JIPTI
JURNAL INOVASI PENDIDIKAN DAN TEKNOLOGI INFORMASI

Volume 7 No. 2, November (2026)
E-ISSN: 2720-9725
P-ISSN: 2987-8462

AI-Based Decision Support System untuk Pemeringkatan Literatur Ilmiah: Presisi *Semantic Retrieval* Bervariasi antar Topik

Fattachul Huda Aminuddin[1*], Fery Purnama[2], Pariyadi[3], Ahmad Husna Ahadi[4], Teuku Djauhari[5]
Universitas Nurdin Hamzah[1245]
Universitas Jambi[3]
E-mail: fattachulhuda@unh.ac.id[1*], ferypurnama@unh.ac.id[2], pariyadi@unja.ac.id[3], ahmad_husna@unh.ac.id[4], technikom@gmail.com[5]

Abstrak

Peningkatan jumlah publikasi ilmiah menjadi salah satu faktor penyebab peneliti menghadapi kesulitan dalam menemukan, mengevaluasi, dan menentukan literatur yang paling relevan. Sistem pencarian berbasis kata kunci masih memiliki keterbatasan dalam memahami hubungan konsep, kualitas evidence, dan peluang penelitian. Penelitian ini bertujuan untuk mengembangkan AI-DSS Research Assistance sebagai pendukung keputusan berbasis kecerdasan buatan untuk membantu analisis literatur dan rekomendasi arah penelitian. Metode penelitian menggunakan pendekatan *system development* and *evaluation* dengan mengintegrasikan semantic retrieval, analisis research opportunity, dan metode DSS menggunakan SAW, TOPSIS, dan WP. Evaluasi sistem menggunakan *benchmark query*, Precision@5, *relevance score*, analisis *research gap*, dan *robustness analysis*. Hasil penelitian menunjukan bahwa sistem mendapatkan nilai rata-rata Precision@5 sebesar 0,56 dan *Mean Relevance Score* sebesar 3,44. Analisis terhadap 20 artikel evidence menghasilkan 2 artikel dengan opportunity signal dan signal coverage sebesar 10 %. Pada pemeringkatan DSS yang menjadi rekomendasi utama yaitu skor TOPSIS sebesar 90,62. Hasil *roburtness analysis* menunjukan bahwa perubahan bobot tidak mempengaruhi rekemondasi utama secara signifikan. Penelitian ini menunjukan bahwa integrasi AI dan DSS dapat membantu proses *literature review* secara lebih sistematis, objektif, dan berbasis evidence.

Kata Kunci: Kecerdasan Buatan; Sistem Pendukung Keputusan; Pencarian Semantik; Kesenjangan Penelitian; Tinjauan Literatur.

Aminuddin, F, H., Purnawa, F., Pariyadi., Ahadi, A, H., & Djauhari, T. (2026). AI-Based Decision Support System untuk Pemeringkatan Literatur Ilmiah: Presisi Semantic Retrieval Bervariasi antar Topik. *Jurnal Inovasi Pendidikan dan Teknologi Informasi (JIPTI), 7(2), 775-792.* https://doi.org/10.52060/jipti.v7i2.4881

http://ejournal.ummba.ac.id/index.php/JIPTI/

Abstract

The increasing volume of scientific publications makes it challenging for researchers to identify relevant literature, detect research opportunities, and prioritize references. This study develops AI-DSS Research Assistant by integrating semantic retrieval, research opportunity analysis, and multi-criteria decision-making methods, namely Simple Additive Weighting (SAW), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), and Weighted Product (WP). An initial set of 100 articles was retrieved and screened to obtain an evidence corpus of 20 articles. The system was evaluated using five benchmark queries, relevance judgments, research opportunity analysis, DSS ranking comparisons, and weight sensitivity tests. The results show a Precision@5 of 0.56 (14 of 25 Top-5 results) and a Mean Relevance Score of 3.44 on a 1–5 scale. Two of the 20 articles exhibited research opportunity signals, yielding a signal coverage of 10%. The highest-ranked article under TOPSIS obtained a score of 90.62. Sensitivity testing indicated that weight changes of ±10% and ±20% did not significantly alter the primary TOPSIS recommendation. These findings suggest that the system can support structured literature analysis, while final research gap validation remains the responsibility of the researcher.

Keywords**:** *Artificial Intelligence; Decision Support System; Semantic Retrieval; Research Gap; Literature Review*.

Submitted: *2026-06-30.* ***Revision:*** *2026-08-11.* ***Accepted:*** *2026-10-21.* ***Publish:*** *2026-11-07.*

PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (AI) telah mendorong peningkatan produksi dan akses terhadap publikasi ilmiah. Berbagai sumber digital memberikan peluang bagi peneliti untuk memperoleh artikel dari beragam ilmu dalam waktu yang relatif singkat. Namun, semakin banyaknya publikasi juga menghadirkan tantangan dalam menemukan dan menentukan literatur yang benar-benar relevan dengan kebutuhan penelitian. Penelitian tidak hanya perlu menemukan artikel berdasarkan kesesuaian topik, namun juga memahami kontribusi penelitian sebelumnya, menilai kualitas bukti, dan mengidentifikasi peluang pengembangan penelitian. Oleh karena itu, proses analisis literatur membutuhkan pendekatan yang tidak hanya membantu pencarian dokumen, tetapi juga mendukung penilaian dan pengambilan keputusan secara sistematis (Gusenbauer & Gauster, 2025).

Pencarian literatur secara konvensional umumnya mengandalkan pencocokan kata kunci. Pendekatan ini berguna untuk menemukan dokumen yang menggunakan istilah tertentu, namun hal ini belum mampu menangkap hubungan makna di antara konsep yang berbeda. Sebagaimana contoh studi kasus dua artikel dapat membahas permasalahan yang sama dengan menggunakan istilah atau terminologi yang berbeda. Dalam kondisi tersebut, pencarian berbasis kata kunci berpotensi tidak menampilkan seluruh artikel yang relevan. Sebaliknya, dokumen yang memiliki kesamaan kata belum tentu membahas permasalahan yang sesuai dengan kebutuhan peneliti. Keterbatasan ini menunjukan perlunya pendekatan pencarian yang mampu memahami konteks dan hubungan semantic dalam dokumen ilmiah (Thakur et al., 2021).

Pemanfaaatan AI, khususnya *Large Languange Models* (LLM), *embedding*, dan model berbasis *transformer*, membuka peluang untuk meningkatkan proses penemuan, penyaringan dan analisis literatur ilmiah (Delgado-Chaves et al., 2025).

Namun, kinerja LLM maish bervariasi dalam proses seleksi artikel dan ekstraksi informasi (Khraisha et al., 2024). Teknologi *semantic retrieval* memungkinkan sistem menemukan artikel berdasarkan kesamaan makna antara kueri dan isi dokumen, sedangakn LLM membantu proses penyaringan, ekstraksi informasi, dan peringkasan artikel. Namun, relevansi hasil pencarian dan ketepatan informasi yang dihasilkan AI tetap perlu evaluasi agar rekomendasi yang diberikan dapat dipertanggungjawabkan secara ilmiah (Bolanos et al., 2024).

