

A SYSTEMATIC LITERATURE REVIEW OF THE USE OF RETRIEVAL-AUGMENTED GENERATION FOR DATA DESA PRESISI-BASED CHATBOTS

KAJIAN LITERATUR SISTEMATIS PENGGUNAAN *RETRIEVAL-AUGMENTED GENERATION* UNTUK *CHATBOT* BERBASIS DATA DESA PRESISI

Ikhlasul Amal^{1†}, Irman Hermadi², and Yani Nurhadryani³

^{1,2,3}Computer Science Study Program, IPB University, Indonesia
†corresponding author: ikhlaslamal@apps.ipb.ac.id

Copyright © 2026 Ikhlasul Amal, Irman Hermadi, and Yani Nurhadryani. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Data Desa Presisi (DDP) is a village data approach that integrates spatial, census-based, and participatory methods to produce contextual and detailed village information. However, accessing DDP-based information remains challenging because the data may consist of village monographs, tables, images, maps, structured census data, regional metadata, and socio-economic indicators. This study presents a systematic literature review on the use of Retrieval-Augmented Generation (RAG) for Data Desa Presisi-based chatbots. The review aims to identify research trends, literature gaps, and conceptual directions for developing chatbots that can answer questions based on heterogeneous village data. Literature was collected through Google Scholar using Publish or Perish and managed using Zotero. A PRISMA-inspired selection flow was adapted to screen 581 initial records, resulting in 501 deduplicated candidates and 24 selected studies for synthesis. The reviewed literature was grouped into five themes: Data Desa Presisi, digital village and smart village, AI/chatbots for villages, cross-domain RAG chatbots, and Hybrid RAG or advanced retrieval. The findings show that RAG has been widely applied in education, healthcare, law, customer support, and document-based question answering. However, within the selected literature, no study was identified that specifically integrates DDP, village monographs, structured census data, and RAG-based chatbots. This study recommends Hybrid RAG as a conceptual approach for future DDP-based chatbot development.

Keywords: Chatbot, Data Desa Presisi, Hybrid RAG, Retrieval-Augmented Generation, Systematic Literature Review.

Abstrak

Data Desa Presisi (DDP) merupakan pendekatan data desa yang mengintegrasikan metode spasial, sensus, dan partisipatif untuk menghasilkan informasi desa yang kontekstual dan rinci. Namun, akses terhadap informasi berbasis DDP masih menjadi tantangan karena data dapat berbentuk monografi desa, tabel, gambar, peta, data sensus terstruktur, metadata wilayah, serta indikator sosial-ekonomi. Penelitian ini melakukan kajian literatur sistematis mengenai penggunaan *Retrieval-Augmented Generation* (RAG) untuk *chatbot* berbasis Data Desa Presisi. Kajian ini bertujuan mengidentifikasi tren penelitian, kesenjangan literatur, dan arah konseptual pengembangan *chatbot* yang mampu menjawab pertanyaan berbasis data desa heterogen. Literatur diperoleh melalui Google Scholar menggunakan Publish or Perish dan dikelola dengan Zotero. Alur seleksi yang terinspirasi PRISMA digunakan untuk menyaring 581 rekaman awal menjadi 501 kandidat setelah deduplikasi dan 24 artikel terpilih untuk sintesis. Literatur dikelompokkan ke dalam lima tema, yaitu Data Desa Presisi, desa digital dan *smart village*, *AI/chatbot* desa, RAG *chatbot* lintas domain, serta *Hybrid RAG* atau *retrieval* lanjutan. Hasil kajian menunjukkan bahwa RAG telah banyak diterapkan pada pendidikan, kesehatan, hukum, layanan pelanggan, dan sistem tanya jawab berbasis dokumen. Namun, berdasarkan literatur terpilih, belum ditemukan studi yang secara spesifik mengintegrasikan DDP, monografi desa, data sensus terstruktur, dan *chatbot* berbasis RAG. Kajian ini merekomendasikan *Hybrid RAG* sebagai pendekatan konseptual untuk pengembangan *chatbot* berbasis DDP.

Kata Kunci: *Chatbot*, Data Desa Presisi, *Hybrid RAG*, Kajian Literatur Sistematis, *Retrieval-Augmented Generation*.

1. Pendahuluan

Data desa yang akurat dan kontekstual diperlukan untuk mendukung perencanaan serta pengambilan keputusan pada tingkat lokal. Dalam konteks desa digital, pemanfaatan teknologi informasi tidak hanya diarahkan pada digitalisasi administrasi, tetapi juga pada penguatan pengelolaan pengetahuan, distribusi informasi, dan adopsi inovasi digital desa (Nurhadryani *et al.*, 2025).

