

Научная статья

УДК 004.9

EDN IXAVJL

DOI 10.17150/2713-1734.2024.6(3).367-376



А.В. Родионов

*Байкальский государственный университет,
г. Иркутск, Российская Федерация*

Д.А. Корж

*Байкальский государственный университет,
г. Иркутск, Российская Федерация*

Выбор инструментов построения чат-бота приемной комиссии учебного заведения

Аннотация. В условиях стремительной цифровизации образовательных процессов, автоматизация взаимодействия с абитуриентами становится приоритетной задачей для вузов. Данная статья посвящена анализу различных подходов к разработке чат-ботов для приемной комиссии. Рассматриваются как традиционные методы, основанные на правилах и ключевых словах, так и современные технологии, использующие машинное обучение и большие языковые модели. Особое внимание уделяется сравнению точности, релевантности и естественности ответов, а также затратам на реализацию и поддержку. Исследование показывает, что современные нейросетевые модели значительно превосходят классические подходы по ряду ключевых параметров, но требуют больших ресурсов и постоянного интернет-доступа. Комбинирование классических и нейросетевых методов может стать оптимальным решением, позволяющим эффективно использовать сильные стороны обоих подходов, обеспечивая при этом высокое качество и оперативность взаимодействия с абитуриентами.

Ключевые слова. Чат-бот, большие языковые модели, машинное обучение, взаимодействие с абитуриентами, метод анализа иерархии.

Информация о статье. Дата поступления: 23 мая 2024 г.; дата принятия к публикации: 1 октября 2024 г.; дата онлайн-размещения: 17 октября 2024 г.

Original article

A.V. Rodionov

*Baikal State University,
Irkutsk, Russian Federation*

D.A. Korzh

*Baikal State University,
Irkutsk, Russian Federation*

Tools for Building a Chatbot for the Admissions Commission of an University

Abstract. In the context of the rapid digitalization of educational processes, automation of interaction with applicants is becoming a priority task for universities. This article is devoted to the analysis of various approaches to the development of chatbots for the admissions committee. Both traditional methods based on rules and keywords, and modern technologies using machine learning and large language models

are considered. Particular attention is paid to comparing the accuracy, relevance and naturalness of responses, as well as implementation and support costs. The study shows that modern neural network models are significantly superior to classical approaches in a number of key parameters, but require large resources and constant Internet access. Combining classical and neural network methods can be the optimal solution, allowing you to effectively use the strengths of both approaches, while ensuring high quality and efficiency of interaction with applicants.

Keywords. Chatbot, large language models, machine learning, interaction with applicants, hierarchy analysis method.

Article info. Received 23 May, 2024; Accepted 1 October, 2024; Available online 17 October, 2024.

Введение

В условиях цифровизации образовательного процесса и повышения требований к оперативности обработки запросов абитуриентов, роль автоматизированных систем общения становится все более значимой. Примером такой системы являются чат-боты — программное обеспечение, осуществляющее коммуникацию с пользователями системы в текстовом формате через различные платформы и мессенджеры. Чат-боты могут иметь различную функциональность, в зависимости от поставленных задач, например способны давать ответы на вопросы пользователей или выполнять действия в других системах по запросам.

В процессе работы приемной комиссии университета чат-бот может выполнять ряд ключевых задач, направленных на улучшение взаимодействия с абитуриентами, повышения качества и скорости выполнения следующих рабочих процессов: предоставление информации об университете; предоставление информации о процедуре поступления; получение ответов на часто задаваемые вопросы; помощь в выборе программы обучения; содействие в регистрации на события; предоставление персонализированных рекомендаций; предобработка заявок на поступление; обработка заявок на поступление; предоставление поддержки на протяжении всего процесса поступления.