Permasalahan muncul ketika peneliti menentukan literatur yang paling berkontribusi terhadap penelitian. Kesamaan topik tidak selalu mencerminkan kualitas bukti ilmiah atau peluang penelitian baru. Oleh karena itu, diperlukan mekanisme pengambilan keputusan yang mempertimbangkan beberapa kriteria secara stimultan, meliputi kesesuaian topik (*topik fit*), peluang penelitian (*research opportunity*), kualitas bukti (*evidence quality*), kebaruan publikasi (*publication recency*), kesesuaian kompetensi, dan aksesbilitas literatur (Hoekstra & Lhachimi, 2025).

Salah satu pendekatan untuk meningkatkan kualitas rekomendasi literatur adalah mengintegrasikan *semantic retrieval* dengan *decision support system* (DSS). Integrasi ini memungkinkan sistem menilai relevansi topik sekaligus mempertimbangkan kualitas bukti, kebaruan publikasi, peluang penelitian, kesesuaian kompetensi, dan aksesbilitas literatur. Metode *Multi-Criteria Decision Making* (MCDM), seperti TOPSIS *Weighted Product* (WP) dan SAW, dapat digunakan untuk mengevaluasi dan memeringkat alternatif berdasarkan sejumlah kriteria secara sistematis (Ciardiello & Genovese, 2023a).

Sejumlah sistem berbasis AI telah dikembangkan untuk mendukung penelusuran literatur ilmiah. Semantic Scholar menyediakan graf literatur terbuka dengan lebih dari 200 juta makalah yang dilengkapi ringkasan dan vektor *embedding* untuk mendukung pencarian semantik (Kinney et al., 2025). Elicit memanfaatkan LLM untuk menyaring judul dan abstrak serta mengekstraksi data dari teks lengkap, dengan rekomendasi *include/exclude* yang disertai penjelasan (Elicit, 2025). Namun, evaluasi independen pada tahap ekstraksi data menyimpulkan bahwa Elicit dapat melengkapi, tetapi tidak menggantikan, pengekstrak manusia, dan hasilnya dapat berubah akibat pembaruan algoritma sehingga mengurangi reproduktibilitas (Lagisz et al., 2026). *ResearchRabbit* memetakan literatur sebagai jaringan sitasi dari artikel *seed* melalui fitur *Similar Work*, *Earlier Work*, dan *Later Work*, tetapi rekomendasinya bersifat algoritmik, tidak selalu dapat direproduksi, dan bukan merupakan basis data terkurasi (Board, 2026)

Pendekatan *Retrieval-Augmented Generation* (RAG) menggabungkan pencarian dokumen dengan pembangkitan teks oleh LLM (Lewis et al., 2020). Pada domain ilmiah, OpenScholar menunjukkan bahwa pemanfaatan RAG berbasis korpus publikasi ilmiah menghasilkan jawaban yang disertai sitasi dan didukung sumber literatur (Asai et al., 2026). Pendekatan tersebut memperlihatkan potensi RAG dalam sintesis informasi ilimah. Namun pengambilan dokumen dan pembangkitan jawaban tidak dengan sendirinya menghasilkan pemeringkatan prioritas literatur berdasarkan sejumlah kriteria keputusan yang ditetapkan pengguna. Perbedaan kapabilitas dan fokus dapat terlihat pada Tabel 1 berikut :

Tabel 1. Perbandingan Pendekatan dalam Analisis Literatur berbasis AI.

Pendekatan	**Kapabilitas utama**	**Fokus yang terdokumentasi**	**Posisi dan kontribusi penelitian yang diteliti**
Elicit	Pencarian Literatur berbantuan AI, penyaringan artikel, ekstraksi data, dan sintesis informasi	Mendukung tahapan systematic review dan pengelolaan bukti ilmiah	Mengintegrasikan pemeringkatan multikriteria menggunakan SAW, TOPSIS, dan WP berdasarkan enam kriteria penilaian
Semantic Scholar	Pencarian publikasi, ringkasan berbantuan AI, rekomendasi dan *Research Feeds*	Penemuan publikasi dan rekomendasi artikel ilmiah	Menggabungkan *semantic retrieval* dengan analisis *research opportunity* dan pemeringkatan berbasis DSS
ResearchRabbit	Eksplorasi hubungan sitasi, referensi, dan jejaring penulis	Penemuan literatur terkait jejaring ilmiah	Menggunakan kriteria kesesuaian topik, kualitas bukti dan kebutuhan pengguna untuk penentuan prioritas artikel
Sistem Berbasis RAG	Pengambilan dokumen relevan yang diikuri sintesis LLM	Pengambilan dan pemanfaatan informasi dari korpus dokumen	Menambahkan pemeringkatan alternatif secara multikriteria dan evaluasi pen

Bedasarkan Tabel 1, Perbandingan ini digunakan untuk memperjelas posisi penelitian yang dikembangkan, bukan untuk menyatakan bahwa sistem tersebut memiliki keterbatasan yang telah dibuktikan melalui pengujian komparatif langsung. Meskipun AI telah dimanfaatkan dalam kajian literatur, integrasi *semantic retrieval*, analisis peluang penelitian, dan pemeringkatan literatur dalam satu sistem terpadu masih perlu dikaji lebih lanjut.

Sistem rekomendasi literatur juga perlu mendukung pengambilan keputusan yang transparan dan berbasis bukti, tidak hanya berfokus pada pencarian dan peringkasan informasi. Selain itu, analisis sensitivitas terhadap perubahan bobot kriteria diperlukan untuk mengevaluasi kestabilan hasil pemeringkatan (Azzini & Munda, 2025).

Berdasarkan tinjauan fitur yang terdokumentasi tersebut, terdapat tiga aspek yang menjadi fokus penelitian ini. Pertama, pencarian, penyaringan, dan sintesis literatur belum secara otomatis menyediakan mekanisme pemeringkatan prioritas artikel

berdasarkan kriteria dan bobot eksplisit yang dapat disesuaikan dan ditelusuri oleh peneliti. Kedua, indikasi peluang penelitian perlu dioperasionalkan sebagai keluaran terstruktur yang dapat ditelusuri kembali ke artikel sumber, mengingat indentifikasi *research gap* belum memiliki pendekatan standar yang berlaku untuk seluruh bidang penelitian (Nyanchoka et al., 2019a). Ketiga, kestabilan peringkat terhadap perubahan bobot kriteria dan perbedaan metode pemeringkatan perlu dievaluasi untuk mengetahui konsistensi rekomendasi yang dihasilkan. Dengan demikian, penelitian ini menempati posisi pada integrasi *semantic retrieval*, analisis peluang penelitian berbasis sinyal research gap, dan pemeringkatan MCDM menggunakan SAW, TOPSIS, dan WP yang dilengkap pengujian *robustness*. Perbandingan tersebut bersifat konseptual berdasarkan fitur dan pendekatan yang terdokumentasi, bukan hasil pengujian kinerja langsung antarperangkat.