Salah satu pendekatan yang relevan dalam konteks Indonesia adalah Data Desa Presisi (DDP). DDP mengintegrasikan pendekatan spasial, sensus, dan partisipatif sehingga data yang dihasilkan lebih dekat dengan realitas masyarakat (Sjaf *et al.*, 2020; Sjaf *et al.*, 2022). Data yang dihasilkan DDP tidak hanya berupa data numerik, tetapi juga mencakup data spasial, data partisipatif, dan dokumentasi naratif dalam bentuk monografi desa. Oleh karena itu, DDP dapat dipahami sebagai domain data yang heterogen karena memuat data terstruktur, semi-terstruktur, dan tidak terstruktur.

Perkembangan *Large Language Models* (LLM) membuka peluang untuk membangun *chatbot* yang dapat menjawab pertanyaan pengguna dalam bahasa alami. Namun, LLM memiliki keterbatasan ketika digunakan pada domain khusus karena jawaban yang dihasilkan perlu ditopang oleh sumber eksternal yang dapat diverifikasi. *Retrieval-Augmented Generation* (RAG) menjadi pendekatan yang relevan karena menghubungkan LLM dengan sumber pengetahuan eksternal sehingga jawaban dapat dihasilkan berdasarkan konteks yang ditelusuri dari dokumen atau basis data tertentu (Pokhrel et al., 2025; Swacha dan Gracel, 2025).

RAG telah digunakan pada berbagai domain, seperti pendidikan, kesehatan, hukum, dan layanan informasi (Pokhrel et al., 2025; Swacha dan Gracel, 2025; Xi et al., 2025; Aljohani dan Alsanoosy, 2026). Pada bidang hukum agraria, Sururi dan Hermadi (2025) mengembangkan *chatbot* berbasis RAG untuk menjawab pertanyaan berdasarkan sumber hukum agraria. Studi-studi tersebut menunjukkan bahwa RAG relevan untuk domain yang membutuhkan jawaban berbasis bukti dan sumber tertentu.

Berdasarkan kondisi tersebut, artikel ini melakukan kajian literatur sistematis mengenai penggunaan RAG untuk *chatbot* berbasis Data Desa Presisi. Kajian ini memetakan literatur terkait DDP, desa digital, AI/*chatbot* desa, RAG, dan *Hybrid* RAG untuk mengidentifikasi kesenjangan serta merumuskan arah konseptual pengembangan *chatbot* berbasis DDP.

2. Metodologi

Penelitian ini menggunakan pendekatan kajian literatur sistematis untuk memetakan hubungan antara *Large Language Models* (LLM), *Retrieval-Augmented Generation* (RAG), *chatbot*, *smart village*, dan Data Desa Presisi (DDP). Kajian ini tidak bertujuan mengevaluasi performa sistem, tetapi mengidentifikasi tren penelitian, kesenjangan literatur, dan arah konseptual pengembangan *chatbot* berbasis DDP. Proses seleksi literatur mengadaptasi alur PRISMA untuk melaporkan tahapan identifikasi, deduplikasi, penyaringan, penilaian kelayakan, dan sintesis akhir secara transparan (Page et al., 2021). Kajian ini tidak mengklaim penerapan PRISMA secara penuh karena pencarian literatur menggunakan Google Scholar melalui Publish or Perish dan proses telaah dilakukan oleh penulis utama.

Kajian ini disusun berdasarkan empat pertanyaan penelitian: (1) Pada domain apa saja RAG dan *Hybrid* RAG telah diterapkan; (2) Aspek apa yang dominan dikaji dalam penelitian *smart village* atau desa digital; (3) Apakah penelitian desa digital telah mengadopsi *chatbot* berbasis RAG; dan (4) Bagaimana arah konseptual *Hybrid* RAG yang sesuai untuk menangani heterogenitas data DDP. Keempat pertanyaan ini digunakan untuk membatasi proses pencarian, penyaringan, dan sintesis literatur.

2.1 Bahan dan Data

Data dalam penelitian ini berupa metadata dan artikel ilmiah yang diperoleh melalui Google Scholar menggunakan Publish or Perish. Metadata hasil pencarian dikelola menggunakan Zotero untuk penghapusan duplikasi dan manajemen referensi. Literatur yang dikaji mencakup artikel jurnal, prosiding, tesis/disertasi, buku ilmiah, dan *preprint* akademik yang relevan dengan RAG, *chatbot*, *smart village*, sistem informasi desa, dan Data Desa Presisi. Rentang publikasi yang digunakan adalah 2019–2026, kecuali referensi fondasional yang diperlukan untuk menjelaskan konsep dasar *e-governance*, Data Desa Presisi, dan RAG.