При проектировании и разработке чат-ботов можно использовать несколько подходов. К условно «классическому» можно отнести подход, в рамках которых алгоритм работы чат-ботов основан на заранее заданных правилах (сценариев работы) или на поиске и анализе ключевых слов. При использовании правил система реагирует на ввод пользователя в соответствии с заранее определенным набором правил или команд.

Несмотря на относительную простоту данных алгоритмов они хорошо зарекомендовали себя в решении базовых задач. Однако нельзя не отметить факт слабой естественности ответов, получаемых пользователем, а также низкий уровень гибкости, ввиду заранее заданных путей поведения и ответов. Развитие ис-

кусственного интеллекта, в частности, машинного обучения и нейросетевых моделей, позволило использовать при разработке чат-ботов как техники машинного обучения для классификации пользовательских запросов и определения их намерений, создания иерархической структуры правил, позволяющих реализовать соответствующее задаче поведение, так и большие языковые модели (LLM), обладающих способностью распознавать и генерировать текст, подобный человеческому общению, способные обрабатывать сложные запросы, вести контекстные диалоги и предоставлять персонализированные ответы [1].

При использовании нейросетевых моделей нужно учитывать не только их способность к обработке запросов, но и различные способы их развертывания. Два основных подхода к использованию нейросетевых моделей: в качестве облачного сервиса и в локальном развертывании. Облачные сервисы, такие как предоставляемые YandexGPT, OpenAI GPT, Claude AI, предлагают удобство и масштабируемость, но сопровождаются периодическими расходами и требуют постоянного доступа к интернету, а также имеют ограниченные возможности для дообучения моделей в рамках конкретных поставленных задач. Локальное развертывание открытых моделей требует значительных вычислительных ресурсов и затрат на начальную настройку, но предоставляет полный контроль над данными и процессом обработки запросов, в том числе с возможностью создания дополнительных моделей в результате процесса дообучения.

Цель данного исследования заключается в сравнении классических подходов к разработке чат-ботов и использовании нейросетевых моделей на примере языковых моделей большого размера (LLM), для определения наиболее подходящего инструмента реализации чат-бота приемной комиссии университета.

Модели и методы исследования

В рамках исследования применимости рассматриваемых подходов является целесообразным применение метода построения дерева проблем [2], позволяющее определить подпроблемы, которые могут быть существенны как в решении поставленной задачи, позволяя детализировать требования к рассматриваемым техникам, так и в процессе выбора рассматриваемых способов к решению, определяя критерии сравнения решений.

В процессе выбора предлагается использование метода анализа иерархий (the Analytic Hierarchy Process, АНП), предложенный Т. Саати [3]. Метод позволяет обеспечивать согласованность между критериями и альтернативами, что позволяет избежать противоречий в процессе принятия решений. Суть метода анализа иерархий для случая нескольких критериев заключается в следую-

щем. Пусть существует множество альтернатив, которые необходимо ранжировать по важности, руководствуясь множеством критериев. Предполагается, что эксперт способен попарно сравнить критерии и альтернативы по критериям.

Рассмотрим численный метод расчета МАИ для одного критерия [4; 5]. Результатом опроса экспертов является матрица попарных сравнений альтернатив A .

Элементы матрицы $A = \{a_{ij}\}$ определяются по следующим правилам:

1. Если $a_{ij} = \alpha$, то $a_{ji} = 1 / \alpha$.
2. Если в результате сравнения альтернативы признаны равнозначными, то $a_{ij} = a_{ji} = 1$.