Berdasarkan permasalahan dan kesenjangan tersebut, penelitian ini mengembangan AI-DSS *Research Assistant* untuk mendukung analisis literatur dan arah penelitian. Sistem ini mengintegrasikan *semantic retrieval* untuk menemukan literatur berdasarkan pemahaman konteks, *research opportunity analysis* untuk mengindentifikasi peluang dan celah penelitian, serta metode MCDM (SAW, TOPSIS, dan WP) untuk menentukan prioritas literatur berdasarkan berbagai kriteria. Pengujian *robustness* dilakukan melalui perubahan bobot kriteria untuk mengevaluasi kestabilan rekomendasi yang dihasilkan. Kontribusi penelitian terletak pada integrasi proses penemuan literatur, analisis peluang penelitian dan pengambilan keputusan multikriteria dalam satu alur kerja yang dapat dievaluasi secara empiris. Dengan demikian, sistem tidak hanya mengidentifikasi literatur yang relevan, tetapi menghasilkan *candidate research gap* berbasis bukti yang dapat ditelusuri dan divalidasi oleh peneliti. Sistem ini berperan sebagai asisten pengambilan keputusan untuk membantu menentukan peluang dan prioritas penelitian, bukan sebagai pengganti peneliti.

METODE PENELITIAN

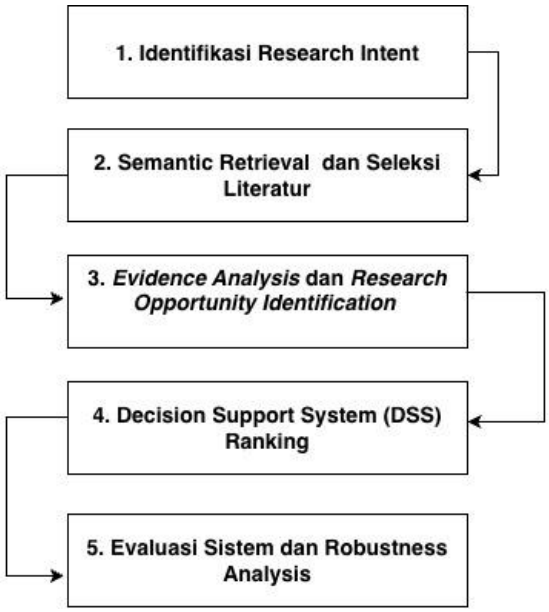
Penelitian ini menggunakan pendekatan eksperimental berbasis pengembangan sistem (*system development and evaluation)* untuk membangun AI-DSS *Research Assistant* sebagai pengambilan keputusan dalam analisis literatur ilmiah. sebanyak 100 artikel diperoleh melalui proses *retrieva*l, kemudian diseleksi menjadi 20 artikel *evidence* berdasarkan relevansi topik, ketersediaan metadata, peluang penelitian dan kualitas dan akses sumber.

Fokus penelitian berfokus pada pengembangan dan evaluasi alur kerja yang mengintegrasikan pencarian literatur, analisis bukti, identifikasi peluang penelitian (*research opportunity analysis*), serta rekomendasi prioritas literatur menggunakan metode MCDM. Pendekatan ini menekankan evaluasi sistematis dan transparansi dalam pemanfataan AI untuk tinjauan literatur (Bolanos et al., 2024). Sementara itu, identifikasi *research gap* dan penentuan prioritas penelitian memerlukan analisis bukti yang terstruktur serta kriteria evaluasi yang jelas (Dewidar et al., 2026).

Penelitian ini terdiri atas lima tahapan utama, yaitu (1) identifikasi kebutuhan penelitian (*research intent analysis*), (2) pengumpulan dan pemeringkatan literatur berbasis *semantic retrieval*, (3) analisis evidence dan peluang, (4) pemeringkatan literatur menggunakan metode MCDM serta (5) evaluasi performa dan kestabilan sistem melalui pengujian benchmark dan *robustness analysis* terlihat pada Gambar 1.

Adapun alur penelitian yang dilakukan terdiri atas lima tahapan utama, yaitu (1) identifikasi kebutuhan penelitian (*research intent analysis*), (2) pengumpulan dan pemeringkatan literatur berbasis *semantic retrieval*, (3) analisis evidence dan peluang, (4) pengambilan keputusan menggunakan metode MCDM, serta (5) evaluasi performa dan kestabilan sistem melalui pengujian benchmark dan *robustness analysis*.

Evaluasi sistem dilakukan dengan empat indikator: yaitu Precision@5, untuk mengukur proporsi artikel relevan pada lima hasil teratas. *Relevance score*, untuk menilai kesesuaian artikel hasil retrieval terhadap kebutuhan penelitian. *DSS Consistency*, untuk membandingkan kesesuaian peringkat yang dihasilkan metode SAW, TOPSIS dan WP. *Robustness analysis*, untuk mengevaluasi kestabilan peringkat terhadap perubahan bobot kriteria. Adapun tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1



Gambar 1. Tahapan Penelitian.

Tahap 1 yaitu *identifikasi research Intent*. Peneliti memasukan topik atau pertanyaan penelitian sebagai *research intent* . Sistem ini menganalisis konteks untuk menentukan domain, fokus, dan kebutuhan pencarian literatur. Tahapan 2 yaitu *semantic retrieval* dan seleksi literatur. Sistem menggunakan *Sentence Transformer* untuk menghasilkan representasi teks (*embedding*) dan *cosine similarity* untuk mengukur kemiripan semantic antar teks. Artikel kandidat kemudian diseleksi berdasarkan relevansi topik dan kelengkapan informasi untuk membentuk *evidence corpus*. Strategi pencarian dan seleksi perlu dirancang secara sistematis agar proses penelusuran dapat dipertanggungjawabkan.

Tahap 3 adalah *evidence analysis* dan *research opportunity identification*. Artikel dianalisis berdasarkan kontribusi, keterbatasan, dan peluang penelitian lanjutan dengan memanfaatkan abstrak , metadata,

serta informasi yang tersedia dalam dokumen. Hasilnya berupa kandidat peluang penelitian yang akan dievaluasi pada tahapan berikutnya.(Gusenbauer & Gauster, 2025a).

Dalam penelitian ini, peluang penelitian (*research opportunity*) didefinisikan secara operasional dengan mengacu pada kerangka *research gap* (Nyanchoka et al., 2019) (Dewidar et al., 2026). Penelitian oleh (Nyanchoka et al., 2019b) menegaskan bahwa belum ada metode standar untuk mengidentifikasi dan memprioritaskan *research gap*, sehingga definisi operasional yang eksplisit diperlukan. Pada hasil penelitian oleh (Dewidar et al., 2026a) membedakan tiga tingkat: *research gap*, yaitu keterbatasan, kekurangan, atau arah yang belum dieksplorasi dalam bukti yang tersintesis; *research need*, yaitu *gap* yang dinilai relevan untuk pengambilan keputusan; dan *research priority*, yaitu *need* yang cukup penting dan mendesak untuk diteliti. Sistem hanya menghasilkan tingkat pertama, yaitu sinyal *candidate research gap* dari abstrak dan metadata berdasarkan tiga indikator: keterbatasan studi terdahulu, rekomendasi penelitian lanjutan, dan perbedaan pendekatan antarstudi. Setiap sinyal diklasifikasikan menurut tipenya (misalnya *Comparative Gap*) dan diberi *opportunity confidence* pada skala 0 sampai 100. Penilaian apakah suatu *gap* menjadi *need* atau *priority* memerlukan pertimbangan nilai dan tetap menjadi tanggung jawab peneliti.

