan Vokasional (TVET) merupakan tunjang penyediaan tenaga kerja mahir negara. Dasar TVET Negara 2030 meletakkan penjajaran kurikulum dengan keperluan industri dan penguasaan kemahiran digital sebagai teras utama transformasi TVET Malaysia (Majlis TVET Negara, 2024). Dalam bidang Teknologi Automotif, tuntutan tersebut bermakna pelajar bukan sahaja perlu menguasai prosedur kerja bengkel, tetapi juga perlu mampu menaakul secara sistematik apabila berhadapan dengan kerosakan yang puncanya tidak dinyatakan kepada mereka.

Kemahiran diagnostik kerosakan enjin bukan sekadar keupayaan mengenal nama atau fungsi komponen. Jonassen dan Hung (2006) menjelaskan bahawa kompetensi penyelesaian masalah troubleshooting menuntut integrasi tiga jenis pengetahuan, iaitu pengetahuan konseptual tentang sistem, pengetahuan strategik tentang urutan pemeriksaan, dan pengetahuan pengalaman berasaskan kes yang pernah ditemui. Abele (2018), melalui analisis fail log 339 orang perantis mekatronik automotif, mendapati bahawa kejayaan diagnostik paling banyak ditentukan oleh keupayaan menguji hipotesis secara sistematik, dan bukan oleh kelajuan atau jumlah maklumat yang dikumpulkan. Kedua-dua kajian tersebut menunjukkan bahawa pembelajaran diagnostik perlu melatih proses penaakulan, bukan sekadar menyampaikan jawapan.

Konteks kolej vokasional mempunyai ciri yang tersendiri. Kurikulum Standard Kolej Vokasional (KSKV) dibina berasaskan proses kerja sebenar industri, dengan penguasaan kompetensi diukur melalui pelaksanaan tugasan amali yang berperingkat (Abd Samad et al., 2017). Waktu amali bengkel adalah terhad dan sukar diganti, manakala seorang guru lazimnya perlu menyelia satu kumpulan besar pelajar yang bekerja pada beberapa stesen enjin secara serentak. Dalam keadaan itu, seorang pelajar yang tersekat dalam proses diagnostik sering terpaksa menunggu giliran untuk mendapatkan bimbingan, dan tempoh menunggu tersebut merupakan masa pembelajaran yang hilang.

Keadaan ini menjadi lebih ketara bagi pelajar Semester 1 yang masih membina asas pengetahuan tentang komponen, fungsi sistem dan prosedur pemeriksaan. Chernikova et al. (2020), melalui meta-analisis 145 kajian pembelajaran berasaskan simulasi, mendapati bahawa pelajar berpengetahuan awal rendah memperoleh manfaat terbesar daripada sokongan berasaskan contoh dan panduan berstruktur, manakala pelajar berpengetahuan awal tinggi lebih mendapat manfaat daripada peluang refleksi terbuka. Dapatan tersebut mencadangkan bahawa sebarang alat sokongan yang diperkenalkan kepada pelajar Semester 1 perlu berstruktur dan bukan sekadar menyediakan ruang pertanyaan bebas.

Dari aspek amalan pengajaran pula, kajian tempatan menunjukkan bahawa penggunaan alat bantu mengajar berasaskan teknologi dalam bilik darjah vokasional Malaysia masih terhad (Abdul Rahman et al., 2020). Pada masa yang sama, Ab. Jalil dan Lee (2025) melaporkan bahawa institusi TVET Malaysia merekodkan tahap kesediaan yang tinggi terhadap penggunaan teknologi kecerdasan buatan, walaupun masih wujud jurang dari segi infrastruktur dan pendedahan. Md Yazid dan Ali (2025), dalam kajian terhadap 374 orang pelajar Kolej Vokasional Zon Selatan, mendapati minat dan sikap merupakan peramal signifikan kepada motivasi pelajar. Gabungan dapatan ini menunjukkan bahawa terdapat ruang dan kesediaan untuk memperkenalkan sokongan digital dalam pembelajaran kolej vokasional.

Beberapa kajian dalam jurnal ARSVOT telah merintis pembangunan alat bantu digital untuk konteks kolej vokasional. Sangin dan Azman (2024) membangunkan aplikasi mudah alih bagi topik keselamatan bengkel sebagai nilai tambah kepada kursus vokasional, manakala Norazizan et al. (2025) membangunkan aplikasi mudah alih dalaman di Kolej Vokasional Sungai Buloh. Jefferi et al. (2025) pula membangunkan alat bantu bengkel khusus untuk pelajar Teknologi Automotif kolej vokasional. Ismail et al. (2025) menilai kebolehgunaan komik interaktif berasaskan amalan 5S dalam kalangan 260 orang pelajar vokasional dan merekodkan min kepuasan keseluruhan 4.502 (SP = 0.403). Kajian-kajian tersebut menunjukkan pola yang konsisten, iaitu pembangunan alat bantu diikuti penilaian penerimaan pengguna.

Penggunaan kecerdasan buatan generatif dalam pendidikan turut disokong oleh kerangka dasar. Dasar Pendidikan Digital menggariskan hala tuju penyepaduan teknologi digital dalam pengajaran dan pembelajaran di institusi pendidikan Kementerian Pendidikan Malaysia (Kementerian Pendidikan Malaysia, 2023). Garis panduan Kementerian Pendidikan Tinggi (2024) pula menegaskan bahawa kecerdasan buatan generatif hendaklah digunakan sebagai alat sokongan pembelajaran dan bukan sebagai pengganti kefahaman pelajar. Pada peringkat antarabangsa, Miao dan Holmes (2023) menekankan pendekatan berpusatkan manusia, iaitu keputusan akhir kekal berada pada pengguna dan pendidik.

Dari sudut teori, sokongan yang diberikan oleh chatbot boleh difahami melalui konsep zon perkembangan proksimal, iaitu jarak antara apa yang mampu dilakukan pelajar secara bersendirian dengan apa yang mampu dilakukannya dengan bantuan (Vygotsky, 1978). Collins et al. (1989) memperincikan mekanisme tersebut melalui kerangka perantisan kognitif, iaitu pemodelan, pembimbingan, perancah, artikulasi dan refleksi, yang bertujuan menjadikan proses pemikiran pakar kelihatan kepada pelajar baharu. Zhou et al. (2025), dalam kajian terhadap 29 orang pelajar, mendapati perancah berbantukan ChatGPT meningkatkan prestasi penyelesaian masalah, tetapi analisis jujukan menunjukkan pelajar berprestasi tinggi beralih kepada soalan bersifat metakognitif manakala pelajar berprestasi rendah kekal pada soalan bersifat prosedur.

Bukti empirikal tentang chatbot dalam pendidikan berkembang pesat. Albadarin et al. (2024) dan Lo (2023) merekodkan pertumbuhan mendadak kajian empirikal berkaitan ChatGPT, manakala Mohebi (2024) dan McGrath et al. (2025) memetakan bidang tersebut sebagai bidang penyelidikan yang masih baharu. Guan et al. (2025) menunjukkan bahawa chatbot pendidikan berpotensi menyokong proses pembelajaran terarah kendiri, manakala Lo et al. (2024)

meneliti hubungannya dengan penglibatan pelajar. Dalam konteks kejuruteraan, Al Hakim et al. (2024) mendapati chatbot yang dibina khusus dan diikat kepada kandungan kursus mengatasi penggunaan ChatGPT generik dan pengajaran konvensional dalam kalangan 106 orang pelajar kejuruteraan elektrik, manakala Sell et al. (2025) melaporkan peningkatan akses pelajar dalam kursus mekanik kejuruteraan sambil menegaskan pengurusan halusinasi AI sebagai prasyarat pelaksanaan.

Walau bagaimanapun, bukti tersebut tidak seragam dan tidak wajar dibaca sebagai jaminan. Sailer et al. (2024), melalui tinjauan sistematik meta-analisis dan meta-analisis tahap kedua, menyimpulkan bahawa kesan pembelajaran bergantung kepada aktiviti pembelajaran yang disokong oleh teknologi dan bukan kepada teknologi itu sendiri. Hillmayr et al. (2020) turut mendapati kesan alat digital bergantung kepada latihan guru dan cara alat itu digunakan bersama pengajaran sedia ada. Jin dan Sercu (2025), dalam tinjauan 21 kajian eksperimen, mendapati intervensi jangka sederhana memberikan hasil terbaik sementara kesan jangka pendek adalah terhad, dan peningkatan pengetahuan melebihi peningkatan kemahiran. Deng et al. (2025) sendiri, walaupun melaporkan kesan positif ChatGPT merentas 69 kajian eksperimen, memperingatkan tentang saiz sampel yang kecil, pengukuran yang dibuat sejurus selepas intervensi, dan kesan kebaharuan.

