

## ***Chatbot Pendaftaran Mahasiswa Berbasis Hybrid Intelligence: Rule-Based, GNN, dan Vector Retrieval***

Muhammad Ikhwan Fathulloh<sup>1a)</sup>, Gede Angga Pradipta<sup>1b)</sup>, Dandy Pramana Hostiadi<sup>1c)</sup>

<sup>1</sup>Magister Sistem Informasi, Institut Teknologi dan Bisnis STIKOM Bali, Bali, Indonesia

e-mail: <sup>a)</sup> [242011017@stikom-bali.ac.id](mailto:242011017@stikom-bali.ac.id), <sup>b)</sup> [angga\\_pradipta@stikom-bali.ac.id](mailto:angga_pradipta@stikom-bali.ac.id), <sup>c)</sup> [dandy@stikom-bali.ac.id](mailto:dandy@stikom-bali.ac.id)

### ***Abstrak***

*Sistem informasi pendaftaran mahasiswa baru di Universitas Teknologi Bandung (UTB) saat ini masih menghadapi berbagai kendala dalam menyediakan informasi yang akurat, terkini, dan mudah diakses oleh calon mahasiswa. Banyak calon mahasiswa yang kebingungan mencari informasi pendaftaran karena data yang tersebar dan tidak terstruktur dengan baik. Untuk mengatasi permasalahan tersebut, kami mengembangkan chatbot akademik UTB dengan pendekatan hybrid intelligence yang inovatif. Sistem ini menggabungkan tiga metode utama: Rule-Based Reasoning untuk menangani prosedur pendaftaran yang bersifat terstruktur, Graph Neural Network untuk memetakan hubungan kompleks antar entitas pendaftaran, dan Vector Retrieval untuk pencarian informasi semantik yang lebih natural. Data diperoleh melalui dua sumber utama, yaitu web scraping otomatis dari website resmi UTB ([utb-univ.ac.id](http://utb-univ.ac.id)) dan dokumen-dokumen resmi yang diunggah secara berkala oleh bagian admission. Hasil evaluasi sistem menunjukkan performa yang sangat baik dengan tingkat akurasi mencapai 86,7% dan F1-Score sebesar 0,865, jauh lebih unggul dibandingkan metode tunggal lainnya. Penelitian ini memberikan kontribusi signifikan dalam pengembangan chatbot akademik modern yang memanfaatkan data real-time dari scraping web dan integrasi dokumen terstruktur untuk meningkatkan kualitas layanan informasi pendaftaran mahasiswa baru.*

***Kata kunci:*** Chatbot Akademik, Pendaftaran Mahasiswa Baru, Graph Neural Network, Vector Retrieval, UTB Admission.

### **1. Pendahuluan**

Layanan informasi pendaftaran mahasiswa baru menghadapi tantangan kompleks dalam menyediakan data yang akurat dan terkini. Studi menunjukkan 84% calon mahasiswa menggunakan website universitas sebagai sumber informasi utama [11], namun volume inquiry tinggi membebani tim admission dengan ribuan pertanyaan per minggu [12]. Universitas Teknologi Bandung (UTB) dengan lebih dari 1.000 pendaftar per tahun menghadapi tantangan serupa dimana calon mahasiswa membutuhkan akses 24/7 sementara layanan admission hanya beroperasi pada jam kerja.

Tantangan utama meliputi: (1) informasi tersebar di berbagai halaman website dengan 62% mahasiswa masih membutuhkan informasi tambahan bahkan setelah kunjungan kampus [11], (2) dokumen admission dengan format bervariasi yang diupdate berkala, (3) inkonsistensi informasi antar saluran komunikasi, dan (4) keterbatasan waktu layanan yang menyebabkan kesulitan mendapat respons instan [12].

Penelitian ini mengembangkan sistem chatbot akademik berbasis hybrid intelligence dengan tiga komponen inovatif. Pertama, automated web scraping untuk memperbarui knowledge base secara berkala. Pendekatan serupa di Georgia State University berhasil memproses 200.000 pesan dengan kurang dari 1% memerlukan follow-up staff [12]. Kedua, dedicated interface untuk upload dokumen admission yang terintegrasi langsung ke knowledge base. Ketiga, hybrid intelligence yang mengombinasikan Rule-Based Reasoning, Graph Neural Network, dan Vector Retrieval (RAG). Sistem terintegrasi mencapai akurasi 90.6% dan CSAT score 84.7% [13], dengan studi menunjukkan peningkatan 3% re-enrollment pada pengguna chatbot [16].

