

# **UsERQA: An LLM-Driven User-Aware Community Question Answering System**

**Seyyede Zahra Aftabi<sup>1</sup>, Saeed Farzi<sup>1\*</sup>**

<sup>1</sup> Faculty of Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran

Received: 09 September 2024, Revised: 05 February 2025, Accepted: 16 February 2025  
Paper type: Research

## **Abstract**

In the present era, question-and-answer communities have become vibrant platforms for sharing knowledge. Every year, millions of questions are posted on these forums with the hope of receiving answers from human experts. Nonetheless, many of these questions fail to receive timely or accurate answers due to experts' limited time or being duplicates. In recent years, a large body of research has focused on identifying entailed questions within community archives and using their accepted answers to fulfill the information needs of newly posed questions. Most of these studies match questions syntactically and semantically while resorting to external knowledge injection or increased model complexity to enhance question understanding. However, the critical role that the topics typically explored by questioners play in disambiguating their queries has been overlooked. This research addresses this gap by introducing UsERQA, a novel retrieval-augmented generation (RAG)-based question-answering system incorporating user knowledge. UsERQA utilizes large language models to represent the questioner's knowledge as a sequence of topical tags. In addition, it employs a question entailment recognition process as a post-retrieval strategy, with a new constraint, mandating the alignment between entailed questions and the questioner's knowledge. Afterward, another large language model generates the final answer using the accepted answers of top entailed questions as context. The goal is to imitate human writing patterns and leverage the knowledge contained in human responses to produce high-quality answers. Experimental results on the CQAD-ReQuEST dataset indicate the efficiency of UsERQA in modeling user knowledge and producing more accurate responses than its user-agnostic counterpart.

**Keywords:** Community question answering systems, Answer generation, Recognizing question entailment, User modeling, Query-focused multi-document summarization.

---

\* Corresponding Author's email: saeedfarzi@kntu.ac.ir

## UsERQA: سیستم پاسخ‌گویی به پرسش‌های انجمن آگاه به کاربر مبتنی بر مدل‌های زبانی بزرگ

سیده زهرا آفتابی<sup>۱</sup>، سعید فرضی<sup>\*۱</sup>

<sup>۱</sup> دانشکده مهندسی کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

تاریخ دریافت: ۱۴۰۳/۰۶/۱۹ تاریخ بازبینی: ۱۴۰۳/۱۱/۱۷ تاریخ پذیرش: ۱۴۰۳/۱۱/۲۸

نوع مقاله: پژوهشی

### چکیده

در عصر حاضر، انجمن‌های پرسش و پاسخ، به بستریایی پویا برای تبادل دانش بدل شده‌اند. سالانه میلیون‌ها پرسش به امید دریافت پاسخ از متخصصین، راهی این انجمن‌ها می‌شوند. اما شمار زیادی از آنها به دلیل محدودیت زمان و منابع متخصصین یا داشتن محتوای تکراری، از دریافت پاسخ صحیح و زودهنگام بی‌نصیب می‌مانند. در این راستا، مطالعات بسیاری به شناسایی پرسش‌های با مضمون مشابه پرسش ورودی در بایگانی انجمن و بهره‌مندی از پاسخ‌های تأییدشده آنها جهت رفع نیاز اطلاعاتی پرسش پرداخته‌اند. عمده این پژوهش‌ها، از تطابق ویژگی‌های نحوی و معنایی زوج پرسش استفاده نموده و برای درک بهتر پرسش‌ها، به تزریق دانش خارجی یا افزایش پیچیدگی مدل متوسل شده‌اند. در این میان، نقش کلیدی دایره موضوعات مورد مطالعه پرسشگر در رفع ابهام از محتوای پرسش مغفول مانده‌است. پژوهش حاضر با ارائه یک سیستم مولد بازبازی افزوده برای پاسخگویی به پرسش‌ها موسوم به UsERQA که مبتنی بر مدل‌سازی دانش پرسشگر است، به رفع این شکاف تحقیقاتی می‌پردازد. UsERQA با بهره‌گیری از مدل‌های زبانی بزرگ، دانش پرسشگر را به صورت دنباله‌ای از برچسب‌های موضوعی توصیف نموده و شرط همسو بودن پرسش‌های متضمن با دانش پرسشگر را به فرآیند شناسایی پرسش‌های متضمن که یک فرآیند پسابازی است می‌افزاید. سپس، یک مدل زبانی دیگر، به تولید پاسخی واحد بر پایه برترین پاسخ‌ها مبادرت می‌ورزد. به موجب این فرآیند می‌توان از سبک نوشتاری و دانش نهفته در پاسخ‌های انسانی، جهت تولید پاسخ باکیفیت، الگوبرداری کرد. نتایج آزمایش‌ها روی دادگان CQAD-ReQuEST، عملکرد موفق UsERQA در مدل‌سازی کاربر و بهبود کیفیت پاسخ‌ها نسبت به مدل مستقل از کاربر را نشان داد.

**کلیدواژگان:** سیستم‌های پاسخگویی به پرسش‌های انجمن، تولید پاسخ، شناسایی پرسش‌های متضمن، مدل‌سازی کاربر، خلاصه‌سازی چندانندی متمرکز بر پرسمان.

\* رایانامه نویسنده مسؤول: saeedfarzi@kntu.ac.ir

## ۱- مقدمه

شناسایی پرسش‌های متضمن در میان انبوه پرسش‌های بایگانی، به این دلیل که افراد مختلف، پرسش‌های مشابه را با واژگان و ساختارهای زبانی متفاوت بیان می‌کنند [۱۰، ۱۷، ۱۸]، فرآیندی سخت و پیچیده است. مثال ۲ در جدول ۱، دو پرسش را نمایش می‌دهد که فاقد همپوشانی واژگانی<sup>۶</sup> بوده اما از نظر معنایی، متضمن یکدیگر می‌باشند. در حالتی دیگر، مطابق مثال ۳، دو پرسش ممکن است با وجود همپوشانی واژگانی، متضمن یکدیگر نباشند. در این راستا، اکثر پژوهش‌ها از مقایسه واژگان و ساختارهای نحوی زوج پرسش فراتر رفته و تمرکز خود را بر درک عمیق‌تر معانی، با تکیه بر اطلاعات بافتاری<sup>۷</sup> [۱۹-۲۲] و منابع دانش خارجی [۲۳-۲۶] معطوف نموده‌اند. با این وجود، مروری بر مطالعات انجام شده در این حوزه، وجود یک شکاف تحقیقاتی، مبنی بر نادیده گرفته شدن دانش و علایق کاربران در فرآیند بازیابی پرسش‌های متضمن را هویدا می‌سازد.

## جدول ۱. نمونه‌هایی از زوج پرسش‌های متضمن و غیرمتضمن

مثال ۱. برچسب: متضمن

1. Why did they cut out the expanding granuloma which was due to Coccidiomycosis immitis, rather than just treat it medically?
2. Why did they cut out the Coccidiomycosis granuloma rather than treat it medically?

مثال ۲. برچسب: متضمن

1. Is this statement: "the letters have been troubling me since over a year" grammatically correct? Or should I instead use: "...for over a year".
2. Since is used to refer to a point of time in the past and for is used with time interval. However I've come across this phrase: "I haven't seen you since ages." Is this a correct usage of "since" since ages denotes an interval of time?

مثال ۳. برچسب: غیرمتضمن

1. Is this statement: "the letters have been troubling me since over a year" grammatically correct? Or should I instead use: "...for over a year".
2. My student writes a statement like this: "She hasn't gone out with him since one year ago." so, I want to know if this sentence is correct or wrong grammatically.

پژوهش حاضر بر آن است تا از زمینه‌های موضوعی مورد علاقه پرسشگر که به‌طور مکرر به جستجو پیرامون آنها می‌پردازد، به عنوان دانش کمکی جهت شناسایی دقیق‌تر پرسش‌های متضمن بهره‌مند گشته و بدین وسیله، صحت پاسخ‌های تولید شده برای کاربر را بهبود بخشد. برای این منظور، یک سیستم جامع پاسخگویی به پرسش با نام UserQA<sup>۸</sup>، متشکل از زنجیره‌ای از مدل‌های زبانی بزرگ، ارائه شده است. این سیستم در چهار مؤلفه اصلی خود به مدلسازی دانش پرسشگر براساس سوابق فعالیت‌های وی در انجمن، بازیابی فهرستی از پرسش‌های مرتبط با پرسش ورودی، شناسایی برترین پرسش‌های متضمن در میان پرسش‌های بازیابی شده و در پایان، ارائه یک پاسخ

در دو دهه گذشته، با ظهور و توسعه انجمن‌های پرسش و پاسخ، تحول شگرفی در عرصه بازیابی اطلاعات و تبادل دانش به وجود آمده است. از جمله انجمن‌های محبوب در زبان انگلیسی می‌توان Stack Overflow، Quora و iCliniq را نام برد. کاربران این انجمن‌ها قادرند نیاز اطلاعاتی خود را در قالب پرسش‌هایی به زبان طبیعی مطرح نموده و پاسخ‌ها و دیدگاه‌های مختلف را از سایر کاربران انجمن، که از این پس با عنوان متخصص از آنها یاد می‌شود، دریافت نمایند [۱]. به محض آنکه پرسشگر، یکی از پاسخ‌های دریافتی را به‌عنوان بهترین پاسخ تأیید نماید، پرسش، حل شده<sup>۱</sup> تلقی شده و در بایگانی<sup>۲</sup> انجمن ذخیره می‌شود [۱-۴]. به این ترتیب، با گذر زمان، بایگانی انجمن به مخزنی غنی از اطلاعات مبدل می‌گردد.

بقای یک انجمن پرسش و پاسخ در گرو جلب رضایت کاربران و ترغیب آنها به مشارکت بیشتر است. در این راستا، سرعت حصول پاسخ و صحت پاسخ، دو رکن اصلی برای رضایت کاربران محسوب می‌شوند. این در حالیست که مطابق آمار منتشر شده از انجمن‌های مختلف، سالانه شمار زیادی از پرسش‌ها از دریافت پاسخ بی‌بهره می‌مانند [۵، ۶]. حال آنکه برای بسیاری از پرسش‌های جدیدالورود، پرسش‌های متضمنی<sup>۳</sup> در بایگانی انجمن وجود دارند که پیش‌تر، پاسخ صحیحی دریافت نموده‌اند [۷، ۸]. یک پرسش متضمن، پرسشی با معنای مشابه است که پاسخ صحیح ارائه شده برای آن می‌تواند به عنوان یک پاسخ صحیح کامل یا بخشی از پاسخ صحیح برای پرسش جدید نیز ارائه گردد [۹]. از این رو، انجمن‌ها، کاربرانشان را به جستجو برای پرسش‌های متضمن و مطالعه پاسخ‌های آنها پیش از اقدام به ثبت یک پرسش جدید تشویق می‌کنند. با این وجود، بسیاری از کاربران، علیرغم صرف زمان قابل توجه برای جستجو، در دستیابی به پاسخ ناکام مانده و با این باور که پرسش آنها منحصر به فرد است، به ثبت پرسش‌های تکراری<sup>۴</sup> مبادرت می‌ورزند [۱۰، ۱۱]. انباشت تدریجی پرسش‌های تکراری نیز تلاش مضاعف متخصصین برای پاسخگویی و افزایش زمان انتظار کاربران برای دریافت پاسخ را در پی دارد. به همین علت، ساخت و توسعه سیستم‌های شناسایی پرسش‌های متضمن<sup>۵</sup> (RQE) که قادر باشند در سریع‌ترین زمان ممکن، فهرستی از پرسش‌های متضمن را بازیابی نمایند، به یکی از حوزه‌های تحقیقاتی پرمخاطب مبدل گشته است [۸، ۱۰-۱۶].

<sup>۵</sup> Recognizing question entailment (RQE)<sup>۶</sup> Word overlap<sup>۷</sup> Contextual<sup>۸</sup> User-aware Entailment Recognition-based Question Answering<sup>۱</sup> Resolved<sup>۲</sup> Archive<sup>۳</sup> Entailed questions<sup>۴</sup> Duplicate questions

هستند. در دسته اول، مجموعه‌ای از ویژگی‌ها به صورت دستی از پرسش‌ها استخراج می‌شوند [۲۸، ۲۹]. مهمترین نقطه ضعف این رویکردها را می‌توان ناتوانی معیارهای شباهت سنتی، در تشخیص هم‌ارزی‌های معنایی عنوان کرد [۳۰، ۳۱]. در دسته دوم، مهندسی ویژگی‌ها به مدل‌های شبکه عصبی سپرده می‌شود. تعدادی از پژوهش‌ها، بردارهای تعبیه را در سطح کلمه<sup>۹</sup> [۳۱، ۳۲] و تعدادی در سطح جمله<sup>۱۰</sup> [۳۳] تولید نموده‌اند. در این میان، تحلیل خطاها از ناتوانی مدل‌های تعبیه ایستا<sup>۱۱</sup> در مواجهه با چالش‌هایی نظیر چندمعنایی<sup>۱۲</sup>، اصطلاحات تخصصی و واژگان خارج از لغت‌نامه حکایت داشته‌است. از این رو، با پیدایش مدل‌های تعبیه بافتاری، بسیاری از محققین به استفاده از آنها گرایش یافتند.

گروهی از پژوهش‌ها تلاش نمودند تا با تزریق دانش از طرق مختلف، بازنمایی‌های برداری را غنی‌تر نمایند. استفاده از مدل‌های پیش‌آموزش‌دیده<sup>۱۳</sup> روی داده‌های خاص-دامنه<sup>۱۴</sup>، یکی از راهکارهایی است که محققین برای تزریق دانش تخصصی پیشنهاد کردند. نگوین و همکاران [۳۴] با هدف رفع ابهام از اصطلاحات تخصصی و اختصارات<sup>۱۵</sup> در دامنه پزشکی، بهره‌برداری همزمان از تعبیه خاص-دامنه Bio-BERT و تعبیه حاصل از مدل دامنه-باز BERT<sup>۱۶</sup> را پیشنهاد نمودند. محققین در [۳۵]، کیفیت تعبیه Word2vec در یک شبکه مبتنی بر Bi-LSTM<sup>۱۷</sup> را با تعبیه حاصل از Bio-BERT در یک شبکه عصبی چندلایه مقایسه نموده و عملکرد برتر Word2vec را شاهد بودند. داده‌افزایی<sup>۱۸</sup>، دیگر راهکاری است که پژوهشگران به منظور افزایش تعمیم‌پذیری بازنمایی‌ها، جلوگیری از بیش‌برازش<sup>۱۹</sup> و تأمین داده برای آموزش مدل‌های زبانی به آن رجوع کردند. مرینی و همکاران [۳۶] با اثبات رابطه هم‌ارزی میان وظایف RQE و خلاصه‌سازی پرسش، رویکردی برای داده‌افزایی جهت تسهیل یادگیری چندوظیفه‌ای<sup>۲۰</sup> پیشنهاد دادند. آنها از هر زوج پرسش در دادگان خلاصه‌سازی، یک زوج معادل برای دادگان RQE تولید نمودند. برخی محققین نیز استفاده از یادگیری انتقالی<sup>۲۱</sup> را توصیه کردند. زو و همکاران [۳۷]، تنظیم دقیق<sup>۲۲</sup> مدل MT-DNN<sup>۲۳</sup> برای وظیفه استنتاج زبان طبیعی<sup>۲۴</sup> را در بهبود عملکرد آن در وظیفه RQE مؤثر یافتند.