Dua kaveat reka bentuk khususnya relevan kepada kajian seperti ini. Pertama, Rodrigues et al. (2022), dalam kajian longitudinal 14 minggu terhadap 756 orang pelajar, menunjukkan bahawa manfaat motivasi sesuatu inovasi tidak stabil sepanjang masa dan boleh menurun pada peringkat awal sebelum pulih semula. Kesannya, sesuatu intervensi yang singkat berkemungkinan merekodkan kesan yang dipengaruhi kebaharuan. Kedua, Deslauriers et al. (2019) menunjukkan bahawa persepsi pelajar terhadap pembelajaran mereka sendiri boleh bergerak berlawanan arah dengan pencapaian yang diukur secara objektif. Oleh itu, min persepsi yang tinggi tidak boleh ditafsirkan sebagai ukuran pembelajaran.

Terdapat juga risiko yang khusus kepada domain teknikal. Walters dan Wilder (2023) mendapati sebahagian besar petikan bibliografi yang dijana model bahasa besar adalah rekaan atau mengandungi ralat, manakala Kasneci et al. (2023) menyenaraikan ketidaktepatan maklumat sebagai cabaran utama penggunaan model bahasa besar dalam pendidikan. Zhai et al. (2024) melaporkan bahawa pergantungan berlebihan kepada sistem dialog AI dikaitkan dengan pelemahan keupayaan membuat keputusan dan penaakulan analitik, manakala Fan et al. (2025) mendapati sokongan ChatGPT memperbaik hasil tugasan serta-merta tetapi tidak memperbaik pemindahan pengetahuan, disertai kecenderungan pelajar menyerahkan proses regulasi kendiri kepada alat tersebut. Dalam bidang automotif, penerimaan respons AI tanpa pengesahan bukan sekadar risiko akademik kerana diagnosis yang salah boleh mengakibatkan kerja pembaikan yang tidak selamat.

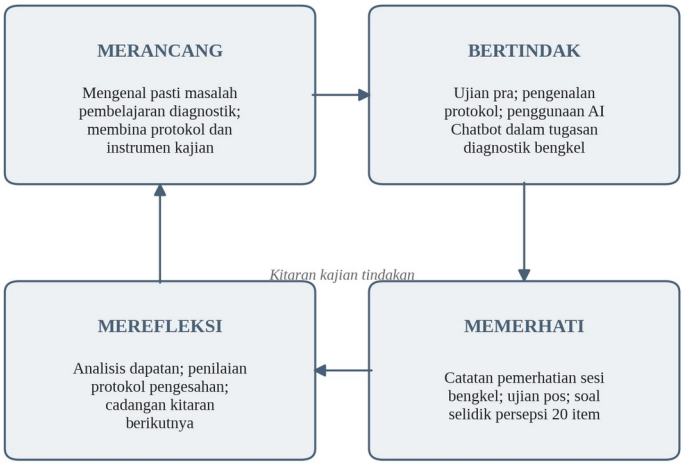
Justeru, terdapat satu jurang yang jelas. Kebanyakan kajian sedia ada meneliti penggunaan chatbot dalam konteks pendidikan tinggi dan tugasan berasaskan teks, manakala kajian empirikal yang boleh disahkan tentang penggunaan kecerdasan buatan generatif dalam pembelajaran diagnostik automotif di peringkat sijil vokasional masih sukar ditemui. Lebih penting lagi, kajian sedia ada jarang memerihalkan prosedur operasi yang mengikat respons AI kepada sumber teknikal yang sah. Kajian tindakan ini cuba menangani jurang tersebut dengan memerihalkan dan menilai satu protokol penggunaan AI Chatbot yang setiap responsnya disahkan melalui manual servis, pemeriksaan fizikal dan bimbingan guru.

Sehubungan itu, kajian ini mempunyai tiga objektif, iaitu (i) merekodkan perubahan pencapaian diagnostik pelajar sebelum dan selepas pelaksanaan intervensi AI Chatbot; (ii) menentukan tahap peningkatan pembelajaran menggunakan nilai peningkatan ternormal (N-Gain); dan (iii) meneliti persepsi pelajar terhadap penggunaan AI Chatbot dalam pembelajaran diagnostik automotif. Persoalan kajian yang sepadan ialah (i) apakah perubahan pencapaian diagnostik peserta selepas intervensi; (ii) apakah tahap peningkatan pembelajaran yang direkodkan; dan (iii) bagaimanakah peserta menanggapi penggunaan AI Chatbot sebagai alat sokongan. Perlu ditegaskan bahawa kajian ini tidak mendakwa membuktikan keberkesanan kausal AI Chatbot. Reka bentuk satu kumpulan tanpa kumpulan kawalan tidak membenarkan dakwaan sedemikian, dan dapatan dilaporkan sebagai perubahan yang direkodkan dalam satu konteks kelas yang khusus.

2. Metodologi

2.1 Reka Bentuk Kajian

Kajian ini merupakan kajian tindakan yang dilaksanakan mengikut model Kemmis dan McTaggart, iaitu satu kitaran berulang yang merangkumi fasa merancang, bertindak, memerhati dan merefleksi (Kemmis et al., 2014). Model tersebut dipilih kerana pengkaji merupakan guru kepada kelas yang terlibat, dan tujuan kajian adalah menambah baik amalan pengajaran sendiri dalam konteks bilik darjah yang sebenar. Rajah 1 memaparkan kitaran kajian tindakan berserta tindakan sebenar yang dilaksanakan dalam setiap fasa.



Rajah 1: Kitaran Kajian Tindakan Kemmis dan McTaggart Berserta Tindakan Sebenar dalam Setiap Fasa

Dalam kitaran tersebut, satu reka bentuk kuantitatif satu kumpulan ujian pra dan ujian pos digunakan bagi merekodkan perubahan pencapaian peserta. Kajian ini tidak menggunakan kumpulan kawalan dan tidak melibatkan pengagihan rawak kerana keseluruhan kelas menerima intervensi yang sama sebagai sebahagian daripada pengajaran biasa. Implikasi metodologi pilihan tersebut dinyatakan secara terperinci dalam bahagian 4.9.

2.2 Peserta Kajian

Peserta kajian terdiri daripada 30 orang pelajar Sijil Vokasional Malaysia (SVM) Teknologi Automotif Semester 1 di Kolej Vokasional Sungai Buloh, Selangor. Pensampelan bertujuan digunakan, iaitu peserta dipilih kerana mereka merupakan kelas yang diajar oleh pengkaji dan sedang mengikuti modul yang melibatkan pemeriksaan dan diagnosis kerosakan enjin. Kesemua peserta mengikuti intervensi yang sama dan menduduki ujian pra serta ujian pos yang sama. Tiada peserta tercicir dalam tempoh kajian.

2.3 Instrumen Kajian

Tiga instrumen digunakan dalam kajian ini. Pertama, ujian pra dan ujian pos yang mengukur pencapaian diagnostik peserta dengan skor penuh 100 mata. Kedua, soal selidik persepsi yang mengandungi 20 item berskala Likert lima mata. Item soal selidik disusun dalam lima konstruk empat item, iaitu kemudahan penggunaan (A1 hingga A4), kefahaman diagnostik (B1 hingga B4), keyakinan dan penglibatan (C1 hingga C4), pembelajaran kendiri (D1 hingga D4) dan penerimaan keseluruhan (E1 hingga E4). Ketiga, catatan pemerhatian guru sepanjang sesi bengkel, yang merekodkan tingkah laku pembelajaran peserta.

Temu bual tidak dilaksanakan dan tidak digunakan sebagai sumber data. Catatan pemerhatian yang digunakan dalam kajian ini merupakan catatan naratif guru dan bukan instrumen berskala yang dikod menggunakan rubrik berkekerapan. Kekangan ini dinyatakan semula dalam bahagian 4.9.