Kata kunci pencarian dikelompokkan ke dalam empat kelompok, yaitu RAG dan *chatbot*, *Hybrid* RAG, desa digital, serta Data Desa Presisi. Ringkasan kelompok kata kunci ditampilkan pada Tabel 1.

Tabel 1: Kelompok kata kunci pencarian literatur.

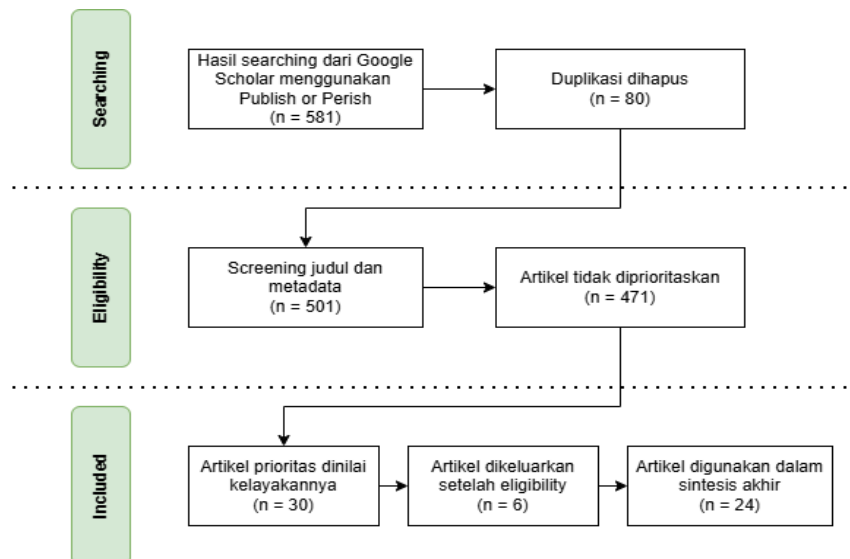
Kelompok	Kata Kunci Pencarian	Tujuan
RAG dan <i>chatbot</i>	“retrieval augmented generation” AND <i>chatbot</i> ; “large language model” AND <i>chatbot</i>	Menelusuri studi <i>chatbot</i> berbasis RAG.
<i>Hybrid</i> RAG	“hybrid retrieval” AND “retrieval augmented generation”; “GraphRAG”; “parent document retrieval”	Menelusuri teknik <i>retrieval</i> lanjutan dan <i>Hybrid</i> RAG.
Desa Digital	“smart village” AND “artificial intelligence”; “village information system” AND <i>chatbot</i>	Menelusuri AI, <i>chatbot</i> , dan sistem informasi pada konteks desa.
Data Desa Presisi	“Data Desa Presisi”; “Data Desa Presisi” AND <i>chatbot</i> ; “Data Desa Presisi” AND “retrieval augmented generation”	Menelusuri literatur DDP dan keterkaitannya dengan <i>chatbot</i> /RAG.

Literatur dimasukkan apabila membahas LLM, RAG, *Hybrid* RAG, *chatbot*, *question answering*, *smart village*, sistem informasi desa, *e-governance*, atau Data Desa Presisi (DDP). Literatur diprioritaskan untuk sintesis akhir apabila memiliki keterkaitan langsung dengan pertanyaan penelitian, menjelaskan metode atau pendekatan yang digunakan, serta memiliki kontribusi yang relevan terhadap *chatbot*, data heterogen, desa digital, atau DDP. Literatur dikeluarkan apabila tidak berkaitan dengan AI, *chatbot*, RAG, sistem informasi, desa digital, atau DDP. Literatur juga dikeluarkan apabila hanya berupa berita atau opini populer, merupakan duplikasi, atau tidak memiliki informasi metodologis dan kontribusi yang dapat diidentifikasi.

2.2 Metode Penelitian

Hasil pencarian awal menghasilkan 581 rekaman. Setelah penghapusan duplikasi menggunakan Zotero, diperoleh 501 kandidat literatur. Penyaringan awal terhadap 501 kandidat dilakukan berdasarkan judul, metadata, dan relevansi terhadap empat kelompok kata kunci, yaitu RAG dan *chatbot*, *Hybrid* RAG, desa digital, serta Data Desa Presisi (DDP). Artikel tidak diprioritaskan apabila tidak berkaitan langsung dengan AI, *chatbot*, RAG, sistem informasi desa, desa digital, atau DDP; hanya berupa berita, opini populer, atau dokumen non-akademik, tidak memiliki abstrak atau informasi metodologis yang dapat diidentifikasi, atau memiliki cakupan yang terlalu jauh dari pertanyaan penelitian.