В качестве оценки решений α рассматривается суждение о паре объектов, принимающее значения от 1 до 9 (от «несравнимость» до «абсолютная значимость»).

$$A = \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{pmatrix}.$$

Получения собственного вектора матрицы выполняется следующим образом:

1. Для каждой строки матрицы A вычисляется ее среднее геометрическое получая вектора $a = (a_1, \dots, a_n)$

$$A = \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{pmatrix} \begin{matrix} a_1 = \sqrt[n]{1 \cdot a_{12} \cdot \cdots \cdot a_{1n}} \\ a_2 = \sqrt[n]{1/a_{12} \cdot 1 \cdot \cdots \cdot a_{2n}} \\ \vdots \\ a_n = \sqrt[n]{1/a_{1n} \cdot 1/a_{2n} \cdot \cdots \cdot 1} \end{matrix}.$$

2. Элементы вектора a нормируют, получая нормированный вектор s

$$S_i = \frac{a_i}{\sum_i a_i}.$$

3. Составляется вектор b , координатами которого служат суммы столбцов матрицы A .

4. Вычисляется скалярное произведение векторов b и s . Результатом является сумма λ_{max} , используемая при расчете индекса и отношения согласованности

5. Рассчитывается индекс согласованности и отношение согласованности, где CC — индекс случайной согласованности, представленный в табл. 1.

$$\text{ИС} = \frac{\lambda_{\max} - n}{n - 1}, \text{ОС} = \frac{\text{ИС}}{\text{СС}}.$$

Таблица 1

Индексы случайной согласованности

Порядок матрицы	3	4	5	6	7	8	9	10
Коэффициент связи	0,58	0,9	1,12	1,24	1,32	1,41	1,45	1,49

Если отношение согласованности $\text{ОС} < 0,15$, то составленная матрица попарных сравнений приемлемо согласована, а полученный вектор s есть вектор приоритетов сравниваемых альтернатив w .

При многокритериальном анализе, в отличие от рассмотренного способа с одним критерием итоговый вес альтернативы вычисляется путем произведения вектора весов критериев на матрицу, столбцами которой являются вектора приоритетов альтернатив, соответствующих критериям. Получение вектора весов критериев s осуществляется способом, аналогичным поиску вектора приоритетов сравниваемых альтернатив.

$$(w_1, \dots, w_i, \dots, w_n) = (s_1, \dots, s_i, \dots, s_n) \times$$

$$\times \begin{pmatrix} S_1(K_1) & \dots & S_i(K_1) & \dots & S_n(K_1) \\ \dots & \dots & \dots & \dots & \dots \\ S_1(K_i) & \dots & S_i(K_i) & \dots & S_n(K_i) \\ \dots & \dots & \dots & \dots & \dots \\ S_1(K_n) & \dots & S_i(K_n) & \dots & S_n(K_n) \end{pmatrix}.$$

Результаты

Корневой проблемой в рамках деятельности приемной комиссии является потеря потенциальных студентов, подходящих под критерии вступительных испытаний направлений высшего учебного заведения. Для более детального разбора выявленной проблемы было составлено визуальное представление в виде дерева проблем (рис. 1).

Полученное дерево проблем позволяет определить критерии. Ввиду наличия проблемы некачественных ответов на запросы абитуриентов важными критериями является их точность и релевантность, а также общее понимание контекста чат-бота. Так как выделяемые ресурсы на разработку конечны, необходима предусмотреть возможность интеграции чат-бота с существующими системами высшего учебного заведения и средствами связи, а также обеспе-