Tahapan 4 yaitu *Decision Support System* (DSS) *Ranking.* Prioritas literatur ditentukan menggunakan metode MCDM , yaitu TOPSIS, SAW, dan WP, berdasarkan enam kriteria evaluasi yang disajikan pada Tabel 1. Perbandingan hasil pemeringkatan diperlukan karena metode MCDM yang berbeda dapa meghasilkan urutan alternatif yang berbeda (Li et al., 2024).

Tabel 2. Kriteria Evaluasi Literatur

Kriteria	Bobot
Topic Fit	40%
Research Opportunity	20%
Evidence Quality	15%
Publication Recency	10%
Competency Fit	10%
Evidence Access	5%

Penelitian ini menggunakan enam kriteria untuk menilai dan memeringkat artikel ilmiah. Kriteria *Topic Fit* mengukur kesesuaian artikel dengan kebutuhan pencarian berdasarkan tiga aspek, yaitu *intent fit*, *semantic relevance*, dan *query domain relevance.* Penggunaan beberapa aspek relevansi ini mengacu pada penelitian (Peikos & Pasi, 2024a) yaitu menjelaskan bahwa relevansi dalam *information retrieval* dipengaruhi oleh pengguna, tujuan pencarian, dan konteks domain. Sementara itu, pengukuran kemiripan semantik menggunakan *sentencene embedding* mengacu kepada pendekatan Sentence-BERT yang dikembangkan oleh (Reimers & Gurevych, 2019).

Kriteria *Research Opportunity* mengindentifikasi indikasi kesenjangan penelitian berdasarkan keterbatasan penelitian sebelumnya, rekomendasi penelitian lanjutan, dan pola kesenjangan dalam abstrak. Definisi operasional ini mengacu kepada kajian (Nyanchoka et al., 2019a) dan (Dewidar et al., 2026) yang

menunjukan bahwa identifikasi *research gap* memiliki beragram pendekatan. Evidence Quality meniliai kelengkapan informasi artikel, seperti judul, abstrak. DOI, jurnal, tahun, sumber, status akses, serta informasi metode dan fitur penelitian. *Publication Recency* mempertimbangkan usia publikasi untuk memprioritaskan informasi yang lebih baru, sejalan denagn aspek timeliness dalam peniliaian relevansi multidimensi (Peikos & Pasi, 2024b) *Competency Fit* menilai keseluruhan artikel dengan bidang keahlian pengguna, sebagaimana kebutuhan personalisasi dibahas dalam sistem rekomendasi artikel ilimiah. Semenyara, Evidencene Access mempertimbangkan kemudahan memperoleh artikel, khususnya berdasarkan status open acess. Hal ini menunjukan pentingnya akses terbuka dalam penyebaran dan pemanfaatan publikasi ilimiah (Piwowar et al., 2018).

Pendekatam *Multiple Criteria Decision Making* (MCDM) digunakan karena pemeringkatan literatur melibatkan beberapa kriteria yang memiliki tingkat kepentingan berbeda dan dapat menghasilkan penilaian yang saling bertentangan. Sebagai contoh, artikel yang memiliki relevansi semantic tinggi belum tentu memiliki informasi yang lengkap, indikasi kesenjangan penelitian, atau akses teks penuh. MCDM memungkinkan berbagai kriteria tersebut digabungkan untuk menilai dan memeringkat alternatif berdasarkan prioritas yang ditetapkan (Greco et al., 2025). Penelitian ini menerapkan tiga metode yaitu, SAW, TOPSIS, dan WP untuk membandingkan hasil pemeringkatan berdasarkan mekanisme agregasi yang berbeda, yaitu penjumlahan berbobot, kedekatan terhadap solusi ideal, dan perkalian berbobot. (Ciardiello & Genovese, 2023).

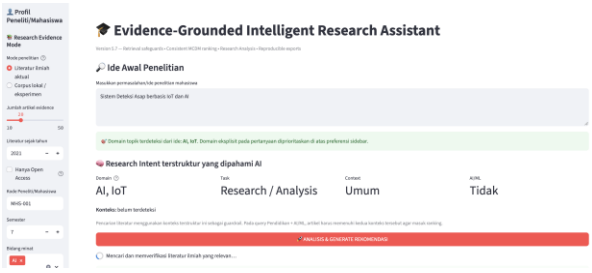
Bobot kriteria ditetapkan oleh peneliti berdasarkan prioritas pengembangan sistem, dengan *Topic Fit* memperoleh bobot tertinggi sebesar 40 %, diikuti *Research Opportunity* 20 %, *Evidence Quality* 15 %, *Publication Recency* 10%, *Competency Fit* 10%, dan *Evidence Access* 5 %. Pembobotan ini merupakan konfigurasi penelitian, bukan hasil penilaian pakar. Secara operasional, Research Opportunity didefinisikan sebagai indikasi kesenjangan penelitian yang terdeteksi melalui pencocokan pola kata dan frasa pada abstrak artikel. Sinyal tersebut diklasifikasikan melalui pencocokan pola kata dan frasa pada asbtrak artikel. SInyal tersebut diklasifikasikan ke dalam delapan kategori: *Methodology Gap*, *Context/Population Gap*, *Feature/Data Gap*, *Temporal Gap*, *Comaparative Gap*, *Application Gap*, *Evaluation Gap*, dan *Contradictionary Evidence*. Kategori tersebut merupakan klasifikasi operasional yang dikembangkan dalam penelitian ini berdasarkan kerangka indentifikasi research gap (Nyanchoka et al., 2019a) dan (Dewidar et al., 2026b).

HASIL DAN PEMBAHASAN

A. Evaluasi Kinerja Retrieval Literatur

Evauasi tahap pertama dilakukan untuk menilai kemampuan AI-DSS *Research Assistant* dalam mencari dan merekomendasikan literatur ilmiah berdasarkan pengguna. Sistem mengintegrasikan pendekatan *semantic*

retrieval menggunakan model *Sentence Transformer* dengan mekanisme *Decision Suport System* (DSS) untuk melakukan pemeringkatan artikel berdasarkan sejumlah kriteria evaluasi. Pendekatan ini memungkinkan pencarian berdasarkan kemiripan makna, bukan hanya kesamaan kata kunci, sehingga konsep yang berkaitan dengan secara semantik tetap dapat teridentifikasi meskipun menggunakan istilah berbeda. Kemampuan *Sentence Transformer* dalam mempresentasikan makna kalimat mendukung penerapannya pada pencarian informasi berbasis semantic . Gambar 2 menunjukan tampilan awal AI-*DSS research assistant* dalam menganalisis pertanyaan penelitian pengguna. Pada tahap ini, sistem mengindetifikasi konteks penelitian, bidang kajian, tujuan pencarian dan karakteristik topik sebelum melakukan retrieval literatur. Analisis awal tersebut bertujuan mengarahkan pencarian agar artikel yang ditemukan tidak hanya memiliki kemiripan istilah, namun tetap relevan dengan kebutuhan pengguna seperti pada Gambar 2 berikut.