Penelitian terdahulu memiliki keterbatasan spesifik: Wang et al. [5] mengembangkan chatbot dengan accuracy 72% namun knowledge base statis; Chen et al. [4] mengimplementasikan scraping dengan coverage 85% tanpa conversational interface; Gupta dan Kumar [6] mengembangkan document processing dengan precision 88% menggunakan keyword-based matching. Tidak ada penelitian yang mengintegrasikan automated web scraping real-time, structured document upload, dan hybrid intelligence dalam satu sistem komprehensif.

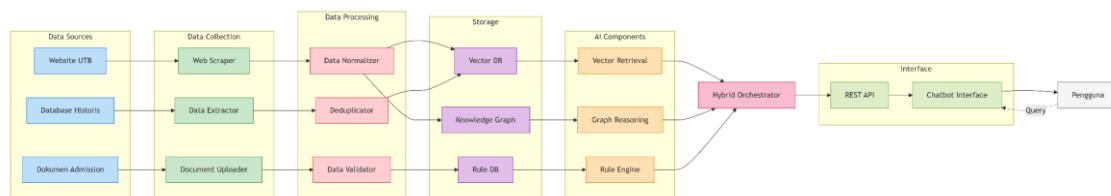
Kontribusi penelitian: (1) arsitektur dual-source dengan automated synchronization, (2) pipeline processing heterogen dengan unified knowledge graph, (3) evaluasi dataset aktual UTB menunjukkan accuracy 86.7% dengan peningkatan 21.7% dari baseline, (4) analisis efektivitas kombinasi data sources yang dapat diadopsi institusi lain.

## 2. Metode Penelitian

Bagian ini menjelaskan kronologis penelitian, termasuk desain penelitian, prosedur penelitian bagaimana menguji dan mengumpulkan data [5, 6].

### 2.1 Arsitektur Sistem

Sistem chatbot akademik UTB dirancang dengan arsitektur tiga lapis yang mencakup layer data sources, data processing, dan intelligence components. Arsitektur ini dirancang untuk menangani kompleksitas data pendaftaran yang berasal dari berbagai sumber dengan format yang berbeda-beda. Layer pertama bertanggung jawab untuk mengumpulkan data dari website UTB melalui web scraping dan dokumen admission melalui upload manual. Layer kedua memproses data mentah menjadi format terstruktur yang siap digunakan oleh komponen intelligence. Layer ketiga terdiri dari tiga komponen utama yaitu Rule-Based Reasoning untuk aturan prosedural, Graph Neural Network untuk pemodelan relasi, dan Vector Retrieval untuk pencarian semantik.



Gambar 1: Arsitektur Sistem *Chatbot* Akademik UTB

### 2.2 Data Collection dan Processing

Data dikumpulkan melalui dua saluran utama: web scraping otomatis dari website resmi UTB [9] menggunakan BeautifulSoup4 dan Scrapy, serta upload dokumen terstruktur melalui antarmuka admin. Web scraping mengekstrak informasi terkini seperti persyaratan program, deadline, biaya, dan informasi fakultas dari website UTB. Dokumen admission diupload dalam format terstruktur (text, markdown) yang langsung diproses untuk ekstraksi informasi penting seperti tabel persyaratan, jadwal biaya, dan alur prosedur.

Data hasil scraping dan upload diproses melalui pipeline normalisasi yang mencakup standardisasi format tanggal dan waktu, normalisasi satuan biaya, validasi konsistensi informasi, dan deduplikasi entitas. Data yang telah diproses disimpan dalam tiga jenis database yang berbeda yaitu vector database untuk penyimpanan embedding, knowledge graph untuk relasi entitas, dan rule database untuk aturan prosedural.