Dari tahap penyaringan awal, 30 artikel prioritas diperiksa lebih lanjut melalui abstrak, pendahuluan, metode, hasil/pembahasan, dan kesimpulan. Penilaian kelayakan dilakukan dengan mempertimbangkan empat kriteria, yaitu kesesuaian dengan pertanyaan penelitian, kejelasan kontribusi, ketersediaan informasi metode atau pendekatan, serta relevansi terhadap karakteristik data heterogen atau konteks desa. Sebanyak 6 artikel dikeluarkan karena memiliki relevansi lebih rendah, bersifat redundan dengan artikel lain, atau tidak cukup mendukung sintesis terhadap DDP, desa digital, *chatbot*, RAG, dan *Hybrid* RAG. Dengan demikian, 24 artikel digunakan dalam sintesis akhir. Alur seleksi literatur ditampilkan pada Gambar 1.



Gambar 1: Alur seleksi literatur adaptasi PRISMA.

Informasi dari 24 literatur terpilih diekstraksi menggunakan *spreadsheet* dengan komponen penulis, tahun, fokus literatur, metode, dan relevansi terhadap kajian. Pengelompokan artikel ke dalam lima tema dilakukan melalui kombinasi pendekatan deduktif dan induktif. Secara deduktif, tema awal diturunkan dari pertanyaan penelitian, yaitu Data Desa Presisi, desa

digital, *AI/chatbot* desa, RAG *chatbot* lintas domain, serta *Hybrid* RAG dan *retrieval* lanjutan. Secara induktif, setiap artikel dikodekan berdasarkan fokus utama yang muncul pada judul, abstrak, metode, dan kontribusi penelitian. Artikel kemudian ditempatkan pada tema yang paling dominan, yaitu DDP, desa digital, *AI/chatbot* desa, RAG *chatbot* lintas domain, atau *Hybrid* RAG dan *retrieval* lanjutan, berdasarkan fokus utama dan kontribusi masing-masing artikel. Apabila sebuah artikel memiliki lebih dari satu fokus, artikel ditempatkan pada tema yang paling dominan berdasarkan kontribusi utamanya. Hasil kategorisasi tersebut diringkas pada Tabel 2.

Tabel 2: Ringkasan ekstraksi informasi dari literatur terpilih.

Kelompok Literatur	Jumlah	Contoh Referensi	Makna Sintesis
Data Desa Presisi	6	Sjaf <i>et al.</i> (2020, 2024); Daryanto <i>et al.</i> (2022)	DDP merupakan domain data spasial, sensus, partisipatif, dan kontekstual.
Desa digital dan <i>smart village</i>	3	Trapp dan Hess (2020); Lindawaty dan Sartika (2025)	Kajian desa digital masih banyak berfokus pada sistem informasi dan layanan digital.
AI dan <i>chatbot</i> desa	3	Arafat <i>et al.</i> (2025); Saputro <i>et al.</i> (2025)	<i>AI/chatbot</i> desa mulai berkembang, tetapi masih dominan pada administrasi dan layanan publik.
RAG <i>chatbot</i> lintas domain	5	Sukhwal (2024); Pokhrel <i>et al.</i> (2025); Swacha dan Gracel (2025)	RAG efektif untuk sistem tanya jawab berbasis dokumen dan domain spesifik.
<i>Hybrid</i> RAG dan <i>retrieval</i> lanjutan	7	Omrani <i>et al.</i> (2024); Nguyen (2025); Wan <i>et al.</i> (2025); Lee dan Lim (2026)	<i>Hybrid</i> RAG relevan untuk data heterogen yang membutuhkan <i>retrieval</i> dokumen, kueri terstruktur, dan penyaringan metadata.

Sintesis dilakukan secara tematik, bukan melalui perbandingan metrik kuantitatif antarstudi. Tema sintesis meliputi distribusi dan karakteristik literatur, penerapan RAG/*Hybrid* RAG pada berbagai domain, perkembangan desa digital dan *chatbot*, serta kesenjangan dan arah konseptual *chatbot* berbasis DDP.