2.4 Kesahan dan Kebolehpercayaan

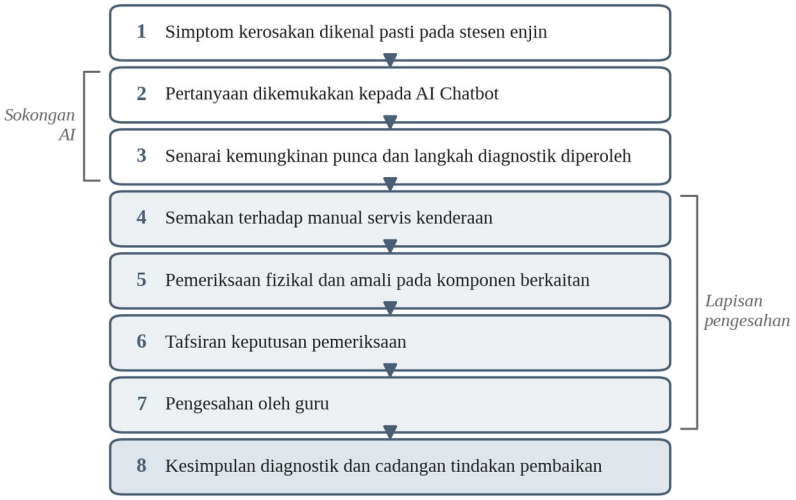
Item ujian dan item soal selidik melalui proses semakan dan pemurnian sebelum digunakan. Semakan tersebut memberi tumpuan kepada keselarasan item dengan objektif kajian, kesesuaian aras kesukaran dengan pelajar Semester 1, kejelasan arahan dan kesesuaian bahasa.

Analisis kebolehpercayaan bagi keseluruhan 20 item merekodkan nilai alfa Cronbach .897, dengan selang keyakinan 95 peratus [.835, .943] mengikut kaedah Feldt. Nilai tersebut menunjukkan konsistensi dalaman yang tinggi bagi skor jumlah, tetapi tiga kaveat perlu dinyatakan. Pertama, selang keyakinan itu luas kerana sampel hanya 30 orang, iaitu saiz yang mencukupi untuk menguji alfa berbanding sifar tetapi tidak untuk menganggarkan nilai alfa yang tinggi dengan tepat (Bujang et al., 2018). Kedua, alfa merupakan ukuran konsistensi dalaman dan bukan bukti tunggal kesahan instrumen (Taber, 2018). Ketiga, nilai alfa yang tinggi bagi skor jumlah tidak bermakna setiap konstruk boleh dilaporkan secara berasingan, dan analisis dalam bahagian 3.4 menunjukkan bahawa alfa peringkat konstruk jauh lebih rendah.

2.5 Protokol Penggunaan AI Chatbot

Intervensi dilaksanakan dalam tempoh kajian 22 Julai 2026 hingga 9 September 2026. AI Chatbot yang digunakan ialah perkhidmatan chatbot generatif awam yang dicapai melalui telefon pintar pelajar sendiri di dalam bengkel. Penggunaan chatbot tidak bertujuan membolehkan pelajar menerima diagnosis secara terus. Sebaliknya, chatbot digunakan sebagai sumber rangsangan kognitif apabila pelajar tersekat dalam proses penaakulan, dan setiap respons yang diperoleh wajib melalui langkah pengesahan sebelum diterima. Rajah 2 memaparkan protokol lapan langkah yang digunakan.

Protokol tersebut menetapkan bahawa pelajar hanya boleh membuat kesimpulan diagnostik selepas respons chatbot dibandingkan dengan manual servis kenderaan, diuji melalui pemeriksaan fizikal pada stesen enjin, dan disahkan oleh guru. Susunan ini penting kerana model bahasa besar berkemungkinan menjana maklumat yang tidak tepat atau tidak sesuai dengan spesifikasi kenderaan tertentu (Kasneci et al., 2023; Walters & Wilder, 2023).



Rajah 2: Protokol Lapan Langkah Penggunaan AI Chatbot dalam Proses Diagnostik Kerosakan Enjin

Jadual 1 memperincikan fasa pelaksanaan intervensi mengikut tempoh sebenar kajian, iaitu daripada pemerhatian awal sehingga penulisan laporan.

Jadual 1: Fasa Pelaksanaan Intervensi AI Chatbot

Tempoh (2026)	Fokus	Aktiviti utama
22 hingga 26 Julai	Pemerhatian awal dan ujian pra	Mengenal pasti masalah pembelajaran diagnostik dan merekodkan pencapaian awal peserta
27 Julai hingga 2 Ogos	Pengenalan protokol	Penerangan fungsi AI Chatbot, batasannya dan protokol pengesahan lapan langkah
3 hingga 9 Ogos	Intervensi	Penggunaan AI Chatbot dalam tugasan diagnostik bengkel dengan semakan manual servis
10 hingga 16 Ogos	Intervensi dan penilaian	Penggunaan berterusan, diikuti ujian pos dan soal selidik persepsi
17 hingga 30 Ogos	Analisis	Analisis data ujian dan soal selidik
31 Ogos hingga 9 September	Refleksi	Refleksi amalan pengajaran dan penulisan laporan kajian

Nota: Tempoh keseluruhan kajian ialah tujuh minggu, iaitu 22 Julai hingga 9 September 2026. Fasa penggunaan AI Chatbot dalam tugasan diagnostik bengkel berlangsung selama tiga minggu, iaitu 27 Julai hingga 16 Ogos 2026.

Sepanjang fasa intervensi, guru berperanan sebagai fasilitator dan pengesah. Guru tidak memberikan jawapan diagnostik secara terus, sebaliknya menyemak kesahihan penaakulan pelajar, membetulkan salah faham konsep, dan mengesahkan keputusan akhir sebelum sebarang tindakan pembaikan dicadangkan.

2.6 Analisis Data

Skor ujian dianalisis dengan mengira min, sisihan piawai, skor minimum dan skor maksimum bagi ujian pra dan ujian pos. Kenormalan taburan perbezaan skor berpasangan diuji menggunakan ujian Shapiro-Wilk sebelum ujian inferensi dijalankan. Perbezaan antara ujian pra dan ujian pos diuji menggunakan ujian-t sampel berpasangan dan disemak semula

menggunakan ujian Wilcoxon bertanda pangkat sebagai semakan bukan berparameter. Saiz kesan dilaporkan sebagai Cohen dz, iaitu min perbezaan dibahagi sisihan piawai perbezaan, berserta selang keyakinan 95 peratus berasaskan taburan-t bukan berpusat.

Tahap peningkatan pembelajaran ditentukan menggunakan nilai peningkatan ternormal mengikut formula Hake (1998), dan ditafsirkan mengikut kategori yang sama, iaitu tinggi bagi nilai 0.70 ke atas, sederhana bagi nilai 0.30 hingga 0.69, dan rendah bagi nilai di bawah 0.30. Nilai tersebut dikira dengan dua cara, iaitu daripada skor min kumpulan dan sebagai purata nilai peningkatan ternormal setiap peserta, kerana kedua-dua kaedah boleh menghasilkan nilai yang berbeza apabila taburan skor tidak sekata.

Data soal selidik dianalisis pada tiga peringkat. Peringkat item melaporkan min, sisihan piawai, taburan penuh respons dan korelasi item dengan jumlah baki item. Peringkat konstruk melaporkan min, sisihan piawai dan alfa Cronbach bagi setiap set empat item. Peringkat instrumen melaporkan min keseluruhan dan alfa Cronbach bagi kesemua 20 item, berserta selang keyakinan Feldt. Kaedah tafsiran min mengikut skala Likert yang digunakan di sini melaporkan min bersama sisihan piawai dan taburan respons, dan tidak menggunakan jadual pembahagian tahap yang lazim dalam tesis tempatan kerana sumber asal jadual tersebut tidak dapat disahkan (Sullivan & Artino, 2013). Catatan pemerhatian dirumuskan secara tematik.

2.7 Pertimbangan Etika

Kebenaran melaksanakan kajian diperoleh daripada pihak pentadbiran kolej, dan peserta dimaklumkan tentang tujuan kajian serta penggunaan data sebelum pengumpulan data bermula. Semua data dilaporkan secara agregat. Nama, nombor pendaftaran dan sebarang maklumat pengenalan peserta tidak disertakan.

Gambar pelaksanaan program yang memaparkan wajah pelajar tidak disertakan dalam artikel ini kerana kebenaran bertulis penjaga bagi penerbitan imej tersebut tidak diperoleh. Peserta kajian merupakan pelajar Semester 1 yang sebahagian besarnya berumur di bawah 18 tahun, dan penerbitan imej mereka dalam jurnal akses terbuka memerlukan kebenaran bertulis penjaga yang khusus.

