

SKRIPSI
PENERAPAN METODE *RETRIEVAL-AUGMENTED GENERATION (RAG)*
PADA *CHATBOT* HUMBET AI UNTUK SISTEM INFORMASI KELAS
DIGITAL HUMA BETANG



DISUSUN OLEH:

NURDIN

213130503177

JURUSAN/PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS PALANGKA RAYA

2026

**PENERAPAN METODE RETRIEVAL-AUGMENTED GENERATION (RAG) PADA CHATBOT
HUMBET AI UNTUK SISTEM INFORMASI KELAS DIGITAL HUMA BETANG**

SKRIPSI

Sebagai salah satu syarat untuk menyelesaikan Program Strata-1 pada Jurusan Teknik Informatika Fakultas
Teknik Universitas Palangka Raya

Oleh

NURDIN

213130503177

Telah dipertahankan didepan tim penguji, pada :

Hari/Tanggal : Selasa, 14 Juli 2026

Waktu : 12:05-13:35 WIB

Tempat : Ruang Ujian 1

1. **FELICIA SYLVIANA, S.T., M.M.**
NIP. 19760118 200312 2 003

..... (Ketua Penguji)

2. **WIDIATRY, S.T., M.T.**
NIP. 19820717 200312 2 002

..... (Pembimbing Utama/Pertama)

3. **NOVERA KRISTIANI, S.T., M.T.**
NIP. 19931116 202203 2 012

..... (Pembimbing Pendamping/Kedua)

4. **NOVA NOOR KAMALA SARI, S.T., M.Kom.**
NIP. 19890407 201504 2 004

..... (Anggota)

Mengetahui :

**Jurusan / Program Studi Teknik Informatika
Fakultas Teknik Universitas Palangka Raya
Ketua Jurusan,**



ARIELA LESTARI, S.Kom., M.Cs., Ph.D.
NIP. 19800322 200501 2 004

HALAMAN PERNYATAAN

Dengan ini saya menyatakan dengan sebenar-benarnya bahwa dalam Skripsi ini tidak terdapat karya ilmiah yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, serta tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam Skripsi ini dan disebutkan dalam Tinjauan Pustaka serta Daftar Pustaka.

Palangka Raya, 16 Juli 2026



NURDIN

213130503177

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan penyusunan skripsi yang berjudul "**Penerapan Metode *Retrieval-Augmented Generation (RAG)* pada *Chatbot* Humbet AI untuk Sistem Informasi Kelas Digital Huma Betang**" sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik pada Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya. Skripsi ini bertujuan untuk menerapkan metode *Retrieval-Augmented Generation (RAG)* pada *chatbot* Humbet AI guna meningkatkan kualitas layanan informasi pada Sistem Informasi Kelas Digital Huma Betang melalui penyajian jawaban yang lebih akurat, relevan, dan kontekstual. Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan belum sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi penyempurnaan penelitian ini di masa mendatang. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi referensi bagi pengembangan penelitian di bidang kecerdasan buatan, khususnya penerapan *Retrieval-Augmented Generation (RAG)* pada sistem *chatbot*.

Palangka Raya, 16 Juli 2026



NURDIN

213130503177

**IMPLEMENTASI *CHATBOT* BERBASIS *RETRIEVAL-AUGMENTED
GENERATION (RAG)* UNTUK LAYANAN INFORMASI PADA SISTEM
KELAS DIGITAL HUMA BETANG**

NURDIN (213130503177)

