Computers and Electronics in Agriculture 242 (2026) 111317

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Human versus machine: Can generative AI anticipate insect biological control outcomes?☆

Kris A.G. Wyckhuys [a,b,c,d,*], Komivi S. Akutse [e,f], Divina M. Amalin [g], Salah-Eddin Araj [h], Marie Joy B. Beltran [i], Ibtissem Ben Fekih [j], Paul-André Calatayud [e,k], Lizette Cicero [l], Marcellin C. Cokola [j], Yelitza C. Colmenarez [m], Kenza Dessauvages [j], Thomas Dubois [e], Léna Durocher-Granger [n], José L. Fernández-Triana [o], Frederic Francis [j], Khalid Haddi [p], Rhett D. Harrison [q], Muhammad Haseeb [r], Natasha S.A. Iwanicki [s], Lara R. Jaber [h], Fathiya M. Khamis [e], Jesusa C. Legaspi [t], Refugio J. Lomeli-Flores [u], Baoqian Lyu [v], James Montoya-Lerma [w], Ihsan Nurkomar [x], James E. O'Hara [o], Jermaine D. Perier [y], Ricardo Ramírez-Romero [z], Francisco J. Sanchez-Garcia [aa], Ann Marie S. Robinson-Baker [r], Luis C.P. Silveira [p], Larisner Simeon [r], Leellen F. Solter [ab], Oscar F. Santos-Amaya [ac], Wagner de Souza Tavares [ad], Rogelio Trabanino [ae], Fernando H. Valicente [af], Carlos Vásquez [ag], Zhenying Wang [b], Lian-Sheng Zang [ah], Wei Zhang [ah], Kennedy J. Zimba [ai], Kongming Wu [b], Yubak D. GC [d]

[a] *Chrysalis Consulting, Danang, Viet Nam*
[b] *Institute for Plant Protection, China Academy of Agricultural Sciences (CAAS), Beijing, China*
[c] *School of the Environment, University of Queensland, Saint Lucia, Australia*
[d] *Food and Agriculture Organization (FAO), Bangkok, Thailand*
[e] *International Centre of Insect Physiology and Ecology (icipe), Nairobi, Kenya*
[f] *Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa*
[g] *Institute of Biological Control, De La Salle University, Taft Avenue, Manila, Philippines*
[h] *School of Agriculture, The University of Jordan, Amman, Jordan*
[i] *National Crop Protection Center, University of the Philippines Los Baños, Laguna, Philippines*
[j] *Functional and Evolutionary Entomology, Gembloux Agro-Bio Tech, University of Liege, Gembloux, Belgium*
[k] *Institut Diversité Ecologie et Evolution du Vivant (IDEEV), Université Paris-Saclay, CNRS, IRD, UMR Evolution, Génomes, Comportement et Ecologie, Gif-sur-Yvette, France*
[l] *Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Yucatán, Mexico*
[m] *CAB International Latin America, FEPAF-UNESP-FCA. Fazenda Exp. Lageado. Botucatu, São Paulo, Brazil*
[n] *CAB International, Kalundu, Lusaka, Zambia*
[o] *Canadian National Collection of Insects, Arachnids and Nematodes, Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada*
[p] *Laboratory of Conservation Biological Control, Department of Entomology, Universidade Federal de Lavras, Brazil*
[q] *CIFOR-ICRAF, 12-14 St Eugene Office Park, Lake Road, Lusaka, Zambia*
[r] *Center for Biological Control, Florida A&M University, Tallahassee, FL, USA*
[s] *Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil*
[t] *United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Center for Medical, Agricultural and Medical Entomology, Tallahassee, FL, USA*
[u] *Posgrado en Fitosanidad, Colegio de Postgraduados, Montecillo, Texcoco, Mexico*
[v] *Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China*
[w] *Department of Biology, Universidad del Valle, Cali, Colombia*
[x] *Universitas Muhammadiyah Yogyakarta, Indonesia*
[y] *University of Georgia, Tifton, GA, USA*
[z] *Biological Control Laboratory (LabCB-AIFEN), University of Guadalajara, Guadalajara, Mexico*
[aa] *Independent Researcher, Murcia, Spain*
[ab] *Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL, USA*
[ac] *Universidad of Pamplona, Pamplona, Colombia*

☆ *© FAO, 2025. Kris Wyckhuys, Yubak Dhoj GC. Food and Agriculture Organization of the United Nations. The views expressed in this publication are those of the author(s) and do not necessarily reflect the views or policies of the Food and Agriculture Organization of the United Nations.

* Corresponding author at: University of Queensland, Australia.
E-mail address: k.wyckhuys@uq.edu.au (K.A.G. Wyckhuys).

https://doi.org/10.1016/j.compag.2025.111317
Received 28 October 2025; Received in revised form 4 December 2025; Accepted 5 December 2025
0168-1699/© 2025 Food and Agriculture Organization of the United Nations. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND IGO license (http://creativecommons.org/licenses/by-nc-nd/3.0/igo/).

[ad] Riau Andalan Pulp and Paper (RAPP), Pangkalan Kerinci, Riau, Sumatra, Indonesia
[ae] Zamorano, Apartado Postal 93, Tegucigalpa, Honduras
[af] EMBRAPA Maize and Sorghum Research Station, Sete Lagoas, Minas Gerais, Brazil
[ag] Facultad de Ciencias Agropecuarias, Universidad Técnica de Ambato, Campus Querochaca, Cevallos, Province of Tungurahua, Ecuador
[ah] State Key Laboratory of Green Pesticides, Guizhou University, Guiyang, China
[ai] School of Agricultural Sciences, University of Zambia, Lusaka, Zambia

ARTICLE INFO

Keywords:
Evidence synthesis
Agroecology
Biological control
Pesticide reduction
Artificial intelligence
Chatbot
Systematic literature review
Large language models

ABSTRACT

Generative artificial intelligence (AI) could transform evidence synthesis and revolutionize the global scientific enterprise, yet its agricultural applications are understudied. Here, we systematically assess the performance of three web-grounded AI engines (ChatGPT, ScholarAI and DeepSeek) in synthesizing the global literature on biological control of the fall armyworm *Spodoptera frugiperda*, and benchmark their outputs against a recent, near-exhaustive human review. Though all engines rapidly screened vast literature corpora, they exhibited shortcomings in factual accuracy, reporting reliability and data consistency. In machine-run syntheses, natural enemy prevalence and performance data often diverged from published records while the level of agreement in enumerating top-performing taxa was evenly low. Meanwhile, internal consistency between laboratory and field-level parasitism data for ScholarAI and DeepSeek was similar to that in human-run reviews. All models tended towards faulty data extrapolation, hallucination and data fabrication, and a sporadic exclusion of key species. While autonomous, machine-only efforts accurately capture coarse-grained patterns in natural enemy identity, abundance, and impacts, they carry limited utility for (living) evidence syntheses or rigorous decision-support. Yet, handled with prudence and due human oversight, machine power might eventually revitalize underfunded disciplines and advance nature-friendly farming.

1. Introduction

Following upon the 2022 launch of ChatGPT, the field of generative artificial intelligence (AI) has rapidly expanded into nearly every domain of human society. Large language models (LLMs), i.e., AI systems that use self-supervised machine learning to autonomously identify, analyze, and interpret vast corpora of digitized text, are reshaping global education, healthcare, logistics and manufacturing (Bommasani et al., 2021; Baidoo-Anu & Ansah, 2023). The unveiling of the open-source, web-grounded model DeepSeek in early 2025 intensified this pace of innovation and spurred competitive improvements in ChatGPT. As early versions of ChatGPT were non-grounded, they solely operated using their initial training databases and were unable to access online information in real time. Further, the structured reasoning and reinforcement learning of DeepSeek make it adept at solving technical problems (Conroy & Mallapaty, 2025; Jin et al., 2025; Normile, 2025). Some capabilities such as deep technical reasoning and scientific proficiency also feature in the web-grounded version of ChatGPT i.e., GPT-5 as launched in early August 2025. The model's performance is further enforced through ScholarAI, a plug-in that connects the subscriber version of ChatGPT to academic databases. As such, the three models hold varying potential to transform global scientific inquiry (Gibney, 2025).

To date, the scientific contribution of LLMs has almost exclusively been explored in human healthcare and medicine. Specifically, both ChatGPT and DeepSeek have proven proficient in tasks such as answering medical licensing exams, diagnosing pathologies, and supporting clinical decision-making (Kung et al., 2023; Van Veen et al., 2024; Kaygisiz & Teke, 2025; Tordjman et al., 2025). In performing these tasks, either model presents distinct strengths and weaknesses. LLMs could also transform scientific literature reviews (SLRs) – which tend to be costly, labor-intensive, and hampered by the exponential growth in scientific output (Bolaños et al., 2024; Scherbakov et al., 2025; Clark et al., 2025). Further, in fast-evolving fields, traditional human-run SLRs tend to be instantaneously outdated (Elliott et al., 2017; Turner et al., 2023). Even though web-grounded, general-purpose LLMs may not truly replicate the SLR methodology or access information that is locked behind paywalls, they may still be useful in democratizing scientific knowledge. They can enable fast data retrieval or information synthesis (Khraisha et al., 2024), fill knowledge gaps in near-real time (Gartlehner et al., 2020) and distil key insights from extensive databases (Joos et al., 2024).