Satu batasan etika perlu dinyatakan secara terbuka. Dalam kajian ini, pengkaji merupakan guru kelas, pereka bentuk intervensi, pemerhati proses pembelajaran dan pentadbir soal selidik yang menilai intervensi tersebut. Keadaan ini menimbulkan risiko kecenderungan pengkaji dan risiko peserta memberikan jawapan yang dianggap diingini oleh guru. Bagi mengurangkan risiko tersebut, soal selidik dijawab tanpa nama dan peserta dimaklumkan bahawa jawapan mereka tidak mempengaruhi pentaksiran kursus. Risiko ini dinyatakan semula dalam bahagian 4.9.

3. Dapatan Kajian

3.1 Pencapaian Ujian Pra dan Ujian Pos

Jadual 2 memaparkan statistik deskriptif bagi kedua-dua ujian.

Jadual 2: Statistik Deskriptif Ujian Pra dan Ujian Pos (n = 30)

Ukuran	Ujian pra	Ujian pos	Perbezaan
Min	39.10	82.70	43.60
Sisihan piawai	3.42	5.91	5.98
Skor minimum	32	66	26
Skor maksimum	46	92	56

Nota: Skor penuh ujian ialah 100 mata. Selang keyakinan 95 peratus bagi perbezaan min ialah [41.37, 45.83]. Korelasi antara skor ujian pra dan ujian pos ialah r = .27. Ujian Shapiro-Wilk terhadap perbezaan berpasangan: W = .943, p = .109. Ujian-t sampel berpasangan: t(29) = 39.93, p < .001; ujian Wilcoxon bertanda pangkat: p < .001. Cohen dz = 7.29, 95% SK [5.39, 9.19]; lihat bahagian 4.2 tentang had tafsiran nilai ini. Peningkatan ternormal: 0.716 berdasarkan min kumpulan dan 0.717 (SP = 0.094) sebagai purata nilai individu.

Skor min peserta meningkat daripada 39.10 (SP = 3.42) pada ujian pra kepada 82.70 (SP = 5.91) pada ujian pos, iaitu perbezaan min sebanyak 43.60 mata dengan selang keyakinan 95 peratus [41.37, 45.83]. Kesemua 30 peserta merekodkan peningkatan, dengan peningkatan individu berjulat antara 26 hingga 56 mata. Tiada peserta merekodkan penurunan skor.

Ujian Shapiro-Wilk terhadap taburan perbezaan skor berpasangan tidak menolak andaian kenormalan, W = .943, p = .109, maka ujian-t sampel berpasangan digunakan. Perbezaan antara ujian pra dan ujian pos adalah signifikan secara statistik, t(29) = 39.93, p < .001, dan disahkan semula melalui ujian Wilcoxon bertanda pangkat, p < .001. Saiz kesan Cohen dz ialah 7.29 dengan selang keyakinan 95 peratus [5.39, 9.19].

Dua ciri taburan skor perlu diberikan perhatian kerana kedua-duanya menentukan cara nilai di atas boleh ditafsirkan. Pertama, sisihan piawai ujian pra hanya 3.42 mata pada skala 100 mata, dengan kesemua skor berhimpun dalam julat sempit 32 hingga 46 mata. Kedua, korelasi antara skor ujian pra dan skor ujian pos hanya r = .27, iaitu kedudukan relatif

peserta banyak berubah antara kedua-dua ujian. Implikasi kedua-dua ciri ini terhadap tafsiran saiz kesan dibincangkan dalam bahagian 4.2.

3.2 Tahap Peningkatan Pembelajaran

Nilai peningkatan ternormal yang dikira daripada skor min kumpulan ialah 0.716, iaitu berada dalam kategori tinggi mengikut Hake (1998). Nilai yang dikira sebagai purata peningkatan ternormal setiap peserta hampir sama, iaitu 0.717 (SP = 0.094), yang menunjukkan taburan peningkatan agak sekata dan kedua-dua kaedah pengiraan tidak menghasilkan gambaran yang bercanggah.

Walau bagaimanapun, nilai purata itu menyembunyikan satu perbezaan yang bermakna. Peningkatan ternormal individu berjulat antara 0.433 hingga 0.857. Apabila setiap peserta dikategorikan secara berasingan, hanya 19 daripada 30 orang peserta mencapai kategori tinggi, manakala 11 orang lagi berada dalam kategori sederhana dan tiada seorang pun dalam kategori rendah. Dengan kata lain, pernyataan bahawa kajian ini merekodkan peningkatan kategori tinggi adalah tepat pada peringkat kumpulan tetapi tidak tepat bagi lebih satu pertiga peserta secara individu.

3.3 Persepsi Pelajar Mengikut Item

Soal selidik 20 item dijawab oleh kesemua 30 orang peserta, menghasilkan 600 respons. Min keseluruhan instrumen ialah 4.31 daripada 5.00 (SP = 0.32, julat min responden 3.75 hingga 5.00). Jadual 3 memaparkan analisis mengikut item.

Jadual 3: Analisis Soal Selidik Persepsi Mengikut Item (n = 30)

Item	Min	SP	3	4	5	r item-jumlah
A1	4.30	0.60	2	17	11	0.151
A2	4.30	0.47	0	21	9	0.728
A3	4.37	0.56	1	17	12	0.406
A4	4.30	0.47	0	21	9	0.216
B1	4.33	0.55	1	18	11	0.581
B2	4.33	0.55	1	18	11	0.581
B3	4.23	0.63	3	17	10	0.597
B4	4.27	0.58	2	18	10	0.528
C1	4.23	0.63	3	17	10	0.519
C2	4.40	0.50	0	18	12	0.447
C3	4.27	0.52	1	20	9	0.625
C4	4.27	0.64	3	16	11	0.625
D1	4.30	0.53	1	19	10	0.635
D2	4.43	0.50	0	17	13	0.460
D3	4.30	0.60	2	17	11	0.570
D4	4.20	0.41	0	24	6	0.647
E1	4.30	0.60	2	17	11	0.684
E2	4.40	0.50	0	18	12	0.435
E3	4.27	0.52	1	20	9	0.707
E4	4.47	0.57	1	14	15	0.372

Nota: Skala Likert lima mata. Lajur 3, 4 dan 5 menunjukkan bilangan responden yang memilih setiap skala; tiada satu pun respons direkodkan pada skala 1 atau 2, maka kedua-dua lajur tersebut digugurkan. Lajur terakhir ialah korelasi item dengan jumlah baki item. Kod item merujuk konstruk dalam Jadual 4.

Min item berjulat sempit antara 4.20 dan 4.47. Item yang merekodkan min tertinggi ialah E4 (M = 4.47), manakala min terendah direkodkan oleh D4 (M = 4.20), diikuti B3 dan C1 (kedua-duanya M = 4.23). Julat 0.27 mata antara item tertinggi dan terendah adalah terlalu kecil untuk menyokong sebarang kesimpulan bahawa satu aspek dinilai lebih baik daripada aspek yang lain.

Taburan respons memberikan gambaran yang lebih bermakna daripada min. Daripada 600 respons yang dikumpulkan, 212 respons (35.3 peratus) berada pada skala 5, 364 respons (60.7 peratus) pada skala 4 dan 24 respons (4.0 peratus) pada skala 3. Tiada satu pun respons direkodkan pada skala 1 atau 2. Pola ini bermakna tiada seorang peserta pun,

pada mana-mana item, menyatakan sebarang tahap ketidaksetujuan terhadap penggunaan AI Chatbot. Implikasi pola tersebut dibincangkan dalam bahagian 4.4.

Analisis korelasi item dengan jumlah baki item menunjukkan kebanyakan item berfungsi dengan baik, dengan nilai antara .37 dan .73. Dua item merupakan pengecualian. Item A1 merekodkan korelasi .151 dan item A4 merekodkan .216, iaitu kedua-duanya di bawah nilai .30 yang lazimnya dijadikan penanda aras minimum. Pengguguran item A1 sahaja akan menaikkan alfa Cronbach instrumen daripada .897 kepada .903, dan A1 merupakan satu-satunya item yang pengggugurannya menaikkan nilai alfa.

3.4 Persepsi Mengikut Konstruk dan Kebolehpercayaan

Jadual 4 memaparkan analisis mengikut lima konstruk soal selidik berserta nilai alfa Cronbach bagi setiap konstruk.