3. Hasil dan Pembahasan

3.1 Distribusi Fokus Literatur

Berdasarkan hasil ekstraksi pada Tabel 2, literatur terpilih dapat dikelompokkan ke dalam lima fokus utama, yaitu Data Desa Presisi, desa digital atau *smart village*, AI/chatbot desa, RAG chatbot lintas domain, serta Hybrid RAG dan *retrieval* lanjutan. Fokus Hybrid RAG dan *retrieval* lanjutan menjadi kategori dengan jumlah studi terbanyak, yaitu tujuh artikel. Fokus Data Desa Presisi terdiri dari enam artikel, RAG chatbot lintas domain terdiri dari lima artikel, sedangkan desa digital/*smart village* dan AI/chatbot desa masing-masing terdiri dari tiga artikel. Distribusi ini menunjukkan bahwa kajian yang dilakukan tidak hanya meninjau aspek teknologi RAG, tetapi juga menghubungkannya dengan konteks data desa dan transformasi digital desa.

3.2 Penerapan RAG dan Hybrid RAG pada Berbagai Domain

Subbagian ini menjawab pertanyaan penelitian (1), yaitu domain apa saja yang telah menerapkan RAG dan Hybrid RAG. Hasil kajian menunjukkan bahwa RAG telah banyak diterapkan pada domain pendidikan, layanan akademik, layanan pelanggan, kesehatan, hukum, kueri situs web, dan sistem tanya jawab berbasis dokumen. Pada domain pendidikan dan layanan akademik, RAG digunakan untuk meningkatkan akses terhadap dokumen resmi dan mengurangi risiko jawaban yang tidak sesuai konteks (Pokhrel et al., 2025; Swacha dan Gracel, 2025). Pada domain layanan pelanggan dan situs web, RAG digunakan untuk menghubungkan chatbot dengan basis pengetahuan eksternal sehingga respons yang diberikan lebih relevan dengan kebutuhan pengguna (Radhakrishnan et al., 2024; Sukhwal, 2024; Farnaz dan Huyck, 2025). Pada domain kesehatan dan hukum, RAG menjadi penting karena jawaban yang diberikan harus berbasis sumber dan dapat diverifikasi (Xi et al., 2025; Aljohani dan Alsanoosy, 2026).

Kajian ini juga menunjukkan bahwa Hybrid RAG dan *retrieval* lanjutan digunakan pada domain yang memiliki struktur data lebih kompleks. Beberapa studi menggabungkan *sparse retrieval*, *dense retrieval*, pemeringkatan ulang, *parent-document retrieval*, GraphRAG, *Reciprocal Rank Fusion* (RRF), dan penalaran berbasis struktur (Omrani et al., 2024; Mudet dan Bakkali, 2025; Nguyen, 2025; Wan et al., 2025; Lee dan Lim, 2026). Temuan ini menjawab pertanyaan penelitian (1) bahwa RAG tidak hanya berkembang pada domain umum berbasis dokumen, tetapi juga mulai diarahkan pada domain yang membutuhkan penggabungan beberapa strategi pengambilan informasi. Hal ini relevan dengan DDP karena data DDP tidak hanya berupa teks naratif, tetapi juga mencakup tabel, data sensus terstruktur, metadata wilayah, dan konteks spasial.

3.3 Desa Digital, AI Desa, dan Chatbot

Subbagian ini menjawab pertanyaan penelitian (2), yaitu aspek dominan dalam penelitian *smart village* atau desa digital. Hasil kajian menunjukkan bahwa penelitian desa digital masih banyak berfokus pada sistem informasi desa, layanan administrasi, otomasi surat, ekosistem digital pedesaan, dan tata kelola digital (Trapp dan Hess, 2020; Siman dan Wiratama, 2023; Lindawaty dan Sartika, 2025). Studi tentang desa digital menunjukkan bahwa teknologi berperan dalam mendukung pelayanan publik dan pengelolaan informasi desa, tetapi belum banyak membahas integrasi data presisi yang bersumber dari monografi, sensus, dan data spasial.

Pertanyaan penelitian (3) menanyakan apakah penelitian desa digital telah mengadopsi *chatbot* berbasis RAG. Hasil kajian menunjukkan bahwa AI dan *chatbot* mulai digunakan pada konteks desa, terutama untuk layanan administrasi, sistem informasi desa, dan otomasi dokumen (Arafat *et al.*, 2025; Saputro *et al.*, 2025; Priambodo *et al.*, 2026). Namun, *chatbot* desa yang ditemukan dalam literatur terpilih umumnya masih menggunakan data situs web desa atau data layanan administratif. Belum ditemukan studi dalam literatur terpilih yang secara spesifik menggunakan Data Desa Presisi (DDP) sebagai basis pengetahuan *chatbot*. Dengan demikian, pertanyaan penelitian (2) dan (3) menunjukkan bahwa meskipun desa digital dan AI desa telah berkembang, pemanfaatan *chatbot* berbasis RAG untuk mengakses data DDP masih menjadi ruang penelitian yang belum banyak dikaji.