صحیح و یکپارچه به پرسشگر بر پایه پاسخ‌های تأییدشده<sup>۱</sup> برای پرسش‌های متضمن می‌پردازد. در واقع، سیستم پیشنهادی پژوهش حاضر، یک سیستم مولد بازبازی‌افزوده<sup>۲</sup> (RAG) توسعه‌یافته است که به یک مؤلفه پسابازی<sup>۳</sup> مجهز گشته است. این سیستم نه تنها از دانش جهانی<sup>۴</sup> مدل‌های زبانی بزرگ برای تولید پاسخ‌هایی مشابه پاسخ‌های انسانی بهره می‌برد، بلکه از چالش‌هایی نظیر تولید پاسخ‌های غیرواقعی<sup>۵</sup> که گریبان‌گیر این مدل‌ها است مصون می‌ماند. افزون بر این، با تکیه بر دانش و علایق پرسشگر، از ورود پاسخ‌های کاندید نامرتبط با نیاز اطلاعاتی به مرحله تولید پاسخ ممانعت نموده و اعتبار، دقت و صحت بالاتر را در پاسخگویی به ارمغان می‌آورد.

به‌طور کلی، نوآوری‌های مطالعه حاضر به شرح زیر است:

۱. ارائه سیستمی برای پاسخگویی خودکار و سریع به پرسش‌های انجمن و بهبود صحت و کیفیت پاسخ‌ها با کاهش خطا در مؤلفه‌های بازبازی پرسش و شناسایی پرسش‌های متضمن
۲. بهبود فرآیند شناسایی پرسش‌های متضمن از طریق افزودن شرط همسویی پرسش‌های کاندید با دانش و علایق پرسشگر
۳. مدل‌سازی دانش پرسشگر در قالب دنباله‌ای از برجسب‌های موضوعی با تکیه بر سوابق پرسش‌های وی در انجمن

ادامه این مقاله به شرح زیر سازمان‌دهی شده است. در بخش دوم مروری بر کارهای پیشین در حوزه شناسایی پرسش‌های متضمن ارائه می‌گردد. در بخش سوم، سیستم پیشنهادی و عملکرد هر یک از مؤلفه‌های آن تشریح می‌شود. بخش چهارم به معرفی دادگان، تنظیمات سیستم، نتایج آزمایش‌ها و بحث پیرامون آنها پرداخته و بخش پنجم، با جمع‌بندی و ارائه پیشنهادات به آن خاتمه می‌دهد.

## ۲- کارهای مرتبط

با معرفی وظیفه تشخیص تضمین پرسش‌ها (RQE) توسط آباچا و همکاران در سال ۲۰۱۶ [۲۷]، مطالعات بسیاری، به ارائه مدل‌های RQE و ارتقاء عملکرد آنها از زوایای مختلف پرداختند.

بطور کلی، مدل‌های RQE از منظر روش تولید بازنمایی‌های<sup>۶</sup> برداری به دو دسته «مبتنی بر ویژگی<sup>۷</sup>» و «مبتنی بر تعبیه<sup>۸</sup>» قابل تقسیم

<sup>13</sup> Pre-trained

<sup>14</sup> Domain-specific

<sup>15</sup> Abbreviations

<sup>16</sup> Bidirectional Encoder Representations from Transformers

<sup>17</sup> Bidirectional Long Short-Term memory

<sup>18</sup> Data augmentation

<sup>19</sup> Overfitting

<sup>20</sup> Multi-task learning

<sup>21</sup> Transfer learning (TL)

<sup>22</sup> Fine-tune

<sup>23</sup> Multi-task Deep Neural Network

<sup>24</sup> Natural Language Inference (NLI)

<sup>1</sup> Accepted answers

<sup>2</sup> Retrieval-augmented generation (RAG)

<sup>3</sup> Post-retrieval

<sup>4</sup> (Parametric) world knowledge

<sup>5</sup> Hallucination

<sup>6</sup> Representation

<sup>7</sup> Feature-based approaches

<sup>8</sup> Embedding-based approaches

<sup>9</sup> Word embedding

<sup>10</sup> Sentence-level embedding

<sup>11</sup> Static

<sup>12</sup> Polysemy

یا عمومی از طریق گراف‌های دانش<sup>۱۲</sup> پرداخته‌اند. گودوین و همکاران [۲۶] از گراف دانش ConceptNet برای کسب دانش مستقل از وظیفه<sup>۱۳</sup> بهره‌مند گشتند. یاداو و همکاران [۲۵] یک روش رمزگذاری دوگانه<sup>۱۴</sup> مبتنی بر گراف دانش را معرفی کردند که در آن، همزمان با استخراج اطلاعات بافتاری محلی توسط مدل‌های زبانی، اطلاعات عمومی نیز در سه گام، شامل ساخت گراف هم‌آیندی<sup>۱۵</sup> کلمات، تکمیل آن به کمک گراف‌های دانش خاص-دامنه و تولید تعبیه گرافی با استفاده از شبکه‌های پیچشی گرافی، استخراج می‌شد. آنگاه، یک مدل ترانسفورمر<sup>۱۶</sup>، دو تعبیه مذکور را در هم آمیخته و بازنمایی حاصل به یک شبکه عصبی ارسال می‌شد. نتایج آنها توانایی این مدل در درک بهتر پرسش‌ها در دامنه پزشکی را نشان داد. اگرچه، پیچیدگی، تعمیم‌پذیری و مقیاس‌پذیری را می‌توان سه چالش مهم در مطالعه آنها دانست. جیانگ و همکاران [۴۱] از گراف ConceptNet، پایگاه دانش Wiktionary و مدل تعبیه BERT برای تولید بازنمایی دانش، و از یک مدل متشکل از چندین شبکه CNN و لایه‌های توجه<sup>۱۷</sup> برای تولید بازنمایی زوج پرسش استفاده نمودند. سپس بازنمایی‌های حاصل را با بهره‌گیری از یک مکانیزم گیت<sup>۱۸</sup> با یکدیگر تلفیق کردند. آزمایش‌ها اثربخشی و مقیاس‌پذیری روش پیشنهادی آنها روی دادگان مختلف را تأیید نموده است اما تحلیلی بر پیچیدگی مدل ارائه نشده است.

گروهی از محققین از فراداده‌های<sup>۱۹</sup> پرسش به عنوان یک منبع دانش تکمیلی استفاده نمودند. برای مثال، از دیدگاه سان و همکاران [۴۲]، کیفیت پاسخ پرسش‌های کاندید، عامل مهمی است که باید در کنار شباهت‌های نحوی و معنایی، برای رتبه‌بندی پرسش‌ها لحاظ گردد. آنها پس از بازیابی پرسش‌های مرتبط به کمک نمایه معکوس<sup>۲۰</sup> و TF-IDF<sup>۲۱</sup>، کیفیت پاسخ هر پرسش مرتبط را توسط یک مدل مبتنی بر ELMo<sup>۲۲</sup> سنجیده و این امتیاز را به همراه سایر امتیازات، به الگوریتم رتبه‌بندی ارائه نمودند. یافته‌های آنها، صرف‌نظر از افزایش حجم محاسبات زمان واقعی، مؤثر بودن ویژگی‌های منتخب را به اثبات رساند. قاسمی و شاکری [۲۳] یک مدل زبانی مبتنی بر ترجمه را برای وظیفه بازیابی پرسش ارائه دادند بطوریکه، علاوه بر بدنه پرسش، موضوع، عنوان و پاسخ‌ها نیز در آن مشارکت داشتند.

گروه دیگری از پژوهش‌ها، تقویت معماری مدل را در پیش گرفتند. به همین علت، گذار تدریجی از مدل‌های سنتی یادگیری ماشین [۲۸، ۲۹، ۳۲-۳۴] به سمت مدل‌های پیشرفته یادگیری عمیق را شاهد هستیم. به عنوان نمونه، فصیح و همکاران [۲۱] از ظرفیت شبکه‌های حافظه طولانی کوتاه‌مدت<sup>۱</sup> و شبکه‌های عصبی پیچشی<sup>۲</sup> (CNN) برای دریافت همزمان وابستگی‌های بلندمدت و محلی بهره‌مند شدند. بازنمایی‌های زوج پرسش، پس از تولید توسط مدل‌های فوق، برای دسته‌بندی به یک مدل پرسپترون<sup>۳</sup> ارائه شدند.

برخی مطالعات با ارائه مدل‌های پیچیده چندوظیفه‌ای، درصد تزیق دانش به مدل برآمدند. یادگیری چندوظیفه‌ای، از طریق لایه‌های اشتراکی، امکان تبادل دانش بین وظایف مرتبط را فراهم می‌آورد [۳۸]. ژو و همکاران [۳۹] یک شبکه تخصصی<sup>۴</sup> برای وظایف RQE و پاسخ‌گویی به پرسش، شامل یک رمزگذار<sup>۵</sup> مشترک، یک متمایزگر<sup>۶</sup> وظیفه و دو دسته‌بند طراحی نمودند. در این شبکه، متمایزگر با تشخیص نوع وظیفه، رمزگذار را بر حذف ویژگی‌های خاص-وظیفه<sup>۷</sup> از بازنمایی‌ها ترغیب می‌نمود. ناتوانی در درک پرسش‌های مبهم و اختصارات را می‌توان از نقاط ضعف روش پیشنهادی آنها برشمرد. شانگ و همکاران [۴۰] از یادگیری بدون نظارت<sup>۸</sup> چندوظیفه‌ای با یک رمزگذار مشترک، به منظور ارتقاء کیفیت بازنمایی‌ها بهره جستند. آنها وظایف تولید کلیدواژه برای پرسش و بازنمایی پرسش را به صورت توأم آموزش دادند. اگرچه آزمایش‌ها از عملکرد مؤثر مدل آنها در برخورد با پرسش‌های کوتاه حکایت دارد، اما کارایی آن برای پرسش‌های طولانی و پیچیده، مورد سؤال است. مرینی و همکاران [۱۹] یک مدل مبتنی بر BART<sup>۹</sup> شامل وظایف خلاصه‌سازی پرسش و RQE پیشنهاد نمودند. نوآوری اصلی آنها تعریف دو مؤلفه رمزگشا<sup>۱۰</sup> با اشتراک تدریجی-نرم<sup>۱۱</sup> در کنار یک رمزگذار مشترک بود. نتایج تجربی آنها مؤید نقش مؤثر خلاصه‌سازی پرسش در تقویت عملکرد RQE بود.

به عقیده برخی محققین، مدل‌های عمیق موجود در زمینه RQE، به دلیل اتکاء صرف بر دانش داخلی، فاقد جامعیت و ناتوان در استدلال در سطوح عمیق هستند [۴۱]. به همین علت، در ادبیات پژوهش می‌توان مطالعاتی را نیز یافت که به تزیق دانش دامنه‌ای

<sup>12</sup> Knowledge Graph (KG)

<sup>13</sup> Task-agnostic

<sup>14</sup> Dual-encoding

<sup>15</sup> Co-occurrence

<sup>16</sup> Transformer

<sup>17</sup> Attention layers

<sup>18</sup> Gating mechanism

<sup>19</sup> Metadata

<sup>20</sup> Inverted index

<sup>21</sup> Term Frequency-Inverse Document Frequency

<sup>22</sup> Embeddings from Language Models

<sup>1</sup> Long Short-Term Memory network (LSTM)

<sup>2</sup> Convolutional Neural Network (CNN)

<sup>3</sup> Multi-Layer Perceptron (MLP)

<sup>4</sup> Adversarial

<sup>5</sup> Encoder

<sup>6</sup> Discriminator

<sup>7</sup> Task-specific

<sup>8</sup> Unsupervised

<sup>9</sup> Bidirectional Auto-Regressive Transformers

<sup>10</sup> Decoder

<sup>11</sup> Gradually-soft

بنای اصلی رویکرد آنها بر دو مدل BERT و GPT، به ترتیب برای پیش‌بینی دسته موضوعی پرسش جدید و مبسوط نمودن آن از طریق تولید و الحاق پاسخ، استوار بود. به این ترتیب، مدل مبتنی بر ترجمه، پرسش‌های کاندید را براساس شباهتشان به پرسش

بسط‌یافته رتبه‌بندی می‌کرد. اگرچه مدل آنها در اختصاص رتبه‌های بالاتر به پرسش‌های مرتبط، موفق عمل کرد، اما سربار محاسباتی برخط، یکی از معایب آن به شمار می‌آید. جدول ۲ به مرور اجمالی مطالعات پیشین می‌پردازد.