Gambar 2. Tampilan analisis awal AI DSS *Research Assistant*

B. Desain *Benchmark Query* dan Metode Evaluasi

Evaluasi dilakukan menggunakan lima *benchmark query* yang mewakili variasi bidang dan hubungan antarkonsep penelitian, yaitu: (Q1) *Artificial Intelligence* dalam Auditing; (Q2) *Enterprise Risk Management* dan *Financial Resilience*; (Q3) *Risk Culture* dan *Risk Management Effectiveness*; (Q4) Canva *Interactive Learning* dan Motivasi Belajar; (Q5) *Retrieval Augmented Generation* untuk Pencarian Literatur Akademik.

Setiap query menghasilkan lima rekomendasi artikel teratas (*Top-5 retrival*) yang dievaluasi mealui *human relevance judgement*, yaitu penilaian manusia terhadap kesesuaian artikel dengan kebutuhan penelitian. Penilaian ini mempertimbangkan relevansi sebagai aspek multidimensi yang dipengaruhi oleh konteks dan tujuan pencarian . Kinerja sistem diukur menggunakan Precision@5 dan *Mean Relevance Score*. Hasil evaluasi disajikan pada Tabel 2 berikut.

Tabel 2. Hasil Analisis Benchmark DSS

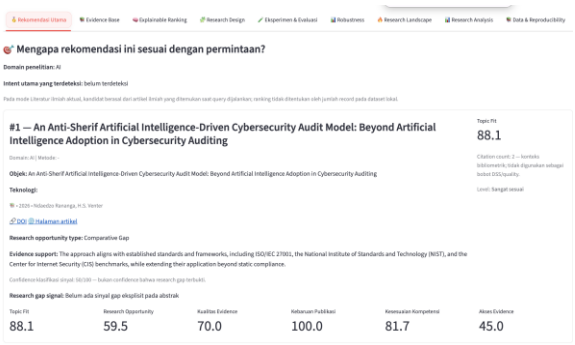
Kode	Benchmark Query	Bidang Penelitian	Dokumen Relevan Top-5	Precision @5	Mean Relevance Score
Q1	Artificial Intelligence in Auditing	Accounting & AI	5/5	1,00	4,80
Q2	Enterprise Risk Management terhadap Financial Resilience	Risk Management	1/5	0,20	2,40

Kode	Benchmark Query	Bidang Penelitian	Dokumen Relevan Top-5	Precision @5	Mean Relevance Score
Q3	Risk Culture terhadap Risk Management Effectiveness	Organization & Governance	3/5	0,60	3,20
Q4	Canva Interactive Learning terhadap Benchak Query	Education Technology Bidang Penelitian	3/5 Dokumen Relevan Top-5	0,60 Precision @5	4,00 Mean Relevance Score
Q5	Retrieval Augmented Generation untuk Academic Literature Search	AI & NLP	2/5	0,40	2,80
Rata-rata			14/25	0,56	3,44

Berdasarkan Tabel 2, sistem memperoleh nilai rata-rata Precision@5 sebesar 0,56. Nilai tersebut menunjukan bahwa 56% artikel dalam lima rekomendasi teratas dinilai relevan terhadap kueri penelitian. Rata-rata *Mean Relevance Score* sebesar 3,44 menunjukan tingkat relevansi rekomendasi secara keseluruhan berdasarkan skala penilaian yang digunakan. Kinerja terbaik diperolah pada Q1 dengan Precision@5 sebesar 1.00 dan skor relevansi 4,80, sedangkan Q2 menghasilkan Precision@5 terendah, yaitu 0,20. Hasil ini menunjukan bahwa kinerja retrieval bervariasi menurut topik yang dicari. Perbedaan tersebut mengindikasikan perlunya evaluasi lebih lanjut terhadap kemampuan sistem dalam menangani hubungan antarkonsep dan konteks penelitian yang lebih kompleks.

C. Analisis Hasil Retrieval Berdasarkan Query

Hasil pengujian Q1 menunjukan bahwa AI-DSS Research Asistant mampu merekomendasikan artikel yang sesuai dengan topik articial intelligence dalam auditing. Atikel dengan peringkat tertinggi memperoleh skor *Topik Fit* sebesar 88,1 kebaruan publikasi 100,0 dan kesesuaian kompetensi adalah 81,7. Skor tersebut menunjukan kesesuaian topik, kebaruan publikasi dan kecocok dengan bidang kompetensi pengguna yang realtif tinggi. Sebaliknya, skor *Evidence Acces* sebesar 45,0 menunjukan bahwa akses terhadap sumber masih menjadi keterbatasan. Temuan ini memperlihatkan bahwa pemeringkatan DSS membantu mengidentifikasi kelebihan dan keterbatasan artikel berdasarkan beberapa kriteria evaluasi, sejalan dengan penelitian (Mitrov et al., 2026) mengenai optimalisasi pencarian dokumen ilmiah menggunakan *text embedding* dan model bahasa besar seperti terlihat pada Gambar 3.

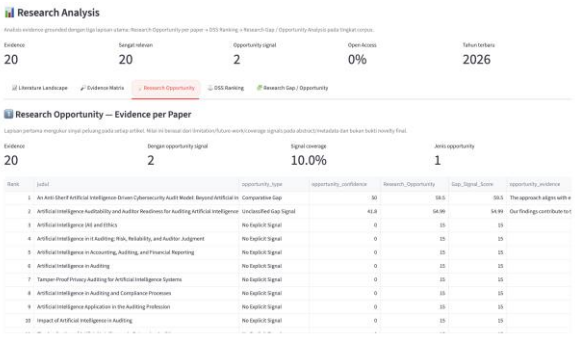


Gambar 3. Hasil Retrieval dan Ranking pada Query AI driven Cybersecurity Auditing

Pada Gambar 3 menunjukan bahwa pemeringkatan berbasis DSS mengintegrasikan beberapa kriteria untuk menilai rekomendasi literatur, sehingga artikel tidak hanya dievakuasi berdasarkan kesesuaian semantik, namun juga kebaruan , kompetensi pengguna , dan aksesbilitas sumber. Pendekatan ini melengkapi proses *semantic retrieval* dalam mendukung penelusuran literatur ilmiah.

D. Analisis *Research Opportunity* dan *Research Gap Indentification*

Setelah proses *retrieval literatur*, AI-DSS *Research Asistance* menganalisis 20 artikel untuk mengidentifikasi indikasi peluang dan celah penelitian berdasarkan keterbatasan studi terdahulu, rekomendasi penelitian lanjutan, serta perbedaan pendekatan. Analisis ini bertujuan untuk membantu peneliti menemukan potensi pengembangan riset dan literatur yang diperoleh. Pendekatan tersebut sejalan dengan kajian (Gusenbauer & Gauster, 2025) yang menekankan pentingnya strategi pencarian dan penyaringan literatur yang sistematis, transparan, dan dapat direproduksi. Hasil analisis disajikan pada Gambar 4.