Jurusan Teknik Informatika Fakultas Teknik Universitas Palangka Raya Kampus

Tanjung Nyaho Jl. Yos Sudarso, Palangka Raya 73112

Email: nudriin@mhs.eng.upr.ac.id

ABSTRAK

Sistem *chatbot* Humbet AI pada platform Kelas Digital Huma Betang sebelumnya menggunakan pendekatan *rule-based* yang bergantung pada pencocokan kata kunci, sehingga mengalami kesulitan dalam memahami variasi bahasa alami, kueri ambigu, dan pertanyaan kompleks. Penelitian ini bertujuan mengimplementasikan metode *Retrieval-Augmented Generation (RAG)* untuk meningkatkan kemampuan *chatbot* dalam menghasilkan jawaban. Arsitektur yang dikembangkan menerapkan konsep *Advanced RAG* dengan mengintegrasikan *Query Refinement (RQ-RAG)*, *Dynamic Retrieval (DRAGIN)*, *semantic chunking*, *embedding*, dan *vector database*. Basis pengetahuan dibangun dari 6 dokumen panduan Kelas Digital Huma Betang yang diproses menjadi 82 *chunk* dan disimpan dalam basis data vektor. Model *Google Gemini 2.5 Flash* digunakan sebagai *Large Language Model* tanpa proses *fine-tuning*.

Evaluasi kualitas respons dilakukan menggunakan metode *G-EVAL* dan *Win Rate Analysis (MT-Bench)*, sedangkan evaluasi *pipeline RAG* dilakukan menggunakan *RAGAS*. Hasil penelitian menunjukkan bahwa sistem memperoleh rata-rata skor *G-EVAL* sebesar 4,825, meningkat 40,6% dibandingkan sistem lama yang memperoleh skor 3,431. Pada evaluasi *Win Rate Analysis*, sistem *RAG* memenangkan 87,5% pengujian *head-to-head* terhadap sistem *baseline*. Evaluasi menggunakan *RAGAS* menghasilkan skor rata-rata *faithfulness* sebesar 0,797, *answer relevancy* sebesar 0,476, dan *context utilization* sebesar 0,709, yang menunjukkan bahwa sistem mampu memanfaatkan konteks secara efektif serta menghasilkan jawaban yang akurat dengan tingkat halusinasi yang rendah. Hasil menunjukkan bahwa *Advanced RAG* berhasil meningkatkan kualitas layanan informasi pada *chatbot* Humbet AI.

Kata kunci: *Retrieval-Augmented Generation, Advanced RAG, Chatbot, Large Language Model, Kelas Digital Huma Betang.*

**IMPLEMENTASI CHATBOT BERBASIS RETRIEVAL-AUGMENTED
GENERATION (RAG) UNTUK LAYANAN INFORMASI PADA SISTEM
KELAS DIGITAL HUMA BETANG**

NURDIN (213130503177)

*Department of Informatics Engineering, Faculty of Engineering, Palangka Raya
University Campus*

Tanjung Nyaho Jl. Yos Sudarso, Palangka Raya 73112

Email: nudriin@mhs.eng.upr.ac.id

ABSTRACT

The Humbet AI chatbot in the Huma Betang Digital Learning platform previously employed a rule-based approach relying on exact keyword matching, which limited its ability to understand natural language variations, ambiguous queries, and complex questions. This study aims to implement the Retrieval-Augmented Generation (RAG) method to improve the chatbot's capability in generating contextual and relevant responses. The proposed architecture adopts an Advanced RAG framework by integrating Query Refinement (RQ-RAG), Dynamic Retrieval (DRAGIN), semantic chunking, embedding, and a vector database. The knowledge base was constructed from six Huma Betang user manual documents, processed into 82 semantic chunks and indexed in a vector database. Google Gemini 2.5 Flash was employed as the Large Language Model without additional fine-tuning.

Response quality was evaluated using G-EVAL and Win Rate Analysis (MT-Bench), while the RAG pipeline was assessed using RAGAS. The experimental results demonstrate that the proposed system achieved an average G-EVAL score of 4.825, representing a 40.6% improvement over the baseline system, which obtained a score of 3.431. In the head-to-head comparison, the proposed RAG system achieved an overall 87.5% win rate against the existing rule-based chatbot. Furthermore, the RAGAS evaluation produced average scores of 0.797 for faithfulness, 0.476 for answer relevancy, and 0.709 for context utilization, indicating that the retrieval mechanism effectively provides relevant context and supports the generation of accurate responses with a low level of hallucination. These findings demonstrate that the implementation of the Advanced RAG architecture significantly improves the quality of information services provided by the Humbet AI chatbot compared with the previous rule-based approach.