جدول ۲. مروری بر کارهای پیشین در حوزه شناسایی متون متضمن

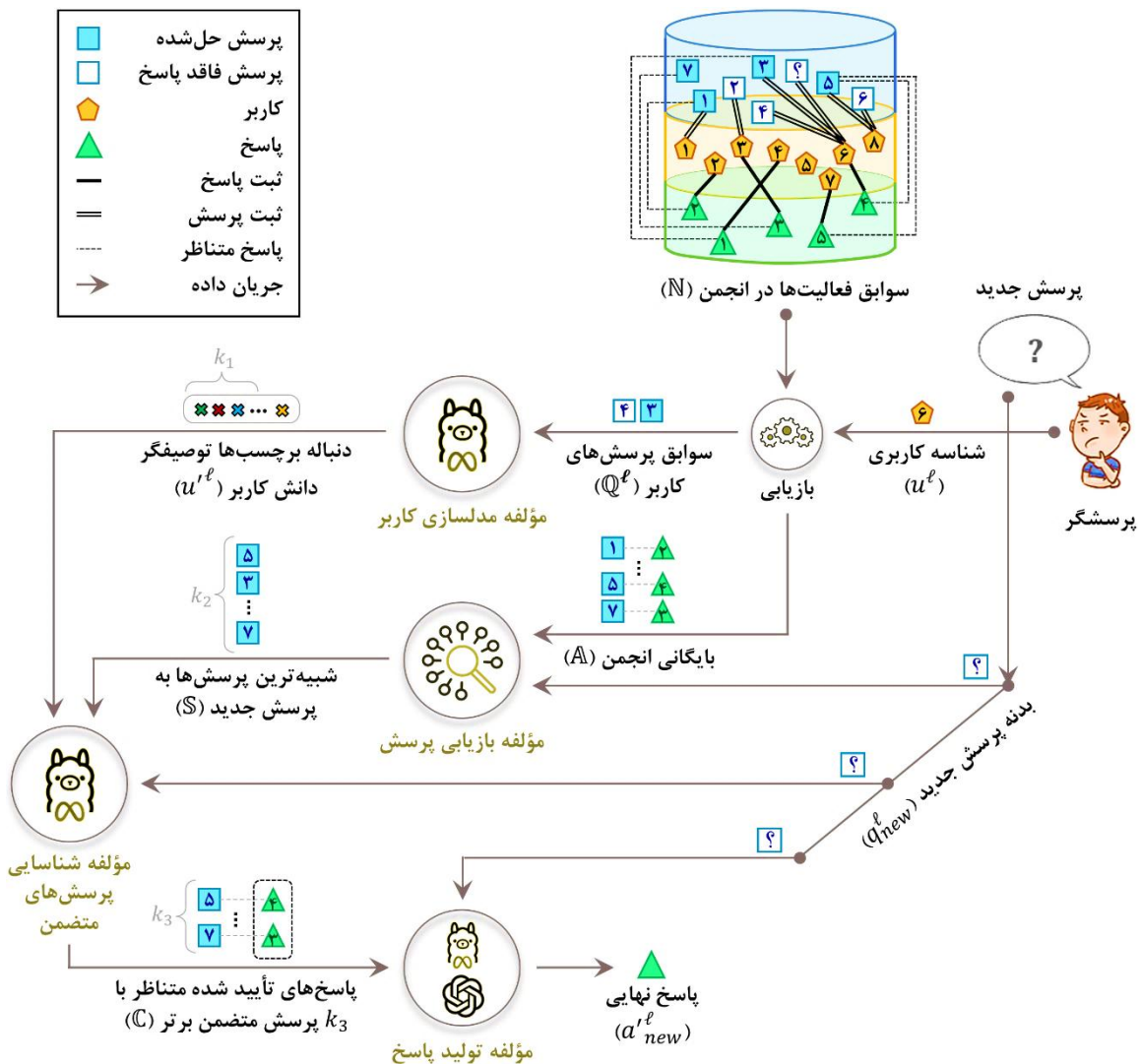
مرجع وظیفه	پیش‌پردازش	روش بازنمایی	مدل یادگیری	دادگان	دانش	نتایج
[۹] RQE	حذف کلمات توقف، ریشه‌یابی	Feature-based, GloVe	LR, MLP	SNLI, Multi-NLI, ClinicalQE, Quora	پرسش	برتری مدل LR، اثربخشی روش ترکیبی شامل بازیابی با روش‌های سنتی و رتبه‌بندی مجدد براساس نتایج RQE
[۳۲] RQE	حذف کلمات توقف، ریشه‌یابی	Feature-based, BioBERT	.NB, SVM, AdaBoost, KNN, LR, XGBoost	MediQA-RQE	پرسش	برتری مدل‌های ترکیبی نظیر AdaBoost و XGBoost بر مدل‌های محض
[۲۸] ISQ	بن‌یابی	Feature-based, Word2vec	SVM, MLP, CNN, NB	SemEval (2016)	پرسش، عنوان	عملکرد برتر مدل CNN نسبت به مدل‌های سنتی، افت نتایج در استفاده از رای اکثریت
[۳۱] DQD	-	GloVe	MLP, Siamese-CNN	Quora	پرسش	برتری قابل توجه CNN بر MLP
[۳۳] RQE	-	USE	LR, MLP	MediQA-RQE	پرسش	کیفیت قابل قبول تعبیه‌های USE با وجود سادگی مدل MLP
[۲۰] RQE	-	.ELMO, FastText, GloVe, InferSent, SciBERT, BioBERT	LR	MediQA-RQE	پرسش	برتری تعبیه در سطح جمله نسبت به تعبیه در سطح کلمه، برتری مدل‌های تعبیه خاص دامنه پزشکی بر مدل‌های عمومی
[۳۴] RQE	-	BioBERT, BERT	SVM	MediQA-RQE	پرسش	تأثیر مثبت ترکیب مدل‌های خاص-دامنه با مدل‌های عمومی در رفع نقاط ضعف آنها
[۳۵] RQE	-	BioBERT, Word2vec	Siamese-BiLSTM, MLP	MediQA-RQE	پرسش	عملکرد برتر مدل Siamese با تعبیه‌های word2vec نسبت به مدل‌های خاص-دامنه
[۳۷] RQE	گسترش اختصارات	SciBERT, BioBERT, MT-DNN, GloVe	.ESIM, BERT, RCNN-NLI, سایر	MediQA-RQE	پرسش، UMLS	حذف ویژگی‌های خاص-وظیفه و ارتقاء کیفیت بازنمایی‌ها با یادگیری تخصصی
[۲۳] DQD	حذف کلمات توقف، ریشه‌یابی	BERT	GPT-2, BERT, GPT-4, GPT-3	SemEval (2017)	پرسش، پاسخ، رده موضوعی	اثبات اهمیت فراداده‌ها، کارآمدی گسترش پرسش توسط LLMها در دقت بازیابی مؤثر بودن الحاق بازنمایی‌های LSTM و CNN در مدل‌سازی الگوهای پیچیده در داده و برتری این روش نسبت به BERT
[۲۱] DQD	حذف کلمات توقف و علائم نشانه‌گذاری، ریشه‌یابی	GloVe, Word2vec	.BiLSTM + CNN, BERT, AdaBoost, SVM, MLP, سایر	Stack Overflow (2019)	پرسش	تولید بازنمایی‌های برداری غنی‌تر از طریق تولید کلید واژه برای پرسش‌ها و یادگیری چندوظیفه‌ای بدون ناظر بر پایه کلیدواژه‌ها
[۴۰] QR	حذف کلمات توقف	.RoBERTa, BERT, LDA, سایر	-	.FAQIR, TREC-50, Stack Overflow, StackFAQ	پرسش	کارایی ادغام دانش خارجی در بازنمایی‌های برداری و بهره‌گیری از مکانیزم توجه در ارتقاء درک معانی، ضمن حفظ قابلیت تعمیم مدل
[۴۱] QR	-	attention + CNN, BERT	MLP	.MedNLI, SNLI, SICK, Quora, سایر	پرسش، ConceptNet, Wiktionary	مؤثر بودن تزریق دانش دامنه‌ای و جهانی درباره موجودیت‌ها به بازنمایی موجودیت‌ها
[۲۶] RQE	-	BERT	BERT	MediQA-RQE, CoPA, سایر	پرسش، ConceptNet	بهبود درک محلی و سراسری با ترکیب مدل‌های عمومی و منابع دانش خاص-دامنه
[۲۵] RQE	-	attention + GCN + BERT, ClinicalBERT, BioBERT	MLP	MediQA-RQE, ICD-10, سایر	پرسش، UMLS	بهبود رتبه‌بندی پرسش‌های کاندید با نگاه به کیفیت پاسخ‌های آنها
[۴۲] QR	حذف کلمات توقف و علائم نشانه‌گذاری	.Word2vec, Feature-based, attention + BiLSTM	Ranking SVM	SemEval (2016, 2017)	پرسش، پاسخ	تولید بازنمایی‌های غنی‌تر از طریق یادگیری توأمان وظایف مرتبط
[۱۹]	-	BART	BART	MeQSum, iCliniq, سایر	پرسش	

اختصارات: BERT: بازنمایی‌های رمزگذار دوسویه از ترانسفورمرها، BART: ترانسفورمر دوسویه خودبازگشتی، CNN: شبکه عصبی پیچشی، DQD: شناسایی پرسش‌های تکراری، GCN: شبکه پیچشی گرافی، GPT: ترانسفورمر پیش‌آموزش‌دیده مولد، ISQ: شناسایی پرسش‌های مشابه، k-NLI: نزدیک‌ترین همسایه، LR: رگرسیون لجستیک، BiLSTM: شبکه حافظه طولانی کوتاه‌مدت دوسویه، MAP: میانگین متوسط دقت، MLP: پرسپترون چند لایه، MRR: میانگین معکوس رتبه، NB: بیز محض، QR: بازیابی پرسش، RCNN-NLI: شبکه عصبی بازگشتی پیچشی برای استنتاج زبان طبیعی، RQE: شناسایی پرسش‌های متضمن، SVM: ماشین بردار پشتیبان، UMLS: سیستم یکپارچه زبان پزشکی، USE: رمزگذار جهانی جملات

### ۳- سیستم پیشنهادی

گردد. به‌طور هم‌زمان، مؤلفه بازیابی پرسش نیز با دریافت بدنه پرسش، اقدام به سنجش شباهت آن با پرسش‌های حل‌شده در بایگانی انجمن نموده و فهرستی از پرسش‌های مشابه را بازیابی می‌نماید. خروجی‌های این دو مؤلفه در گام بعد در اختیار مؤلفه شناسایی پرسش‌های متضمن قرار می‌گیرند. این مؤلفه با آگاهی از دانش و علایق پرسشگر، پرسش‌های مشابه را غربال<sup>۲</sup> نموده و پرسش‌هایی که پتانسیل ارائه یک پاسخ صحیح کامل یا جزئی به پرسش جدید را دارند شناسایی می‌کند. در گام آخر، پاسخ‌های تأییدشده برای برترین پرسش‌های متضمن، به مؤلفه تولید پاسخ ارائه می‌گردند تا در یک پاسخ واحد به زبان طبیعی، متمرکز بر نیاز پرسش، تجمیع گردند. در ادامه، عملکرد هر مؤلفه از سیستم، با جزئیات تشریح می‌شود.

UsERQA یک سیستم هوشمند مولد پاسخ برای انجمن‌های پرسش‌وپاسخ است که با استفاده از زنجیره‌ای از مدل‌های زبانی بزرگ، به تحلیل علاقمندی‌های پرسشگر پرداخته و پاسخی جامع به زبان طبیعی برای تأمین نیاز اطلاعاتی وی ارائه می‌دهد. این سیستم، مطابق شکل ۱، از چهار مؤلفه اصلی شامل مدلسازی کاربر، بازیابی پرسش، شناسایی پرسش‌های متضمن و تولید پاسخ تشکیل شده است. در نخستین گام پس از ثبت یک پرسش جدید در انجمن، شناسه کاربری پرسشگر در اختیار مؤلفه مدلسازی کاربر قرار می‌گیرد تا با تحلیل سوابق فعالیت‌های این کاربر، دانش و علایق وی در قالب یک دنباله متنی، حاوی تعدادی عنوان<sup>۱</sup> کلیدی، توصیف



شکل ۱. نمای کلی از سیستم پیشنهادی پژوهش (UsERQA)

<sup>2</sup> Filter

<sup>1</sup> Topic

### ۳-۱- مدلسازی کاربر

مؤلفه مدلسازی کاربر را می‌توان اصلی‌ترین نوآوری پژوهش جاری دانست. این مؤلفه مبتنی بر یک مدل زبانی بزرگ است که با بررسی سوابق فعالیت‌های یک کاربر، بر استنباط دانش پیش‌زمینه و موضوعات مورد علاقه وی اهتمام می‌ورزد. گام‌های اتخاذ شده در این فرآیند، در الگوریتم (۱) ارائه شده است.

---

#### Algorithm 1: User knowledge modeling

---

**Inputs:** User id  $u^l$ , Community questions  $\mathbb{N}$ , Number of topical tags  $k_1$

**Output:** User knowledge  $u'^l \in \mathbb{U}$

---

```

1.  $\square^l \leftarrow \{\}$ 
2. foreach  $\langle q_i, u_i \rangle$  in  $\mathbb{N}$  do
3.   if  $u_i = u^l$ : append  $q_i$  to  $\square^l$ 
4. end
5.  $\mathbb{T} \leftarrow \{\}$ 
6. foreach  $q_i$  in  $\mathbb{Q}^l$  do
7.    $p_i \leftarrow \text{prompt}(q_i)$ 
8.    $t'_i \leftarrow \text{Llama}(p_i)$ 
9.    $\mathbb{T} \leftarrow \text{union}(\mathbb{T}, t'_i.\text{split}(', '))$ 
10. end
11.  $u' \leftarrow \text{unique}(\mathbb{T})$ 
12.  $u'.\text{sort}(\text{frequency}(\mathbb{T}), \text{order}(\mathbb{T}))$ 
13.  $u'^l \leftarrow \text{top}(u', k_1)$ 
14. return  $u'^l$ 

```

---

متابiq الگوریتم (۱)، انجمنی متشکل از مجموعه کاربران  $\mathbb{U}$  را در نظر بگیرید بطوریکه مجموعه  $\mathbb{N}$ ، مجموعه تمام پرسش‌های ثبت‌شده در این انجمن باشد. چنانچه کاربر  $u^l \in \mathbb{U}$  پرسش جدید  $q_{new}^l$  را در انجمن مطرح سازد، مؤلفه مدلسازی کاربر، در نخستین گام، کلیه پرسش‌هایی را که وی از بدو عضویتش تا به اکنون ثبت کرده است واکنشی می‌نماید (خطوط ۱ تا ۴). در اینجا،  $\mathbb{Q}^l$  معرف تاریخچه پرسش‌های ثبت‌شده توسط کاربر  $u^l$  است. پس از بازیابی سوابق کاربر، هریک از پرسش‌های  $\mathbb{Q}^l$ ، به یک اعلان  $p_i^l$  با الگویی مشخص تبدیل شده و متعاقباً، به یک مدل زبانی Llama2 ارسال می‌شوند تا به ازای هر پرسش، یک دنباله متنی از برچسب‌ها<sup>۲</sup> که توصیفگر موضوعات محوری آن پرسش هستند تولید گردد (خطوط ۵ تا ۸). لازم به ذکر است که مدل Llama2، پیش‌تر برای وظیفه تولید برچسب‌های موضوعی، تنظیم دقیق می‌شود. در گام بعد، دنباله‌های تولیدشده که با نماد  $t'_i^l$  نمایش داده شده‌اند، به برچسب‌هایشان تفکیک شده و از اجتماع آنها، مجموعه  $\mathbb{T}$  از

برچسب‌ها شکل می‌گیرد (خط ۹). این برچسب‌ها براساس بسامد<sup>۳</sup> و ترتیب وقوعشان مرتب شده و  $k_1$  برچسب نخست، به عنوان مدل نهایی دانش کاربر  $(u'^l)$  انتخاب می‌گردند (خطوط ۱۱ تا ۱۴).

عدم وابستگی مؤلفه مدلسازی کاربر به پرسش ورودی را می‌توان از مهمترین نقاط قوت آن برشمرد. به عبارت دیگر، انجمن قادر خواهد بود با استفاده از این مؤلفه، زمینه‌های دانش همه کاربران را به صورت کاملاً برون‌خط<sup>۴</sup>، استنتاج نموده و در صورت نیاز، به طور دوره‌ای آن را بروزرسانی کند. بعلاوه، با تعریف بازه‌های زمانی روی سوابق کاربر یا محدود نمودن این فرآیند به استفاده از جدیدترین پرسش‌های کاربر، می‌توان مسئله گذار تدریجی<sup>۵</sup> علایق کاربران را نیز مدیریت نمود.

### ۳-۲- بازیابی پرسش

به مجموعه تمامی پرسش‌های دارای پاسخ تأییدشده در انجمن، بایگانی انجمن می‌گویند. هدف اصلی پژوهش جاری، تولید پاسخ‌هایی صحیح و جامع برای پرسش‌های ورودی، برگرفته از پاسخ‌های تأییدشده برای پرسش‌های متضمن آنها در بایگانی انجمن است. از این رو، سرعت و کیفیت پاسخگویی، در گرو شناسایی سریع و دقیق پرسش‌های متضمن در بایگانی می‌باشد. این در حالی است که بایگانی انجمن‌ها، غالباً حجمی بالغ بر هزاران یا میلیون‌ها پرسش را در بر می‌گیرند. به این ترتیب، ضروریست که فضای جستجو، پیش از اقدام به شناسایی پرسش‌های متضمن، محدود گردد. در سیستم پیشنهادی پژوهش جاری، این وظیفه بر عهده مؤلفه بازیابی پرسش گذاشته شده است. الگوریتم (۲) به شرح وظایف این مؤلفه می‌پردازد.

---

#### Algorithm 2: Similar question retrieval

---

**Inputs:** Newly submitted question  $q_{new}^l$ , Community archive  $\mathbb{A} \subset \mathbb{N}$ , Number of similar questions  $k_2$

**Output:** Top most similar questions  $\mathbb{S}$

---

```

1.  $s \leftarrow []$ 
2.  $e_{new}^l \leftarrow \text{represent}(q_{new}^l)$ 
3. foreach  $q_i$  in  $\mathbb{A}$  do
4.    $e_i \leftarrow \text{represent}(q_i)$ 
5.    $S[i] \leftarrow \text{similarity}(e_{new}^l, e_i)$ 
6. end
7.  $\mathbb{M} \leftarrow \text{top}(\text{argsort}(s), k_2)$ 
8.  $\mathbb{S} \leftarrow \{\}$ 
9. foreach  $m$  in  $\mathbb{M}$  do
10.   append  $\langle q_m, s[m] \rangle$  to  $\mathbb{S}$ 
11. end
12. return  $\mathbb{S}$ 

```

---

<sup>4</sup> Offline

<sup>5</sup> Drift

<sup>1</sup> Prompt

<sup>2</sup> Tags

<sup>3</sup> Frequency



این وظیفه آموزش دیده است ارسال می‌شوند. به این ترتیب، طبق رابطه (۴)، چنانچه پرسش  $q_j$ ، مرتبط با دامنه تخصصی پرسشگر بوده و از پاسخ‌های صحیح آن بتوان برای پاسخگویی به پرسش  $q_{new}^l$  استفاده کرد، عبارت "متضمن" و درغیراینصورت، عبارت "غیرمتضمن" توسط مدل زبانی تولید می‌گردد.

$$\mathcal{L}(q_j, q_{new}^l) = \text{Llama}(p_j) = \begin{cases} \text{Entailed} & (q_{new}^l \rightarrow q_j) \ \& \ (q_j \in u^l) \\ \text{Not-entailed} & (\text{otherwise}) \end{cases} \quad (4)$$

### ۳-۴- تولید پاسخ

بقای یک انجمن پرسش‌وپاسخ در گرو جلب حداکثری رضایت کاربران آن است. برای این منظور، در سیستم پیشنهادی این پژوهش، مؤلفه تولید پاسخ مطابق الگوریتم (۳)، تعبیه شده است تا به جای آنکه پرسشگر، با ارائه چند پیوند<sup>۵</sup>، به ریشه<sup>۶</sup> پرسش‌های متضمن هدایت شود، یک پاسخ صحیح، جامع و موجز را به‌طور مستقیم دریافت نماید.