Beyond the medical domain, the comparative performance and usefulness of LLMs in evidence synthesis has only sporadically been assessed. This especially applies to global agriculture and food production. Though the diffusion of (digital) innovations in the agri-food sector is relatively slow compared to other fields, agricultural science, policy, and practice are steadily being infiltrated by generative AI (De Clercq et al., 2024). This includes the advent of LLM-powered chatbots that offer on-demand expertise and decision-support to farmers or pest management professionals (Silva et al., 2023; Tzachor et al., 2023; Yang et al., 2024; Shepherd et al., 2025) as well as disease forecasting systems (Calone et al., 2025). This exemplifies how generative AI is poised to transform agricultural pest management and can help to sustainably mitigate the food losses due to biotic stressors (Oerke, 2006). LLMs could help promote best practice (Wyckhuys et al., 2021), prioritize agroecological and biodiversity-based solutions (Deguine et al., 2023) and ultimately alleviate farmers' overreliance upon synthetic pesticides (Shattuck et al., 2023). This, in turn, can slow or reverse pesticide-induced biodiversity loss, environmental pollution and human health hazards at local, regional, and global scales (Schaffner et al., 2024; Tang et al., 2025). Yet, regardless of the societal benefits of an AI-enabled progress in agri-food science and practice, the utility of LLMs in pest or pest management knowledge synthesis remains unexamined. Specifically, it is unclear to what extent this knowledge can be accessed by web-grounded, general-purpose bots.

The fall armyworm, *Spodoptera frugiperda* J.E. Smith (FAW; Lepidoptera: Noctuidae) is a cosmopolitan crop pest that has exerted significant impacts on global agri-food production (Kenis et al., 2023). Native to the Americas, FAW invaded western Africa in 2016 and has since spread across the world's tropics and subtropics, triggering indiscriminate use of pesticides (Tambo et al., 2020; Yang et al., 2021). Yet, across its native and invasive range, a diverse set of vertebrate, invertebrate and microbial organisms naturally suppress FAW populations (Abbas et al., 2022; Kenis et al., 2023). In a conventional (i.e., human-run) synthesis of the global FAW literature up to mid-2023, Wyckhuys et al. (2024) reported 46, 304 and 215 entomopathogen, parasitoid and predator taxa that are affecting FAW, respectively. Those biological control agents sporadically inflict mortality levels up to 80–90 %, especially in diversified settings in the FAW native range (e.g., Meagher et al., 2016) and thereby lessen the need for chemical interventions (Janssen & van Rijn, 2021). Conversely, in the FAW invasive

range, biological control agents are continually being discovered, described, and evaluated, with field-level evidence on their contributions to FAW control accruing on a daily basis (Kenis et al., 2023). Also, as local natural enemies adapt to a new invader and form 'new associations' (Carlsson et al., 2009; Bradicich et al., 2024), their impacts intensify over time. Yet, given that human-run reviews on FAW and other alien species are only periodically updated, evidence on the true merit of biological control becomes uncertain. This slows the advancement of science, policy, and practice (Turner et al., 2023); a shortcoming that can possibly be remediated by LLMs (Gurr et al., 2024). However, whether human-run syntheses of nature-friendly crop protection can be supplanted or enriched by machine-run ones requires scrutiny.

In this study, we methodically assess the relative performance of three popular web-grounded LLMs in synthesizing the FAW biological control literature. First, our work uses either AI tool to consolidate and interpret data on the laboratory- and field-level performance of FAW biological control agents across its native and invasive range. Next, we systematically assess the factual accuracy, reporting reliability, breadth of knowledge and data consistency for each of the three LLMs and compare this to data published by Wyckhuys et al. (2024) in a similar way as done by Khraisha et al. (2024). By carefully scrutinizing AI outputs, this study lays the groundwork for a reliable and ultimately scalable use of AI-powered research assistants or decision-support tools in sustainable crop protection.

2. Materials & methods

This study compared reporting data accuracy, reliability, consistency and review coverage or completeness between a recent human-run literature synthesis and outputs of three LLMs or AI search engines or chatbots, i.e., the subscription version of ChatGPT-5 (OpenAI, San Francisco, USA) with an assumed cut-off date for its initial training data of October 1, 2024, its ScholarAI plug-which provides real-time queries, and DeepSeek-R1 (DeepSeek AI, Hangzhou, China). Though all three bots can be classified as web-grounded models, they possibly offer varying degrees of real-time access to only information in less popular domains such as crop protection. Specifically, we compared AI-generated data on FAW natural enemy abundance or performance with those published by Wyckhuys et al. (2024). This was achieved by submitting standardized search prompts to each of the three engines between August 15, 2025, and September 5, 2025.

2.1. Human-run review

The initial literature corpus of Wyckhuys et al. (2024) was built in March 2023 by querying the Web of Science Core Collection database (1900–2022) through a University of Queensland staff subscription. The corpus comprised 710, 215 and 320 initial publications on entomopathogens, predators and parasitoids, respectively, from across the global distributional range of FAW. This database was further expanded until October 2023 by adding in studies that were cited in the original publications. From this database, we then extracted data on FAW entomopathogens, parasitoids or invertebrate predators from 127, 86 or 64 laboratory studies and 35, 102 and 26 field survey reports, respectively. As a result, data on 34, 129 and 82 taxa (i.e., species or genera) of pathogens, parasitoids and canopy-dwelling predators were generated. Ground-dwelling predators were not taken into consideration.

2.2. Prompt development

Next, field- and laboratory-level abundance, prevalence, or performance data of the prevailing FAW entomopathogens, parasitoids and invertebrate predators were used as point of departure for machine-run syntheses. Standardized prompts were developed that were aimed at reproducing the data for either natural enemy guild and observational context i.e., laboratory or field reported in Wyckhuys et al. (2024) (Supplementary Table 1). Specifically, five sets of prompts were developed: 1) prevalence and induced mortality (average, maximum) for 10 FAW entomopathogens under field conditions in the FAW native or invasive range; 2) average peak parasitism rates for a subset of egg or larval parasitoids under laboratory conditions; 3) average and maximum field parasitism rate of the 25 best-performing parasitoids in the FAW native or invasive range; 4) average peak lifetime predation for 10 and 39 egg or larval predators under laboratory conditions; and 5) average and maximum field-level densities of the 25 most abundant invertebrate predators in the FAW native or invasive range. In the latter prompt, predator density was expressed as the number of individuals per plant – assuming optimum planting densities of seven maize plants per square meter. Also, in prompts relating to field-level (invertebrate) natural enemy abundance or performance, instructions were given to solely consider natural densities and to omit any records from experimental fields where biological control releases had been performed. This restriction however was relaxed in the human-run review by Wyckhuys et al. (2024), e.g., for *Telenomus* or *Trichogramma* spp. Further, zero values were not taken into consideration when computing the average field abundance or parasitism rate of specific natural enemies. No additional instructions were provided regarding the language, type, or source of the scientific publication, which allowed for a comprehensive screening of the all-time, published literature in the delineated geography i.e., field studies in the FAW native or invasive range. In the meantime, such relatively concise, simple, and unstructured prompts may lead to less reliable, generic, and possibly biased outcomes (Jin et al., 2025).

2.3. Data extraction and benchmarking

As a next step, laboratory and field abundance, prevalence or performance data were tabulated per natural enemy species, guild, and target geography. For each AI engine, data accuracy was assessed by systematically comparing the species-specific performance or abundance data with those reported by Wyckhuys et al. (2024). This was done by computing data disparity i.e., the degree (%) to which AI-reported results deviated from the above published data. Reliability in reporting was assessed by comparing the identities of top-ranked predators or parasitoids for each AI engine with those listed by Wyckhuys et al. (2024), computing the share of species matches or mismatches across engines. For each of these two natural enemy guilds, data reliability was expressed as the overall level of agreement (0–4) in reporting across the four literature syntheses (i.e., three machine-run efforts and the published human review) and plotted in a histogram. As such, a taxon that was reported by all three Ai engines and the human review received a value of 4. Lastly, we assessed data consistency both between and within human- and machine-run syntheses i.e., external, and internal consistency. External consistency was determined by regressing species-specific (field or laboratory-level) performance or abundance data generated by either AI engine or reported in the human-run synthesis. Meanwhile, internal consistency for both human and machine-run syntheses was assessed by regressing field-level performance data against those reported at the laboratory-level for different parasitoid species.

2.4. Literature screening completeness

Lastly, we determined the literature review coverage or completeness for the three AI search engines. For each set of ten queries covering different FAW natural enemy taxa, performance parameters, or geographies, we instructed each AI engine to disclose its underlying literature review process using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Based upon the (self-reported) list of literature resources, we then drew up a PRISMA flow diagram for each AI engine. Next, at three specific steps i.e., initial records identified (step #1), full texts assessed (#2), and number of papers included in the final

analysis (#3), we compared the extent of literature review coverage or completeness between each engine.