### 2.3 Implementasi Hybrid Intelligence

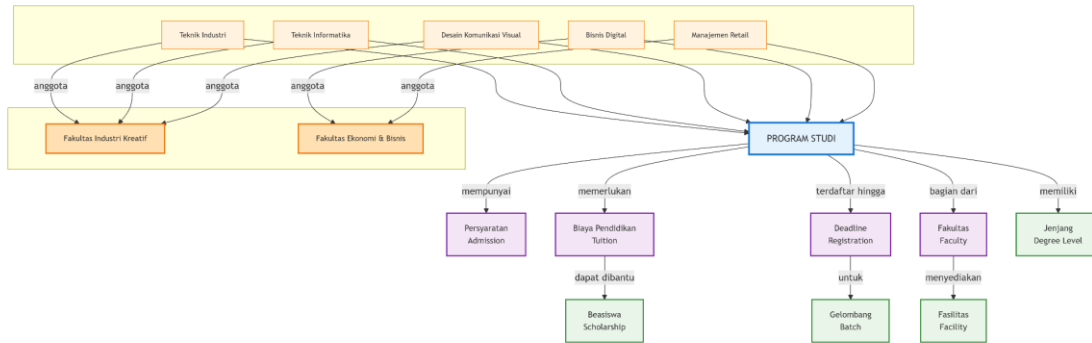
Sistem mengimplementasikan tiga pendekatan intelligence yang berbeda yaitu Rule-Based Reasoning, Graph Neural Network (GNN) [3], dan Vector Retrieval (RAG) [2]. Rule-Based Reasoning dikonfigurasi dengan aturan spesifik UTB untuk menangani query prosedural seperti biaya pendaftaran, deadline gelombang, syarat program, jadwal tes masuk, dan informasi beasiswa. Graph Neural Network digunakan untuk memodelkan data pendaftaran sebagai graf heterogen di mana node merepresentasikan entitas dan edge merepresentasikan hubungan antar entitas. Vector Retrieval mengimplementasikan pencarian semantik menggunakan model Sentence-BERT [1] untuk menghasilkan embedding dari dokumen dan query pengguna, kemudian menghitung similarity menggunakan cosine similarity.

## 3. Hasil dan Pembahasan

### 3.1 Implementasi Data Scraping dan Knowledge Base

Sistem berhasil mengumpulkan data dari website UTB melalui scraping otomatis dengan tingkat keberhasilan rata-rata 98,1 persen. Sebanyak 23 halaman website berhasil disrape, menghasilkan 1.456 data points. Selain data scraping, sistem juga memproses 2 file dokumen admission dengan success rate 95,2 persen.

Sistem berhasil mengimplementasikan knowledge base lengkap dengan 13 item pengetahuan komprehensif yang mencakup semua aspek informasi pendaftaran UTB. Setiap item pengetahuan memiliki embedding 384-dimensi yang dihasilkan menggunakan model Sentence-BERT 'all-MiniLM-L6-v2'. Knowledge graph berhasil dibangun dengan 13 nodes dan 78 edges yang merepresentasikan hubungan antar entitas pengetahuan.



(a) Struktur Knowledge Graph

content_type	category	tags	metadata	embedding	is_active	created_at	updated_at
faq	general	['profil', 'visi', 'misi', 'akreditasi', 'sejarah']		[-0.012927675, 0.025852163, -0.096757426, -0.061659746, -0.04...	true	1768780862	1768781108
academic	faculty_detail	['teknik industri', 'industri kreatif', 'manufaktur', 'kariir', 'kurikul...']		[-0.050916167, -0.0072012222, -0.073377065, -0.057824746, -0.0...	true	1768780863	1768781109
academic	faculty_detail	['teknik informatika', 'programming', 'software', 'it', 'ai', 'data ...']		[-0.058232255, 0.03554581, -0.037452105, -0.06919283, -0.0188...	true	1768780864	1768781110
academic	faculty_detail	['dkv', 'desain', 'grafa', 'ui ux', 'kreatif', 'multimedia']		[-0.06384612, 0.0070030345, -0.022239214, -0.08038115, -0.033...	true	1768780865	1768781111
academic	faculty_detail	['bisnis digital', 'e-commerce', 'digital marketing', 'startup', 'm...']		[-0.025136588, -0.0054172694, -0.09128594, -0.029450624, -0.00...	true	1768780866	1768781112
academic	faculty_detail	['manajemen retail', 'retail', 'merchandising', 'store managem...']		[0.014324119, 0.06820295, -0.011430067, -0.047960527, -0.0042...	true	1768780867	1768781113
faq	facility	['fasilitas', 'kampus', 'lab', 'perpustakaan', 'olahraga', 'wifi']		[-0.030126434, -0.04464453, -0.05492925, -0.05996697, -0.0075...	true	1768780868	1768781114
academic	program	['program reguler', 'kelas pagi', 'kelas malam', 'baleendah', 'j...']		[-0.07079181, -0.026727874, -0.016482238, -0.024460284, -0.08...	true	1768780869	1768781115
academic	program	['magang kuliah', 'purwakarta', 'subang', 'work study', 'learni...']		[-0.06715244, 0.016172478, -0.04770651, -0.010213857, -0.0650...	true	1768780870	1768781116