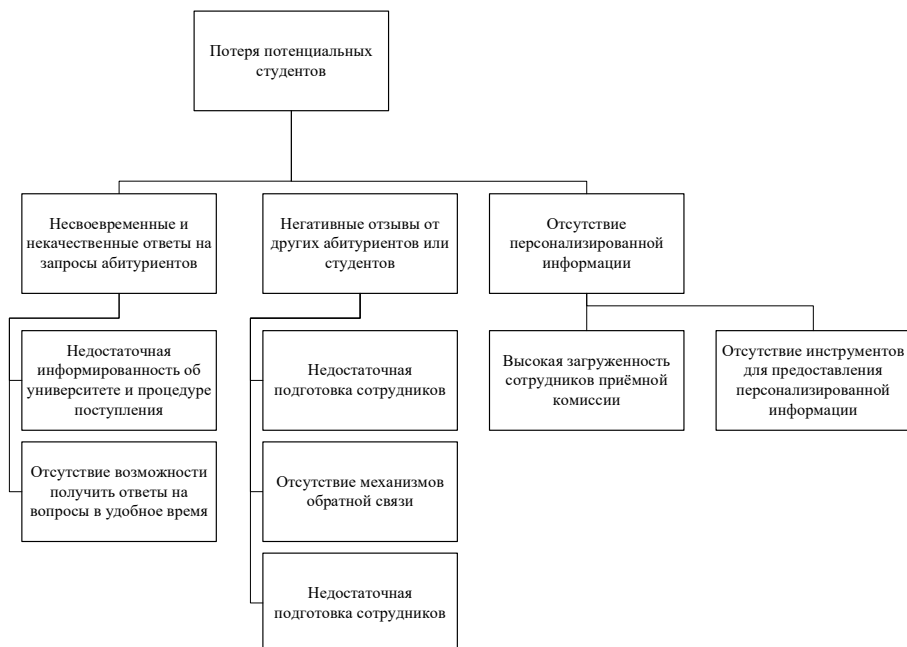


Рис. 1. Дерево проблем

чение приемлемый уровень стоимости реализации и поддержки, с приемлемым временем на разработку и внедрение. Решение должно быть достаточно масштабируемо, чтобы справляться с изменяющимся уровнем нагрузки на протяжении всего процесса работы приемной комиссии. В связи с этим основными критериями выбраны:

1. Точность и релевантность ответов — критерий определяет, насколько точно и соответствующе бот может отвечать на вопросы пользователей. Чат-бот должен быть способен понять вопросы и предоставлять релевантные и точные ответы. Высокая точность и релевантность ответов обеспечивают удовлетворение пользователей и повышают эффективность работы бота.

2. Понимание контекста — критерий связан с способностью чат-бота учитывать контекст предыдущих вопросов и ответов, а также уточнять непонятные или неоднозначные запросы пользователя. Чат-бот должен быть способен поддерживать беседу и учитывать предыдущий контекст для предоставления более точных и релевантных ответов.

3. Естественность языка — чат-бот должен обладать способностью вести диалог с пользователем естественным образом, так чтобы пользователь мог задавать вопросы и получать ответы, как если бы он общался с другим человеком. Важно, чтобы чат-бот использовал понятный и естественный язык, а также умел обрабатывать сленг, синонимы и контекстуальные нюансы.

4. Время на разработку и внедрение — критерий относится к затратам времени, необходимым для разработки и внедрения чат-бота. Различные подходы к разработке чат-ботов могут иметь разную сложность и требовать различного объема времени на разработку. Выбор подхода должен учитывать возможности команды разработчиков и требуемый срок внедрения.

5. Стоимость реализации и поддержки — критерий оценивает затраты на реализацию и последующую поддержку чат-бота. Разные подходы к разработке могут иметь различную стоимость. Некоторые подходы могут требовать значительных инвестиций в инфраструктуру и обучение моделей, в то время как другие могут быть более доступными с точки зрения бюджета.

В результате работы составлена анкета и проведен опрос экспертов в рамках предметной области: представители приемной комиссии, эксперты в области машинного обучения и работы с нейросетями, программисты и аналитики решающие похожие задачи, представители кафедры математических методов и информационных технологий ФГБОУ БГУ. Критерии и альтернативы представлены на рис. 2, а частные оценки альтернатив приведены в табл. 2.