2.5. Data analysis

All data were checked for normality and homoscedasticity prior to analysis. Data that did not meet these assumptions were either transformed or analyzed using non-parametric tests. As such, data disparity and the level of agreement in natural enemy listing were compared between AI engines using a Kruskal-Wallis non-parametric test. Spearman rank correlation analysis was used to relate pathogen prevalence or mortality and laboratory-level invertebrate data between the four different literature syntheses, whereas linear regression analysis was employed to assess data consistency for field-level performance data. Species-level parasitism rates under field and laboratory conditions were also related using linear regression analysis. Lastly, a Repeated Measures ANOVA was used to compare literature coverage at three consecutive steps of the review process between individual AI engines and queries. GraphPad Prism version 10.6.1 (Boston, Massachusetts, USA) was used for both data visualization and statistical analysis.

3. Results

3.1. Data accuracy

AI-generated natural enemy abundance or performance data diverged markedly from those published by Wyckhuys et al. (2024). Across all three AI engines, disparity in entomopathogen prevalence and mortality data ranged from −84.5 % to 2,400.0 % and −99.2 % to 51.0 %, respectively (Fig. 1A). Data disparity did not differ between the three AI engines for neither pathogen prevalence (Kruskal-Wallis; $\chi^2 = 2.91$, p = 0.245) nor mortality ($\chi^2 = 0.09$, p = 0.961). Laboratory-level parasitism rates, as reported by AI engines, also differed from published data i.e., Wyckhuys et al. (2024) – exhibiting disparity levels for egg and larval parasitoids that ranged from −50.0 % to 1,200.0 % and −42.7 % to 650.0 %, respectively (Fig. 1B). Data disparity did not differ between AI engines for neither egg nor larval parasitoids ($\chi^2 = 2.41$, p = 0.300; $\chi^2 = 0.45$, p = 0.799, respectively). Lastly, AI-reported laboratory-level performance data for egg and larval predators also exhibited disparities ranging from −19.0 % to 8,439.0 % and −65.7 % to 5,802.0 %, respectively (Fig. 1C). In contrast to the above natural enemy guilds, data disparity for both predator groups differed between the three AI engines ($\chi^2 = 8.84$, p = 0.012; $\chi^2 = 7.93$, p = 0.019, respectively). Specifically, for FAW egg and larval predators alike, ChatGPT reported data that diverted least from Wyckhuys et al. (2024). Overall, when assessing patterns across natural enemy taxa and performance metrics,

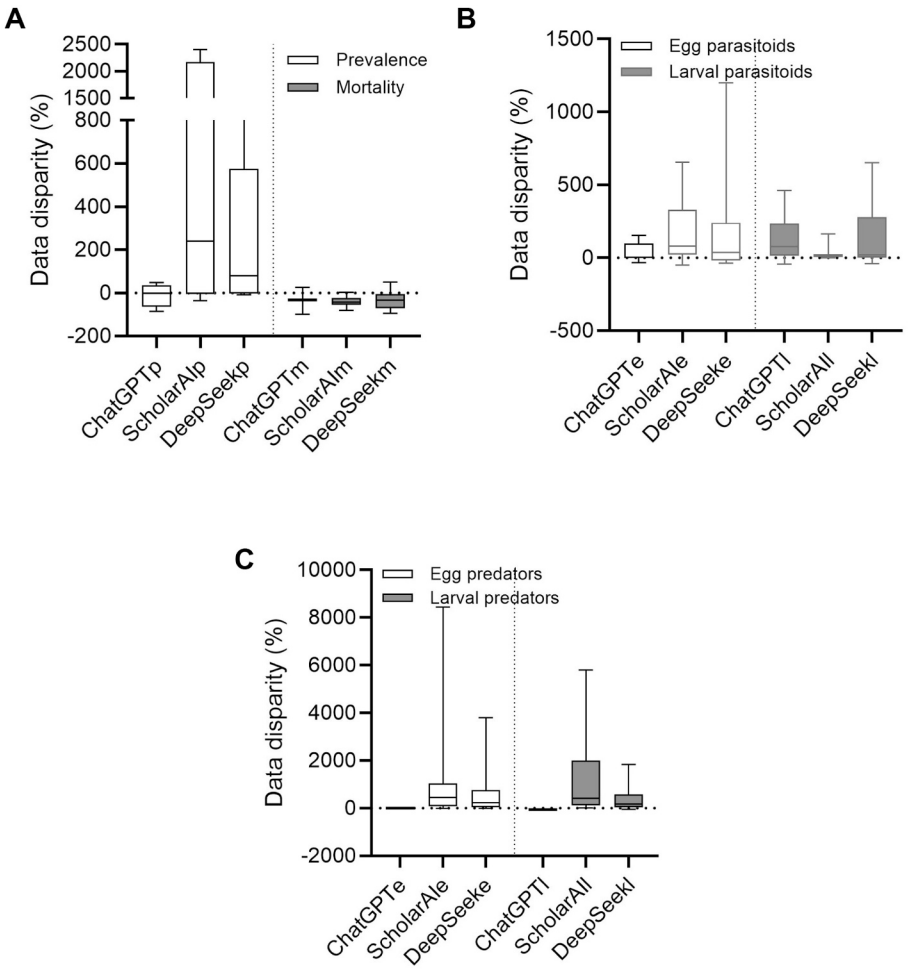

Fig. 1. Disparity in natural enemy prevalence or performance data between human- and machine-run literature syntheses. Patterns are plotted for the maximum field-level prevalence or mortality of FAW entomopathogen (A) and laboratory-level peak parasitism (B) or predation rates (C). Data disparity is expressed as % deviation of AI-reported data from those in Wyckhuys et al. (2024). Inter-engine differences are only statistically significant in panel C, with the underlying results reported in the main text.

ChatGPT yielded data that were most in line with those of the human-run review (ANOVA; $F_{2,174} = 7.06$, $p = 0.001$). If the patterns reported by Wyckhuys et al. (2024) indeed reflected reality, then ChatGPT exhibited the highest level of accuracy.

3.2. Data reliability

Overall reliability in the reporting of top-performing FAW predators or parasitoids was generally low. For average data, the level of agreement in natural enemy listings between the four syntheses equaled 1.7 ± 1.1 (average ± SD; parasitoids in FAW native range), 1.6 ± 1.0 (parasitoids, invasive range), 1.6 ± 0.9 (predators, native range) and 1.1 ± 0.5 (predators, invasive range). For maximum values of parasitism rate or predator abundance, the level of agreement attained 1.9 ± 1.0, 1.7 ± 1.0, 1.5 ± 0.8 and 1.1 ± 0.5, respectively (Fig. 2). Data reliability differed between natural enemy taxa and geographies for both performance averages ($\chi^2 = 17.21$, $p < 0.001$) and maxima ($\chi^2 = 25.81$, $p < 0.001$). Overall, the extent of species mismatches among the four syntheses was highest for invertebrate predators in the FAW invasive range.

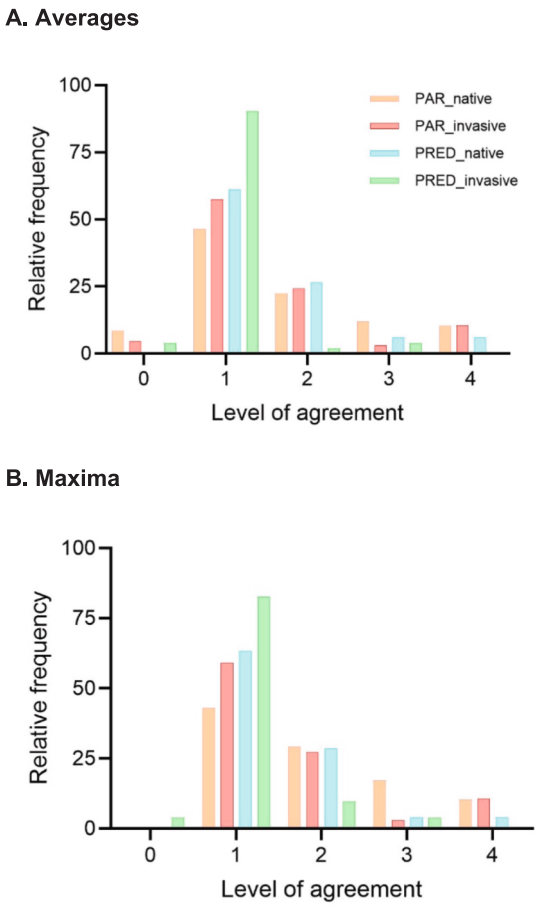

Fig. 2. Histograms showing the level of inter-engine agreement in the listing of top FAW predator and parasitoid taxa. Patterns are plotted for top-ranked invertebrate predators (PRED) and parasitoids (PAR) either in the FAW native or invasive range. In the Y axis, we report the frequency of taxa for which average or maximum field performance is reported by varying numbers of human- or machine-run syntheses (out of 4). A value of 0 indicates instances in which average performance is reported in one synthesis, but maximum values are absent across all four syntheses or vice versa. Taxa that are solely reported by Wyckhuys et al. (2024) were not included.