Jadual 4: Analisis Mengikut Konstruk dan Kebolehpercayaan (n = 30)

Konstruk	Bil. item	Min	SP	Alfa Cronbach
Kemudahan penggunaan (A1-A4)	4	4.32	0.33	0.481
Kefahaman diagnostik (B1-B4)	4	4.29	0.42	0.692
Keyakinan dan penglibatan (C1-C4)	4	4.29	0.41	0.680
Pembelajaran kendiri (D1-D4)	4	4.31	0.38	0.706
Penerimaan keseluruhan (E1-E4)	4	4.36	0.37	0.600
Keseluruhan instrumen	20	4.31	0.32	0.897

Nota: Min dan sisihan piawai dikira daripada min setiap responden bagi konstruk berkenaan. Selang keyakinan 95 peratus bagi alfa instrumen keseluruhan ialah [.835, .943] mengikut kaedah Feldt. Empat daripada lima konstruk merekodkan alfa di bawah .70, maka min konstruk dilaporkan sebagai maklumat deskriptif sahaja dan tidak digunakan untuk membandingkan antara konstruk.

Min konstruk berjulat antara 4.29 dan 4.36, iaitu julat yang lebih sempit daripada julat item. Dapatan yang lebih penting terletak pada lajur alfa. Walaupun instrumen 20 item secara keseluruhan merekodkan alfa .897, tiada satu pun konstruk empat item mencapai nilai .70 kecuali konstruk pembelajaran kendiri yang merekodkan .706. Konstruk kemudahan penggunaan merekodkan nilai paling rendah, iaitu .481, iaitu konstruk yang sama yang mengandungi kedua-dua item berkorelasi rendah A1 dan A4.

Dapatan ini bermakna instrumen boleh dipercayai sebagai satu skala tunggal yang mengukur penerimaan secara umum, tetapi tidak boleh dipercayai sebagai lima subskala yang berasingan. Min konstruk dalam Jadual 4 oleh itu dilaporkan sebagai maklumat deskriptif sahaja dan tidak digunakan untuk membandingkan satu aspek dengan aspek yang lain.

3.5 Pemerhatian Proses Pembelajaran

Catatan pemerhatian guru sepanjang sesi bengkel merekodkan tiga pola yang berulang. Pertama, dari aspek penglibatan, peserta didapati lebih kerap mengemukakan pertanyaan dan meneroka beberapa kemungkinan punca sebelum menetapkan satu diagnosis. Kedua, dari aspek keyakinan, peserta lebih bersedia menyatakan alasan bagi sesuatu cadangan diagnostik apabila ditanya. Ketiga, dari aspek pembelajaran kendiri, peserta didapati lebih cenderung merujuk manual servis dan membuat pemeriksaan awal sendiri sebelum meminta pengesahan guru.

Pola ketiga merupakan catatan yang paling konsisten dengan tujuan asal intervensi, iaitu mengurangkan kebergantungan segera kepada guru. Namun begitu, catatan ini merupakan pemerhatian naratif guru yang juga merupakan pengkaji, tanpa rubrik berkekerapan, tanpa pemerhati kedua dan tanpa pengiraan persetujuan antara penilai. Catatan tersebut dilaporkan sebagai konteks kepada dapatan kuantitatif dan bukan sebagai data yang berdiri sendiri.

4. Perbincangan

4.1 Perubahan Pencapaian dan Apa yang Boleh Disimpulkan Daripadanya

Peningkatan min sebanyak 43.60 mata dengan selang keyakinan [41.37, 45.83] merupakan perubahan yang besar dan konsisten, dan hakikat bahawa kesemua 30 peserta meningkat tanpa seorang pun menurun menunjukkan perubahan itu bukan kesan beberapa peserta yang menonjol. Namun, kekuatan kesimpulan yang boleh dibuat daripada angka tersebut terikat kepada reka bentuk kajian. Knapp (2016) menyenaraikan empat ancaman yang tidak dapat diasingkan dalam reka bentuk satu kumpulan ujian pra dan ujian pos, iaitu sejarah, kematangan, kesan ujian dan regresi ke arah min. Dalam kajian ini, keempat-empat ancaman tersebut kekal terbuka.

Ancaman kesan ujian dan kematangan khususnya relevan. Peserta menduduki ujian pra pada awal semester ketika mereka baru mula mempelajari sistem enjin, dan skor min 39.10 mencerminkan tahap pengetahuan permulaan tersebut. Sepanjang tujuh minggu berikutnya, peserta turut mengikuti pengajaran modul yang sama seperti biasa. Oleh itu,

sebahagian daripada peningkatan sewajarnya dikaitkan dengan pembelajaran kurikulum yang berlaku secara semula jadi dalam tempoh tersebut. Ketiadaan kumpulan kawalan bermakna bahagian sumbangan setiap faktor tidak dapat diasingkan.

4.2 Mengapa Saiz Kesan yang Besar Bukan Ukuran Kesan Intervensi

Nilai Cohen dz sebanyak 7.29 jauh melebihi julat yang lazim dilaporkan dalam penyelidikan pendidikan, dan nilai sebegini tidak wajar dibaca sebagai bukti bahawa AI Chatbot mempunyai kesan yang luar biasa. Terdapat tiga sebab teknikal yang menjelaskan nilai tersebut, dan ketiga-tiganya berpunca daripada ciri sampel dan reka bentuk, bukan daripada kekuatan intervensi.

Pertama, Cohen dz membahagikan min perbezaan dengan sisihan piawai perbezaan, bukan dengan sisihan piawai skor. Dalam kajian ini sisihan piawai perbezaan hanya 5.98 mata sedangkan min perbezaan 43.60 mata, iaitu peningkatan yang hampir seragam merentas peserta. Nisbah itulah yang menghasilkan nilai 7.29. Kedua, sisihan piawai ujian pra hanya 3.42 mata pada skala 100 mata, dengan kesemua peserta berhimpun dalam julat 32 hingga 46 mata. Kesempitan itu merupakan kesan lantai yang dijangka apabila satu kelas menduduki ujian tentang kandungan yang belum diajar kepada mereka. Menukar formula tidak menyelesaikan masalah ini, kerana penyeragaman menggunakan sisihan piawai ujian pra menghasilkan nilai yang lebih besar lagi. Ketiga, dan paling penting, saiz kesan dalam reka bentuk satu kumpulan mengukur magnitud perubahan dalam tempoh kajian, bukan perbezaan antara satu keadaan dengan keadaan yang lain. Tanpa kumpulan kawalan, ia merangkumi kesan intervensi, kesan pengajaran kurikulum biasa dan kesan kematangan secara sekali gus.

Oleh itu, nilai 7.29 dilaporkan di sini demi ketelusan dan bukan sebagai sokongan kepada dakwaan keberkesanan. Pembaca yang ingin membandingkan kajian ini dengan kajian eksperimen berkumpulan kawalan digalakkan merujuk nilai peningkatan ternormal dan taburan individu dalam bahagian 3.2, yang memberikan gambaran yang lebih boleh dibandingkan.

4.3 Intervensi Bersepadu, Bukan Kesan Chatbot Semata-mata

Satu ciri penting reka bentuk intervensi ini ialah AI Chatbot tidak digunakan secara terasing. Setiap respons chatbot melalui semakan manual servis, pemeriksaan fizikal dan pengesahan guru. Kesannya, apa yang dilaksanakan sebenarnya ialah satu pakej pembelajaran diagnostik berstruktur yang mengandungi empat komponen, dan chatbot hanyalah satu daripadanya.

Pembacaan ini selari dengan kesimpulan Sailer et al. (2024) bahawa kesan pembelajaran terletak pada aktiviti pembelajaran yang disokong oleh teknologi dan bukan pada teknologi itu sendiri, serta dengan dapatan Hillmayr et al. (2020) bahawa kesan alat digital bergantung kepada cara ia disepadukan dengan pengajaran sedia ada. Dapatan Al Hakim et al. (2024) menguatkan lagi bacaan ini, iaitu chatbot yang diikat kepada kandungan kursus mengatasi penggunaan chatbot generik. Dalam kajian ini, ikatan tersebut dibuat melalui manual servis dan stesen enjin, bukan melalui pembangunan chatbot khusus.

4.4 Kepekatan Respons dan Risiko Jawapan yang Diingini

Min persepsi 4.31 daripada 5.00 menunjukkan penerimaan yang tinggi, dan nilai ini setanding dengan min kepuasan 4.502 yang dilaporkan Ismail et al. (2025) bagi alat bantu digital dalam kalangan 260 orang pelajar vokasional. Namun, taburan respons yang mendasari min itu memerlukan perhatian yang lebih daripada min itu sendiri.