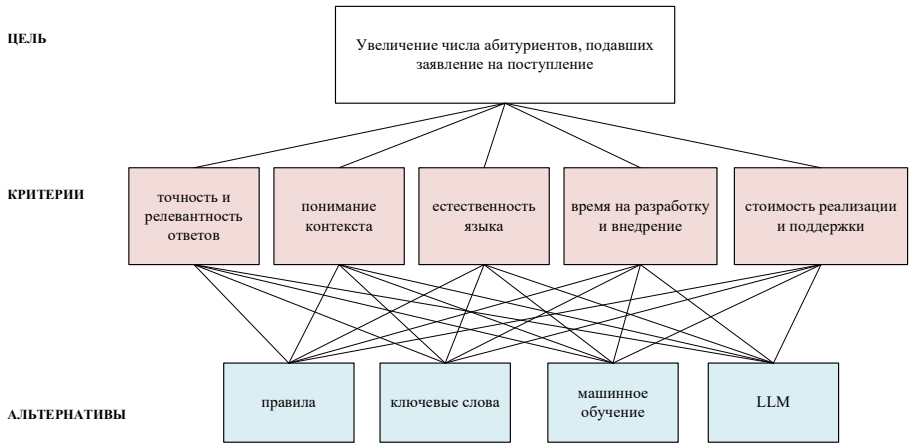


Рис. 2. Метод анализа иерархий, критерии и альтернативы

Таблица 2

Частные оценки альтернатив

Критерии	Значи- мость критерия	Оценки приоритетов альтернатив по критериям			
		Прави- ла	Ключевые слова	ML	LLM GPT
Точность и релевант- ность ответов	0,3859	0,2144	0,0968	0,2144	0,4744
Понимание контекста	0,2697	0,0753	0,2320	0,2493	0,4433

Критерии	Значи- мость критерия	Оценки приоритетов альтернатив по критериям			
		Прави- ла	Ключевые слова	ML	LLM GPT
Естественность языка	0,0607	0,2865	0,2833	0,2026	0,1703
Время на разработку и внедрение	0,1057	0,3369	0,2833	0,2382	0,1416
Стоимость реализа- ции и поддержки	0,1779	0,4794	0,2166	0,1957	0,1083
<i>Глобальный приоритет</i>		<i>0,2414</i>	<i>0,1856</i>	<i>0,2223</i>	<i>0,3473</i>

Результаты экспертного опроса и расчеты согласно методу анализа иерархий показали, что оценка глобального приоритета наиболее высока у подхода, использующего большие языковые модели. Анализ показал, что нейросетевые модели обеспечивают значительно более высокую точность и релевантность ответов по сравнению с «классическими» подходами.

Несмотря на высокий общий уровень естественности языка у больших языковых моделей, по результатам опроса, он находится в среднем на более низком уровне чем у «классических» подходов к разработке чат-ботов, ввиду большого количества отклонений от заданных базовых и требуемых шаблонов и стандартов поведения, снижающих общий уровень пользовательского опыта в результате некорректной обработки запроса.

Обсуждение результатов и заключение

В процессе исследования были рассмотрены различные методы разработки чат-ботов для приемной комиссии университета. Изучались как «классические» методы, базирующиеся на правилах и ключевых словах, так и современные подходы, использующие машинное обучение и большие языковые модели (LLM). Проведенный в работе анализ показал, что современные методы на основе нейросетей превосходят «классические» по точности, релевантности и естественности ответов. Однако внедрение и поддержка таких систем требуют значительных ресурсов. Высокие затраты и необходимость постоянного доступа к интернету для облачных решений могут стать ограничивающими факторами.

Несмотря на эти ограничения, нейросетевые модели предоставляют мощные возможности для разработки чат-ботов. Они позволяют гибко настраивать и масштабировать системы. Большинство современных платформ для создания чат-ботов уже поддерживают интеграцию с LLM, что упрощает их внедрение.

Комбинирование «классических» методов с нейросетевыми моделями может стать оптимальным решением, позволяющим ис-

пользовать сильные стороны каждого подхода. Например, применение LLM для генерации ответов в рамках заранее определенных правил может значительно улучшить взаимодействие чат-бота с пользователями.