Keywords: *Retrieval-Augmented Generation, Advanced RAG, Chatbot, Large Language Model, Huma Betang Digital Learning System.*

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN.....	iii
RIWAYAT PENYUSUN.....	iv
HALAMAN PERSEMBAHAN	v
ABSTRAK	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.5.1 Manfaat Teoritis	5
1.5.2 Manfaat Praktis	6
1.6 Sistematika Penulisan.....	6
1.7 Jadwal Skripsi.....	7
BAB II LADASAN TEORI	8
2.1 Tinjauan Pustaka.....	8
2.2 Sistem Informasi.....	10
2.3 <i>Chatbot</i> (Agen Percakapan)	11
2.4 <i>Large Language Model (LLM)</i>	12

2.5	<i>Retrieval-Augmented Generation (RAG)</i>	12
2.6	Arsitektur dan Komponen <i>RAG</i>	14
2.6.1	<i>Retriever</i>	14
2.6.1.1	<i>Indexing</i> (Pengeindeksan).....	14
2.6.1.2	<i>Retrieval</i> (Pengambilan).....	14
2.6.1.3	Jenis <i>Retriever</i>	14
2.6.2	<i>Generator</i>	15
2.7	Paradigma dan Peningkatan <i>RAG</i>	15
2.7.1	<i>Naive RAG</i>	15
2.7.2	<i>Advanced RAG</i> dan <i>Modular RAG</i>	15
2.7.2.1	Peningkatan <i>Pre-Retrieval</i> (Pra-Pencarian)	16
2.7.2.1.1	<i>Query Transformation/Refinement</i>	16
2.7.2.1.2	<i>Query Expansion</i>	16
2.7.2.1.3	<i>Hypothetical Document Embeddings (HyDE)</i>	16
2.7.2.2	Peningkatan <i>Post-Retrieval</i> (Pasca-Pencarian)	17
2.7.2.2.1	<i>Re-Ranking</i>	17
2.7.2.2.2	<i>Compression</i> (Kompresi)	17
2.7.2.3	Peningkatan Proses Augmentasi (Iteratif & Adaptif).....	17
2.7.2.3.1	<i>Iterative Retrieval</i> (Pencarian Iteratif)	18
2.7.2.3.2	<i>Adaptive Retrieval</i> (Pencarian Adaptif)	18
2.8	<i>Query Refinement</i> (RQ-RAG).....	18
2.8.1.	<i>Explicit Rewriting</i> (Penulisan Ulang Eksplisit).....	19
2.8.2.	<i>Decomposition</i> (Dekomposisi).....	19
2.8.3.	<i>Disambiguation</i> (Disambiguasi)	19
2.9	<i>Dynamic Decision</i> (DRAGIN)	19

2.9.1.	<i>Uncertainty</i> (Ketidakpastian)	19
2.9.2.	<i>Informativeness</i> (Keinformatifan)	20
2.10	Evaluasi Berbasis <i>Large Language Model (LLM-as-a-Judge)</i>	20
2.10.1.	G-EVAL	21
2.10.2.	<i>Win-Rate Analysis (LLM-as-a-Judge)</i>	25
2.10.3.	<i>RAGAS Evaluation</i>	28
2.11	<i>RAG</i> untuk Adaptasi Domain	32
BAB III METODE PENELITIAN		33
3.1	Studi Literatur dan Identifikasi Masalah	34
3.1.1	Analisis Kesenjangan	34
3.1.2	Identifikasi Keterbatasan Cakupan Pengetahuan (<i>Knowledge Gap</i>) Sistem Lama	34
3.1.3	Eksplorasi Paradigma	35
3.2	Persiapan dan <i>Preprocessing</i> Data	35
3.2.1	Sumber Data	35
3.2.2	Teknik Pengumpulan Data	36
3.2.3	<i>Adaptive Chunking</i>	37
3.2.4	<i>Embedding dan Indexing</i>	37
3.3	Perancangan Arsitektur <i>Advanced RAG</i>	37
3.3.1	<i>Module RQ (Query Rewriter - Berbasis Konsep RQ-RAG)</i>	37
3.3.2	<i>Module DRAGIN (Dynamic Trigger - Berbasis Konsep Dynamic Retrieval)</i>	38
3.4	Implementasi dan Integrasi Sistem	42
3.4.1	<i>Dataset</i> dan Sumber Pengetahuan	42
3.4.2	Implementasi Large Language Model	42