Daripada 600 respons, tiada satu pun berada di bawah skala 3, dan hanya 24 respons berada pada skala 3. Dalam kalangan 30 orang pelajar yang menjawab 20 item, tiada seorang pun pernah menyatakan ketidaksetujuan terhadap sebarang aspek penggunaan AI Chatbot. Kepekatan sebegini jarang berlaku dalam data sikap yang sebenar, dan penjelasan yang paling munasabah bukan bahawa intervensi itu sempurna, tetapi bahawa keadaan pentadbiran soal selidik menggalakkan jawapan yang positif. Soal selidik ditadbir oleh guru kelas yang juga pereka intervensi dan pengkaji, di dalam bengkel yang sama, sejurus selepas intervensi. Walaupun jawapan dikumpulkan tanpa nama dan peserta dimaklumkan bahawa jawapan tidak mempengaruhi pentaksiran, susunan itu tetap meletakkan peserta dalam kedudukan menilai guru mereka sendiri.

Deslauriers et al. (2019) menunjukkan bahawa persepsi pelajar terhadap pembelajaran mereka sendiri boleh bergerak berlawanan arah dengan pencapaian yang diukur secara objektif. Digabungkan dengan kepekatan respons di atas, min persepsi dalam kajian ini paling selamat ditafsirkan sebagai petunjuk bahawa pendekatan ini tidak ditolak oleh pelajar, dan bukan sebagai ukuran manfaat pembelajaran. Terdapat juga risiko kebaharuan. Rodrigues et al. (2022) menunjukkan bahawa tindak balas motivasi terhadap sesuatu inovasi tidak stabil sepanjang masa, manakala Jin dan Sercu (2025) mendapati intervensi jangka pendek memberikan kesan yang lebih terhad berbanding intervensi jangka sederhana.

4.5 Kualiti Instrumen dan Had Tafsiran Peringkat Konstruk

Analisis peringkat item mendedahkan satu perkara yang tidak kelihatan pada nilai alfa keseluruhan. Alfa .897 bagi 20 item menunjukkan skor jumlah boleh dipercayai, tetapi alfa bagi empat daripada lima konstruk empat item berada di

bawah .70, dengan konstruk kemudahan penggunaan serendah .481. Konstruk itu juga mengandungi kedua-dua item yang berkorelasi lemah dengan jumlah baki item, iaitu A1 (.151) dan A4 (.216).

Pola ini lazim berlaku apabila satu skala pendek dibina dengan item yang sangat serupa antara satu sama lain: alfa keseluruhan naik kerana bilangan item bertambah, sedangkan setiap subskala empat item terlalu pendek untuk mencapai kebolehpercayaan yang mencukupi. Implikasinya jelas. Min konstruk dalam Jadual 4 tidak boleh digunakan untuk mendakwa bahawa pelajar menilai satu aspek lebih tinggi daripada aspek yang lain, dan kajian lanjutan perlu menyemak semula perkataan item A1 dan A4 sebelum instrumen ini digunakan semula. Taber (2018) mengingatkan bahawa alfa yang tinggi bukan bukti kesahan, dan kes ini merupakan contoh yang jelas.

4.6 Sokongan kepada Penaakulan dan Pembelajaran Kendiri

Catatan pemerhatian yang menunjukkan peserta lebih cenderung merujuk manual servis dan membuat pemeriksaan awal sendiri sebelum meminta pengesahan guru merupakan pola yang paling selari dengan tujuan asal intervensi. Pola ini konsisten dengan dapatan Guan et al. (2025) bahawa chatbot pendidikan berpotensi menyokong proses pembelajaran terarah kendiri, dan dengan dapatan Lo et al. (2024) tentang hubungan antara penggunaan chatbot dengan penglibatan pelajar.

Dari sudut teori, protokol yang digunakan boleh difahami sebagai satu bentuk perancah yang menyediakan sokongan pada saat pelajar tersekat, iaitu dalam zon antara apa yang mampu dilakukannya sendiri dengan apa yang mampu dilakukannya dengan bantuan (Vygotsky, 1978). Langkah pengesahan yang wajib pula menuntut pelajar mengartikulasikan dan menguji penaakulan mereka, iaitu dua unsur yang ditekankan dalam kerangka perantisan kognitif (Collins et al., 1989). Walau bagaimanapun, dapatan Zhou et al. (2025) mengingatkan bahawa bentuk interaksi pelajar dengan chatbot berbeza mengikut tahap pencapaian, iaitu pelajar berprestasi rendah cenderung kekal pada soalan bersifat prosedur. Perbezaan itu berkemungkinan menjelaskan mengapa 11 orang peserta hanya mencapai kategori peningkatan sederhana, tetapi kajian ini tidak merekodkan kandungan interaksi pelajar dengan chatbot, maka andaian tersebut tidak dapat diuji.

4.7 Risiko Pergantungan Berlebihan dan Fungsi Langkah Pengesahan

Langkah pengesahan dalam protokol ini bukan sekadar formaliti prosedur. Walters dan Wilder (2023) menunjukkan kadar rekaan dan ralat yang tinggi dalam keluaran model bahasa besar, manakala Kasneci et al. (2023) menyenaraikan ketidaktepatan maklumat sebagai cabaran utama penggunaannya dalam pendidikan. Dalam bidang automotif, penerimaan diagnosis yang salah boleh membawa kepada kerja pembaikan yang tidak selamat, dan bukan sekadar jawapan yang salah dalam tugasan.

Terdapat juga risiko yang lebih halus. Zhai et al. (2024) melaporkan bahawa pergantungan berlebihan kepada sistem dialog AI dikaitkan dengan pelemahan penaakulan analitik, manakala Fan et al. (2025) mendapati sokongan ChatGPT memperbaik hasil tugasan serta-merta tanpa memperbaik pemindahan pengetahuan, disertai kecenderungan pelajar menyerahkan proses regulasi kendiri kepada alat tersebut. Kajian ini mengukur pencapaian sejurus selepas intervensi tanpa ujian pengekalan, dan ujian pos ditadbir dalam keadaan pelajar masih mempunyai akses kepada chatbot sepanjang tempoh kajian. Oleh itu, persoalan sama ada peserta mampu mendiagnosis secara berdikari tanpa chatbot kekal tidak terjawab, dan itulah persoalan yang paling penting untuk kajian lanjutan.

4.8 Implikasi kepada Amalan Pengajaran TVET

Sumbangan kajian ini bukan penemuan bahawa chatbot berkesan, tetapi pemerihalan satu prosedur operasi yang boleh diguna semula oleh guru lain. Protokol lapan langkah dalam Rajah 2 menetapkan kedudukan chatbot dalam aliran kerja diagnostik, iaitu sebagai penjana kemungkinan pada peringkat awal dan bukan sebagai sumber jawapan pada peringkat akhir. Kedudukan ini selari dengan pendirian Kementerian Pendidikan Tinggi (2024) dan Miao dan Holmes (2023) bahawa kecerdasan buatan generatif merupakan alat sokongan dan bukan pengganti kefahaman pengguna.

Pendekatan ini juga menunjukkan bahawa penyepaduan teknologi digital yang digariskan dalam Dasar Pendidikan Digital (Kementerian Pendidikan Malaysia, 2023) dan Dasar TVET Negara 2030 (Majlis TVET Negara, 2024) boleh dilaksanakan tanpa mengurangkan masa pemeriksaan amali. Dalam protokol ini, chatbot digunakan untuk memendekkan tempoh pelajar tersekat, manakala masa bengkel kekal digunakan untuk pemeriksaan fizikal. Dapatan Ab. Jalil dan Lee (2025) tentang kesediaan institusi TVET terhadap teknologi kecerdasan buatan mencadangkan bahawa pendekatan seumpama ini boleh diperluas, dengan syarat guru dilatih untuk menguruskan langkah pengesahan dan bukan sekadar membenarkan penggunaan chatbot.

Penemuan ini perlu dibaca bersama kedudukan literatur tempatan. Abdul Rahman et al. (2020) menunjukkan penggunaan alat bantu berasaskan teknologi dalam bilik darjah vokasional Malaysia masih terhad, manakala kajian pembangunan alat bantu digital dalam jurnal ARSVOT lazimnya berakhir pada peringkat penilaian penerimaan pengguna (Ismail et al., 2025; Jefferi et al., 2025; Norazizan et al., 2025; Sangin & Azman, 2024). Kajian ini menambah satu dimensi kepada pola tersebut, iaitu pemerihalan prosedur penggunaan, bukan sekadar pembangunan alat.