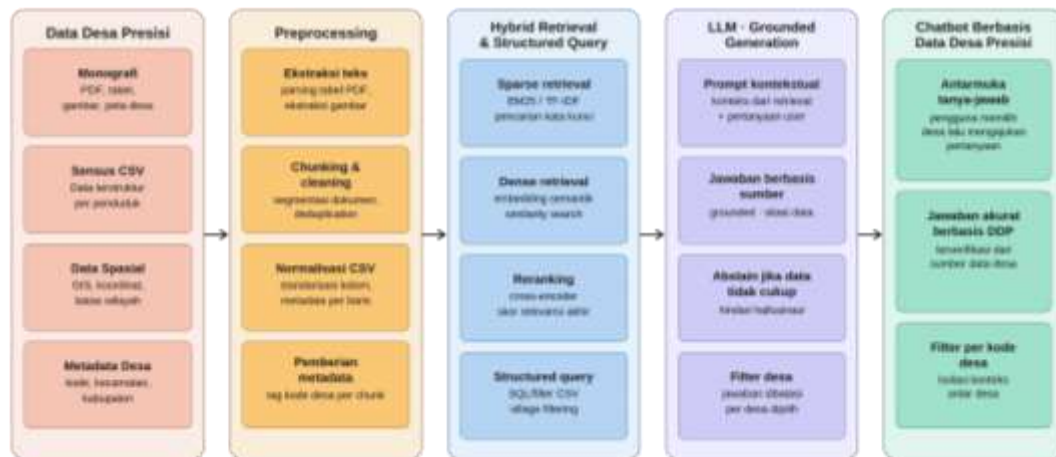
3.4 Kesenjangan dan Arah Konseptual *Chatbot* Berbasis DDP

Pertanyaan penelitian (4) menanyakan arah konseptual *Hybrid* RAG yang sesuai untuk menangani heterogenitas data DDP. Berdasarkan sintesis literatur, kesenjangan utama dalam kajian ini adalah belum ditemukannya studi dalam literatur terpilih yang secara spesifik mengintegrasikan Data Desa Presisi, monografi desa, data sensus terstruktur, dan *chatbot* berbasis RAG/LLM. Studi AI desa umumnya masih berfokus pada layanan administrasi dan sistem informasi, sedangkan studi RAG umumnya berfokus pada dokumen atau domain khusus seperti pendidikan, kesehatan, hukum, dan layanan pelanggan.

Data Desa Presisi memiliki struktur data yang lebih kaya karena menggabungkan monografi naratif, tabel, gambar, peta, data sensus terstruktur, metadata wilayah, serta indikator sosial-ekonomi (Sjaf *et al.*, 2020; Muhammad *et al.*, 2021; Sjaf *et al.*, 2021; Daryanto *et al.*, 2022; Sjaf *et al.*, 2022; Sjaf *et al.*, 2024). Oleh karena itu, *chatbot* berbasis DDP tidak cukup dirancang sebagai *chatbot* LLM biasa atau RAG dokumen tunggal. Dalam kajian ini, *hybrid retrieval* dipahami sebagai strategi pengambilan konteks yang menggabungkan lebih dari satu metode pencarian, seperti *sparse retrieval*, *dense retrieval*, pemeringkatan ulang, atau *graph retrieval*. Sementara itu, *Hybrid* RAG dipahami sebagai konfigurasi sistem yang lebih luas karena menggabungkan *retrieval* dokumen, kueri terstruktur, penyaringan metadata, dan generasi jawaban berbasis LLM.

Berdasarkan kesenjangan tersebut, arah konseptual *chatbot* berbasis DDP perlu mencakup beberapa lapisan utama, yaitu lapisan data, praproses, *hybrid retrieval* dan kueri terstruktur, generasi jawaban berbasis

konteks menggunakan LLM, serta lapisan aplikasi berbasis *chatbot*. Lapisan data mencakup monografi desa, data sensus CSV, data spasial, dan metadata desa. Lapisan praproses mencakup ekstraksi teks, pembersihan data, *chunking* dokumen, normalisasi data sensus, dan pemberian metadata desa. Lapisan *retrieval* menggabungkan pencarian dokumen, kueri terstruktur, dan penyaringan metadata berbasis desa. Lapisan generasi menggunakan LLM untuk menghasilkan jawaban berdasarkan konteks yang ditelusuri. Kerangka konseptual *chatbot Hybrid RAG* berbasis Data Desa Presisi ditampilkan pada Gambar 2.