3.3. Data consistency

Field- and laboratory-level abundance or performance data of specific natural enemies varied markedly between and among human or machine-run syntheses (Figs. 3, 4). External consistency exhibited variability between natural enemy taxa. Among FAW pathogens, only the prevalence data reported by ScholarAI correlated with those of human-run syntheses (Spearman Rank; $r = 0.85$, $p = 0.025$) whereas prevalence and mortality data were consistent for ScholarAI and DeepSeek ($r = 0.91$, $p < 0.001$; $r = 0.95$, $p < 0.001$; Fig. 3). For laboratory-derived data on FAW parasitoids, none of the average performance data reported by the three AI engines correlated with those of Wyckhuys et al. (2024) even though ScholarAI-reported data were consistent with those of DeepSeek ($r = 0.66$, $p < 0.001$). Lastly, for laboratory-derived data on FAW predators, average egg or larval predation data reported by ChatGPT and DeepSeek were consistent with those of Wyckhuys et al. (2024) ($r = 0.90$, $p = 0.002$; $r = 0.40$, $p = 0.018$, respectively). As above, reporting consistency was also high between ScholarAI and DeepSeek ($r = 0.55$, $p < 0.001$). Similar patterns were detected for field data of invertebrate natural enemies. Published average parasitoid performance data were consistent with those reported by ChatGPT (ANOVA, $F_{1,46} = 9.57$, $p = 0.003$, $R^2 = 0.172$) and DeepSeek ($F_{1,35} = 15.97$, $p < 0.001$, $R^2 = 0.313$), and this also held for maximum values ($F_{1,41} = 20.63$, $p < 0.001$, $R^2 = 0.335$; $F_{1,24} = 12.14$, $p = 0.002$, $R^2 = 0.336$, respectively; Fig. 4). Notably, AI-generated parasitism averages aligned better with those in the human review than the maximum values – which may hint at eventual inaccuracies in Wyckhuys et al. (2024). Further, published predator abundance data were in line with those reported by ScholarAI ($F_{1,12} = 6.72$, $p = 0.024$, $R^2 = 0.359$) but this only held for average values and not for maximum values (Fig. 4).

In contrast, internal consistency tended to be high for Wyckhuys et al. (2024) and for two of the three machine-run syntheses. Specifically, maximum field-level parasitism related to the average peak parasitism of FAW parasitoids in the laboratory for human-run syntheses ($F_{1,15} = 8.94$, $p = 0.009$, $R^2 = 0.373$), ScholarAI ($F_{1,26} = 13.12$, $p = 0.001$, $R^2 = 0.335$) and DeepSeek ($F_{1,18} = 6.34$, $p = 0.022$, $R^2 = 0.261$). Internal data consistency was highest for the human-run literature synthesis (Fig. 5).

3.4. Literature screening completeness

Literature review coverage or completeness differed notably between the three AI engines at three consecutive PRISMA steps, with DeepSeek consistently covering larger literature corpora (Fig. 6). This possibly can be due to the inclusion of duplicates, which have been disclosed to varying extent by either of AI engine. Across the ten different queries, literature coverage differed between AI search engine at step #1 (Repeated Measures ANOVA, $F_{2,18} = 62.25$, $p < 0.001$), step #2 ($F_{2,18} = 42.77$, $p < 0.001$) and step #3 ($F_{2,18} = 41.17$, $p < 0.001$). At all steps, the type of query e.g., pathogen prevalence vs. predator abundance did not exert a significant effect. Surprisingly, screening completeness differed from the self-reported coverage of literature sources by either search engine, with ChatGPT routinely covering more than twice as many databases than DeepSeek (Supplementary Fig. 1).

4. Discussion

Generative AI could propel agri-food science by offloading mundane, repetitive, and resource-intensive tasks and by freeing researchers' time for more strategic work (Van Veen et al., 2024). If LLMs can indeed reliably assist in real-time evidence synthesis (Khraisha et al., 2024; Bolaños et al., 2024; Scherbakov et al., 2025), their impact on the scientific enterprise could be transformative. This especially applies to chronically underfunded and understaffed disciplines such as agroecology or biological control (Van Lenteren, 2012; Pimbert & Moeller, 2018; Pavageau et al., 2020). Here, we show how AI-run literature

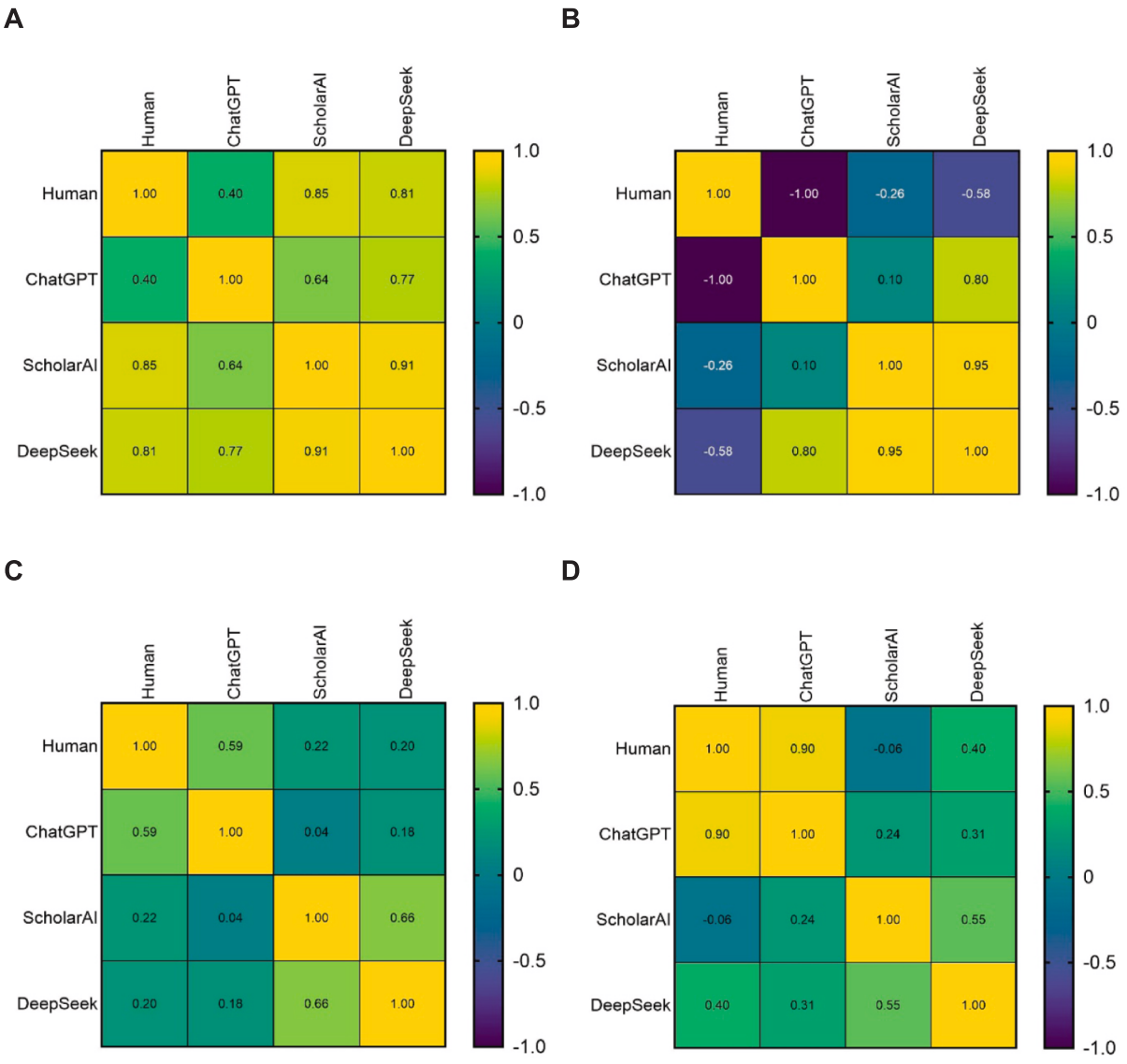

Fig. 3. Correlograms showing the degree of correlation between natural enemy performance data as reported through human- or machine-run literature syntheses. Patterns are plotted for field-level pathogen prevalence (A) or mortality (B) and laboratory-level peak parasitism (C) or peak predation (D). Each panel depicts the results of correlation analyses between the respective average values as reported by Wyckhuys et al. (2024) and those generated by either ChatGPT, ScholarAI or DeepSeek.

syntheses currently exhibit important shortcomings in data accuracy, reporting reliability and data consistency. Across natural enemy taxa and performance metrics, ChatGPT reported data that diverted least from those in human-run reviews with average disparities ranging from −65.7 % to 126.4 % (versus −39.8 % to 1,435.0 % for ScholarAI and –33.7 % to 734.5 % for DeepSeek). The latter two LLMs consistently overestimated natural enemy prevalence and performance or, alternatively, mixed up taxa from the FAW native and invasive range. On the other hand, AI-generated syntheses could have picked up inaccuracies in the human-run review e.g., for maximum field parasitism where Wyckhuys et al. (2024) may not have filtered out all data from experimental releases. Human-machine agreement in enumerating top-ranked FAW predators or parasitoids was invariably low, averaging between 1.1 and 1.9 on a 0–4 scale. While data consistency between human and machine-reviews varied between natural enemy taxa and performance metrics, outputs from ScholarAI and DeepSeek were generally well-aligned. Lastly, internal consistency between field- and laboratory-level parasitism data – as determined using linear regression analysis – for the two engines above was on par with that of the published data. Nevertheless, all three LLMs exhibited important shortcomings that constrain their utility for agri-food science, policy, and practice.