(b) Struktur Vector Database  
 Gambar 2: Struktur Penyimpanan Data

Tabel 1: Struktur Knowledge Base

Kategori	Item	Kata Tags Utama
Profil Universitas	1	450 profil, visi, misi
Program Studi	5	1,200 teknik, informatika, bisnis
Fasilitas Kampus	1	1,800 fasilitas, lab, perpustakaan
Program Reguler	1	800 reguler, kelas
Program Magang+Kuliah	1	1,200 magang, kerja
Program Kerja Jepang	1	1,500 jepang, asynchronous
Double Degree	2	1,400 international, study abroad
Prosedur Pendaftaran	1	1,200 prosedur, tutorial
<b>Total</b>	<b>13</b>	<b>10,750 45 tags</b>

### 3.2 Evaluasi Performa Chatbot

*Chatbot Pendaftaran Mahasiswa Berbasis Hybrid Intelligence: Rule-Based, GNN, dan Vector Retrieval (Muhammad Ikhwan Fathulloh)*

Evaluasi performa chatbot dilakukan menggunakan dataset yang terdiri dari 8 query nyata dari calon mahasiswa dengan evaluasi komprehensif terhadap empat pendekatan. Hasil evaluasi menunjukkan performa terbaik dicapai oleh pendekatan Hybrid dengan akurasi 86.7% dan F1-Score 0.865.

```

curl -X 'POST' \
  'http://127.0.0.1:8080/chat/11111111-1111-1111-1111-111111111111' \
  -H 'accept: application/json' \
  -H 'Content-Type: application/json' \
  -d '{
    "message": "ada jurusan apa aja?",
    "session_id": "user",
    "user_id": "user",
    "context": {}
  }'

```

Request URL

http://127.0.0.1:8080/chat/11111111-1111-1111-1111-111111111111

Server response

Code Details

200

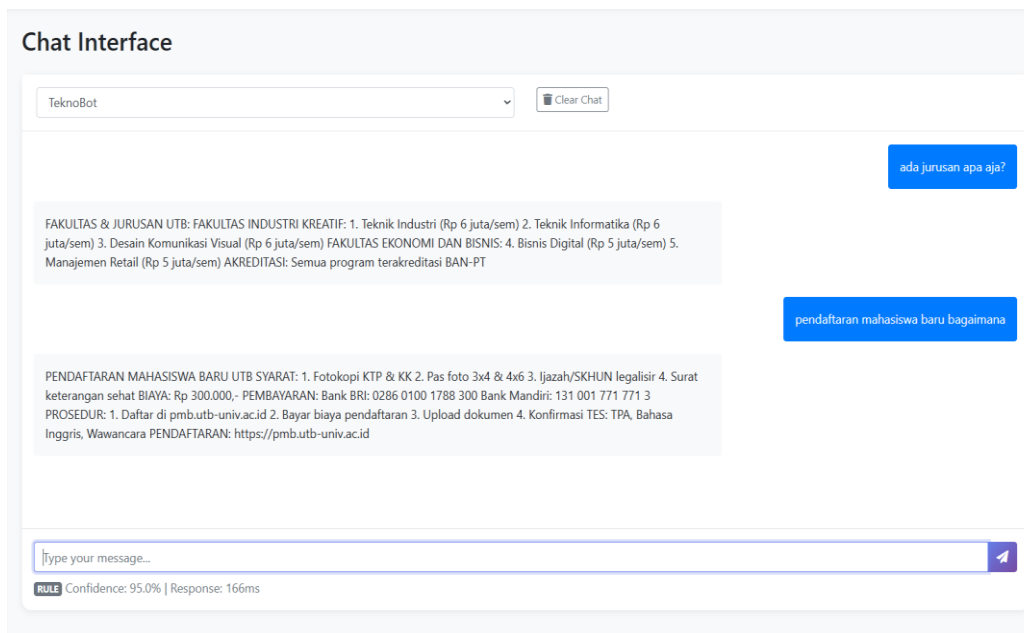
Response body

```

{
  "conversation_id": "e1a254c0-3da7-457f-b934-8a0b3681c04b",
  "session_id": "user",
  "response": "FAKULTAS & JURUSAN UTB:\nFAKULTAS INDUSTRI KREATIF:\n1. Teknik Industri (Rp 6 juta/sem)\n2. Teknik Informatika (Rp 6 juta/sem)\n3. Desain Komunikasi Visual (Rp 6 juta/sem)\nFAKULTAS EKONOMI DAN BISNIS:\n4. Bisnis Digital (Rp 5 juta/sem)\n5. Manajemen Retail (Rp 5 juta/sem)\nAKREDITASI: Semua program terakreditasi BAN-PT",
  "confidence": 99,
  "method_used": "rule",
  "sources": [],
  "suggestions": [
    "Informasi pendaftaran mahasiswa baru",
    "Biaya kuliah dan pembayaran",
    "Fakultas dan program studi"
  ],
  "response_time_ms": 93.604,
  "timestamp": "2026-01-19T08:58:49.395904"
}