Gambar 2: Kerangka konseptual *chatbot Hybrid RAG* berbasis Data Desa Presisi.

Dengan demikian, hasil kajian menjawab pertanyaan penelitian (4) bahwa pendekatan *Hybrid RAG* lebih sesuai untuk DDP karena mampu menggabungkan pengambilan informasi dari dokumen naratif dan kueri terstruktur terhadap data sensus. Namun, kesenjangan yang dimaksud dalam kajian ini dibatasi pada literatur yang berhasil ditemukan dan disintesis melalui proses pencarian yang digunakan, bukan sebagai klaim mutlak terhadap seluruh penelitian yang pernah dilakukan.

4. Simpulan dan Saran

Kajian literatur sistematis ini menunjukkan bahwa RAG telah banyak diterapkan pada domain pendidikan, layanan akademik, layanan pelanggan, kesehatan, hukum, kueri situs web, dan sistem tanya jawab berbasis dokumen. Sementara itu, penelitian desa digital dan AI desa masih banyak berfokus pada sistem informasi desa, layanan administrasi, otomasi surat, situs web desa, dan tata kelola digital. Berdasarkan literatur terpilih yang dianalisis, belum ditemukan studi yang secara spesifik mengintegrasikan Data Desa Presisi, monografi desa, data sensus terstruktur, dan *chatbot* berbasis RAG/LLM.

Data Desa Presisi memiliki karakteristik kompleks karena menggabungkan data spasial, sensus, partisipatif, monografi desa, tabel, gambar, peta, metadata wilayah, serta data sensus terstruktur.

Karakteristik ini menunjukkan bahwa *chatbot* berbasis DDP tidak cukup dirancang sebagai *chatbot* LLM biasa atau RAG dokumen tunggal. Kajian ini merekomendasikan pendekatan *Hybrid* RAG yang menggabungkan *retrieval* dokumen, kueri terstruktur, pemeringkatan ulang, penyaringan metadata berbasis desa, dan generasi jawaban berbasis konteks.

Penelitian selanjutnya perlu membandingkan *pipeline* RAG standar atau *Naive* RAG dengan *pipeline Hybrid* RAG pada data DDP. RAG standar dapat menjadi pembanding dasar untuk pengambilan konteks dari monografi desa, sedangkan *Hybrid* RAG dapat menggabungkan *hybrid retrieval* dan kueri terstruktur terhadap monografi PDF, tabel, gambar, serta data sensus CSV. Evaluasi perlu mencakup akurasi *retrieval*, kesetiaan jawaban terhadap sumber, kemampuan menjawab pertanyaan numerik, penyaringan berbasis desa, efisiensi waktu respons, serta aspek privasi dan pembatasan akses data.

Daftar Pustaka

- Aljohani B, Alsanoosy T. 2026. Enhancing Medical Question Answering with LLMs via a Hybrid Retrieval-Augmented Generation Framework. *Information (Switzerland)*. 17(2). doi:10.3390/info17020133.
- Arafat A, et al. 2025. Designing an AI-Based Village Information System Using Research and Development Approach for Public Governance Modernization in Popalia Village. *Jurnal Teknik Informatika (Jutif)*. 6(5):5251–5269. doi:10.52436/1.jutif.2025.6.5.5257.
- Daryanto TJ, et al. 2022. Integrasi Data Melalui Pemetaan dengan Metode Drone Participatory Mapping (DPM) dan Pendataan Sensus Sosial dalam Proses Digitalisasi Desa Songbledeg. *Jurnal Ilmu Sosial, Agama, Budaya, dan Terapan*. 2(2):53–60.
- Farnaz A, Huyck C. 2025. TravQuery: A Customer Support Chatbot Based on Retrieval Augmented Generation (RAG). SPRINGER NATURE.
- Lee H, Lim S. 2026. Hybrid Retrieval-Augmented Generation: Semantic and Structural Integration for Large Language Model Reasoning. *Applied Sciences (Switzerland)*. 16(5). doi:10.3390/app16052244.
- Lindawaty DS, Sartika I. 2025. Pemanfaatan Artificial Intelligence (AI) dalam Pengembangan Desa Digital Utilization of Artificial Intelligence (AI) in Digital Village Development. 16(2). doi:10.22212/jp.v16i2.5240.
- Mudet A, Bakkali S. 2025 Des 14. Hybrid Retrieval-Augmented Generation for Robust Multilingual Document Question Answering.
- Muhammad B, et al. 2021. Partisipasi Pemuda Dalam Membangun Data Desa Presisi di Kelurahan Situ Gede, Kota Bogor. *Jurnal Pusat Inovasi Masyarakat*. 3(1).
- Nguyen TH. 2025. Optimizing Retrieval Augmented Generation Chatbots: A

Comparative Analysis. Technische Hochschule Mittelhessen.