First and foremost, all AI engines comprehensively screened the global literature, extracting, processing and tabulating data for tens of natural enemy taxa in a matter of minutes. For instance, in a query on parasitoids in the FAW native range, GPT-5, ScholarAI and DeepSeek rapidly screened a (self-reported) corpus of 260, 132 and 387 papers, respectively. This was accomplished by accessing 18 different literature databases or repositories. This rapid, extensive screening has immediate implications for scientists' work. Such achievement alone could cut their workload by 88–98 % (Scherbakov et al., 2025) and may also help pick up the large share of papers that are overlooked when using manual filtering (Gartlehner et al., 2020). Considering how formal systematic reviews account for a quarter of the world's scientific output (Hanson et al., 2024), prove costly and are labor-intensive, generative AI *in principle* can accelerate the process of scientific inquiry. In doing so, it can free up vital resources for more strategic activities that directly contribute to sustainable development research (Candel, 2022; Schneider et al., 2023; Gohr et al., 2025). In addition, AI tools can also

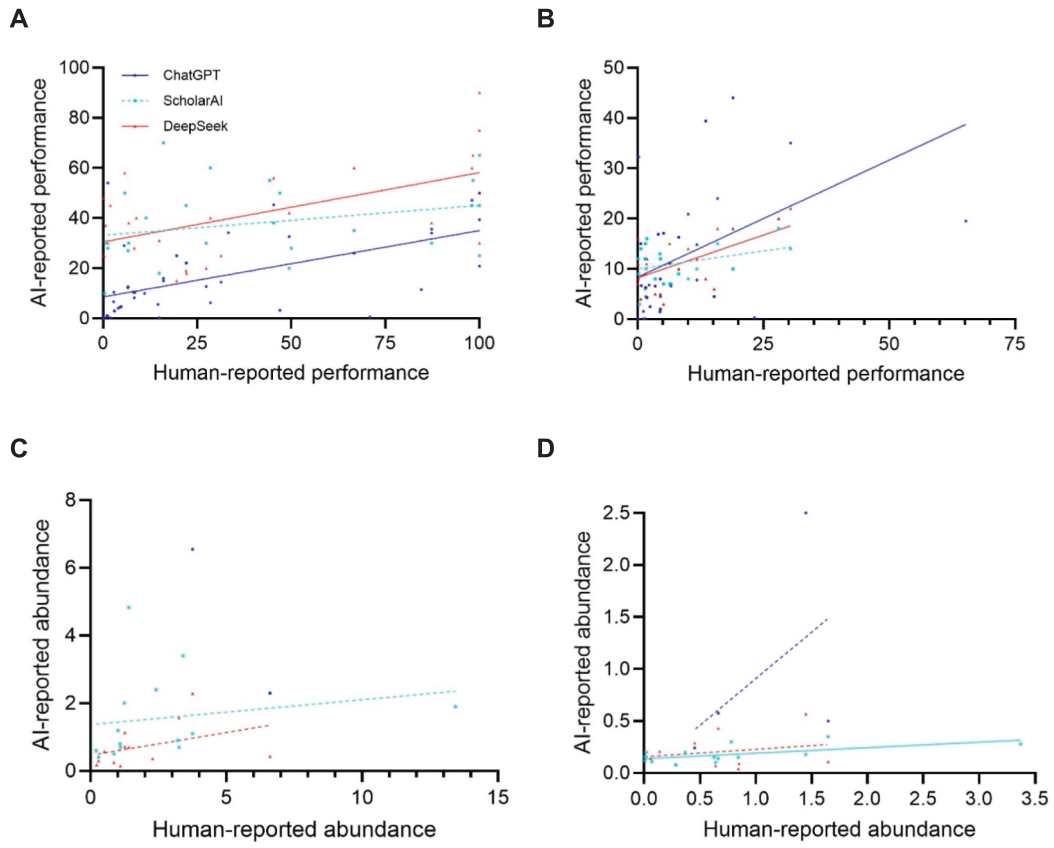

Fig. 4. Consistency between human- and AI-reported performance or abundance data for FAW parasitoids (A, B) and predators (C, D). In-field performance of FAW parasitoids is expressed as % parasitism. Regression patterns reflect the extent to which AI-generated maximum (A,C) or average (B,D) values correspond with those reported in Wyckhuys et al. (2024). Linear regression lines are drawn for each of three AI chatbots (ANOVA, $p < 0.05$), whereas non-significant patterns are indicated with a dotted line. In panel C, ChatGPT trends are not plotted due to lack of sufficient data. Statistical details are presented in the main text.

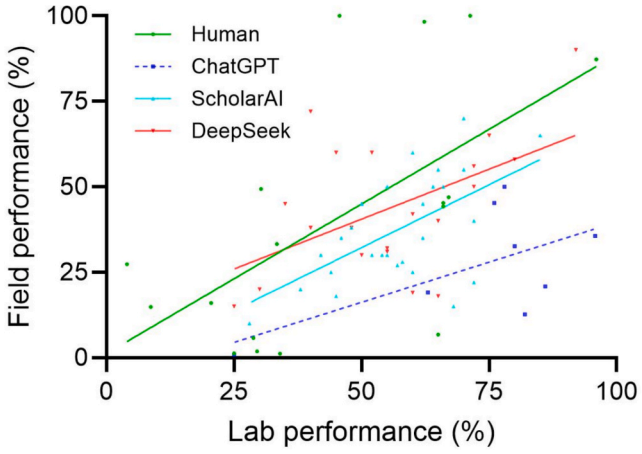

Fig. 5. Correspondence between field- and laboratory-level parasitoid performance, as reported through human- and machine-run syntheses. Parasitoid performance is expressed as % parasitism. For each type of literature synthesis or AI engine, average peak parasitism levels (%) under laboratory conditions are plotted against field-level maxima. Linear regression lines are drawn using data from Wyckhuys et al. (2024) or either of three AI chatbots (ANOVA, $p < 0.05$). Non-significant patterns are indicated with a dotted line. Statistical details are reported in the main text.

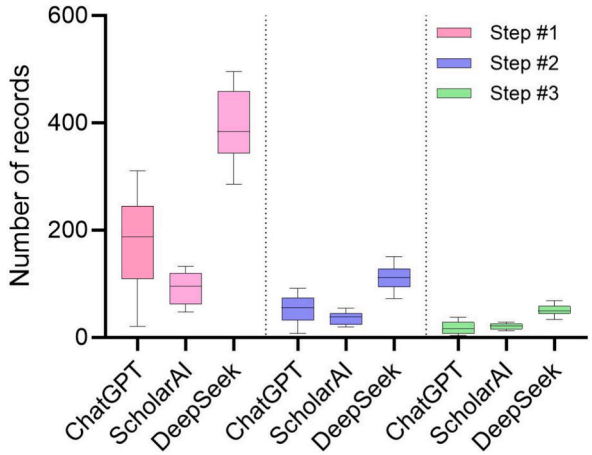

Fig. 6. Literature coverage at three consecutive steps of the screening process conducted by either of three AI chatbots. Patterns are depicted for the following PRISMA steps, self-reported by either chatbot: initial records identified (1), full texts assessed (2), and number of papers included in the analysis (3). Statistical details are presented in the main text. Patterns for the human-run review are not plotted, because literature coverage in the three successive PRISMA steps was not reported by Wyckhuys et al. (2024).

provide decision-support to non-scientific actors. For fast-moving fields of inquiry, such as biological control research of newly invasive pests, LLMs can distill new evidence as it is generated (Elliott et al., 2017). For instance, by anticipating the strength and stability of biological control during times of precipitous biodiversity decline (Wyckhuys et al., 2023), LLMs potentially can help to move ecosystem services into adaptive decision-making (Daily et al., 2009; Gonzalez-Chang et al., 2020; EC,

2021). Regardless of the above, present AI tools are unlikely to fully replace human-run reviews due to data inaccuracies and other shortcomings (Mostafapour et al., 2024).

In an earlier comparative assessment of web-grounded versus non-grounded models (Wyckhuys et al., in press), the former ones exhibited higher data consistency, breadth of knowledge and factual accuracy. Such inter-model differences were not evident in this study; all three web-grounded models carried merit. Overall, as models such as ChatGPT are continually updated, any cross-model comparisons and perceived advantages of one LLM over another prove temporally instable. While GPT-generated data aligned best with those in human-run reviews in our assessment, the remaining two LLMs performed well in terms of internal and external data consistency. Hence, outputs of all three AI engines carried equivalent value. This is regardless of DeepSeek reportedly screening 128–328 % more initial records and extracting data from 144-183 % more final publications or of ChatGPT consulting twice as many databases as the other two models. Some deficiencies reported in Wyckhuys et al. (in press) could be ascribed to the free-tier version of GPT-4o, which may shift down to older models when subject to extensive search tasks or complex queries. The more extensive literature corpus accessed by DeepSeek can possibly be attributed to its superior proficiency in text analysis and technical reasoning in Chinese language (Jin et al., 2025; Luo et al., 2025; Yuan et al., 2025), especially as 60–65 % of the world's scientific outputs in agricultural or ecological disciplines are currently generated by Chinese researchers (Anonymous, 2024). For all three engines, some outputs were seemingly generated by chance agreement or proved entirely hypothetical – as also observed by Khraisha et al. (2024) and Wyckhuys et al. (in press). The phenomenon of hallucination, in which fictitious content is generated, and false references are fabricated, is well documented (Emsley, 2023; Kacena et al., 2024; Susjnak et al., 2025; Bolaños et al., 2025). The exact underlying determinants of data fabrication and reference hallucination are unclear but potentially relate to retrieval failures, synthesis errors, or eventual flaws in the underlying training data for entomological knowledge. Besides affecting the veracity and credibility of AI-generated outputs in our study, it limits the usability of LLMs for knowledge synthesis in and beyond the crop protection domain.