Gambar 4. Hasil Analisis Research Opportunity dan Research Gap Indentification

Berdasarkan Gambar 4, sebanyak 20 artikel dikategorikan relevan, namun hanya 2 artikel yang menunjukan opportunity signal, dengan signal coverage sebesar 10 %. Hasil ini menunjukan bahwa relevansi literatur tidak selalu disertai indikasi *research gap* yang dapat diidentifikasi secara eksplisit. Artikel peringkat pertama dikategorikan sebagai Comparative Gap dengan opportunity confidence sebesar 50, yang mengindikasikan potensi pengembangan penelitian melalui perbandingan pendekatan dalam *artificial intelligence for cybersecurity auditing*. Namun, skor tersebut merupakan keluaran penilaian sistem dan perlu diverifikasi melalui pembacaan isi artikel.Hal ini juga sejalan dengan penelitian yang dilakukan oleh (Dewidar et al., 2026). Berikut adalah hasil temuan yang dipaparkan pada Tabel 3.

Tabel 3. Detail *Research Opportunity* pada Artikel yang terpilih.

Rank	Judul Artikel	Opportunity Type	Opportunity Confidence
1	An Anti-Sheriff Artificial Intelligence-Driven Cybersecurity Audit Model: Beyond Artificial Intelligence Adoption in Cybersecurity Auditing	Comparative Gap	50
2	Artificial Intelligence Auditability and Auditor Readiness for Auditing Artificial Intelligence Systems	Unclassified Gap Signal	41,8
3	Artificial Intelligence (AI) and Ethics	No Explicit Signal	0
4	Artificial Intelligence in Auditing: Risk, Reliability, and Auditor Judgment	No Explicit Signal	0
5	Artificial Intelligence in Accounting, Auditing, and Financial Reporting	No Explicit Signal	0

Berdasarkan Tabel 3, artikel pertama memperoleh kategori *Comparative Gap* dengan *opportunity confidence* sebesar 50, sedangkan artikel kedua dikategorikan sebagai *Unclassified Gap Signal* dengan nilai 41,8.

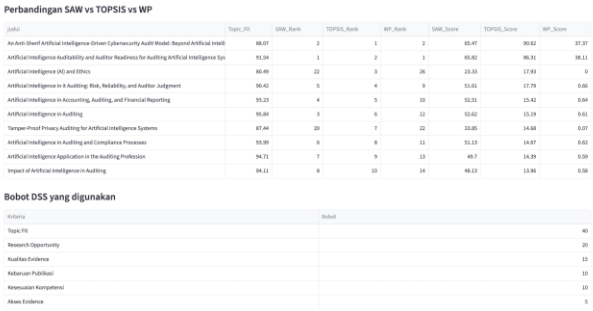
Tiga artikel lainnya tidak menunjukkan sinyal *research gap* secara eksplisitpada informasi yang dianalisis. Hasil ini menunjukan bahwa sistem dapat membantu mengidentifikasi indikasi peluang penelitian, namun tetap memerlukan klarifikasi melalui telaah isi artikel. Hal ini sejalan dengan penelitian yang dilakukan oleh (Nyanchoka et al., 2019a) yang menunjukan bahwa metode identifikasi *research gap* bervariasi dan belum memiliki pendekatan standar yang berlaku secara universal.

E. Analisis *DSS Ranking* dan Prioritas Literatur

AI-DSS *Research Assistant* menerapkan tiga metode *Multi-Criteria Decision Making* (MCDM), yaitu *Simple Additive Weighting* (SAW), *Technique for Order Preference by Similarity to Ideal Solution* (TOPSIS), dan Weighted Product (WP) untuk melakukan peringkatan artikel berdasarkan enam kriteria: *Topic Fit*, *Research Opportunity*, *Evidence Quality*, *Publication Recency*, *Competency Fit*, dan *Evidence Access.*

Ketiga metode tesebut memiliki mekanisme pemeringkatan yang berbeda sehingga memungkinkan perbandingan prioritas literatur dari berbagai perspektif. Perbedaan hasil antar metode merupakan hal yang wajar dalam MCDM, hal ini dijelaskan oleh (Li et al., 2024). Hasil

perbandingan peringkat disajikan pada Gambar 5.



Gambar 5. Hasil perbandingan peringkat

Berdasarkan Gambar 5, artikel *An Anti-Sheriff Artificial Intelligence-Driven Cybersecurity Audit Model: Beyond Artificial Intelligence Adoption in Cybersecurity Auditing"* menempati peringkat tertinggi pada TOPSIS dengan skor 90,62, sedangkan artikel *"Artificial Intelligence Auditability and Auditor Readiness for Auditing Artificial Intelligence Systems"*memperoleh peringkat pertama pada SAW dan WP, masing-masing dengan skor 65,82 dan 38,11.

Perbedaan hasil ini menunjukan bahwa prioritas artikel dipengaruhi oleh mekanisme perhitungan setiap metode MCDM. Oleh karena itu, hasil perbandingan tersebut dibandingkan untuk memberikan penilaian yang lebih komprehensif terhadap prioritas literatur. Ringkasan intepretasi hasil pemeringkatan disajikan pada Tabel 4.

Tabel 4. Interpretasi Hasil DSS Ranking Literatur

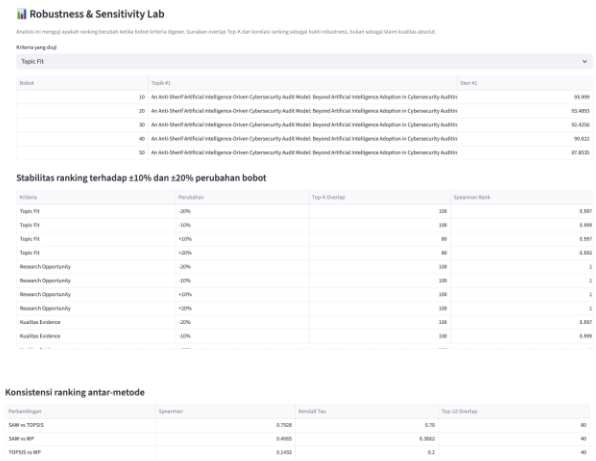
Artikel Prioritas	Hasil DSS	Interpretasi
An Anti-Sheriff Artificial Intelligence-Driven Cybersecurity Audit Model: Beyond Artificial Intelligence Adoption in Cybersecurity Auditing	Prioritas utama (TOPSIS Rank 1; SAW & WP Rank 2)	Referensi utama dengan relevansi, evidence, dan peluang penelitian terbaik.
Artificial Intelligence Auditability and Auditor Readiness for Auditing Artificial Intelligence Systems	Prioritas tinggi (SAW & WP Rank 1; TOPSIS Rank 2)	Sangat relevan untuk kajian audit AI dan kesiapan auditor.
Artificial Intelligence (AI) and Ethics	Prioritas pendukung	Mendukung aspek etika dan tata kelola AI.
Artificial Intelligence in Auditing: Risk, Reliability, and Auditor Judgment	Prioritas pendukung	Mendukung kajian risiko dan pengambilan keputusan auditor.
Artificial Intelligence in Accounting, Auditing, and Financial Reporting	Prioritas pendukung	Memberikan landasan penerapan AI dalam akuntansi dan audit.

Berdasarkan Tabel 4, pemeringkatan DSS menunjukkan bahwa prioritas literatur ditentukan oleh kombinasi berbagai kriteria, bukan hanya kesesuain topik. Pendekatan ini membantu peneliti mengevaluasi alternatif

secara sistematis berdasarkan kebutuhan penelitian. Oleh karena itu, sangatlah penting mempertimbangkan atribut penilaian dan ketidakpastian bobot dalam pemeringkatan alternatif dengan metode MCDM (Novoa-Hernandez et al., 2025). Dengan demikian, penerapan DSS dapat mendukung proses seleksi literatur yang lebih terstruktur, walaupun hasilnya tetap bergantung kepada kriteria dan bobot yang digunakan.