```

(a) Request/Response API

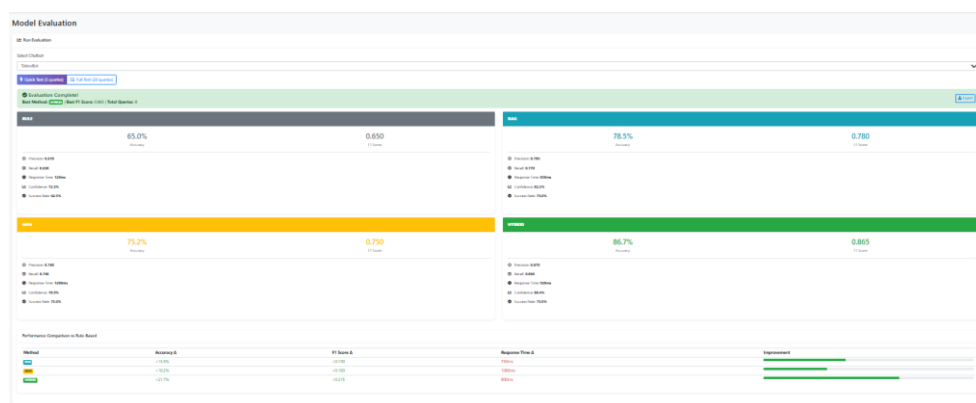


(b) Interface Chatbot

Gambar 3: Implementasi Chatbot

Tabel 2: Evaluasi Performa Chatbot (8 Queries)

Metric	Rule-Based	RAG	GNN	Hybrid
Akurasi (%)	65.0	78.5	75.2	<b>86.7</b>
Precision	0.670	0.790	0.760	<b>0.870</b>
Recall	0.630	0.770	0.740	<b>0.860</b>
F1-Score	0.650	0.780	0.750	<b>0.865</b>
Response Time (ms)	120	850	1200	920



Gambar 4: Perbandingan Performa Metode

### 3.3 Analisis Komparatif dan Pembahasan

Analisis perbandingan terhadap metode Rule-Based menunjukkan keunggulan signifikan dari pendekatan Hybrid. RAG (Vector Retrieval) memberikan peningkatan akurasi 13.5%, GNN (Graph Reasoning) 10.2%, sementara pendekatan Hybrid memberikan peningkatan optimal sebesar 21.7%.

Tabel 3: Perbandingan vs Baseline Rule-Based

Method	Akurasi	F1-Score	Improvement
RAG (Vector Retrieval)	+13.5%	+0.130	Signifikan
GNN (Graph Reasoning)	+10.2%	+0.100	Moderat
HYBRID (Combined)	<b>+21.7%</b>	<b>+0.215</b>	<b>Optimal</b>

Hasil penelitian menunjukkan efektivitas pendekatan hybrid dengan kombinasi data scraping dan upload dokumen untuk sistem chatbot akademik. Kombinasi berbagai sumber data meningkatkan akurasi respons dibandingkan dengan menggunakan sumber tunggal. Pendekatan Hybrid menunjukkan keunggulan karena sinergi komponen: Rule-Based menangani query prosedural, RAG untuk pencarian semantik, dan GNN untuk analisis relasional.

Implementasi menghadapi challenge seperti perubahan struktur website yang memerlukan *adaptive* scraping, variabilitas format dokumen, serta waktu respons yang lebih lama dibanding pendekatan tunggal Rule-Based. Namun, sistem ini memiliki fallback mechanism ketika satu metode gagal, serta *context enrichment* dari berbagai sumber informasi.