#### Algorithm 3: Answer generation

**Inputs:** Newly submitted question  $q_{new}^l$ , Similar questions  $\mathcal{S}$ , Number of candidate answers  $k_3$

**Output:** Final answer  $a_{new}^l$

1.  $\mathcal{C} \leftarrow \{\}$
2. **foreach**  $q_j$  in  $\mathcal{S}$  **do**
3.     **if**  $\mathcal{L}(q_j, q_{new}^l) = \text{'Entailed'}$  **do**
4.         append  $a_j$  to  $\mathcal{C}$
5.     **end**
6. **end**
7.  $\mathcal{C} \leftarrow \text{top}(\mathcal{C}, k_3)$
8.  $p \leftarrow \text{prompt}(\mathcal{C}, q_{new}^l)$
9.  $a_{new}^l \leftarrow \text{LLM}(p)$
10. **return**  $a_{new}^l$

سومین مدل زبانی بزرگ در سیستم پیشنهادی، در مؤلفه تولید پاسخ به کار گرفته می‌شود. در اولین گام از الگوریتم (۳)، پاسخ‌های تأییدشده متناظر با  $k_3$  پرسشی که بیشترین شباهت را به پرسش ورودی داشته و برچسب متضمن دریافت کرده‌اند، واکنشی می‌شوند (خطوط ۱ تا ۷). در اینجا،  $a_j$  معرف پاسخ تأییدشده برای پرسش  $q_j \in \mathcal{S}$  می‌باشد. همچنین، یادآور می‌گردد که پرسش‌های مجموعه  $\mathcal{S}$ ، قبلاً براساس میزان شباهتشان به پرسش ورودی، به صورت نزولی مرتب شده‌اند (الگوریتم ۲، خطوط ۷ تا ۱۱). در گام دوم، پرسش ورودی همراه با پاسخ‌های کاندید، به یک اعلان با الگوی مشخص

مطابق الگوریتم (۲)، مؤلفه بازیابی پرسش موظف است بازنمایی‌های برداری متناظر با پرسش ورودی  $q_{new}^l$  و یکایک پرسش‌های بایگانی  $q_i \in \mathcal{A}$  را تولید نموده و شباهت میان آنها را در فضای برداری یکسان، اندازه‌گیری نماید (خطوط ۱ تا ۶). در پژوهش جاری، معیار شباهت کسینوسی، که توسط رابطه (۱) فرموله شده است، مورد استفاده قرار می‌گیرد.

$$\text{cosine-similarity}(e_{new}^l, e_i) = \frac{e_{new}^l \cdot e_i}{\|e_{new}^l\| \|e_i\|} \quad (1)$$

پس از سنجش شباهت،  $k_2$  پرسشی که بیشترین شباهت را به پرسش ورودی دارند در خروجی ارائه می‌شوند (خطوط ۷ تا ۱۲).

### ۳-۳- شناسایی پرسش‌های متضمن

مطابق تعریف ارائه شده توسط آباچا و همکاران در [۹]، پرسش  $B$  یک پرسش متضمن برای پرسش  $A$  نامیده می‌شود ( $A \rightarrow B$ ) اگر هر پاسخ صحیح ارائه شده برای پرسش  $B$ ، یک پاسخ صحیح کامل یا جزئی برای پرسش  $A$  نیز محسوب شود. لذا، شناسایی پرسش‌های متضمن، نیازمند یک مدل دسته‌بند دودویی<sup>۱</sup> است که مطابق رابطه (۲) عمل کند.

$$f(A, B) = \begin{cases} 1 & (A \rightarrow B) \\ 0 & (A \nrightarrow B) \end{cases} \quad (2)$$

مطالعه جاری، مدل زبانی بزرگ Llama2 را جایگزین مدل‌های دسته‌بند سنتی نموده و فرآیند شناسایی پرسش‌های متضمن را در قالب فرآیند تولید متن<sup>۲</sup>، مطابق رابطه (۳)، مدل‌سازی نموده است.

$$f(A, B) = \begin{cases} \text{Entailed} & (A \rightarrow B) \\ \text{Not-entailed} & (A \nrightarrow B) \end{cases} \quad (3)$$

اگرچه بررسی ویژگی‌های متنی پرسش‌های  $A$  و  $B$  می‌تواند تا حد زیادی وجود رابطه تضمین میان آنها را آشکار سازد، اما زوج پرسش‌هایی که خیلی کوتاه، مبهم، فاقد اشتراک واژگانی<sup>۳</sup>، خیلی طولانی یا دربرگیرنده اطلاعات جانبی<sup>۴</sup> باشند ممکن است با تشخیص اشتباه مدل مواجه شوند [۱۵]. پژوهش حاضر، اطلاع از زمینه‌های دانش و علایق کاربران را در برجسته نمودن موضوعات اصلی پرسش‌ها و رفع ابهام از آنها، راهگشا دانسته و دومین نوآوری خود را بر این اساس بنا می‌نهد. در این راستا، مؤلفه شناسایی پرسش‌های متضمن، علاوه بر پرسش جدید  $q_{new}^l$  و پرسش کاندید  $q_j \in \mathcal{S}$ ، دانش مدل‌شده برای پرسشگر، یعنی  $u^l$  را نیز در ورودی دریافت می‌کند. ورودی‌های مذکور پس از تبدیل شدن به یک اعلان با ساختار مشخص ( $p_j$ )، برای یک مدل Llama2 که پیش‌تر برای

<sup>4</sup> Peripheral details

<sup>5</sup> Link

<sup>6</sup> Thread

<sup>1</sup> Binary classification

<sup>2</sup> Text generation

<sup>3</sup> Lexical overlap

دقیق مؤلفه شناسایی پرسش‌های متضمن استفاده می‌شوند. در پژوهش جاری همچنین، یک دادگان اعتبارسنجی، مشتمل بر ۵۰۰ زوج پرسش از دادگان CQADupStack، با هدف تنظیم پارامترها برای آموزش مؤلفه شناسایی پرسش‌های متضمن، استخراج شده است. این مجموعه داده با دادگان آزمون همپوشانی نداشته و از نظر توزیع مشابه آنها می‌باشد. برخی جزئیات آماری در مورد دادگان پژوهش در جدول ۳ ارائه شده است.

جدول ۳. مقدار پارامترهای مورد استفاده

مشخصات آماری	داده	دادگان آموزش	دادگان آزمون	دادگان اعتبارسنجی
میانگین تعداد نشانه‌ها <sup>۴</sup> در پرسش‌های $Q^1$	۱۱۳	۱۱۶	۸۸	
میانگین تعداد نشانه‌ها در پرسش‌های $Q^2$	۱۱۲	۱۰۳	۹۰	
میانگین طول دنباله برچسب‌های $Q^1$	۶	۶	-	
بیشینه طول دنباله برچسب‌های $Q^1$	۲۷	۲۳	-	
میانگین طول پاسخ‌های تأییدشده برای $Q^1$	۲۴۳	۲۵۳	-	
تعداد زوج‌های متضمن	۳۹۸۸	۱۰۱۲	۲۵۰	
تعداد زوج‌های غیرمتضمن	۴۰۱۲	۹۸۸	۲۵۰	

تعداد ۱۱۰۱ مورد از پرسش‌های  $Q^1$  در بخش آزمون، دارای پاسخ تأییدشده هستند که از این تعداد، تنها ۱۴۸ پرسش توسط کاربران انجمن به عنوان پرسش تکراری تشخیص داده شده‌اند. این مجموعه از پرسش‌ها برای ارزیابی عملکرد مؤلفه تولید پاسخ مورد استفاده قرار می‌گیرند. جدول الف در پیوست ۱ دو نمونه از داده‌های مطالعه را نشان می‌دهد.

#### ۴-۲- پیکربندی و پارامترهای اجرا

آزمایش‌های این پژوهش با استفاده از زبان پایتون، بر بستر Google Colab با ۸۳،۵ GB رم سیستم و یک پردازنده گرافیکی NVIDIA A100-SXM4-40GB انجام شده است. برای پیاده‌سازی مدل زبانی Llama2 از نسخه ۷ بیلیون پارامتری آن در کتابخانه Hugging face<sup>۵</sup> و برای GPT-4o mini از پلتفرم برخط آن<sup>۶</sup> استفاده شده است. جدول ۴، مقادیر ابرپارامترهای مورد استفاده در این مطالعه را گزارش می‌کند.

بیشینه طول دنباله ورودی در مؤلفه مدلسازی کاربر، با در نظر گرفتن مجموع سه مقدار، شامل تعداد نشانه‌ها در اعلان ورودی (شکل ۲)، میانگین تعداد نشانه‌ها در پرسش‌های  $Q^1$  (جدول ۳) و بیشینه طول دنباله برچسب‌های  $Q^1$  (جدول ۳)، برابر با ۲۵۶ تنظیم شده است. بیشترین تعداد نشانه‌های جدید نیز براساس بیشینه طول دنباله برچسب‌های  $Q^1$  و با احتساب حداقل سه علامت کاما (,) در

تبدیل شده و برای یک مدل زبانی پیش‌آموزش‌دیده، نظیر Llama2 یا GPT-4o mini، ارسال می‌شوند (خطوط ۸ تا ۱۰). در این مرحله، مهندسی اعلان<sup>۱</sup> به منظور ارائه بهترین دستورالعمل<sup>۲</sup> جهت انجام وظیفه خلاصه‌سازی چند پاسخی متمرکز بر پرسش<sup>۳</sup>، امری ضروری است.

#### ۴- آزمايش‌ها و تحليل نتايج

اصلي‌ترين پرسش‌هاي پژوهشي مطالعه حاضر به شرح زير هستند:

۱. مؤلفه مدلسازی کاربر تا چه اندازه در تشخیص موضوعاتی که کاربر بطور متناوب دنبال می‌کند کارآمد است؟
۲. استفاده از کدام مدل بازنمایی برداری، دقت بالاتری را در بازیابی پرسش‌های مشابه به ارمغان می‌آورد؟
۳. مدلسازی دانش کاربر چه اندازه در بهبود فرآیند شناسایی پرسش‌های متضمن مؤثر است؟
۴. برتری مؤلفه شناسایی پرسش‌های متضمن آگاه به کاربر نسبت به شناسایی پرسش‌های متضمن مستقل از کاربر، در دستیابی به پاسخ‌های صحیح و جامع، چه میزان برآورد می‌شود؟

در ادامه، دادگان پژوهش، تنظیمات اولیه سیستم و معیارهای ارزیابی تعیین می‌گردند. سپس، شرح کاملی از نتایج آزمایش‌های انجام‌شده جهت ارزیابی عملکرد سیستم پیشنهادی ارائه می‌شود.

#### ۴-۱- مجموعه داده

آموزش مؤلفه‌های سیستم پیشنهادی و ارزیابی عملکرد آنها با استفاده از دادگان CQADupStack [۴۳] و نسخه کوچکتر و تغییر یافته آن، CQAD-ReQuEST [۴۴] انجام می‌پذیرد. دادگان CQADupStack، اطلاعات ۴۷۰،۳۴۴ پرسش، نظیر مشخصات پرسشگر، عنوان، برچسب‌ها، تاریخ ثبت، آراء مثبت و منفی، پاسخ‌های دریافتی و پاسخ تأییدشده برای هر کدام (در صورت وجود) را در ۱۲ شاخه موضوعی مختلف، شامل شده است. از این پرسش‌ها، ۲۴۸،۴۲۶ مورد (حدود ۵۲/۸٪) دارای پاسخ تأییدشده می‌باشند که در سیستم پیشنهادی پژوهش، با عنوان بایگانی انجمن از آنها یاد می‌شود.

دادگان CQAD-ReQuEST شامل ۸۰۰۰ زوج پرسش ( $Q^1, Q^2$ ) در بخش آموزش و ۲۰۰۰ زوج در بخش آزمون است. پرسش‌های  $Q^1$  و برچسب‌های متناظر آنها در دادگان آموزش، برای تنظیم دقیق مؤلفه مدلسازی کاربر و کلیه زوج پرسش‌های آموزش، برای تنظیم

<sup>4</sup> Tokens

<sup>5</sup> <https://huggingface.co/meta-llama/Llama-2-7b-chat-hf>

<sup>6</sup> <https://chat.openai.com/>

<sup>1</sup> Prompt engineering

<sup>2</sup> Instruction

<sup>3</sup> Question-focused multi-answer summarization

دارد. میان برچسب‌ها، برابر با ۳۰ تعیین شده‌است. این دو ابرپارامتر در مؤلفه تولید پاسخ نیز به ترتیب، براساس پیکربندی پیش‌فرض Llama2 و میانگین طول پاسخ‌های تأییدشده برای  $Q^1$  (جدول ۳)، با مقادیر ۴۰۹۶ و ۲۴۰ مقداردهی شده‌اند.

تنظیم ابرپارامترهای مربوط به آموزش مؤلفه شناسایی پرسش‌های متضمن براساس داده‌های اعتبارسنجی انجام شده است که در بخش ۳-۴-۴ به آن پرداخته می‌شود. علاوه‌براین، بیشینه طول دنباله ورودی در این مؤلفه، با در نظر گرفتن طول اعلان ورودی (شکل ۵)، میانگین طول پرسش‌های  $Q^1$  و  $Q^2$  (جدول ۳)، حداکثر ۱۰۰ نشانه برای دانش کاربر و پنج نشانه برای برچسب خروجی، برابر با ۵۱۲ تنظیم شده است. به‌منظور جلوگیری از سوگیری ناشی از تفاوت در طول برچسب‌های خروجی، به‌جای برچسب‌های "Entailed" و "Not-entailed" که به ترتیب دارای سه و چهار نشانه هستند، برچسب‌های "positive" و "negative" که دارای طول یکسان (یک نشانه) هستند به عنوان خروجی مدل استفاده شده‌اند.

جدول ۴. مقدار ابرپارامترهای مورد استفاده

نام پارامتر	شماره مؤلفه			
	۱	۲	۳	۴
max_epochs	۱۰	-	۱۰	-
learning_rate	$2 \times 10^{-4}$	-	$3 \times 10^{-5}$	-
Batch_size	۳۲	-	۳۲	-
max_seq_length	۲۵۶	-	۵۱۲	۴۰۹۶
max_new_tokens	۳۰	-	۱	۲۴۰
lora_r	۶۴	-	۶۴	-
lora_alpha	۱۶	-	۱۶	-
lora_dropout	۰/۱	-	۰/۳	-
temperature	۰/۹۷	-	۰/۹۷	۰
4- or 8-bit quantization	خیر	-	خیر	خیر
k	۱۰	۵۰	-	۳, ۵

$$\text{ROUGE-N} = \frac{\sum_{g_N \in S} \sum_{\hat{g}_N \in \hat{S}} 1(g_N == \hat{g}_N)}{\sum_{g_N \in S} 1} \quad (5)$$

در مواردی که ترتیب  $N$ -گرام‌ها نیز حائز اهمیت است استفاده از ROUGE-L که طبق رابطه (۶)، بزرگترین دنباله مشترک<sup>۲</sup> را ملاک قرار می‌دهد، توصیه می‌گردد.

$$\text{ROUGE-L} = \frac{\text{LCS}(S, \hat{S})}{|\hat{S}|} \quad (6)$$

**BERTScore**: برخلاف معیار Rouge که شباهت واژگانی را اندازه‌گیری می‌کند، معیار BERTScore بر سنجش شباهت معنایی میان دنباله هدف و دنباله تولیدشده متمرکز است. برای این منظور، ابتدا هر دو دنباله، به یک مدل ترانسفورمر مبتنی بر BERT ارائه شده و بردارهای تعبیه  $E^S$  و  $E^{\hat{S}}$  تولید می‌شوند. سپس، براساس رابطه (۷)، امتیاز شباهت کسینوسی بردار تعبیه هر نشانه از دنباله هدف نسبت به یکایک نشانه‌های دنباله تولیدشده محاسبه شده و در پایان، با انجام یک تطابق حریمانه (تابع بیشینه) امتیاز نهایی در خروجی ارائه می‌شود.

$$\text{BERTScore}_{\text{Recall}} = \frac{1}{|\hat{S}|} \sum_{e_i \in E^S} \max_{\hat{e}_j \in E^{\hat{S}}} e_i^T \hat{e}_j \quad (7)$$

در این رابطه،  $e_i$  و  $\hat{e}_i$  مبین بردارهای تعبیه  $i$ -امین نشانه در دنباله‌های  $S$  و  $\hat{S}$  هستند.