4.9 Batasan Kajian

Kajian ini mempunyai sepuluh batasan yang perlu dinyatakan secara terbuka. Pertama, reka bentuk satu kumpulan tanpa kumpulan kawalan dan tanpa pengagihan rawak tidak membenarkan sebarang kesimpulan kausal, kerana sejarah, kematangan, kesan ujian dan regresi ke arah min kekal tidak terkawal (Knapp, 2016). Kedua, dan berkait rapat dengan yang pertama, saiz kesan yang direkodkan tidak boleh dibandingkan dengan saiz kesan daripada kajian berkumpulan kawalan atas sebab yang dinyatakan dalam bahagian 4.2.

Ketiga, kajian hanya melibatkan 30 orang pelajar daripada satu kelas, satu program dan satu institusi, maka generalisasi kepada populasi pelajar TVET yang lebih luas tidak wajar dibuat. Keempat, taburan skor ujian pra yang sempit menunjukkan kesan lantai, iaitu ujian pra berkemungkinan terlalu sukar bagi tahap pengetahuan peserta pada ketika itu, dan keadaan tersebut membesarkan ruang peningkatan yang tersedia. Kelima, ujian pra dan ujian pos tidak disemak untuk kesetaraan kandungan dan aras kesukaran secara formal, maka sebahagian daripada perbezaan skor berkemungkinan mencerminkan perbezaan antara kedua-dua kertas ujian.

Keenam, kepekatan respons soal selidik pada skala 4 dan 5 serta ketiadaan sebarang respons di bawah skala 3 menunjukkan risiko jawapan yang diingini secara sosial, dan min persepsi sewajarnya dibaca dengan berhati-hati. Ketujuh, pengkaji merupakan guru kelas, pereka bentuk intervensi, pemerhati dan pentadbir soal selidik yang menilai intervensi tersebut, iaitu keadaan yang menimbulkan risiko kecenderungan pengkaji dan yang berkemungkinan besar menyumbang kepada batasan keenam.

Kelapan, alfa peringkat konstruk berada di bawah .70 bagi empat daripada lima konstruk, maka instrumen hanya boleh dipercayai sebagai skala tunggal dan bukan sebagai lima subskala. Kesembilan, catatan pemerhatian merupakan catatan naratif tanpa rubrik berkekerapan, tanpa pemerhati kedua dan tanpa pengiraan persetujuan antara penilai. Kesepuluh, ujian pos ditadbir sejurus selepas intervensi tanpa ujian pengekalan dan tanpa ujian prestasi tanpa akses chatbot, manakala kandungan interaksi pelajar dengan chatbot tidak direkodkan, sehingga tidak dapat ditentukan sama ada peserta menggunakan chatbot mengikut protokol atau memintas langkah pengesahan apabila tidak diperhatikan.

4.10 Cadangan Kajian Lanjutan

Cadangan berikut disusun untuk menangani batasan yang dinyatakan di atas secara langsung. Bagi menangani batasan pertama dan kedua, kajian lanjutan wajar menggunakan reka bentuk kuasi eksperimen dengan kumpulan kawalan yang mengikuti pengajaran biasa tanpa chatbot, supaya sumbangan intervensi dapat diasingkan daripada pembelajaran kurikulum biasa dan saiz kesan menjadi boleh dibandingkan.

Bagi menangani batasan ketiga, kajian lanjutan boleh melibatkan beberapa kolej vokasional dalam satu zon. Bagi menangani batasan keempat dan kelima, ujian pra dan ujian pos hendaklah dibina sebagai dua bentuk setara yang disemak oleh panel pakar kandungan, dan aras kesukaran ujian pra dilaraskan supaya taburan skor awal tidak berhimpun pada julat yang sempit.

Bagi menangani batasan keenam dan ketujuh, soal selidik hendaklah ditadbir oleh pihak ketiga yang tidak terlibat dalam pengajaran, di luar waktu bengkel, dan sebaiknya disertai beberapa item bertanda songsang bagi mengesan pola jawapan yang tidak dibaca. Bagi menangani batasan kelapan, bilangan item bagi setiap konstruk perlu ditambah dan perkataan item A1 dan A4 disemak semula, diikuti analisis faktor bagi mengesahkan struktur konstruk. Bagi menangani batasan kesembilan, satu rubrik pemerhatian berskala dengan pemerhati kedua hendaklah digunakan, disertai pengiraan persetujuan antara penilai.

Bagi menangani batasan kesepuluh, satu ujian pengekalan hendaklah ditadbir sekurang-kurangnya lapan minggu selepas intervensi, dengan tempoh intervensi yang dipanjangkan supaya kesan kebaharuan berpeluang reda (Rodrigues et al., 2022), dan transkrip interaksi pelajar dengan chatbot dikumpulkan serta dianalisis untuk meneliti jenis soalan yang dikemukakan, sebagaimana yang dilakukan oleh Zhou et al. (2025). Akhir sekali, satu ujian prestasi diagnostik yang ditadbir tanpa akses kepada chatbot akan menjawab persoalan yang paling penting, iaitu sama ada pelajar mampu mendiagnosis secara berdikari selepas intervensi.

5. Kesimpulan

Kajian tindakan ini memerihalkan dan menilai penggunaan AI Chatbot sebagai alat sokongan pembelajaran diagnostik kerosakan enjin dalam kalangan pelajar Teknologi Automotif Semester Satu di sebuah kolej vokasional. Skor min ujian peserta merekodkan peningkatan yang besar dan signifikan selepas intervensi, dan setiap peserta tanpa kecuali merekodkan peningkatan. Peningkatan ternormal pada peringkat kumpulan berada dalam kategori tinggi, walaupun apabila setiap peserta dikategorikan secara berasingan, lebih satu pertiga daripada mereka hanya mencapai kategori sederhana.

Dapatan tersebut perlu ditafsirkan mengikut batasan reka bentuk kajian. Oleh sebab kajian tidak melibatkan kumpulan kawalan, pengagihan rawak mahupun ujian pengekalan, perubahan yang direkodkan tidak boleh dikaitkan secara kausal dengan penggunaan chatbot. Peserta turut mengikuti pengajaran modul yang sama sepanjang tempoh kajian, dan sebahagian daripada perubahan tersebut sewajarnya dikaitkan dengan pembelajaran kurikulum yang berlaku

secara semula jadi. Saiz kesan yang direkodkan pula mencerminkan kehomogenan sampel dan kesempitan taburan skor awal, dan bukan kekuatan intervensi.

Analisis soal selidik menunjukkan penerimaan yang tinggi, tetapi taburan respons yang hampir sepenuhnya berhimpun pada hujung positif skala, tanpa sebarang respons yang menyatakan ketidaksetujuan, menunjukkan risiko jawapan yang diingini secara sosial. Analisis kebolehpercayaan pula menunjukkan instrumen boleh dipercayai sebagai satu skala tunggal tetapi tidak sebagai subskala yang berasingan. Oleh itu, dapatan persepsi ditafsirkan sebagai petunjuk bahawa pendekatan ini tidak ditolak oleh pelajar, dan bukan sebagai ukuran manfaat pembelajaran.

Sumbangan utama kajian ini terletak pada cara chatbot disepadukan, bukan pada penggunaan chatbot itu sendiri. Protokol yang digunakan menetapkan bahawa setiap respons chatbot wajib disahkan melalui manual servis, pemeriksaan fizikal dan bimbingan guru sebelum sebarang kesimpulan diagnostik dibuat. Susunan ini meletakkan chatbot sebagai penjana kemungkinan pada peringkat awal proses penaakulan, manakala pengesahan kekal berasaskan sumber teknikal dan pemeriksaan sebenar.

Bagi amalan pengajaran TVET, kajian ini mencadangkan bahawa penyepaduan kecerdasan buatan generatif dalam pembelajaran diagnostik berpotensi dilaksanakan tanpa mengurangkan masa pemeriksaan amali, dengan syarat langkah pengesahan dikawal oleh guru. Kajian lanjutan yang menggunakan kumpulan kawalan, ujian pra dan ujian pos yang setara, penilaian oleh pihak ketiga, instrumen dengan subskala yang lebih panjang serta ujian pengekalan diperlukan sebelum sebarang dakwaan keberkesanan boleh dibuat.

Penghargaan

Penulis merakamkan penghargaan kepada pihak pentadbiran Kolej Vokasional Sungai Buloh atas kebenaran dan sokongan melaksanakan kajian ini, serta kepada pelajar Program Teknologi Automotif yang terlibat sebagai peserta kajian.