F. Analisis Robustness Sistem

Analisis *robustness* dilakukan untuk mengevaluasi kestabilan pemeringkatan AI-DSS *Research Assistant* terhadap perubahan bobot kriteria dan perbedaan metode DSS. Pada Gambar 6, perubahan bobot sebesar 10% dan ±20% pada kriteria *Topic Fit*, *Research Opportunity*, dan *Evidence Quality* dilaporkan tidak mengubah artikel rekomendasi utama secara drastis. Temuan ini mengindikasikan bahwa hasil rekomendasi relatif stabil terhadap bobot yang diuji yang terlihat pada Gambar 6.



Gambar 6. Hasil Robusness Analysis terhadap perubahan bobot kriteria DSS.

Selain pengujian bobot, evaluasi dilakukan dengan membandingkan konsistensi peringkat yang disajikan oleh metode SAW, TOPSIS, dan WP. Hasil perbandingan tersebut disajikan pada Tabel 5.

Tabel 5. Konsistensi Ranking antar Metode DSS

Perbandingan Metode	Nilai Konsistensi
SAW vs TOPSIS	0,79
SAW vs WP	0,46
TOPSIS vs WP	0,43

Berdasarkan Tabel 5, pasangan SAW dan TOPSIS menunjukkan nilai konsistensi tertinggi dengan nilai 0,79 dibandingkan kombinasi metode lainnya. Hal ini menunjukkan bahwa kedua metode memiliki pola evaluasi yang relatif serupa dalam menentukan prioritas literatur. Sementara itu, nilai konsistensi antara SAW-WP dan TOPSIS-WP yang lebih rendah menunjukkan adanya perbedaan karakteristik perhitungan, terutama karena metode WP menggunakan pendekatan perkalian berbobot yang lebih sensitif terhadap perubahan nilai kriteria. Secara keseluruhan, hasil *robustness analysis* menunjukkan bahwa AI-DSS Research Assistant memiliki tingkat kestabilan keputusan yang baik. Konsistensi hasil antar metode dan kestabilan terhadap perubahan bobot memperkuat bahwa rekomendasi literatur yang diberikan sistem tidak bersifat acak, tetapi dihasilkan melalui proses evaluasi multi-kriteria yang dapat dipertanggungjawabkan.

KESIMPULAN

Penelitian ini berhasil mengembangkan AI-DSS *Research Assistant*

sebagai sistem pendukung keputusan berbasis kecerdasan buatan untuk membantu proses eksplorasi dan seleksi literatur ilmiah. Hasil evaluasi *benchmark query* menunjukkan bahwa sistem mampu menghasilkan rekomendasi literatur dengan nilai rata-rata Precision@5 sebesar 0,56 dan *Mean Relevance Score* sebesar 3,44, yang menunjukkan kemampuan sistem dalam menemukan artikel yang relevan berdasarkan pendekatan *semantic retrieval*. Analisis *research opportunity* terhadap 20 artikel evidence menunjukkan bahwa seluruh artikel memiliki relevansi tinggi, dengan 2 artikel teridentifikasi memiliki opportunity signal dan nilai *signal coverage* sebesar 10%.

Hasil ini menunjukkan bahwa sistem mampu membantu peneliti menemukan indikasi *research gap* sebagai dasar pengembangan penelitian lanjutan. Pada tahap pengambilan keputusan literatur, sistem menerapkan metode SAW, TOPSIS, dan Weighted Product (WP) dengan enam kriteria evaluasi, yaitu Topic Fit (40%), Research Opportunity (20%), Evidence Quality (15%), Publication Recency (10%), Competency Fit (10%), dan Evidence Access (5%). Hasil DSS menunjukkan bahwa artikel *"An Anti-Sheriff Artificial Intelligence-Driven Cybersecurity Audit Model: Beyond Artificial Intelligence Adoption in Cybersecurity Auditing"* menjadi rekomendasi utama dengan peringkat tertinggi pada metode TOPSIS (90,62) serta tetap berada pada posisi prioritas pada metode SAW dan WP.

Validasi melalui *robustness analysis* menunjukkan bahwa perubahan bobot kriteria sebesar ±10% dan ±20% tidak menyebabkan perubahan signifikan terhadap rekomendasi utama. Selain itu, hasil konsistensi antar metode menunjukkan kesesuaian ranking antara SAW dan TOPSIS sebesar 0,79, yang menunjukkan bahwa keputusan sistem memiliki tingkat kestabilan yang baik. Secara keseluruhan, AI-DSS Research Assistant mampu mengintegrasikan proses pencarian literatur, identifikasi peluang penelitian, dan pemilihan referensi prioritas dalam satu sistem berbasis evidence. Hasil penelitian menunjukkan bahwa pendekatan kombinasi *semantic retrieval* dan DSS dapat menjadi solusi efektif untuk mendukung peneliti dalam melakukan kajian literatur secara lebih sistematis, objektif, dan efisien.

DAFTAR PUSTAKA

Asai, A., He, J., Shao, R., Shi, W., Singh, A., Chang, J. C., Lo, K., Soldaini, L., Feldman, S., D'arcy, M., Wadden, D., Latzke, M., Tian, M., Ji, P., Liu, S., Tong, H., Wu, B., Xiong, Y., Zettlemoyer, L., ... Hajishirzi, H. (2024). OpenScholar: Synthesizing scientific literature with retrieval-augmented LMs. *arXiv*. https://doi.org/10.48550/arXiv.2411.14199

Asai, A., He, J., Shao, R., Shi, W., Singh, A., Chang, J. C., Lo, K., Soldaini, L., Feldman, S., D'Arcy, M., Wadden, D., Latzke, M., Sparks, J., Hwang, J. D., Kishore, V., Tian, M., Ji, P., Liu, S., Tong, H., … Hajishirzi, H. (2026). Synthesizing scientific literature with retrieval-augmented language models. *Nature*, *650*(8103), 857–863. https://doi.org/10.1038/s41586-025-10072-4

Azzini, I., & Munda, G. (2025). Sensitivity and Robustness Analyses in Social Multi-Criteria Evaluation of Public Policies. *Journal of Multi-Criteria Decision Analysis*, *32*(1), e70006. https://doi.org/10.1002/mcda.70006

Board, C. E. (2026, July 17). *Research Rabbit: Features, Pricing, Alternatives*. CASRAI. CASRAI. https://casrai.org/guides/research-rabbit-citation-mapping-literature-discovery-tool

Bolaños, F., Salatino, A., Osborne, F., & Motta, E. (2024a). Artificial intelligence for literature reviews: Opportunities and challenges. *Artificial Intelligence Review*, *57*(10), 259. https://doi.org/10.1007/s10462-024-10902-3

Bolaños, F., Salatino, A., Osborne, F., & Motta, E. (2024b). Artificial intelligence for literature reviews: Opportunities and challenges. *Artificial Intelligence Review*, *57*(10), 259. https://doi.org/10.1007/s10462-024-10902-3