#### ۴-۴- تحلیل نتایج

در این بخش تحلیل نتایج آزمایش‌ها در چهار سناریو، متناظر با چهار پرسش اصلی پژوهش، ارائه می‌گردد.

#### ۴-۴-۱- ارزیابی عملکرد مؤلفه مدل‌سازی کاربر

مدل زبانی Llama2 در مؤلفه مدل‌سازی کاربر موظف است برای هر پرسش از سوابق پرسشگر، دنباله‌ای از برچسب‌های موضوعی را تولید نماید. برای تحقق این امر، ابتدا مدل Llama2 با استفاده از دادگان آموزش CQAD-ReQuEST تنظیم می‌گردد. شکل ۲، الگوی اعلان ورودی و شکل ۳، سیر تغییرات تابع خطا در جریان آموزش را نشان می‌دهد. روند نزولی خطا در این نمودار، شاهدی بر آموزش موفق مدل Llama2 و بهبود تدریجی عملکرد آن است. ضمن اینکه به نظر می‌رسد با افزایش تعداد تکرارهای آموزش بتوان به نتایج دقیق‌تر و پایدارتری دست یافت.

#### ۴-۳- معیارهای ارزیابی

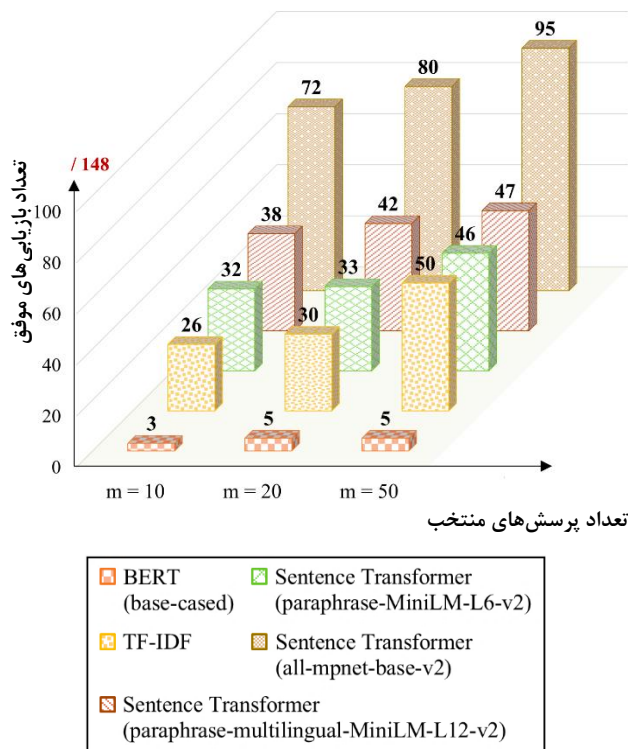
در این مطالعه، شباهت نحوی و معنایی یک دنباله تولیدشده توسط سیستم پیشنهادی ( $\hat{S}$ ) نسبت به دنباله هدف ( $S$ )، به ترتیب براساس معیارهای استاندارد Rouge [۴۵] و BERTScore [۴۶] ارزیابی می‌شود.

**Rouge**: برای محاسبه این معیار کفایست مطابق رابطه (۵) همپوشانی<sup>۱</sup>  $N$ -گرام‌های دنباله تولیدشده و دنباله هدف را بر تعداد کل  $N$ -گرام‌های دنباله هدف تقسیم نمود. خروجی این رابطه، عددی در بازه [۰, ۱] است بطوریکه، نزدیکتر بودن آن به مقدار ۱، بر همپوشانی واژگانی بیشتر و کیفیت بالاتر دنباله‌های تولیدشده دلالت

<sup>۲</sup> Longest common sequence of words (LCS)

<sup>۱</sup> Overlap

بردارهایی در فضای چندبعدی است؛ فرآیندی که به آن بازنمایی برداری متن گفته می‌شود. هرچه این بردارها روابط عمیق معنایی میان واژگان را بهتر منعکس نموده و در مدل‌سازی مفاهیم انتزاعی موفق‌تر عمل کنند، شباهت میان زوج پرسش‌ها نیز با دقت بیشتری سنجیده شده و کیفیت بازیابی بهبود می‌یابد. در این بخش از آزمایشات، ۵ مدل مختلف بازنمایی مورد مقایسه قرار گرفته‌اند. برای این منظور، تعداد پرسش‌های ورودی که حداقل یکی از پرسش‌های تکراری آن‌ها در فهرست  $m$  پرسش برتر بازیابی شده برایشان ظاهر شده است، شمارش شده‌اند. شکل ۴ به گزارش این نتایج می‌پردازد. جدول ۶ نیز میانگین زمان موردنیاز برای تولید بازنمایی برداری یک پرسش توسط هر مدل را ارائه می‌کند. اگرچه در سیستم پیشنهادی، زمان لازم برای تولید بازنمایی‌های برداری پرسش‌های بایگانی، به دلیل امکان انجام آن به صورت کاملاً برون خط، اهمیت چندانی ندارد.



شکل ۴. مقایسه مدل‌های مختلف بازنمایی برداری بر اساس تعداد پرسش‌هایی که حداقل یک پرسش تکراری در میان  $m$  پرسش برتر بازیابی شده دارند

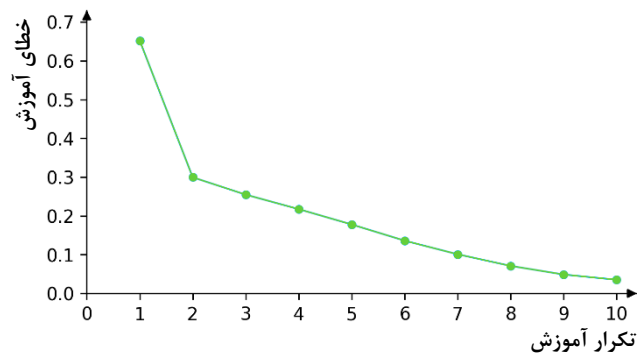
شواهد ارائه شده در شکل ۴، مؤید افزایش تعداد بازیابی‌های موفق، همزمان با افزایش مقدار  $m$  است. بر همین اساس، در ادامه آزمایشات این پژوهش، مقدار  $k_2$  برابر با ۵۰ در نظر گرفته شده است تا از انتشار آبخاری خطا به مؤلفه‌های بعدی جلوگیری شده و امکان بررسی تعداد بیشتری از پرسش‌های تکراری در مؤلفه شناسایی پرسش‌های متضمن فراهم گردد.

```
<s> [INST] <<SYS>>
You are a Tag Generator. Respond only with a list of tags; do not include any additional text or explanations.
<</SYS>>

Please generate at least 5 tags for the provided question. Tags can include multi-word phrases if appropriate and should help hierarchically categorize the question's topics.
### Question:
{Question}
### Tags:
[/INST]

{Tags}</s>
```

شکل ۲. الگوی اعلان مورد استفاده برای تنظیم مؤلفه مدل‌سازی کاربر



شکل ۳. روند تغییرات تابع خطا در زمان آموزش مؤلفه مدل‌سازی کاربر

نتایج حاصل از اعمال مدل Llama2 روی مجموعه آزمون نیز در جدول ۵ گزارش شده است. حصول مقادیر بالای ۹۰٪ برای BERTScore بیانگر آن است که برچسب‌های تولیدشده از نظر معنایی، تا حد زیادی با برچسب‌های هدف مطابقت دارند. بعلاوه، بالا بودن همزمان معیارهای Rouge-L و Rouge-1 بر توانمندی مدل در بازیابی بخش قابل توجهی از نشانه‌ها در برچسب‌های هدف، ضمن رعایت ترتیب رخدادشان دلالت دارد. شایان ذکر است که مدل تولید برچسب دارای ۳۳/۶ میلیون پارامتر قابل آموزش است و متوسط زمان تولید برچسب برای هر پرسش، ۱/۴۲۱ ثانیه می‌باشد.

جدول ۵. ارزیابی عملکرد مدل Llama2 در مدل‌سازی دانش کاربر

روی دادگان آزمون CQAD-ReQuEST		
نام معیار	مبتنی بر فراخوانی	مبتنی بر F1
Rouge-1	۰/۶۲۸	۰/۶۲۴
Rouge-2	۰/۴۱۸	۰/۴۱۶
Rouge-L	۰/۶۲۶	۰/۶۲۲
BERTScore	۰/۹۴۰	۰/۹۴۰

#### ۴-۲- مقایسه کارایی مدل‌های مختلف بازنمایی متن در بازیابی پرسش‌های مشابه

سنجش شباهت میان زوج پرسش‌ها مستلزم نگاشت آن‌ها به

گام بروزرسانی پارامترهای مدل، محاسبه شده است. لازم به ذکر است، تعداد پارامترهای قابل آموزش در مدل شناسایی پرسش‌های متضمن برابر با  $33/6$  میلیون و متوسط زمان پیش‌بینی برچسب برای هر زوج پرسش، حدود  $0/425$  ثانیه است.

```
<s> [INST] <<SYS>>
Given two questions, Q1 and Q2, determine if Q1
entails Q2 or not.
<</SYS>>

Entailment means every answer to Q2 must fully or
partially answer Q1.
Note that, Q2 must align with the user's topics of
interest: ({UserKnowledge}).
Respond with "positive" or "negative" only.
Q1: {Question1}
Q2: {Question2}
### Answer:
[/INST]

{Label}
```

(الف)

```
<s> [INST] <<SYS>>
Given two questions, Q1 and Q2, determine if Q1
entails Q2 or not.
<</SYS>>

Entailment means every answer to Q2 must fully or
partially answer Q1.
Respond with "positive" or "negative" only.
Q1: {Question1}
Q2: {Question2}
### Answer:
[/INST]

{Label}
```

(ب)

شکل ۵. الگوی اعلان مورد استفاده برای تنظیم مؤلفه شناسایی پرسش‌های متضمن (الف) آگاه به کاربر (ب) مستقل از کاربر

نتایج ارائه شده در شکل ۶ (الف)، گویای آن است که افزایش مقادیر  $\alpha$  و  $r$  افزایش ظرفیت مدل برای یادگیری و تسریع همگرایی را به همراه دارد. براساس این نتایج، مقادیر ۶۴ و ۱۶ به عنوان مقادیر بهینه برای این دو پارامتر انتخاب می‌شوند. شکل ۶ (ب) تأثیر مقادیر مختلف نرخ حذف نوروها بر عملکرد مدل را نشان می‌دهد. همانطور که ملاحظه می‌شود، در مراحل اولیه آموزش، مقادیر  $0/1$  و  $0/2$  کمترین خطای اعتبارسنجی را داشته‌اند. با این حال در ادامه، مدل مبتنی بر  $0/3$  توانسته است با شیب نزولی بیشتر، از آنها پیشی گرفته و جایگاه نخست را به خود اختصاص دهد. در مقابل، مقدار  $0/4$  به کم‌برازش مدل انجامیده و در جایگاه آخر قرار گرفته است.

مطابق شکل ۴، مدل BERT با اختلاف زیاد نسبت به سایر رقبا، بدترین عملکرد را ارائه داده و از چرخه آزمایش‌ها خارج می‌شود. در مقابل، مدل ترانسفورمر all-mpnet-base-v2 با تولید تعبیه‌های برداری در سطح جمله، جایگاه نخست را به خود اختصاص داده است. برتری این مدل را می‌توان ناشی از مدیریت بهتر جایگشت<sup>۱</sup> کلمات در جمله و در نتیجه درک بهتر روابط بافتاری دانست [۴۷].

جدول ۶. مقایسه مدل‌های مختلف بازنمایی برداری براساس میانگین

زمان تولید بازنمایی یک پرسش	نام مدل
۱/۰۳ × ۱۰	TF-IDF
۹/۶۳ × ۱۰	paraphrase-MiniLM-L6-v2
۱/۷۱ × ۱۰	paraphrase-multilingual-MiniLM-L12-v2
۲/۸۹ × ۱۰	BERT-base-cased
۳/۹۷ × ۱۰	all-mpnet-base-v2

مدل TF-IDF را نیز می‌توان بهینه‌ترین مدل از نظر زمان اجرا معرفی کرد؛ ضمن اینکه مطابق شکل ۴، کیفیت عملکرد آن نسبت به مدل‌های تعبیه بافتاری، قابل قبول ارزیابی می‌شود. در این میان، مدل‌هایی که برای وظیفه تشخیص بازنویسی<sup>۲</sup> آموزش دیده‌اند از نظر میانگین زمان بازنمایی، سریعتر از مدل mpnet عمل کرده و از نظر کیفیت بازنمایی‌ها، در رده متوسط قرار گرفته‌اند.