LLM-generated content regularly diverged from reality and differed substantially from Wyckhuys et al. (2024). Four major issues could be distinguished. First, AI engines tended to generate data that were overly optimistic (i.e., as per the highly skewed data disparities) or even implausible. For instance, both ScholarAI and DeepSeek reported laboratory-level parasitism data for *Meteorus laphygmae* Viereck and *Meteorus pulchricornis Wesmael* (Hymenoptera: Braconidae). Even though the population genetics of *M. laphygmae* on FAW are well-described (Gonzalez-Maldonado et al., 2019) and *M. pulchricornis* parasitism has been observed on other *Spodoptera* spp. (Nguyen et al., 2005; Guo et al., 2013), the actual laboratory-level performance of either species on FAW remains undocumented. Similarly, two LLMs reported 42–55 % peak pupal parasitism by *Brachymeria ovata* (Say) (Hymenoptera: Chalcididae) and *Diapetimorpha introita* (Cresson) (Hymenoptera: Ichneumonidae) in the laboratory i.e., rates that may have been extrapolated from field observations for *D. introita* (Pair & Gross, 1989). Often, AI claims of high abundance or parasitism of particular taxa could not be corroborated – especially for those with low or modest performance (Fig. 4). Second, often, LLMs poorly distinguished between records from varying geographies. For instance, ScholarAI mistakenly listed *Microplitis manilae* Ashmead (Hymenoptera: Braconidae) and *Arma chinensis* (Fallou) (Hemiptera: Pentatomidae) or *Archytas marmoratus* (Townsend) (Diptera: Tachinidae) as top-ranked natural enemies in the FAW native or invasive range, respectively. Third, all three LLMs routinely pointed towards false literature references. For instance, when prompted to provide details on the performance of FAW predators such as *Euborellia annulipes* (Lucas) (Dermaptera: Anisolabididae), *Mallada basalis* Walker (Neuroptera: Chrysopidae) or *Eriopis connexa* (Coleoptera: Coccinellidae), AI engines pointed towards Molina-Ochoa et al. (2003) or Kenis (2023) i.e., papers that exclusively cover parasitoids. This inability of AI tools to identify the correct source literature may hint at 'hidden' accuracies in data compilation and analysis. Fourth, when queried about predator performance or field-level abundance, ChatGPT expressed an inability to access sufficient, relevant information or only reported data for a handful of taxa. Such intermittent failures are not restricted to the agricultural or entomological domain (Chen et al., 2025; Lamanna et al., 2025). However, in contrast with Wyckhuys et al. (in press), our current exercise detected fewer issues with AI's ability to pick up nomenclatural changes or its hallucination of non-existent taxon names. Only in a few instances were LLMs unable to pick up outdated names and distinguish between synonymized species e.g., listing *Chelonus texanus* and *C. insularis* or *Euplectrus plathyphenae* and *E. platyhyphenae* as distinct species. Some of these issues most certainly will be resolved as AI tools further mature (e.g., Kargupta et al., 2025).

Despite their shortfalls, AI-generated syntheses offer certain advantages. First, AI outputs were variably affected by dataset imbalance and not necessarily skewed by the majority class (Khraisha et al., 2024). All three engines picked up on rare or geographically restricted taxa: for instance, when prompted to report the 25 best-performing natural enemies, all models listed taxa with average peak parasitism rates as low as 1.1–5.0 % and peak field-level abundance of 0.1–0.2 predator individuals per plant. However, the actual listing of rare taxa (i.e., high breadth of knowledge) led to a frequent omission of common ones. In their listing of the 25 top-performing parasitoids in the FAW native range, ChatGPT, ScholarAI and DeepSeek covered ten, eight and seven out of the 12 taxa listed by Kenis (2023), respectively. Surprisingly, the tachinid fly *Lespesia archippivora* (Rley) (Diptera: Tachinidae) – which attains parasitism levels up to 55 % in Central America (Gladstone, 1991) – was overlooked by all three LLMs. This can be partly ascribed to inter-model variability in taxonomic resolution. Yet, overall, AI outputs compared favorably with human-run reviews, in which nine of the above 12 taxa feature in the top 20 list of Wyckhuys et al. (2024). Second, internal consistency between laboratory- and field-level parasitism rates for ScholarAI and DeepSeek – but not for ChatGPT – was nearly on par with that of Wyckhuys et al. (2024). Our work thus confirmed that, even though field performance is impacted by environmental and behavioral parameters, it tends to correspond to laboratory-level observations (Casas et al., 2004; Hoelmer & Kirk, 2005). Third, ScholarAI and DeepSeek listed a more extensive set of predators in the FAW invasive range i.e., where new natural enemies are actively being discovered and described. These records may have been extracted from FAW papers published post-2023 (i.e., as not picked up by Wyckhuys et al., 2024) or extrapolated from insect surveys in maize fields prior to the FAW invasion. To some extent, AI-run analytics could thus foresee future biological control outcomes. This however may only be partially valid, as resident natural enemy populations often shift dramatically following the arrival of new biota (Bradicich et al., 2024; Jara-Chiquito et al., 2024). On the other hand, the ability of AI models to screen the literature in real-time may allow for 'living' evidence syntheses (Elliott et al., 2017). This alone could have transformative impacts on science, policy, and practice in domains such as biological control, invasion biology, or biosecurity.

Overall, we detected high levels of human–machine agreement in gauging the strength of FAW biological control across natural enemy guilds, taxa, and geographies. Though the three dominant AI models exhibit impressive capacity for knowledge synthesis and correctly capture coarse-grained patterns, they suffer drawbacks. Specifically, machine-run syntheses are hampered by factual inaccuracies and data fabrication or omission. Meanwhile, AI-generated species lists, or performance data diverged to varying extent from human reviews. These issues constrain the utility of the current generation of LLMs but possibly will be resolved with the advent of AI-powered research assistants or living evidence syntheses (Bolaños et al., 2024; Susnjak et al., 2025; Scherbakov et al., 2025). Future research could focus on developing domain-specific models that are trained on curated agricultural and

ecological corpora (e.g., Tzachor et al., 2023), rather than relying on general-purpose models. Combining such models with Retrieval-Augmented Generation (RAG) systems (Gupta et al., 2024), which ground the AI's responses in a verified database of scientific literature, could improve factual accuracy.

The present use of generative AI is best restricted to small, focused tasks such as data summarization (Van Veen et al., 2024). For now, vigilance is warranted and human verification remains essential (Van Dis et al., 2023; Clark et al., 2025). Through a so-called 'human-in-the-loop' algorithm, experienced scientists can make inclusion decisions, verify the accuracy of extracted data points, flag suspected data fabrications, or have the final say on AI-drafted judgements (Van Dijk et al., 2023; Tzachor et al., 2023; Khraisha et al., 2024; Ye et al., 2024). Under such modalities, scientists' time could eventually be freed up for actual experimentation, advocacy, farmer extension, and policy engagement. As such, machine power may invigorate underfunded disciplines such as biological control or agroecology and ultimately put agriculture on a greener track.