- Nurhadryani Y, Nuryantika F, Hermadi I, Ahmad HF. 2025. Pengembangan Modul Front-End KMS Desa Digital untuk Meningkatkan Adopsi Inovasi Digital pada Desa di Indonesia. *Jurnal Ilmu Komputer & Agri-Informatika*. 12:72–78.
- Omrani P, Hosseini A, Hooshanfar K, Ebrahimian Z, Toosi R, Ali Akhaee M. 2024. Hybrid Retrieval-Augmented Generation Approach for LLMs Query Response Enhancement. Di dalam: *2024 10th International Conference on Web Research (ICWR)*. IEEE. hal. 22–26.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *J Clin Epidemiol*. 134:178–189. doi:10.1016/j.jclinepi.2021.03.001.
- Pokhrel S, Bina KC, Shah P. 2025. A Practical Application of Retrieval-Augmented Generation for Website-Based Chatbots: Combining Web Scraping, Vectorization, and Semantic Search. *Journal of Trends in Computer Science and Smart Technology*. 6(4):424–442. doi:10.36548/jtcsst.2024.4.007.
- Priambodo G, Dewi RS, Diyah I, Arifah C, Dhenabayu R. 2026. Integration of WhatsApp Business API and Artificial Intelligence in a Assisted-Automated Letter Generation System for Village Governance. *ILKOMNIKA: Journal of Computer Science and Applied Informatics E*. 8(1):165–181. doi:10.28926/ilkomnika.v8i1.876.
- Radhakrishnan KR, Saran Kumar R, Sivarama Krishnan S. 2024. Multifaceted Chatbot: A Retrieval Augmented Generation Approach for Intelligent Website Query Handling. Springer. hal. 513.
- Saputro WT, Jumasa HM, Pasa IY, Chirzah D, Sa'diah H. 2025. Service-Based Village Information System with Artificial Intelligence (AI) Chatbot. *2025 International Conference on ICT for Smart Society (ICISS)*.
- Siman KF, Wiratama J. 2023. A Web-based Village Administrative Information Systems for Improvement Quality of Service towards Smart Village Concept. *G-Tech: Jurnal Teknologi Terapan*. 7(4):1519–1528. doi:10.33379/gtech.v7i4.3196.
- Sjaf S, et. al. 2024. Sustainable Indonesian Rural Development: Utilizing Precision Village Data as a Basis for Socio-Economic Analysis. Di dalam: *IOP Conference Series: Earth and Environmental Science*. Volume ke-1359. Institute of Physics.
- Sjaf S, Elson L, Hakim L, Godya IM. 2020. *Data Desa Presisi*. Ed ke-1. Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) IPB University.

- Sjaf S, Hakim L, Arsyad AA. 2021. Data Desa Presisi untuk Pembangunan Desa Presisi. *Policy Brief Pertanian, Kelautan, dan Biosains Tropika*. 3(1):106–113.
- Sjaf S, Sampean N, Arsyad AA, Elson L, Mahardika AR, Hakim L, Amongjati SA, Gandi R, Barlan ZA, Aditya IMG, *et al.* 2022. Data Desa Presisi: A New Method of Rural Data Collection. *MethodsX*. 9:101868. doi:10.1016/j.mex.2022.101868.
- Sukhwal D. 2024. Retrieval Augmented Generation: An Evaluation of RAG-based Chatbot for Customer Support.
- Sururi AR, Hermadi I. 2025. PENGEMBANGAN CHATBOT BERBASIS RETRIEVAL AUGMENTED GENERATION DENGAN STUDI KASUS DASAR POKOK AGRARIA. Sekolah Sains Data, Matematika, dan Informatika.
- Swacha J, Gracel M. 2025. Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications. *Applied Sciences (Switzerland)*. 15(8). doi:10.3390/app15084234.
- Trapp M, Hess S. 2020. *Digital Villages How Digital Ecosystems Are Structured and What They Can Do*. Springer Vieweg.
- Wan Y, Chen Z, Liu Y, Chen C, Packianather M. 2025. Empowering LLMs by hybrid retrieval-augmented generation for domain-centric Q&A in smart manufacturing. *Advanced Engineering Informatics*. 65. doi:10.1016/j.aei.2025.103212.
- Xi Y, Bai Y, Luo H, Wen W, Liu H, Li H. 2025. Hybrid Retrieval-Augmented Generation Agent for Trustworthy Legal Question Answering in Judicial Forensics.