Konflik Kepentingan

Penulis mengisytiharkan bahawa tiada konflik kepentingan berhubung penyelidikan, penulisan dan penerbitan artikel ini. Kajian ini tidak menerima sebarang geran penyelidikan khusus daripada mana-mana agensi pembiayaan.

Rujukan

Ab. Jalil, A. A., & Lee, M. F. (2025). Kesediaan institusi TVET ke arah penggunaan teknologi kecerdasan buatan dalam era digital. *Online Journal for TVET Practitioners, 10*(2), 34-42. https://doi.org/10.30880/ojtp.2025.10.02.003

Abd Samad, N., Tuan Ahmad, T. A., Ismail, A., Amiruddin, M. H., & Mohd Nor, S. N. F. (2017). Kerangka pembelajaran berasaskan proses kerja Kurikulum Standard Kolej Vokasional (KSKV) Diploma Vokasional Malaysia. *Online Journal for TVET Practitioners, 2*(2).

Abdul Rahman, A. B. W., Mohammad Hussain, M. A., & Mohd Zulkifli, R. (2020). Teaching vocational with technology: A study of teaching aids applied in Malaysian vocational classroom. *International Journal of Learning, Teaching and Educational Research, 19*(7), 176-188. https://doi.org/10.26803/ijlter.19.7.10

Abele, S. (2018). Diagnostic problem-solving process in professional contexts: Theory and empirical investigation in the context of car mechatronics using computer-generated log-files. *Vocations and Learning, 11*(1), 133-159. https://doi.org/10.1007/s12186-017-9183-x

Al Hakim, V. G., Paiman, N. A., & Rahman, M. H. S. (2024). Genie-on-demand: A custom AI chatbot for enhancing learning performance, self-efficacy, and technology acceptance in occupational health and safety for engineering education. *Computer Applications in Engineering Education, 32*(6). https://doi.org/10.1002/cae.22800

Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education, 3*, Article 60. https://doi.org/10.1007/s44217-024-00138-2

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018). A review on sample size determination for Cronbach's alpha test: A simple guide for researchers. *The Malaysian Journal of Medical Sciences, 25*(6), 85-99. https://doi.org/10.21315/mjms2018.25.6.9

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research, 90*(4), 499-541. https://doi.org/10.3102/0034654320933544

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Lawrence Erlbaum Associates.

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences, 116*(39), 19251-19257. https://doi.org/10.1073/pnas.1821936116

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489-530. https://doi.org/10.1111/bjet.13544

Guan, R., Raković, M., Chen, G., & Gašević, D. (2025). How educational chatbots support self-regulated learning? A systematic review of the literature. *Education and Information Technologies, 30*(4), 4493-4518. https://doi.org/10.1007/s10639-024-12881-y

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74. https://doi.org/10.1119/1.18809

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education, 153*, Article 103897. https://doi.org/10.1016/j.compedu.2020.103897

Ismail, Z., Joseph, V. R., & Abdul Mutalib, A. (2025). The development and usability of interactive comics based on 5S practices to enhance students' comprehension of safety and cleanliness in workplaces and labs. *Asian Journal of Vocational Education and Humanities, 6*(2), 21-29. https://doi.org/10.53797/ajvah.v6i2.3.2025

Jefferi, A. A., Mohd Hisham, N. A. F., Megat Aruki, N. F. H., Marjuni, A., Samsuddin, S. N. A., & Ismail, M. A. (2025). Development of a Multi-CartClimb for automotive technology students at vocational colleges. *Asian Journal of Vocational Education and Humanities, 6*(2), 48-53. https://doi.org/10.53797/ajvah.v6i2.6.2025

Jin, Y., & Sercu, L. (2025). ChatGPT interventions in higher education: A systematic review of experimental studies. *Journal of Computer Assisted Learning, 41*(4), Article e70072. https://doi.org/10.1111/jcal.70072

Jonassen, D. H., & Hung, W. (2006). Learning to troubleshoot: A new theory-based design architecture. *Educational Psychology Review, 18*(1), 77-114. https://doi.org/10.1007/s10648-006-9001-8

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kementerian Pendidikan Malaysia. (2023). *Dasar pendidikan digital*. Kementerian Pendidikan Malaysia. https://www.moe.gov.my/dasarmenu/dasar-pendidikan-digital

Kementerian Pendidikan Tinggi. (2024). *Garis panduan penggunaan teknologi kecerdasan buatan generatif (KBG) dalam pengajaran dan pembelajaran (PdP) pendidikan tinggi*. Kementerian Pendidikan Tinggi. https://repositori.mohe.gov.my/id/eprint/109

Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research*. Springer. https://doi.org/10.1007/978-981-4560-67-2

Knapp, T. R. (2016). Why is the one-group pretest-posttest design still used? *Clinical Nursing Research, 25*(5), 467-472. https://doi.org/10.1177/1054773816666280

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), Article 410. https://doi.org/10.3390/educsci13040410

Lo, C. K., Hew, K. F., & Jong, M. S.-y. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education, 219*, Article 105100. https://doi.org/10.1016/j.compedu.2024.105100

Majlis TVET Negara. (2024). *Dasar TVET Negara 2030*. Sekretariat Majlis TVET Negara. https://www.tvet.gov.my/manual/MTVET_DASAR_TVET_NEGARA_2030.pdf

McGrath, C., Farazouli, A., & Cerratto-Pargman, T. (2025). Generative AI chatbots in higher education: A review of an emerging research area. *Higher Education, 89*(6), 1533-1549. https://doi.org/10.1007/s10734-024-01288-w

Md Yazid, N. A., & Ali, A. (2025). Pengaruh minat dan sikap terhadap motivasi pelajar di Kolej Vokasional Zon Selatan. *Online Journal for TVET Practitioners, 10*(2), 92-100. https://doi.org/10.30880/ojtp.2025.10.02.009

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Mohebi, L. (2024). Empowering learners with ChatGPT: Insights from a systematic literature exploration. *Discover Education, 3*, Article 36. https://doi.org/10.1007/s44217-024-00120-y

Norazizan, A. H., Hamzah, M. A. I., Azmi, N. E., Nugroho, D. H., & Mohd Zain, M. (2025). Developing the KVSB EasyBus mobile application. *Asian Journal of Vocational Education and Humanities, 6*(1), 34-38. https://doi.org/10.53797/ajvah.v6i1.5.2025

Rodrigues, L., Pereira, F. D., Toda, A. M., Palomino, P. T., Pessoa, M., Carvalho, L. S. G., Fernandes, D., Oliveira, E. H. T., Cristea, A. I., & Isotani, S. (2022). Gamification suffers from the novelty effect but benefits from the familiarization effect: Findings from a longitudinal study. *International Journal of Educational Technology in Higher Education, 19*(1), Article 13. https://doi.org/10.1186/s41239-021-00314-6

Sailer, M., Maier, R., Berger, S., Kastorff, T., & Stegmann, K. (2024). Learning activities in technology-enhanced learning: A systematic review of meta-analyses and second-order meta-analysis in higher education. *Learning and Individual Differences, 112*, Article 102446. https://doi.org/10.1016/j.lindif.2024.102446

Sangin, S. N. P., & Azman, M. N. A. (2024). Pembangunan aplikasi mudah alih bagi topik keselamatan bengkel sebagai nilai tambah bagi kursus VDC 3013 Pengurusan dan Keselamatan Bengkel. *ANP Journal of Social Sciences and Humanities, 5*(2), 55-64. https://doi.org/10.53797/anp.jssh.v5i2.8.2024

Sell, R., Razdan, R., Kase, K., & Rüütmann, T. (2025). The role of AI chatbots in engineering education: Experimental findings and implementation strategies. *International Journal of Engineering Pedagogy, 15*(5), 4-19. https://doi.org/10.3991/ijep.v15i5.56681

Sullivan, G. M., & Artino, A. R., Jr. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education, 5*(4), 541-542. https://doi.org/10.4300/JGME-5-4-18

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports, 13*(1), Article 14045. https://doi.org/10.1038/s41598-023-41032-5

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*(1), Article 28. https://doi.org/10.1186/s40561-024-00316-7

Zhou, R., He, X., Fan, Q., Li, Y., Li, Y., Xiao, X., & Fang, J. (2025). Exploring ChatGPT-facilitated scaffolding in undergraduates' mathematical problem solving. *Journal of Computer Assisted Learning, 41*(4), Article e70077. https://doi.org/10.1111/jcal.70077