3.4.3	Adaptive Chunking	44
3.4.4	Embedding dan Indexing	44
3.4.5	Pembuatan <i>Retriever</i>	45
3.4.6	Implementasi Query Refinement (RQ-RAG)	46
3.4.7	Implementasi Dynamic Retrieval (DRAGIN)	47
3.4.8	Alur Pipeline RAG	49
3.5	Evaluasi Komparatif <i>End-to-End</i> Berbasis <i>LLM-as-a-Judge</i>	52
3.5.1	Persiapan <i>Dataset</i> Evaluasi	55
3.5.2	Pengumpulan <i>Output</i> Sistem (<i>Inference</i>)	59
3.5.3	Konfigurasi <i>Evaluator LLM-as-a-Judge</i>	60
3.5.4	Pelaksanaan Evaluasi dan Perhitungan Metrik	60
3.5.5	Analisis dan Interpretasi Hasil	60
3.6	Instrumen Penelitian	61
BAB IV HASIL DAN PEMBAHASAN.....		62
4.1	Hasil dan Analisis Penelitian	62
4.2	Simulasi Implementasi <i>Pipeline Advanced RAG</i>	62
4.2.1.	Pengumpulan Data dan <i>Document Parsing</i>	62
4.2.2.	<i>Adaptive Chunking</i>	63
4.2.3.	<i>Input</i> Pengguna	64
4.2.4.	<i>Query Refinement (RQ-RAG)</i>	65
4.2.5.	<i>Retrieval</i> Dokumen	66
4.2.6.	Evaluasi <i>Uncertainty</i> Menggunakan <i>DRAGIN</i>	68
4.2.7.	Generasi Jawaban.....	70
4.3	Evaluasi Sistem	72
4.3.1	Perbandingan Kualitatif	73

4.3.1.1.	<i>Query</i> Sederhana	73
4.3.1.2.	<i>Query</i> Kompleks	75
4.3.1.3.	<i>Query</i> Ambigu.....	77
4.3.2	Perhitungan Evaluasi <i>G-EVAL</i>	81
4.3.3	Perhitungan <i>Win-Rate Analysis</i>	84
4.3.4	Evaluasi Sistem Menggunakan RAGAS.....	87
4.4	Pembahasan	89
BAB V KESIMPULAN.....		91
5.1	Kesimpulan.....	91
5.2	Saran	92
DAFTAR PUSTAKA.....		94
LAMPIRAN.....		97

DAFTAR TABEL

Tabel 1. 1 Jadwal Pelaksanaan Skripsi.....	7
Tabel 2. 1 Contoh Data Mentah Dari <i>G-EVAL</i>	23
Tabel 2. 2 Contoh Hasil Perhitungan Seluruh Metrik Uji <i>G-EVAL</i>	25
Tabel 2. 3 Contoh data <i>faithfulness ragas</i>	30
Tabel 2. 4 Contoh <i>reverse question</i> metrik <i>answer relevance</i>	31
Tabel 2. 5 Contoh data metrik <i>context utilization</i>	32
Tabel 3. 1 Jumlah data <i>knowledge-base</i>	36
Tabel 3. 2 Contoh data probabilitas <i>LLM</i>	40
Tabel 3. 3 <i>Dataset</i> evaluasi	56
Tabel 4. 1 Contoh <i>chunk</i> tersimpan pada <i>database</i>	63
Tabel 4. 2 Hasil Evaluasi Menggunakan <i>G-EVAL</i>	81