CRediT authorship contribution statement

Kris A.G. Wyckhuys: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing, Supervision. **Komivi S. Akutse:** Writing – review & editing, Investigation, Data curation. **Divina M. Amalin:** Writing – review & editing, Investigation, Data curation. **Salah-Eddin Araj:** Writing – review & editing, Investigation, Data curation. **Marie Joy B. Beltran:** Writing – review & editing, Investigation, Data curation. **Ibtissem Ben Fekih:** Writing – review & editing, Investigation, Data curation. **Paul-André Calatayud:** Writing – review & editing, Investigation, Data curation. **Lizette Cicero:** Writing – review & editing, Methodology, Data curation. **Marcellin C. Cokola:** Writing – review & editing, Methodology, Data curation. **Yelitza C. Colmenarez:** Writing – review & editing, Methodology, Data curation. **Kenza Dessauvages:** Writing – review & editing, Methodology, Data curation. **Thomas Dubois:** Writing – review & editing, Methodology, Data curation. **Léna Durocher-Granger:** Writing – review & editing, Methodology, Data curation. **José L. Fernández-Triana:** Writing – review & editing, Methodology, Data curation. **Frederic Francis:** Writing – review & editing, Methodology, Data curation. **Khalid Haddi:** Writing – review & editing, Methodology, Data curation. **Rhett D. Harrison:** Writing – review & editing, Methodology, Data curation. **Muhammad Haseeb:** Writing – review & editing, Methodology, Data curation. **Natasha S.A. Iwanicki:** Investigation, Writing – review & editing. **Lara R. Jaber:** Investigation, Writing – review & editing. **Fathiya M. Khamis:** Writing – review & editing, Methodology, Data curation. **Jesusa C. Legaspi:** Writing – review & editing, Methodology, Data curation. **Refugio J. Lomeli-Flores:** Writing – review & editing, Methodology, Data curation. **Baoqian Lyu:** Writing – review & editing, Methodology, Data curation. **James Montoya-Lerma:** Writing – review & editing, Methodology, Data curation. **Ihsan Nurkomar:** Writing – review & editing, Methodology, Data curation. **James E. O'Hara:** Investigation, Writing – review & editing. **Jermaine D. Perier:** Writing – review & editing, Methodology, Data curation. **Ricardo Ramírez-Romero:** Writing – review & editing, Methodology, Data curation. **Francisco J. Sanchez-Garcia:** Writing – review & editing, Methodology, Data curation. **Ann Marie S. Robinson-Baker:** Writing – review & editing, Methodology, Data curation. **Luis C.P. Silveira:** Writing – review & editing, Methodology, Data curation. **Larisner Simeon:** Writing – review & editing, Methodology, Data curation. **Leellen F. Solter:** Writing – review & editing, Methodology, Data curation. **Oscar F. Santos-Amaya:** Writing – review & editing, Methodology, Data curation. **Wagner de Souza Tavares:** Writing – review & editing, Methodology, Data curation. **Rogelio Trabanino:** Writing – review & editing, Methodology, Data curation. **Fernando H. Valicente:** Writing – review & editing, Methodology, Data curation. **Carlos Vásquez:** Writing – review & editing, Methodology, Data curation. **Zhenying Wang:** Writing – review & editing, Methodology, Data curation. **Lian-Sheng Zang:** Writing – review & editing, Methodology, Data curation. **Wei Zhang:** Writing – review & editing, Methodology, Data curation. **Kennedy J. Zimba:** Writing – review & editing, Methodology, Data curation. **Kongming Wu:** Writing – review & editing, Methodology, Data curation. **Yubak D. Gc:** Funding acquisition, Project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yubak Dhoj GC reports financial support was provided by Food and Agriculture Organization of the United Nations. Kris AG Wyckhuys reports a relationship with Chrysalis Consulting that includes: employment. KAGW is the chief executive officer of Chrysalis Consulting – a firm which provides tailored support to nature-friendly farming and biological control. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded and executed by the United Nations Food and Agriculture Organization (FAO). We wish to thank Marc Kenis for commenting on an earlier draft and for providing suggestions that helped improve the overall quality of our work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2025.111317.

Data availability

All data underlying any of the analyses presented in this manuscript can be facilitated upon request.

References

Abbas, A., Ullah, F., Hafeez, M., Han, X., Dara, M.Z.N., et al., 2022. Biological control of fall armyworm, Spodoptera Frugiperda. Agronomy 12 (11), 2704.
Anonymous, 2024. China has become a scientific superpower. https://www.economist.com/science-and-technology/2024/06/12/china-has-become-a-scientific-superpower, accessed on August 20, 2025.
Baidoo-Anu, D., Ansah, L.O., 2023. Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. J. AI 7 (1), 52–62.
Bolaños, F., Salatino, A., Osborne, F., Motta, E., 2024. Artificial intelligence for literature reviews: opportunities and challenges. Artif. Intell. Rev. 57 (10), 259.
Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., et al., 2021. On the opportunities and risks of foundation models. arXiv 2021. arXiv preprint arXiv: 2108.07258.
Bradicich, P.A., Faris, A.M., Gordy, J.W., Brewer, M.J., 2024. Natural increases in parasitoid and predator abundances and a shift in species dominance point to improved suppression of the sorghum aphid since its invasion into North America. Insects 15 (12), 958.
Calone, R., Raparelli, E., Bajocco, S., Rossi, E., Crecco, L., et al., 2025. Analysing the potential of ChatGPT to support plant disease risk forecasting systems. Smart Agric. Technol. 10, 100824.
Candel, J., 2022. EU food-system transition requires innovative policy analysis methods. Nat. Food 3 (5), 296–298.
Carlsson, N.O., Sarnelle, O., Strayer, D.L., 2009. Native predators and exotic prey–an acquired taste? Front. Ecol. Environ. 7 (10), 525–532.
Casas, J., Swarbrick, S., Murdoch, W.W., 2004. Parasitoid behaviour: predicting field from laboratory. Ecol. Entomol. 29 (6), 657–665.
Chen, H., Jiang, Z., Liu, X., Xue, C.C., Yew, S.M.E., et al., 2025. Can large language models fully automate or partially assist paper selection in systematic reviews? Br. J. Ophthalmol. in press.
Clark, J., Barton, B., Albarqouni, L., Byambasuren, O., Jowsey, T., et al., 2025. Generative artificial intelligence use in evidence synthesis: a systematic review. Res. Synth. Methods 1–19.
Conroy, G., Mallapaty, S., 2025. How China created AI model DeepSeek and shocked the world. Nature 638, 300–301.