Ciardiello, F., & Genovese, A. (2023a). A comparison between TOPSIS and SAW methods. *Annals of Operations Research*, *325*(2), 967–994. https://doi.org/10.1007/s10479-023-05339-w

Ciardiello, F., & Genovese, A. (2023b). A comparison between TOPSIS and SAW methods. *Annals of Operations Research*, *325*(2), 967–994. https://doi.org/10.1007/s10479-023-05339-w

Delgado-Chaves, F. M., Jennings, M. J., Atalaia, A., Wolff, J., Horvath, R., Mamdouh, Z. M., Baumbach, J., & Baumbach, L. (2025). Transforming literature screening: The emerging role of large language models in systematic reviews. *Proceedings of the National Academy of Sciences*, *122*(2), e2411962122. https://doi.org/10.1073/pnas.2411962122

Dewidar, O., Elganga, M., Ringsten, M., Langendam, M., Saif-Ur-Rahman, K. M., Bala, M. M., Arienti, C., Bognanni, A., Darzi, A. J., Jemioło, P., Krawczyk, S., Lazzarini, S. G., Marques-Cruz, M., Meerpohl, J. J., Mollel, G. J., Parmelli, E., Robinson, K. A., Sousa-Pinto, B., Siemens, W., … Schünemann, H. J. (2026a). Frameworks to identify research gaps, frame research needs, and derive research priorities from evidence syntheses: A scoping review. *Journal of Clinical Epidemiology*, *197*. https://doi.org/10.1016/j.jclinepi.2026.112342

Dewidar, O., Elganga, M., Ringsten, M., Langendam, M., Saif-Ur-Rahman, K. M., Bala, M. M., Arienti, C., Bognanni, A., Darzi, A. J., Jemioło, P., Krawczyk, S., Lazzarini, S. G., Marques-Cruz, M., Meerpohl, J. J., Mollel, G. J., Parmelli, E., Robinson, K. A., Sousa-Pinto, B., Siemens, W., … Schünemann, H. J. (2026b). Frameworks to identify research gaps, frame research needs, and derive research priorities from evidence syntheses: A scoping review. *Journal of Clinical Epidemiology*, *197*. https://doi.org/10.1016/j.jclinepi.2026.112342

Greco, S., Słowiński, R., & Wallenius, J. (2025). Fifty years of multiple criteria decision analysis: From classical methods to robust ordinal regression. *European Journal of Operational Research*, *323*(2), 351–377.

https://doi.org/10.1016/j.ejor.2024.07.038

Gusenbauer, M., & Gauster, S. P. (2025a). How to search for literature in systematic reviews and meta-analyses: A comprehensive step-by-step guide. *Technological Forecasting and Social Change*, *212*, 123833. https://doi.org/10.1016/j.techfore.2024.123833

Gusenbauer, M., & Gauster, S. P. (2025b). How to search for literature in systematic reviews and meta-analyses: A comprehensive step-by-step guide. *Technological Forecasting and Social Change*, *212*, 123833. https://doi.org/10.1016/j.techfore.2024.123833

Hoekstra, D., & Lhachimi, S. K. (2025). Optimizing Research Impact: A Toolkit for Stakeholder-Driven Prioritization of Systematic Review Topics. *Cochrane Evidence Synthesis and Methods*, *3*(5), e70039. https://doi.org/10.1002/cesm.70039

Introducing Elicit Systematic Review—Elicit. (2025, February 20). https://elicit.com/blog/systematic-review

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. *Research Synthesis Methods*, *15*(4), 616–626. https://doi.org/10.1002/jrsm.1715

Kinney, R., Anastasiades, C., Authur, R., Beltagy, I., Bragg, J., Buraczynski, A., Cachola, I., Candra, S., Chandrasekhar, Y., Cohan, A., Crawford, M., Downey, D., Dunkelberger, J., Etzioni, O., Evans, R., Feldman, S., Gorney, J., Graham, D., Hu, F., … Weld, D. S. (2025). *The Semantic Scholar Open Data Platform* (arXiv:2301.10140). arXiv. https://doi.org/10.48550/arXiv.2301.10140

Lagisz, M., Mizuno, A., Morrison, K., Pollo, P., Ricolfi, L., Yang, Y., & Nakagawa, S. (2026). Using Elicit AI research assistant for data extraction in systematic reviews: A feasibility study across environmental and life sciences. *Research Synthesis Methods*, 1–18. https://doi.org/10.1017/rsm.2026.10080

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, *33*, 9459–9474.

Li, Y., He, X., Martínez, L., Zhang, J., Wang, D., & Liu, X. A. (2024a). Comparative analysis of three categories of multi-criteria decision-making methods. *Expert Systems with Applications*, *238*, 121824. https://doi.org/10.1016/j.eswa.2023.121824

Li, Y., He, X., Martínez, L., Zhang, J., Wang, D., & Liu, X. A. (2024b). Comparative analysis of three categories of multi-criteria decision-making methods. *Expert Systems with Applications*, *238*, 121824. https://doi.org/10.1016/j.eswa.2023.121824

Mitrov, G., Stanoev, B., Trajkovik, V., Risteska Stojkoska, B., Basnarkov, L., Lameski, P., Kampel, M., &

Zdravevski, E. (2026). Optimizing document retrieval using massive text embeddings and LLM prompt engineering. *Systematic Reviews*, *15*(1), 173. https://doi.org/10.1186/s13643-026-03155-4

Novoa-Hernández, P., Pelta, D. A., & Corona, C. C. (2025). Helping to Choose a Robust Alternative: A Sensitivity Analysis and a Software Tool for Multi-Criteria Decision-Making. *Expert Systems*, *42*(4), e70031. https://doi.org/10.1111/exsy.70031

Nyanchoka, L., Tudur-Smith, C., Thu, V. N., Iversen, V., Tricco, A. C., & Porcher, R. (2019a). A scoping review describes methods used to identify, prioritize and display gaps in health research. *Journal of Clinical Epidemiology*, *109*, 99–110. https://doi.org/10.1016/j.jclinepi.2019.01.005

Nyanchoka, L., Tudur-Smith, C., Thu, V. N., Iversen, V., Tricco, A. C., & Porcher, R. (2019b). A scoping review describes methods used to identify, prioritize and display gaps in health research. *Journal of Clinical Epidemiology*, *109*, 99–110. https://doi.org/10.1016/j.jclinepi.2019.01.005

Peikos, G., & Pasi, G. (2024a). A systematic review of multidimensional relevance estimation in information retrieval. *WIREs Data Mining and Knowledge Discovery*, *14*(5), e1541. https://doi.org/10.1002/widm.1541

Peikos, G., & Pasi, G. (2024b). A systematic review of multidimensional relevance estimation in information retrieval. *WIREs Data Mining and Knowledge Discovery*, *14*(5), e1541. https://doi.org/10.1002/widm.1541

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, *6*, e4375. https://doi.org/10.7717/peerj.4375

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982–3992). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410

Thakur, N., Reimers, N., Daxenberger, J., & Gurevych, I. (2021). Augmented SBERT: Data Augmentation Method for Improving Bi-Encoders for Pairwise Sentence Scoring Tasks. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, & Y. Zhou (Eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 296–310). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.28