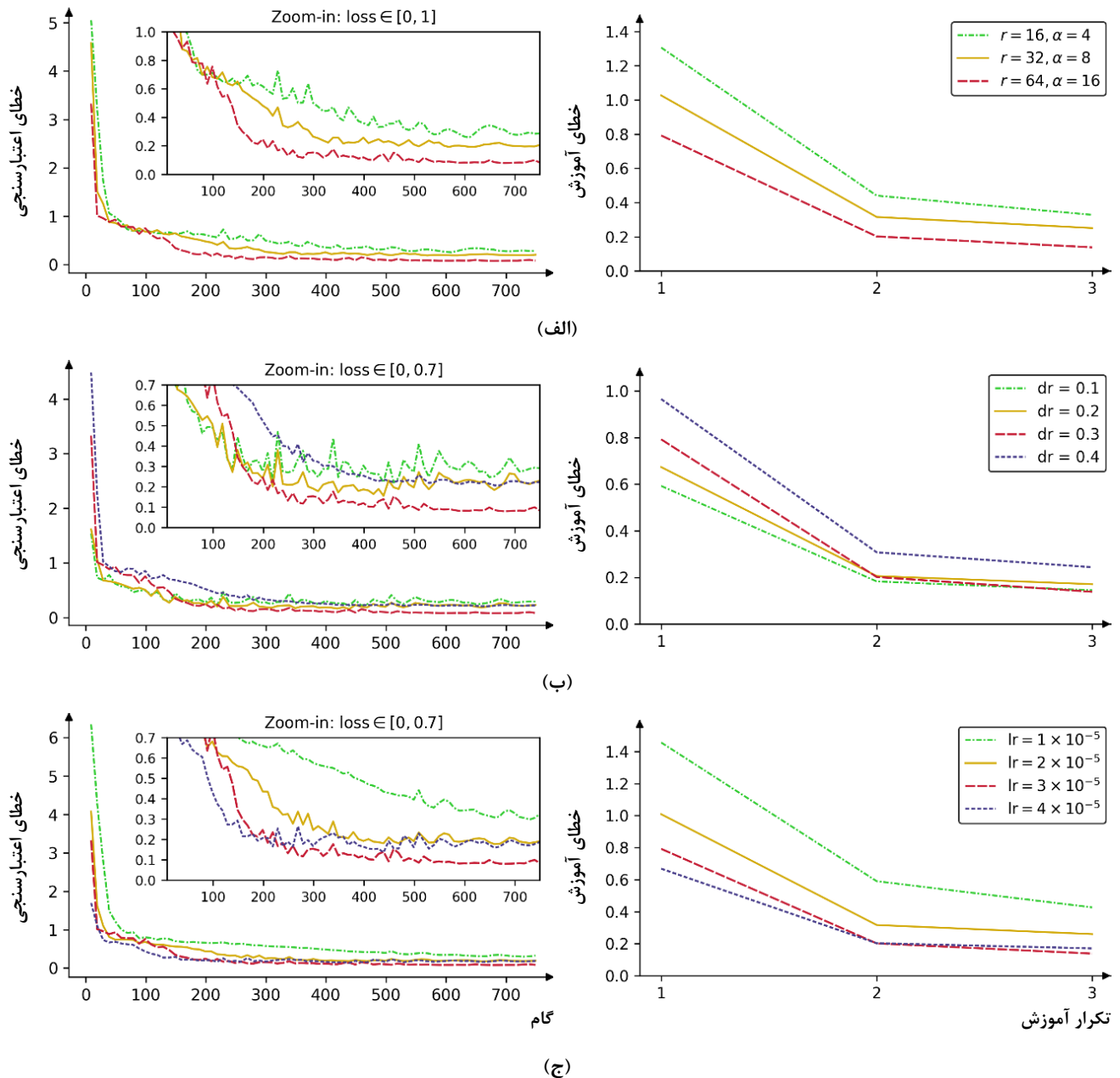
#### ۴-۳-۴- سنجش میزان اثربخشی مدلسازی کاربر در

#### فرآیند شناسایی پرسش‌های متضمن

این بخش از آزمایشات، نخست به تنظیم ابرپارامترها برای آموزش مؤلفه شناسایی پرسش‌های متضمن پرداخته و سپس، با مقایسه عملکرد این مؤلفه در حضور و عدم حضور دانش مدل شده برای کاربر، به سومین پرسش پژوهشی پاسخ می‌دهد. برای این منظور، مدل Llama2 براساس زوج پرسش‌های برچسب‌خورده در دادگان آموزش CQAD-ReQuEST تنظیم دقیق می‌شود. شکل ۵، الگوی اعلان را برای دو حالت آگاه به کاربر و مستقل از کاربر نمایش می‌دهد. در شکل ۶ نیز تغییرات خطای آموزش و اعتبارسنجی به ازای مقادیر مختلف ابرپارامترها به تصویر کشیده شده است. در این آزمایش‌ها، به دلیل محدودیت منابع محاسباتی و گستردگی فضای حالت، فرآیند آموزش به سه تکرار محدود شده است. به‌طور کلی، مقادیر ارائه شده در جدول ۴، به عنوان پیکربندی پایه در نظر گرفته شده و در هر آزمایش، تنها یکی از ابرپارامترها تغییر داده شده است تا تأثیر مستقل آن بر عملکرد مدل ارزیابی گردد. خطای اعتبارسنجی نیز به منظور پایش دقیق‌تر رفتار مدل، پس از هر ۱۰

<sup>2</sup> Paraphrase

<sup>1</sup> Permutation



شکل ۶. مقایسه تغییرات تابع خطای آموزش و اعتبارسنجی در طول فرآیند آموزش مدل برای مقادیر مختلف پارامترهای (الف)  $r$  و  $\alpha$  در  $dora$ ، (ب) نرخ حذف نورون‌ها و (ج) نرخ یادگیری

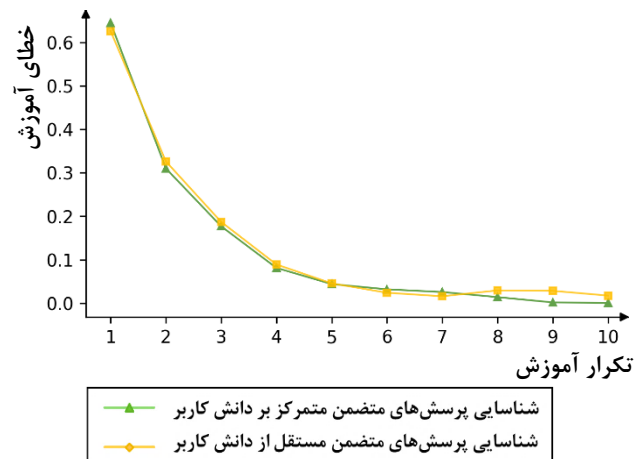
شکل ۷ مقایسه تغییرات تابع خطا در طول فرآیند آموزش در دو اجرای مذکور را نمایش می‌دهد. در این نمودار، محور افقی مبین تکرارهای آموزش و محور عمودی نشان‌دهنده میزان خطای آموزش است. همانطور که ملاحظه می‌شود، رقابت تنگاتنگی میان دو مدل شکل گرفته است. تا تکرار هفتم آموزش، روند همگرایی دو مدل تقریباً یکسان بوده و هیچ‌کدام از آنها نتوانسته‌اند بر رقیب خود فائق آیند. اما با ادامه یافتن فرآیند آموزش و افزایش ناگهانی خطای مدل مستقل از دانش کاربر در تکرار هشتم، نتیجه به نفع مدل متمرکز بر دانش کاربر تغییر می‌کند. در واقع، چنین استنباط می‌شود که

در شکل ۶ (ج)، عملکرد مدل به ازای مقادیر مختلف نرخ یادگیری نشان داده شده است. به‌طور کلی، افزایش نرخ یادگیری سبب تسریع همگرایی در مراحل اولیه آموزش می‌گردد و همزمان، احتمال وقوع بیش‌برازش در مراحل آتی نیز بیشتر می‌شود. در این آزمایش، نرخ یادگیری  $4 \times 10^{-5}$ ، علیرغم آنکه شروع بهتری را رقم زده است با گذشت زمان، ناپایدار گشته و بر مدل مبتنی بر نرخ  $3 \times 10^{-5}$  مغلوب گشته است. بنابراین، نرخ یادگیری  $3 \times 10^{-5}$  با حفظ فاصله معنادار از سایر رقبا، به عنوان مقدار بهینه برای آزمایش‌های بعدی انتخاب می‌شود.

نموده و  $k_3$  پرسش متضمن برتر را پس از شناسایی، به مؤلفه تولید پاسخ ارسال نمودیم. در جدول ۸، تعداد کل پرسش‌هایی که برچسب متضمن یا غیرمتضمن دریافت نموده‌اند، به تفکیک مدل مورد استفاده برای بازنمایی، گزارش شده است. همچنین، به ازای هر یک از مقادیر ۳ و ۵ برای متغیر  $k_3$ ، تعداد پرسش‌هایی که بطور مشابه توسط هر دو مدل آگاه به کاربر و مستقل از کاربر، به عنوان یک پرسش متضمن برتر شناسایی شده‌اند، شمارش شده است. همان‌طور که در جدول ۸ مشاهده می‌شود، مدل آگاه به کاربر در تمام اجراها، نسبت به مدل مستقل از کاربر، تعداد کمتری از پرسش‌ها را به عنوان متضمن شناسایی کرده است. این نتیجه می‌تواند ناشی از افزودن شرط همسویی پرسش‌کنندگان با دانش پرسشگر بوده و تأییدی بر دقت بالاتر مدل پیشنهادی تلقی شود. ذکر این نکته ضروری است که هر دو مدل، دارای محدودیت در طول زوج پرسش ورودی بوده و این محدودیت، در مدل آگاه به کاربر، به دلیل اضافه شدن شرط تطابق با دانش کاربر، بیشتر است. به همین علت است که جمع تعداد پرسش‌های متضمن و غیرمتضمن در تمامی اجراها کمتر از ۷۴۰۰ است.

با مقایسه مدل‌های مختلف بازنمایی از نظر تعداد پرسش‌های متضمن شناسایی‌شده توسط مدل آگاه به کاربر، می‌توان به اهمیت کیفیت بازنمایی در دقت بازیابی پرسش‌های مشابه پی برد. نتایج نشان می‌دهند که هرچه مدل بازنمایی پیچیده‌تر بوده و توانایی بیشتری در درک روابط عمیق معنایی و بافتاری داشته باشد، کاندیداهای شایسته‌تری انتخاب شده و به تبع آن، تعداد بیشتری برچسب متضمن را دریافت می‌کنند. بنابراین، مدل‌های TF-IDF و paraphrase-MiniLM-L6-v2 نیز از چرخه ارزیابی‌ها کنار گذاشته می‌شوند. در تمام اجراها، بیش از ۸۳٪ پرسش‌های منتخب دو مدل، مشترک بوده و این اشتراک، با کاهش پارامتر  $k_3$  افزایش می‌یابد. این شواهد نشان می‌دهند که دو مدل، غالباً در تصمیم‌گیری برای پرسش‌های ابتدا و انتهای فهرست پرسش‌های مشابه، اتفاق نظر دارند و اختلاف آنها، در میانه لیست که پرسش‌ها شباهت کمتری به پرسش جدید دارند برجسته‌تر می‌شود.

هر دو مدل در برخورد با اکثریت زوج پرسش‌ها که فاقد ابهام بوده‌اند، عملکرد یکسانی ارائه داده‌اند؛ اما مدل مبتنی بر دانش کاربر توانسته است با تصمیم‌گیری دقیق‌تر برای اقلیت مبهم پرسش‌ها، برتری خود را به اثبات رساند. ارزیابی عملکرد این دو مدل روی دادگان آزمون که نتایج آن در جدول ۷ ارائه شده است نیز استدلال فوق را تأیید می‌کند. به این ترتیب، مدل متمرکز بر کاربر با کسب ۰.۱٪ دقت بالاتر، به عنوان مؤلفه پیشنهادی برای شناسایی پرسش‌های متضمن در سیستم پژوهش به کار گرفته می‌شود.



شکل ۷. تغییرات خطا در فرآیند آموزش مؤلفه شناسایی پرسش‌های متضمن در دو حالت آگاه به کاربر و مستقل از آن

جدول ۷. مقایسه عملکرد دو مدل شناسایی پرسش‌های متضمن آگاه به کاربر و مستقل از کاربر روی دادگان آزمون CQAD-ReQuEST

مدل	صحت	F1	دقت	فراخوانی
مستقل از کاربر	٪۹۴/۸	٪۹۴/۸	٪۹۴/۷	٪۹۵/۰
متمرکز بر کاربر	٪۹۵/۴	٪۹۵/۴	٪۹۵/۷	٪۹۵/۱

سؤالی که ممکن است به ذهن خطور کند این است که "۰.۱٪ بهبود در دقت مؤلفه پیشنهادی برای شناسایی پرسش‌های متضمن، تا چه اندازه می‌تواند در بهبود کیفیت پاسخ‌های نهایی سیستم اثرگذار باشد؟" برای پاسخ به این پرسش، مدل‌های مذکور را بر روی ۵۰ پرسش مشابهی که برای هر پرسش ورودی بازیابی شده بودند اعمال

جدول ۸. توزیع برچسب‌های پیش‌بینی‌شده توسط دو مدل شناسایی پرسش‌های متضمن آگاه به کاربر و مستقل از کاربر

مدل بازنمایی	برچسب پیش‌بینی‌شده					
	مدل آگاه به کاربر		مدل مستقل از کاربر		اشتراک	
	متضمن	غیرمتضمن	متضمن	غیرمتضمن	$k_3 = 5$	$k_3 = 3$
TF-IDF	۵۰۶۵	۲۲۴۵	۶۰۳۹	۱۳۱۱	۶۱۵ (٪۸۳/۱)	۳۸۷ (٪۸۷/۲)
paraphrase-MiniLM-L6-v2	۵۰۶۱	۱۷۲۹	۶۰۷۲	۸۴۴	۶۲۴ (٪۸۴/۳)	۳۸۳ (٪۸۶/۳)
paraphrase-multilingual-MiniLM-L12-v2	۵۳۰۶	۱۴۸۴	۶۱۲۷	۷۸۲	۶۵۱ (٪۸۸/۰)	۳۹۶ (٪۸۹/۲)
all-mpnet-base-v2	۶۳۲۲	۹۵۴	۶۸۰۷	۴۷۰	۶۹۵ (٪۹۳/۹)	۴۲۰ (٪۹۴/۶)
تعداد کل زوج پرسش‌ها	۷۴۰۰		۷۴۰۰		۷۴۰	۴۴۴

جدول ۹. تحلیل کیفی عملکرد مدل شناسایی پرسش‌های متضمن آگاه به کاربر و مدل مستقل از کاربر

Why do we use their instead of his in this sentence? another driver flashes their lights.	پرسش ۱				
Is there a pronoun I can use as a gender-neutral pronoun? Each student should save his questions until the end. Each student should save her questions until the end. meaning, word choice, grammar, single word requests	پرسش ۲	مثال ۱	دانش کاربر	پیش‌بینی مدل مستقل از کاربر: غیرمتضمن	برچسب هدف
I am writing my thesis and I am experiencing a problem with the space between numbering and the image name. You can see the problem in the picture I include in this question. It starts from 3.10. How can I fix that? I'm preparing a document using the book class. In the table of contents, the indentation before the subsection title is too long. How to reduce it? tikz pgf, tables, table of contents	پرسش ۱ پرسش ۲	مثال ۲	دانش کاربر	پیش‌بینی مدل آگاه به کاربر: متضمن	برچسب هدف
I have the border points of a 2-D body and wish to fill its interior with a solid color. Let us say I have N such points stacked in the N×2 matrix A so that the first column indicates the x co-ordinates and the second column denotes the corresponding y co-ordinates. How can I get the co-ordinates of the convex hull of a piece of Text? graphics, functions, qgis, arctgis desktop	پرسش ۱ پرسش ۲	مثال ۳	دانش کاربر	پیش‌بینی مدل مستقل از کاربر: غیرمتضمن	برچسب هدف
How do i multiply 2 array index variables given the following foo=(1 2) bar=(0.1 0.2) foobar=\$((foo[1]*price[1])) # this is wrong echo "\$foobar" Current output : 0 Correct/Expected output : 0.4 Using echo "20+5" literally produces the text "20+5". What command can I use to get the numeric sum, 25 in this case? Also, what's the easiest way to do it just using bash for floating point? For example, echo \$((3224/3807.0)) prints 0 :( I am looking for answers using either the basic command shell ('command line') itself or through using languages that are available from the command line. linux, bash, shell script, graphics, functions, programming, ubuntu, programming languages	پرسش ۱ پرسش ۲	مثال ۴	دانش کاربر	پیش‌بینی مدل آگاه به کاربر: متضمن	برچسب هدف
For instance: For as long as it takes... but For as long we have a future... I know how to use them because I heard them being used, but I don't know whether there is a rule or not. I would like to know whether or not the following sentences are correct and why: I have been sick since last week. I have been sick since the last week. I have been sick for the last week. meaning, word choice, grammar	پرسش ۱ پرسش ۲	مثال ۵	دانش کاربر	پیش‌بینی مدل مستقل از کاربر: غیرمتضمن	برچسب هدف

Generate a concise, single-paragraph answer to the question below based on the provided answers.