Список использованной литературы

1. Информационно-технологическое обеспечение юридической деятельности (LegalTech) : учебник / ред. А.В. Минбалеев. — Москва : Проспект, 2023. — 368 с.
2. Винокурова Е.Л. Использование методики «Дерево проблем» для определения проблем, влияющих на качество образования студентов технического вуза / Е.Л. Винокурова, А.О. Гайдукова, Г.М. Гринберг. — EDN TAPQTF // Актуальные проблемы авиации и космонавтики. — 2012. — Т. 2, № 8. — С. 475–477.
3. Саати Т.Л. Принятие решений. Метод анализа иерархий / Т.Л. Саати. — Москва : Радио и связь, 1993. — 278 с.
4. Саати Т.Л. Принятие решений при зависимостях и обратных связях: Аналитические сети / Т.Л. Саати. — Москва: Изд-во ЛКИ, 2008. — 360 с.
5. Затеса А.В. Использование метода анализа иерархии для выбора информационной системы / А.В. Затеса. — EDN NYPDDV // Статистика и экономика. — 2010. — № 6. — С. 164–167.

References

1. Minbaleev A.V. (ed.). *Information and technological support for legal activities (LegalTech)*. Moscow, Prospekt Publ., 2023. 368 p.
2. Vinokurova E.L., Gaidukova A.O., Grinberg G.M. Using the “Tree of Problems” methodology to identify problems affecting the quality of education of students at a technical university. *Aktual'nye problemy aviatsii i kosmonavтики = Topical Issues of Airspace and Outer Space*, 2012, vol. 2, no. 8, pp. 475–477. (In Russian). EDN: TAPQTF.
3. Saaty T.L. *Decision Making for Leaders : the Analytic Hierarchy Process for Decisions in a Complex World*. Pittsburgh, University of Pittsburgh, 1988. 306 p. (Russ. ed.: Saaty T.L. *Decision making. Hierarchy analysis method*. Moscow, Radio i svyaz' Publ., 1993. 278 p.).
4. Saaty T.L. *Decision Making with Dependence and Feedback: The Analytic Network Process: the Organization and Prioritization of Complexity*. RWS Publications, 1996. 370 p. (Russ. ed.: Saaty T.L. *Decision Making with Dependence and Feedback : The Analytic Network Process*. Moscow, LKI Publ., 2008. 360 p.).
5. Zatesa A.V. Using the hierarchy analysis method to select an information system. *Statistika i ekonomika = Statistics and Economics*, 2010, no. 6, pp. 164–167. (In Russian). EDN: NYPDDV.

Информация об авторах

Родионов Алексей Владимирович — кандидат технических наук, доцент, кафедра математических методов и цифровых технологий, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: RodionovAV@bgu.ru.

Корж Даниил Андреевич — аспирант, ассистент, кафедра математических методов и цифровых технологий, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: 010628@bgu.ru.

Information about the Authors

Aleksei V. Rodionov — PhD in Technical Sciences, Associate Professor, Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, Russian Federation, e-mail: RodionovAV@bgu.ru.

Daniil A. Korzh — PhD Student, Assistant Lecturer, Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, Russian Federation, e-mail: 010628@bgu.ru.

Вклад авторов

Все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the Authors

The authors contributed equally to this article. The authors declare no conflicts of interests.

Для цитирования

Родионов А.В. Выбор инструментов построения чат-бота приемной комиссии учебного заведения / А.В. Родионов, Д.А. Корж. — DOI 10.17150/2713-1734.2024.6(3).367-376. — EDN IXAVJL // *System Analysis & Mathematical Modeling*. — 2024. — Т. 6, № 3. — С. 367–376.

For Citation

Rodionov A.V., Korzh D.A. Tools for Building a Chatbot for the Admissions Commission of an University. *System Analysis & Mathematical Modeling*, 2024, vol. 6, no. 3, pp. 367–376. (In Russian). EDN: IXAVJL. DOI: 10.17150/2713-1734.2024.6(3).367-376.