Daily, G.C., Polasky, S., Goldstein, J., Kareiva, P.M., Mooney, H.A., et al., 2009. Ecosystem services in decision making: time to deliver. Front. Ecol. Environ. 7 (1), 21–28.
De Clercq, D., Nehring, E., Mayne, H., Mahdi, A., 2024. Large language models can help boost food production, but be mindful of their risks. Front. Artif. Intell. 7, 1326153.
Deguine, J.P., Aubertot, J.N., Bellon, S., Côte, F., Lauri, P.E., et al., 2023. Agroecological crop protection for sustainable agriculture. Adv. Agron. 178, 1–59.
EC, 2021. Evaluating the impact of nature-based solutions – A summary for policy makers. European Commission: Directorate-General for Research and Innovation (DG RTD). Publications Office, Brussels, Belgium.
Elliott, J.H., Synnot, A., Turner, T., Simmonds, M., Akl, E.A., et al., 2017. Living systematic review: 1. Introduction—the why, what, when, and how. J. Clin. Epidemiol. 91, 23–30.
Emsley, R., 2023. ChatGPT: these are not hallucinations–they're fabrications and falsifications. Schizophrenia 9 (1), 52.
Gibney, E., 2025. China's cheap, open AI model DeepSeek thrills scientists. Nature 638, 13–14.
Gartlehner, G., Affengruber, L., Titscher, V., Noel-Storr, A., Dooley, G., et al., 2020. Single-reviewer abstract screening missed 13 percent of relevant studies: a crowdbased, randomized controlled trial. J. Clin. Epidemiol. 121, 20–28.
Gladstone, S.H., 1991. Parasitos del cogollero, *Spodoptera frugiperda* Smith (Lepidoptra: Noctuidae) en maiz sembrado en la epoca seca en Nicaragua. Ceiba 32 (2), 201–206.
Gohr, C., Rodríguez, G., Belomestnykh, S., Berg-Moelleken, D., Chauhan, N., et al., 2025. Artificial intelligence in sustainable development research. Nat. Sustainability 8, 970–978.
González-Chang, M., Wratten, S.D., Shields, M.W., Costanza, R., Dainese, M., et al., 2020. Understanding the pathways from biodiversity to agro-ecological outcomes: a new, interactive approach. Agr. Ecosyst. Environ. 301, 107053.
González-Maldonado, M.B., Correa-Ramírez, M.M., Rosas-García, N.M., Chaírez-Hernández, I., Garzón-Zuñiga, M.A., 2019. Genetic variability of species of the genus *Meteorus* Haliday, 18351, at Durango, Mexico. Southwestern Entomologist 44 (4), 909–918.
Guo, H.F., Fang, J.C., Zhong, W.F., Liu, B.S., 2013. Interactions between *Meteorus pulchricornis* and *Spodoptera exigua* multiple nucleopolyhedrovirus. J. Insect Sci. 13 (1), 12.
Gupta, S., Ranjan, R. and Singh, S.N., 2024. A comprehensive survey of retrieval-augmented generation (rag): Evolution, current landscape and future directions. arXiv preprint arXiv:2410.12837.
Gurr, G.M., Liu, J., Pogrebna, G., 2024. Harnessing artificial intelligence for analysing the impacts of nectar and pollen feeding in conservation biological control. Curr. Opin. Insect Sci. 62, 101176.
Hanson, M.A., Barreiro, P.G., Crosetto, P., Brockington, D., 2024. The strain on scientific publishing. Quant. Sci. Stud. 5 (4), 823–843.
Hoelmer, K.A., Kirk, A.A., 2005. Selecting arthropod biological control agents against arthropod pests: can the science be improved to decrease the risk of releasing ineffective agents? Biol. Control 34 (3), 255–264.
Janssen, A., van Rijn, P.C., 2021. Pesticides do not significantly reduce arthropod pest densities in the presence of natural enemies. Ecol. Lett. 24 (9), 2010–2024.
Jara-Chiquito, J.L., Oliva, F., Lobato-Vila, I., Pujade-Villar, J., 2024. Temporal changes in the composition of parasitoid assemblages associated with the invasive chestnut gall wasp. Ecol. Entomol. 49 (6), 779–797.
Jin, I., Tangsrivimol, J.A., Darzi, E., Hassan Virk, H.U., Wang, Z., et al., 2025. DeepSeek vs. ChatGPT: prospects and challenges. Front. Artif. Intell. 8, 1576992.
Joos, L., Keim, D.A., Fischer, M.T., 2024. Cutting through the clutter: the potential of LLMs for efficient filtration in systematic literature reviews. arXiv preprint arXiv: 2407.10652.
Kacena, M.A., Plotkin, L.I., Fehrenbacher, J.C., 2024. The use of artificial intelligence in writing scientific review articles. Curr. Osteoporos. Rep. 22 (1), 115–121.
Kargupta, P., Zhang, N., Zhang, Y., Zhang, R., Mitra, P., Han, J., 2025. TaxoAdapt: aligning LLM-based multidimensional taxonomy construction to evolving research corpora. arXiv preprint arXiv:2506.10737.
Kaygisiz, Ö.F., Teke, M.T., 2025. Can deepseek and ChatGPT be used in the diagnosis of oral pathologies? BMC Oral Health 25 (1), 638.
Kenis, M., 2023. Prospects for classical biological control of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in invaded areas using parasitoids from the Americas. J. Econ. Entomol. 116 (2), 331–341.
Kenis, M., Benelli, G., Biondi, A., Calatayud, P.A., Day, R., et al., 2023. Invasiveness, biology, ecology, and management of the fall armyworm, Spodoptera Frugiperda. Entomologia Generalis 43 (2), 187.
Khraisha, Q., Put, S., Kappenberg, J., Warraitch, A., Hadfield, K., 2024. Can large language models replace humans in systematic reviews? evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages. Res. Synth. Methods 15 (4), 616–626.
Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., et al., 2023. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digital Health 2 (2), e0000198.
Lamanna, M., Muca, E., Giannone, C., Bovo, M., Boffo, F., Romanzin, A., Cavallini, D., 2025. Artificial intelligence meets dairy cow research: Large language model's application in extracting daily time-activity budget data for a meta-analytical study. J. Dairy Sci. in press.
Luo, P.W., Liu, J.W., Xie, X., Jiang, J.W., Huo, X.Y., et al., 2025. DeepSeek vs ChatGPT: a comparison study of their performance in answering prostate cancer radiotherapy questions in multiple languages. Am. J. Clin. Exp. Urol. 13 (2), 176.
Molina-Ochoa, J., Carpenter, J.E., Heinrichs, E.A., Foster, J.E., 2003. Parasitoids and parasites of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in the Americas and Caribbean Basin: an inventory. Fla. Entomol. 86 (3), 254–289.
Mostafapour, M., Fortier, J.H., Pacheco, K., Murray, H., Garber, G., 2024. Evaluating literature reviews conducted by humans versus ChatGPT: comparative study. Jmir AI 3, e56537.
Nguyen, D.H., Nakai, M., Takatsuka, J., Okuno, S., Ishii, T., et al., 2005. Interaction between a nucleopolyhedrovirus and the braconid parasitoid *Meteorus pulchricornis* (Hymenoptera: Braconidae) in the larvae of *Spodoptera litura* (Lepidoptera: Noctuidae). Appl. Entomol. Zool. 40 (2), 325–334.
Normile, D., 2025. Chinese firm's large language model makes a splash. Science 387, 238.
Oerke, E.C., 2006. Crop losses to pests. J. Agric. Sci. 144 (1), 31–43.
Pair, S.D., Gross, H.R., 1989. Seasonal incidence of fall armyworm (Lepidopera: Noctuidae) pupal parasitism in corn by *Diapetimorpha introita* and *Cryptus albitarsis* (Hymenoptera: Ichneumonidae). J. Entomol. Sci. 24 (3), 339–343.
Pavageau, C., Pondini, S. and Geck, M. 2020. Money flows: what is holding back investment in agroecological research for Africa. Biovision Foundation for Ecological Development & International Panel of Experts on Sustainable Food Systems IPES-Food, Zurich, Switzerland.
Pimbert, M.P., Moeller, N.I., 2018. Absent agroecology aid: on UK agricultural development assistance since 2010. Sustainability 10 (2), 505.
Schaffner, U., Heimpel, G.E., Mills, N.J., Muriithi, B.W., Thomas, M.B., et al., 2024. Biological control for one health. Sci. Total Environ. 951, 175800.
Scherbakov, D., Hubig, N., Jansari, V., Bakumenko, A., Lenert, L.A., 2025. The emergence of large language models as tools in literature reviews: a large language model-assisted systematic review. J. Am. Med. Inform. Assoc. 32 (6), 1071–1086.
Schneider, K., Barreiro-Hurle, J., Rodriguez-Cerezo, E., 2023. Pesticide reduction amidst food and feed security concerns in Europe. Nat. Food 4 (9), 746–750.
Shattuck, A., Werner, M., Mempel, F., Dunivin, Z., Galt, R., 2023. Global pesticide use and trade database (GloPUT): New estimates show pesticide use trends in low-income countries substantially underestimated. Glob. Environ. Chang. 81, 102693.
Shepherd, K.D., Miller, M.A., Kisitu, B., Miles, B.G., Gbedevi, K. et al. 2025. Virtual Agronomist-an AI-assisted chatbot for guiding crop management decisions of smallholder farmers in Africa. agriRxiv, (2025), p.20250269995.
Silva, B., Nunes, L., Esteva, R., Aski, V., Chandra, R., 2023. GPT-4 as an agronomist assistant? Answering agriculture exams using large language models. arXiv [preprint] arXiv.2310.06225. doi: 10.48550/arXiv.2310.06225.
Susnjak, T., Hwang, P., Reyes, N., Barczak, A.L., McIntosh, T., et al., 2025. Automating research synthesis with domain-specific large language model fine-tuning. ACM Trans. Knowl. Discov. Data 19 (3), 1–39.
Tambo, J.A., Kansiime, M.K., Mugambi, I., Rwomushana, I., Kenis, M., et al., 2020. Understanding smallholders' responses to fall armyworm (*Spodoptera frugiperda*) invasion: evidence from five african countries. Sci. Total Environ. 740, 140015.
Tang, F.H., Wyckhuys, K.A.G., Li, Z., Maggi, F., Silva, V., 2025. Transboundary impacts of pesticide use in food production. Nat. Rev. Earth & Environ. 6, 383–400.
Tordjman, M., Liu, Z., Yuce, M., Fauveau, V., et al., 2025. Comparative benchmarking of the DeepSeek large language model on medical tasks and clinical reasoning. Nat. Med. 1.
Turner, T., Lavis, J.N., Grimshaw, J.M., Green, S., Elliott, J., 2023. Living evidence and adaptive policy: perfect partners? Health Research Policy and Systems 21 (1), 135.
Tzachor, A., Devare, M., Richards, C., Pypers, P., Ghosh, A., et al., 2023. Large language models and agricultural extension services. Nat. Food 4 (11), 941–948.
van Dijk, S.H., Brusse-Keizer, M.G., Bucsan, C.C., van der Palen, J., Doggen, C.J., Lenferink, A., 2023. Artificial intelligence in systematic reviews: promising when appropriately used. BMJ Open 13 (7), e072254.
Van Dis, E.A., Bollen, J., Zuidema, W., Van Rooij, R., et al., 2023. ChatGPT: five priorities for research. Nature 614 (7947), 224–226.
Van Lenteren, J.C., 2012. The state of commercial augmentative biological control: plenty of natural enemies, but a frustrating lack of uptake. BioControl 57 (1), 1–20.
Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J.B., Aali, A., et al., 2024. Adapted large language models can outperform medical experts in clinical text summarization. Nat. Med. 30 (4), 1134–1142.
Wyckhuys, K.A.G., Akutse, K.S., Amalin, D.M., Araj, S.E., Barrera, G., et al., 2024. Global scientific progress and shortfalls in biological control of the fall armyworm *Spodoptera frugiperda*. Biol. Control 191, 105460.
Wyckhuys, K.A.G., Leatemia, J.A., Fanani, M.Z., Furlong, M.J., Gu, B., et al., 2023. Generalist predators shape biotic resistance along a tropical island chain. Plants 12 (18), 3304.
Wyckhuys, K.A.G., Sanchez-Bayo, F., Aebi, A., Van Lexmond, M.B., Bonmatin, J.M., et al., 2021. Stay true to integrated pest management. Science 371 (6525), 133.
Yang, X., Wyckhuys, K.A.G., Jia, X., Nie, F., Wu, K., 2021. Fall armyworm invasion heightens pesticide expenditure among chinese smallholder farmers. J. Environ. Manage. 282, 111949.
Yang, S., Yuan, Z., Li, S., Peng, R., Liu, K., Yang, P., 2024. Gpt-4 as evaluator: Evaluating large language models on pest management in agriculture. arXiv preprint arXiv: 2403.11858.
Ye, A., Maiti, A., Schmidt, M., Pedersen, S.J., 2024. A hybrid semi-automated workflow for systematic and literature review processes with large language model analysis. Future Internet 16 (5), 167.
Yuan, X.T., Shao, C.Y., Zhang, Z.Z., Qian, D., 2025. Comparing the performance of ChatGPT and ERNIE Bot in answering questions regarding liver cancer interventional radiology in Chinese and English contexts: a comparative study. Digital Health 11, 20552076251315511.