- Avoid unnecessary elaboration or examples.
- Focus on addressing the question's requirements.
- Ensure every provided answer is analyzed.
- Avoid using your world knowledge for answering. You can only use it for rephrasing answers to match the question requirement.

```
### Question: {Question}
### Answer1: {CandidateAnswer1}
### Answer2: {CandidateAnswer2}
### Answer3: {CandidateAnswer3}
### Response:
```

شکل ۸. الگوی اعلان مورد استفاده برای تولید پاسخ

۴-۴-۴- تحلیل عملکرد سیستم پیشنهادی در تولید پاسخ

در آخرین مؤلفه از سیستم پیشنهادی، پاسخ‌های تأییدشده متناظر با برترین پرسش‌های متضمن، به یک مدل زبانی مولد پاسخ ارسال می‌شوند. شکل ۸ الگوی اعلان ورودی به مدل زبانی را نمایش می‌دهد. جدول ۱۰ نیز نتایج ارزیابی پاسخ‌های تولیدشده نسبت به

مطالعه موردی. اگرچه ارزیابی کمی عملکرد مؤلفه شناسایی

پرسش‌های متضمن، برتری مدل آگاه به کاربر را نشان می‌دهد، تحلیل کیفی نیز می‌تواند مهر تأییدی بر این نتایج باشد. جدول ۹، در مثال‌های ۱ تا ۴، نمونه‌هایی را نشان می‌دهد که در آنها، آگاهی از دانش کاربر، به تشخیص صحیح ارتباط میان زوج پرسش انجامیده است. زوج پرسش مثال ۱، علیرغم عدم اشتراک واژگانی، مضمونی کاملاً مشابه یکدیگر دارند. در این حالت، تأکید بر موضوعاتی نظیر گرامر، معنا و گزینش کلمات در دانش کاربر، مدل را از تمرکز صرف بر تطابق واژگانی فراتر برده و آن را به درک عمیق‌تر مفهوم پرسش‌ها سوق داده است. البته، در موارد نادری نظیر مثال ۵، نادیده گرفتن تطابق واژگانی ممکن است نتیجه‌ای معکوس در پی داشته و موجب گمراهی مدل شود. به عنوان نمونه‌ای دیگر، پرسش اول در مثال ۴، اصلاح یک قطعه برنامه را از متخصصین طلب نموده است. این در حالیست که هیچ یک از کلمات linux, shell یا command به طور مستقیم در پرسش ذکر نشده‌اند. دانش کاربر، این شکاف واژگانی را پر نموده و پل ارتباطی میان زوج پرسش را برقرار کرده است.



و پاسخ‌های دریافتی‌اش را نمایش می‌دهد. در این مثال از یک مدل GPT-4o mini برای تولید بازنمایی و از all-mpnet-base-v2 برای تولید پاسخ استفاده شده و مقدار پارامتر  $k_3$  برابر با ۵ در نظر گرفته شده است. مدل‌های مذکور، در دو مورد از پنج پاسخ کاندید بازایی شده، اشتراک داشته و مابقی پاسخ‌های کاندید متفاوت بوده‌اند. از آنجا که پرسش ورودی شامل چند پرسش کوتاه‌تر در مورد موضوعات مختلف است، می‌توان آن را نسبتاً پیچیده دانست. همانطور که مشاهده می‌شود پاسخ ارائه شده توسط مدل آگاه به کاربر از ساختار منطقی و منسجم‌تری برخوردار بوده و تحلیل صحیحی در مورد ساختار گرامری جمله ارائه داده است. حال آنکه پاسخ ارائه شده توسط مدل مستقل از کاربر، ناکامل بوده و دارای توضیحات سطحی است.

پاسخ‌های تأییدشده برای پرسش‌های ورودی را گزارش می‌کند. این نتایج به وضوح بر برتری مدل آگاه به کاربر دلالت دارند. بطور خاص، اجراهایی که از GPT-4o mini به عنوان مدل تولیدکننده پاسخ و از all-mpnet-base-v2 برای تولید بازنمایی استفاده کرده‌اند، به ازای ۳ پاسخ کاندید به بهبود ۱/۰۷٪ و به ازای ۵ پاسخ کاندید به بهبود ۳/۷۹٪ در معیار Rouge-1 مبتنی بر فراخوانی دست یافته‌اند. این اختلاف در معیارهای Rouge-2 و Rouge-L نیز قابل توجه است.

**مطالعه موردی.** با توجه به آنکه تعداد نمونه‌های آزمون برای مؤلفه تولید پاسخ محدود است، تنها تعداد انگشت‌شماری از پرسش‌ها را می‌توان یافت که پاسخ متفاوتی را از دو مدل آگاه به کاربر و مستقل از کاربر دریافت نموده‌اند. جدول ۱۱، یک نمونه از پرسش‌های آزمون

جدول ۱۰. ارزیابی کیفی پاسخ‌های تولیدشده نسبت به پاسخ‌های تأییدشده در دادگان پژوهش، به تفکیک مدل بازنمایی و مدل مولد پاسخ

		معیار ارزیابی (%)						مدل شناسایی پرسش‌های متضمن		مدل بازنمایی	مدل تولید پاسخ	
BS-f	BS-r	RL-f	RL-r	R2-f	R2-r	R1-f	R1-r					
۸۲/۶۱	۸۲/۷۰	۱۳/۷۶	۲۲/۱۵	۴/۷۰	۷/۴۱	۲۳/۷۵	۳۵/۱۶	آگاه به کاربر	paraphrase-multilingual-MiniLM-L12-v2	LLAMA2	$k_3 = 3$	
۸۲/۵۷	۸۲/۷۳	۱۳/۶۳	۲۱/۹۷	۴/۶۷	۷/۳۸	۲۳/۷۰	۳۵/۰۸	مستقل از کاربر				
▲ ۰/۰۴	▽ ۰/۰۳	▲ ۰/۱۳	▲ ۰/۱۷	▲ ۰/۰۳	▲ ۰/۰۲	▲ ۰/۰۵	▲ ۰/۰۸	تفاوت				
۸۲/۸۹	۸۲/۵۷	۱۴/۳۲	۲۰/۸۶	۵/۰۶	۷/۲۷	۲۳/۹۶	۳۲/۵۴	آگاه به کاربر	all-mpnet-base-v2	LLAMA2	$k_3 = 3$	
۸۲/۸۴	۸۲/۵۱	۱۴/۱۴	۲۰/۵۶	۴/۹۲	۶/۹۹	۲۳/۸۰	۳۲/۲۱	مستقل از کاربر				
▲ ۰/۰۴	▲ ۰/۰۷	▲ ۰/۱۸	▲ ۰/۳۰	▲ ۰/۱۴	▲ ۰/۲۸	▲ ۰/۱۶	▲ ۰/۳۴	تفاوت				
۸۲/۵۶	۸۲/۹۸	۱۲/۵۹	۲۴/۲۹	۴/۵۸	۹/۲۷	۲۲/۸۵	۳۹/۸۶	آگاه به کاربر	paraphrase-multilingual-MiniLM-L12-v2	GPT-4o mini	$k_3 = 3$	
۸۲/۵۶	۸۲/۶۹	۱۲/۵۱	۲۲/۵۴	۴/۲۳	۷/۶۷	۲۳/۱۷	۳۷/۴۱	مستقل از کاربر				
— ۰/۰۰	▲ ۰/۲۹	▲ ۰/۰۸	▲ ۱/۷۴	▲ ۰/۳۵	▲ ۱/۶۱	▽ ۰/۳۲	▲ ۲/۴۵	تفاوت				
۸۲/۹۶	۸۲/۶۴	۱۳/۶۵	۲۶/۶۶	۵/۵۴	۱۰/۰۰	۲۴/۴۱	۴۲/۸۳	آگاه به کاربر	all-mpnet-base-v2	GPT-4o mini	$k_3 = 3$	
۸۲/۸۵	۸۳/۵۵	۱۳/۲۴	۲۵/۷۱	۵/۰۶	۹/۵۲	۲۳/۹۶	۴۱/۷۵	مستقل از کاربر				
▲ ۰/۱۱	▲ ۰/۰۹	▲ ۰/۴۱	▲ ۰/۹۵	▲ ۰/۴۸	▲ ۰/۴۸	▲ ۰/۴۶	▲ ۱/۰۷	تفاوت				
۸۲/۵۲	۸۲/۶۱	۱۳/۵۱	۲۲/۸۷	۴/۵۶	۷/۶۱	۲۳/۰۶	۳۵/۷۶	آگاه به کاربر	paraphrase-multilingual-MiniLM-L12-v2	LLAMA2	$k_3 = 5$	
۸۲/۴۶	۸۲/۶۰	۱۳/۳۹	۲۲/۲۵	۴/۴۶	۷/۰۶	۲۳/۱۱	۳۵/۳۱	مستقل از کاربر				
▲ ۰/۰۶	▲ ۰/۰۱	▲ ۰/۱۱	▲ ۰/۶۲	▲ ۰/۱۰	▲ ۰/۵۵	▽ ۰/۰۵	▲ ۰/۴۵	تفاوت				
۸۲/۵۳	۸۲/۳۰	۱۳/۳۴	۲۰/۹۱	۴/۴۸	۶/۷۵	۲۲/۹۹	۳۳/۲۲	آگاه به کاربر	all-mpnet-base-v2	LLAMA2	$k_3 = 5$	
۸۲/۶۰	۸۲/۳۶	۱۳/۴۵	۲۱/۰۴	۴/۴۷	۶/۷۳	۲۳/۰۰	۳۳/۲۸	مستقل از کاربر				
▽ ۰/۰۷	▽ ۰/۰۶	▽ ۰/۱۱	▽ ۰/۱۳	▲ ۰/۰۱	▲ ۰/۰۲	▽ ۰/۰۲	▽ ۰/۰۶	تفاوت				
۸۲/۲۹	۸۳/۱۶	۱۱/۹۴	۲۶/۵۳	۴/۴۲	۹/۴۹	۲۲/۱۴	۴۳/۲۳	آگاه به کاربر	paraphrase-multilingual-MiniLM-L12-v2	GPT-4o mini	$k_3 = 5$	
۸۲/۵۱	۸۲/۹۰	۱۲/۶۱	۲۴/۲۸	۴/۴۸	۸/۹۶	۲۳/۱۷	۳۹/۸۴	مستقل از کاربر				
▽ ۰/۲۲	▲ ۰/۲۶	▽ ۰/۶۷	▲ ۲/۲۵	▽ ۰/۰۶	▲ ۰/۵۳	▽ ۱/۰۲	▲ ۳/۴۰	تفاوت				
۸۲/۵۳	۸۳/۳۴	۱۲/۸۱	۲۶/۲۳	۴/۸۶	۹/۶۷	۲۳/۵۰	۴۲/۶۸	آگاه به کاربر	all-mpnet-base-v2	GPT-4o mini	$k_3 = 5$	
۸۲/۴۹	۸۲/۹۱	۱۲/۵۹	۲۳/۷۵	۴/۲۷	۷/۵۸	۲۳/۱۵	۳۸/۸۹	مستقل از کاربر				
▲ ۰/۰۴	▲ ۰/۴۳	▲ ۰/۲۲	▲ ۲/۴۸	▲ ۰/۵۹	▲ ۲/۱۰	▲ ۰/۳۵	▲ ۳/۷۹	تفاوت				

جدول ۱۱. مقایسه کیفیت پاسخ‌های ارائه شده در حضور و عدم حضور دانش کاربر

<p>Happy Holi to You and Your Family. Play Safe! Is anything grammatically wrong in above sentences? In which category does the second sentence fall? Can we replace <i>Play Safe!</i> with <i>Play Safely!</i> Are they equally true and depict same meaning? Is the usage of exclamation with both <i>safe</i> and <i>safely</i> correct?</p>	<p>پرسش</p>
<p>The second sentence, "Play Safe!", is grammatically correct but falls under an imperative mood, where it gives a directive. The phrase "Play Safely!" can replace "Play Safe!", but the meaning is slightly different: "Play Safely!" is an adverbial form that explicitly describes the manner of playing, whereas "Play Safe!" uses an adjective and has a more informal, idiomatic feel. Both versions are understood in the context of wishing someone a safe experience, but "Play Safe!" is more colloquial. Regarding the exclamation, both "Safe!" and "Safely!" can be used with an exclamation mark to emphasize urgency or excitement, though "Play Safe!" is the more commonly accepted version.</p>	<p>پاسخ مدل RAG آگاه به کاربر</p>
<p>The second sentence, "Play Safe!" is grammatically correct but uses an imperative verb with an adverbial expression. It falls into the category of an imperative sentence, where a command or suggestion is given. While "Play Safely!" can also be used, it is slightly more formal or precise, as "safely" is an adverb, describing the manner of playing. Both "Play Safe!" and "Play Safely!" convey the same general meaning, but "Play Safe!" may feel more informal. The usage of an exclamation mark with both is acceptable, as it emphasizes the message, although it may be seen as more common in informal contexts with "Play Safe!"</p>	<p>پاسخ مدل RAG مستقل از کاربر</p>
<p>The sentence "Happy Holi to You and Your Family. Play Safe!" is almost grammatically correct, but it can be refined for smoother flow. The second sentence, "Play Safe!" is an imperative sentence, giving a command or instruction. While "Play Safe!" and "Play Safely!" are both grammatically acceptable, there is a subtle difference in usage. "Play Safe!" is more idiomatic and commonly used in casual speech, often implying a general sense of caution or safety, while "Play Safely!" is the technically correct form, with "safely" being the proper adverb to describe how one should play. Both phrases convey nearly the same meaning, with "Play Safe!" being more informal and colloquial. The usage of the exclamation mark with both "safe" and "safely" is appropriate, as both are commands or encouragements, but "Play Safe!" is more often used with the exclamation mark in everyday language. Therefore, while both versions are acceptable, the choice between them depends on the tone you want to convey—informal and idiomatic or more formal and precise.</p>	<p>پاسخ مدل GPT-40 mini</p>
<p>Happy Holi! Well, first of all, in the English language, most words are not capitalized unless they are at the beginning of a sentence, or in the case of a "title", of sorts. Proper names are capitalized, but "you", "your", and "family" are not proper nouns. The one exception to this is if you are making a card of some sort. If you want it to stand out on a greeting card, then the capitalization may be appropriate. In that situation, it is a matter of personal choice. However, in most day-to-day conversations (or a post on someone's Facebook wall), it's generally best to avoid capitalizing words that would not be capitalized in a book. <i>Happy Holi to you and your family!</i> As for the second part, I'm not entirely sure that either of those phrases reflect your intended meaning. First of all, I am assuming by, "In which category does the second sentence fall?" you are asking about what type of sentence it is. The sentence seems to be what is called an <i>imperative</i> sentence. An imperative sentence is, essentially, a command. The subject of the sentence is implied to be "you", the one to whom the sentence is spoken. As for the meanings of the phrases, the imperative sentence, "Play safely," would be the type of warning you would give to a child, to remind him or her to avoid doing anything overly dangerous. The reason I say child, specifically, is because it is generally considered condescending to tell adults to play in most contexts, as "play" is widely considered to be a childish activity. On the other hand, the imperative sentence, "Play safe," is a (slightly-off) variation of an idiomatic expression, "Play it safe," and it has an entirely different meaning. To "play it safe" is to avoid taking risks, rather than to just be safe. Basically, you are telling the individual to go out of their way to make sure that there is as little damage done to them as possible. For the sake of clarity, telling someone to "play it safe" is <i>not</i> considered condescending, as it is an idiom which has little to do with the modern meaning of the word "play", but it is the wrong phrase for the meaning you likely wish to convey. If you wish to offer someone a wish of good health, and safety is the concern (or if you simply want to wish someone luck in staying safe), then the sentence I would recommend would be a simpler phrase: <i>Happy Holi to you and your family! Be safe!</i> This avoids the questionable use of "play" (as well as the grammar issues regarding the word "safe" not being an adverb; only adverbs can describe verbs), and makes the meaning clear. It is another imperative, telling "you" to be safe. Totally lovely.</p>	<p>پاسخ تأیید شده</p>

## ۵- نتیجه‌گیری و کارهای آتی

استفاده از دانش کاربر در شناسایی پرسش‌های متضمن، موضوعی است که تاکنون در ادبیات پژوهش به آن پرداخته نشده است. نتایج آزمایش‌ها روی دادگان CQAD-ReQuEST، عملکرد مطلوب سیستم پیشنهادی و بهبود کیفیت پاسخ‌ها را نشان داد. در مطالعات آتی می‌توان عملکرد سیستم پیشنهادی پژوهش را در پلتفرم‌های پاسخگویی به پرسش‌های خاص-دامنه همچون پزشکی و اقتصاد، که اعتبار و صحت پاسخ‌ها در آنها اهمیت دوچندان می‌یابد، ارزیابی نمود. علاوه بر این، طراحی مؤلفه شناسایی پرسش‌های متضمن بین‌زبانی به منظور بهره‌گیری از پاسخ‌های ارائه شده در سایر زبان‌ها، آموزش مدل‌های زبانی بزرگ برای خلاصه‌سازی چند پاسخ متمرکز

این مقاله، سیستمی با نام UserQA برای پاسخگویی به پرسش‌ها در انجمن‌های پرسش و پاسخ معرفی نموده است. این سیستم بر پایه شناخت پرسشگر از طریق تحلیل سوابق پرسش‌های وی در انجمن استوار است تا به واسطه آن، به درک عمیق‌تری از موضوع اصلی پرسش ورودی دست یافته و فهرست پرسش‌های متضمن را با دقت بالاتری از بایگانی انجمن بازیابی نماید. UserQA به مدد زنجیره‌ای از مدل‌های زبانی بزرگ، دانش پرسشگر را به صورت دنباله‌ای از برچسب‌های موضوعی، مدلسازی نموده و با دریافت یک زوج پرسش، وجود یا عدم وجود رابطه تضمین میان آنها را پیش‌بینی می‌کند. سپس، برترین پاسخ‌های کاندید را با تمرکز بر پرسش ورودی، در یک پاسخ واحد جمع می‌نماید. طبق بهترین دانش ما،

226–234, doi: 10.1137/1.9781611976236.26.