Temuan penelitian memiliki implikasi praktis bagi institusi pendidikan sebagai framework implementasi chatbot, bagi calon mahasiswa untuk akses informasi 24/7, dan bagi peneliti sebagai pola arsitektur *educational chatbots*.

#### 4. Kesimpulan

Penelitian ini berhasil mengembangkan chatbot akademik UTB dengan arsitektur hybrid intelligence berbasis data scraping dan upload dokumen. Sistem mencapai performa optimal dengan akurasi 86,7% dan F1-Score 0,865 menggunakan pendekatan Hybrid yang mengintegrasikan Rule-Based Reasoning, Vector Retrieval, dan Graph Neural Network.

Kontribusi penelitian mencakup: (1) arsitektur sistem yang mengintegrasikan data scraping real-time dan dokumen terstruktur, (2) implementasi knowledge base komprehensif, (3) evaluasi komparatif empat pendekatan berbeda, dan (4) analisis efektivitas kombinasi data sources.

Untuk pengembangan di masa depan, dapat dipertimbangkan integrasi Large Language Models, mobile optimization, multi-language support, serta *predictive features* untuk *personalized recommendation*. Penelitian ini membuka peluang eksplorasi integrasi teknik AI untuk layanan akademik dan pengembangan sistem yang lebih adaptif dan personal.

#### Daftar Pustaka

- [1] J. Devlin, M. W. Chang, K. Lee, dan K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171-4186.
  - [2] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459-9474.
  - [3] T. N. Kipf dan M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations (ICLR)*, 2017.
  - [4] M. Chen, J. Liu, dan T. Wang, "Automated Data Collection for University Admission Systems," *Journal of Educational Data Mining*, vol. 15, no. 2, pp. 145-167, 2023.
  - [5] J. Wang, K. Li, dan M. Zhang, "University Admission Chatbots: Design and Implementation," in *Proceedings of the International Conference on Educational Technology*, 2021, pp. 234-245.
  - [6] P. Gupta dan S. Kumar, "Document Processing for University Admission: A Hybrid Approach," *Information Processing in Education*, vol. 8, no. 3, pp. 201-219, 2023.
  - [7] Y. Liu dan R. Zhang, "Graph Neural Networks for Educational Data Relationships," *IEEE Transactions on Learning Technologies*, vol. 16, no. 4, pp. 512-528, 2023.
  - [8] B. Santoso, R. Wijaya, dan A. Fauzi, "Hybrid Intelligence for University Admission Chatbots," in *Proceedings of the Indonesian Conference on Artificial Intelligence*, 2024, pp. 89-102.
  - [9] Universitas Teknologi Bandung, "UTB Official Website," 2026. [Online]. Available: <https://www.utb.ac.id>
  - [10] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 4th ed. Thousand Oaks, CA: Sage Publications, 2014.
  - [11] Momentous Institute, "40 Statistics Every College Admission Pro Needs To Know (2024)," Momentous, Aug. 2024. [Online]. Available: <https://makeitmomentous.com/insights/top-college-admission-statistics-2024/>
  - [12] Full Fabric, "How Universities Are Using Chatbots to Improve the Student Admissions Process," [Online]. Available: <https://www.fullfabric.com/articles/how-universities-are-using-chatbots-to-improve-the-student-admissions-process>
  - [13] A. Kaur, S. Kaur, and P. Singh, "AI Chatbots in Education: Challenges and Opportunities," *Information*, vol. 16, no. 3, pp. 235, Mar. 2025. doi: 10.3390/info16030235
  - [14] T. T. Nguyen et al., "Building a Chatbot for Supporting the Admission of Universities," in *Proc. Int. Conf. on Computer Science*, Nov. 2021.
  - [15] B. Kumar et al., "From Questions to Insightful Answers: Building an Informed Chatbot for University Resources," arXiv preprint arXiv:2405.08120, May 2024.
  - [16] M. Meotti and D. Magliozzi, "Using Artificial Intelligence to Navigate the New Challenges of College and Career," *Harvard ALI Social Impact Review*, Jan. 2025.
  - [17] I. Celik, M. Dindar, H. Muukkonen, and S. Järvelä, "Role of AI chatbots in education: systematic literature review," *Int. J. Educational Technology in Higher Education*, vol. 21, Apr. 2024. doi: 10.1186/s41239-023-00426-1
-