DAFTAR GAMBAR

Gambar 3. 1 Bagan Alir Penelitian	33
Gambar 3. 2 Arsitektur <i>RAG Modular</i>	39
Gambar 3. 3 Bagan Alir Evaluasi.....	52
Gambar 4. 1 Data profil pada manual book siswa	63
Gambar 4. 2 Input mentah pengguna	65
Gambar 4. 3 Proses query refinement.....	66
Gambar 4. 4 Proses retrieval dokumen	67
Gambar 4. 5 Proses evaluasi menggunakan DRAGIN	69
Gambar 4. 6 Hasil generasi dari chatbot	71
Gambar 4. 7 Uji Query Sederhana Sistem Lama	74
Gambar 4. 8 Uji Query Sederhana Sistem Baru	75
Gambar 4. 9 Query Uji Kompleks Sistem Lama	76
Gambar 4. 10 Uji Query Kompleks Sistem Baru.....	77
Gambar 4. 11 Uji Query Ambigu Sistem Lama	80
Gambar 4. 12 Uji Query Ambigu Sistem Baru	80
Gambar 4. 14 Grafik Batang per Kategori Kueri.....	82
Gambar 4. 14 Grafik Batang Perbandingan RAG dan Sistem Lama.....	83
Gambar 4. 15 Grafik Pie Distribusi Kemenangan	85
Gambar 4. 16 Grafik Batang Kemenangan per Kategori Kueri.....	85
Gambar 4. 17 Heatmap per Kategori Kueri	86
Gambar 4. 18 Grafik Batang hasil RAGAS per Kategori Kueri.....	87
Gambar 4. 19 Grafik Batang RAGAS per Metrik	88

DAFTAR PUSTAKA

- Bai, G., Liu, J., Bu, X., He, Y., Liu, J., Zhou, Z., Lin, Z., Su, W., Ge, T., Zheng, B., & Ouyang, W. (2024). *MT-Bench-101: A Fine-Grained Benchmark for Evaluating Large Language Models in Multi-Turn Dialogues*. <https://doi.org/10.18653/v1/2024.acl-long.401>
- Chan, C.-M., Xu, C., Yuan, R., Luo, H., Xue, W., Guo, Y., & Fu, J. (2024). *RQ-RAG: Learning to Refine Queries for Retrieval Augmented Generation*. <http://arxiv.org/abs/2404.00610>
- Cheng, X., Luo, D., Chen, X., Liu, L., Zhao, D., & Yan, R. (n.d.). *Lift Yourself Up: Retrieval-augmented Text Generation with Self-Memory*. Retrieved <https://github.com/Hannibal046/SelfMemory>
- Colverd, G., Darm, P., Silverberg, L., & Kasmanoff, N. (2023). *FloodBrain: Flood Disaster Reporting by Web-based Retrieval Augmented Generation with an LLM*. <http://arxiv.org/abs/2311.02597>
- Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2025). *Ragas: Automated Evaluation of Retrieval Augmented Generation*. <http://arxiv.org/abs/2309.15217>
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). *Retrieval-Augmented Generation for Large Language Models: A Survey*. <http://arxiv.org/abs/2312.10997>
- Huang, Y., & Huang, J. X. (2018). The Survey of Retrieval-Augmented Text Generation in Large Language Models. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)* (Vol. 1). <https://doi.org/XXXXXXX.XXXXXXX>
- Kirchenbauer, J., & Barns, C. (n.d.). *Hallucination Reduction in Large Language Models with Retrieval-Augmented Generation Using Wikipedia Knowledge*.
- Lewis, C. C., Mettert, K. D., Stanick, C. F., Halko, H. M., Nolen, E. A., Powell, B. J., & Weiner, B. J. (2021). The psychometric and pragmatic evidence rating scale (PAPERS) for measure development and evaluation. *Implementation Research and Practice*, 2. <https://doi.org/10.1177/26334895211037391>
- Li, H., Su, Y., Cai, D., Wang, Y., & Liu, L. (2022). *A Survey on Retrieval-Augmented Text Generation*. <http://arxiv.org/abs/2202.01110>
- Lin, Y.-T., & Chen, Y.-N. (2023). *LLM-EVAL: Unified Multi-Dimensional Automatic Evaluation for Open-Domain Conversations with Large Language Models*.

- Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R., & Zhu, C. (2023). *G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment*. <http://arxiv.org/abs/2303.16634>
- Liu, Y., Pei, C., Wang, H., Dong, X., Feng, Z., Li, Z., Kehang, J., Pei, D., Xu, L., Zheng, L., & Ji, K. (2026). Eagle: Leveraging Operations Documents for Comprehensive Benchmark Question Generation. *34th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (FSE Companion '26), July 5â•fi9, 2026, Montreal, QC, Canada, I*, 11. <https://doi.org/10.1145/3803437.3805214>
- Miao, J., Thongprayoon, C., Suppadungsuk, S., Garcia Valencia, O. A., & Cheungpasitporn, W. (2024). Integrating Retrieval-Augmented Generation with Large Language Models in Nephrology: Advancing Practical Applications. In *Medicina (Lithuania)* (Vol. 60, Number 3). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/medicina60030445>
- Narsupalli, Y., Chandra, A., Muppirala, S., Gupta, M., & Goyal, P. (2024). *ReFeR: Improving Evaluation and Reasoning through Hierarchy of Models*. <http://arxiv.org/abs/2407.12877>
- Neupane, S., Hossain, E., Keith, J., Tripathi, H., Ghiasi, F., Golilarz, N. A., Amirlatifi, A., Mittal, S., & Rahimi, S. (2024). *From Questions to Insightful Answers: Building an Informed Chatbot for University Resources*. <http://arxiv.org/abs/2405.08120>
- Peng, B., Zhu, Y., Zhang, Y., Bo, X., Liu, Y., Shi, H., Hong, C., & Tang, S. (2024). Graph Retrieval-Augmented Generation: A Survey. In *Journal of the ACM* (Vol. 37, Number 4). <https://doi.org/XXXXXXX.XXXXXXX>
- Press, O., Zhang, M., Min, S., Schmidt, L., Smith, N. A., & Lewis, M. (2023). *Measuring and Narrowing the Compositionality Gap in Language Models*. <http://arxiv.org/abs/2210.03350>
- Shao, Z., Gong, Y., Shen, Y., Huang, M., Duan, N., & Chen, W. (2023). *Enhancing Retrieval-Augmented Large Language Models with Iterative Retrieval-Generation Synergy*. <http://arxiv.org/abs/2305.15294>
- Shu, L., Wichers, N., Luo, L., Zhu, Y., Liu, Y., Chen, J., Meng, L., & Deepmind, G. (n.d.). *Fusion-Eval: Integrating Assistant Evaluators with LLMs*.
- Shuster, K., Poff, S., Chen, M., Kiela, D., & Weston, J. (2021). *Retrieval Augmentation Reduces Hallucination in Conversation*. <http://arxiv.org/abs/2104.07567>

- Siledar, T., Nath, S., Muddu, S. S. R. R., Rangaraju, R., Nath, S., Bhattacharyya, P., Banerjee, S., Patil, A., Singh, S. S., Chelliah, M., & Garera, N. (2024). *One Prompt To Rule Them All: LLMs for Opinion Summary Evaluation*. <http://arxiv.org/abs/2402.11683>
- Son, S., Oh, J.-M., Jin, H., Jang, C., Jeong, J., & Kim, K. (n.d.). *Arena-Lite: Efficient and Reliable Large Language Model Evaluation via Tournament-Based Direct Comparisons Work performed primarily while at NC AI*. Retrieved <https://huggingface>.
- Su, W., Tang, Y., Ai, Q., Wu, Z., & Liu, Y. (2024). *DRAGIN: Dynamic Retrieval Augmented Generation based on the Information Needs of Large Language Models*. <http://arxiv.org/abs/2403.10081>
- Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., Yang, L., Zhang, W., Jiang, J., & Cui, B. (2024). *Retrieval-Augmented Generation for AI-Generated Content: A Survey*. <http://arxiv.org/abs/2402.19473>
- Zhao, R., Zhang, W., Ken Chia, Y., Xu, W., Zhao, D., & Bing, L. (2024). *Auto-Arena: Automating LLM Evaluations with Agent Peer Battles and Committee Discussions* (Vol. 1). <https://auto-arena.github.io/>
- Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). *Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena*. <http://arxiv.org/abs/2306.05685>