- [15] D. Hoogeveen, A. Bennett, Y. Li, K. Verspoor, and T. Baldwin, 'Detecting Misflagged Duplicate Questions in Community Question-Answering Archives', *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 12, no. 1, pp. 112–120, Jun. 2018, doi: 10.1609/icwsm.v12i1.15011.
- [16] P. K. Roy, S. Saumya, J. P. Singh, S. Banerjee, and A. Gutub, 'Analysis of community question- answering issues via machine learning and deep learning: State- of- the- art review', *CAAI Trans. Intell. Technol.*, vol. 8, no. 1, pp. 95–117, Mar. 2023, doi: 10.1049/cit2.12081.
- [17] N. Othman, R. Faiz, and K. Smaili, 'Enhancing Question Retrieval in Community Question Answering Using Word Embeddings', *Procedia Comput. Sci.*, vol. 159, pp. 485–494, 2019, doi: 10.1016/j.procs.2019.09.203.
- [18] G. Zhou, L. Cai, J. Zhao, and K. Liu, 'Phrase-based translation model for question retrieval in community question answer archives', *ACL-HLT 2011 - Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol.*, vol. 1, pp. 653–662, 2011.
- [19] K. Mrini et al., 'A Gradually Soft Multi-Task and Data-Augmented Approach to Medical Question Understanding', in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2021, pp. 1505–1515, doi: 10.18653/v1/2021.acl-long.119.
- [20] N. S. Tawfik and M. R. Spruit, 'Evaluating sentence representations for biomedical text: Methods and experimental results', *J. Biomed. Inform.*, vol. 104, no. February, p. 103396, Apr. 2020, doi: 10.1016/j.jbi.2020.103396.
- [21] M. Faseeh, M. A. Khan, N. Iqbal, F. Qayyum, A. Mehmood, and J. Kim, 'Enhancing User Experience on Q&A Platforms: Measuring Text Similarity Based on Hybrid CNN-LSTM Model for Efficient Duplicate Question Detection', *IEEE Access*, vol. 12, no. January, pp. 34512–34526, 2024, doi: 10.1109/ACCESS.2024.3358422.
- [22] V. Nguyen, S. Karimi, and Z. Xing, 'Combining Shallow and Deep Representations for Text-Pair Classification', *ALTA 2021 - Proc. 19th Work. Australas. Lang. Technol. Assoc.*, pp. 68–78, 2021.
- [23] S. Ghasemi and A. Shakeri, 'Harnessing the Power of Metadata for Enhanced Question Retrieval in Community Question Answering', *IEEE Access*, vol. 12, no. May, pp. 65768–65779, 2024, doi: 10.1109/ACCESS.2024.3395449.
- [24] A. M. Monea, 'Medical Question Entailment based on Textual Inference and Fine-tuned BioMed-RoBERTa', in *2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2021, pp. 319–326, doi: 10.1109/ICCP53602.2021.9733687.
- [25] S. Yadav, V. Pallagani, and A. Sheth, 'Medical Knowledge-enriched Textual Entailment Framework', in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 1795–1801, doi: 10.18653/v1/2020.coling-main.161.
- [26] T. R. Goodwin and D. Demner-Fushman, 'Bridging the Knowledge Gap: Enhancing Question Answering with World and Domain Knowledge', *CoRR abs/1910.07429*, Oct. 2019, [Online]. Available: <http://arxiv.org/abs/1910.07429>.
- [27] A. B. Abacha, D. Demner-fushman, and U. S. N. Library, 'Recognizing Question Entailment for Medical Question Answering', in *AMIA Annual Symposium Proceedings*, 2016, pp. 310–318.
- [28] T. Baldwin, H. Liang, B. Salehi, D. Hoogeveen, Y. Li, and L. Duong, 'UniMelb at SemEval-2016 Task 3: Identifying Similar Questions by combining a CNN with String Similarity Measures', in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 851–856, doi: 10.1137/1.9781611976236.26.
- بر پرسش، تنظیم نرم اعلان<sup>1</sup> و مدیریت شروع سرد<sup>2</sup> و پویایی کاربر از موضوعات تحقیقاتی بالقوه برای پژوهش‌های آتی به‌شمار می‌آیند.

## مراجع

- [1] B. Patra, 'A survey of Community Question Answering', *CoRR, abs/1705.04009*, May 2017, [Online]. Available: <http://arxiv.org/abs/1705.04009>.
- [2] D. Hoogeveen, L. Wang, T. Baldwin, and K. M. Verspoor, 'Real and Misflagged Duplicate Question Detection in Community Question-Answering', 2018.
- [3] A. Figueroa, 'Automatically generating effective search queries directly from community question-answering questions for finding related questions', *Expert Syst. Appl.*, vol. 77, pp. 11–19, Jul. 2017, doi: 10.1016/j.eswa.2017.01.041.
- [4] I. Srba and M. Bielikova, 'A Comprehensive Survey and Classification of Approaches for Community Question Answering', *ACM Trans. Web*, vol. 10, no. 3, pp. 1–63, Aug. 2016, doi: 10.1145/2934687.
- [5] M. Sulir and M. Regeci, 'Software Engineers' Questions and Answers on Stack Exchange', in *2022 IEEE 16th International Scientific Conference on Informatics (Informatics)*, Nov. 2022, pp. 304–310, doi: 10.1109/Informatics57926.2022.10083403.
- [6] M. Asaduzzaman, A. S. Mashiyat, C. K. Roy, and K. A. Schneider, 'Answering questions about unanswered questions of Stack Overflow', in *2013 10th Working Conference on Mining Software Repositories (MSR)*, May 2013, pp. 97–100, doi: 10.1109/MSR.2013.6624015.
- [7] S. A. Bhaskar, R. Rungta, J. Route, E. Nyberg, and T. Mitamura, 'Sieg at MEDIQA 2019: Multi-task Neural Ensemble for Biomedical Inference and Entailment', in *Proceedings of the 18th BioNLP Workshop and Shared Task*, 2019, pp. 462–470, doi: 10.18653/v1/W19-5049.
- [8] A. Merchant, N. Shenoy, A. Bharali, and A. K. M., 'Identifying Similar Questions in the Medical Domain Using a Fine-tuned Siamese-BERT Model', in *2022 IEEE 19th India Council International Conference (INDICON)*, Nov. 2022, pp. 1–6, doi: 10.1109/INDICON56171.2022.10040144.
- [9] A. B. Abacha and D. Demner-Fushman, 'A question-entailment approach to question answering', *BMC Bioinformatics*, vol. 20, no. 1, Dec. 2019, doi: 10.1186/s12859-019-3119-4.
- [10] Z. Xu and H. Yuan, 'Forum Duplicate Question Detection by Domain Adaptive Semantic Matching', *IEEE Access*, vol. 8, pp. 56029–56038, 2020, doi: 10.1109/ACCESS.2020.2982268.
- [11] L. Wang, L. Zhang, and J. Jiang, 'Duplicate Question Detection With Deep Learning in Stack Overflow', *IEEE Access*, vol. 8, pp. 25964–25975, 2020, doi: 10.1109/ACCESS.2020.2968391.
- [12] D. D. Koswate and S. Hettiarachchi, 'Optimized Duplicate Question Detection in Programming Community Q&A Platforms using Semantic Hashing', in *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, Aug. 2021, pp. 375–380, doi: 10.1109/ICIAfS52090.2021.9606030.
- [13] V.-T. Nguyen, A.-C. Le, and H.-N. Nguyen, 'A Model of Convolutional Neural Network Combined with External Knowledge to Measure the Question Similarity for Community Question Answering Systems', *Int. J. Mach. Learn. Comput.*, vol. 11, no. 3, pp. 194–201, May 2021, doi: 10.18178/ijmlc.2021.11.3.1035.
- [14] R. Zhang, Q. Zhou, B. Wu, W. Li, and T. Mo, 'What do questions exactly ask? mfae: duplicate question identification with multi-fusion asking emphasis', in *Proceedings of the 2020 SIAM International Conference on Data Mining, SDM 2020*, 2020, pp.

<sup>2</sup> Cold start

<sup>1</sup> Soft prompt tuning

- Transfer Learning with Data Augmentation for Recognizing Question Entailment in the Medical Domain', in *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*, Aug. 2021, pp. 339–346, doi: 10.1109/ICHI52183.2021.00058.
- [39] H. Zhou, X. Li, W. Yao, C. Lang, and S. Ning, 'DUT-NLP at MEDIQA 2019: An Adversarial Multi-Task Network to Jointly Model Recognizing Question Entailment and Question Answering', in *Proceedings of the 18th BioNLP Workshop and Shared Task*, 2019, pp. 437–445, doi: 10.18653/v1/W19-5046.
- [40] A. Shang, X. Zhu, M. Danner, and M. Rättsch, 'Unsupervised question-retrieval approach based on topic keywords filtering and multi-task learning', *Comput. Speech Lang.*, vol. 87, no. March, p. 101644, Aug. 2024, doi: 10.1016/j.csl.2024.101644.
- [41] K. Jiang, G. Jin, Z. Zhang, R. Cui, and Y. Zhao, 'Incorporating external knowledge for text matching model', *Comput. Speech Lang.*, vol. 87, no. March, p. 101638, Aug. 2024, doi: <https://doi.org/10.1016/j.csl.2024.101638>.
- [42] Y. Sun and J. Song, 'Research on question retrieval method for community question answering', *Multimed. Tools Appl.*, vol. 82, pp. 24309–24325, 2023, doi: <https://doi.org/10.1007/s11042-023-14458-2>.
- [43] D. Hoogeveen, K. M. Verspoor, and T. Baldwin, 'CQADupStack: A Benchmark Data Set for Community Question-Answering Research', in *Proceedings of the 20th Australasian Document Computing Symposium*, Dec. 2015, pp. 1–8, doi: 10.1145/2838931.2838934.
- [44] S. Z. Aftabi, S. M. Seyyedi, M. Maleki, and S. Farzi, 'ReQuEST: A Small-Scale Multi-Task Model for Community Question-Answering Systems', *IEEE Access*, vol. 12, pp. 17137–17151, 2024, doi: 10.1109/ACCESS.2024.3358287.
- [45] C. Y. Lin, 'Rouge: A package for automatic evaluation of summaries', *Proc. Work. text Summ. branches out (WAS 2004)*, no. 1, pp. 25–26, 2004, [Online]. Available: [papers2://publication/uuid/5DDA0BB8-E59F-44C1-88E6-2AD316DAEF85](https://arxiv.org/abs/2004.09675).
- [46] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, 'BERTScore: Evaluating Text Generation with BERT', *CoRR abs/1904.09675*, pp. 1–43, Apr. 2019, [Online]. Available: <http://arxiv.org/abs/1904.09675>.
- [47] K. Song, X. Tan, T. Qin, J. Lu, and T.-Y. Liu, 'MPNet: Masked and Permuted Pre-training for Language Understanding', *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, pp. 1–11, Apr. 2020, [Online]. Available: <http://arxiv.org/abs/2004.09297>.
- 10.18653/v1/S16-1131.
- [29] A. B. Abacha and D. Demner-Fushman, 'On the Summarization of Consumer Health Questions', in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 2228–2234, doi: 10.18653/v1/P19-1215.
- [30] C. dos Santos, L. Barbosa, D. Bogdanova, and B. Zadrozny, 'Learning Hybrid Representations to Retrieve Semantically Equivalent Questions', in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2015, vol. 2, no. 1, pp. 694–699, doi: 10.3115/v1/P15-2114.
- [31] D. A. Prabowo and G. Budi Herwanto, 'Duplicate Question Detection in Question Answer Website using Convolutional Neural Network', in *2019 5th International Conference on Science and Technology (ICST)*, Jul. 2019, no. 1, pp. 1–6, doi: 10.1109/ICST47872.2019.9166343.
- [32] A. Agrawal, R. Anil George, S. S. Ravi, S. Kamath S, and A. Kumar, 'ARS\_NITK at MEDIQA 2019: Analysing Various Methods for Natural Language Inference, Recognising Question Entailment and Medical Question Answering System', in *Proceedings of the 18th BioNLP Workshop and Shared Task*, 2019, pp. 533–540, doi: 10.18653/v1/W19-5059.
- [33] N. Tawfik and M. Spruit, 'UU\_TAILS at MEDIQA 2019: Learning Textual Entailment in the Medical Domain', in *Proceedings of the 18th BioNLP Workshop and Shared Task*, 2019, pp. 493–499, doi: 10.18653/v1/W19-5053.
- [34] V. Nguyen, S. Karimi, and Z. Xing, 'ANU-CSIRO at MEDIQA 2019: Question Answering Using Deep Contextual Knowledge', in *Proceedings of the 18th BioNLP Workshop and Shared Task*, 2019, pp. 478–487, doi: 10.18653/v1/W19-5051.
- [35] D. Bandyopadhyay, B. Gain, T. Saikh, and A. Ekbal, 'IITP at MEDIQA 2019: Systems Report for Natural Language Inference, Question Entailment and Question Answering', in *Proceedings of the 18th BioNLP Workshop and Shared Task*, 2019, pp. 517–522, doi: 10.18653/v1/W19-5056.
- [36] K. Mrini, F. Deroncourt, W. Chang, E. Farcas, and N. Nakashole, 'Joint Summarization-Entailment Optimization for Consumer Health Question Understanding', in *Proceedings of the Second Workshop on Natural Language Processing for Medical Conversations*, 2021, pp. 58–65, doi: 10.18653/v1/2021.nlpmc-1.8.
- [37] W. Zhu *et al.*, 'PANLP at MEDIQA 2019: Pre-trained Language Models, Transfer Learning and Knowledge Distillation', in *Proceedings of the 18th BioNLP Workshop and Shared Task*, 2019, pp. 380–388, doi: 10.18653/v1/W19-5040.
- [38] M. Sarrouti, A. B. Abacha, and D. Demner-Fushman, 'Multi-Task

## پیوست ۱

### جدول الف. نمونه‌هایی از زوج پرسش‌های متضمن و غیرمتضمن در دادگان پژوهش

مثال ۱. برچسب: متضمن					
شناسه پرسش	شناسه پرسشگر	تاریخ ثبت	شناسه نظرات	پاسخ‌ها	تعداد علاقمندان
۱۶۹۷۰۷	۱۸۴۹۵	۲۰۱۴-۰۴-۰۵T۰۹:۵۴:۳۴.۹۶۷	۳۹۰۷۳۰-۳۹۰۶۵۶-۳۹۰۶۵۵	۱۶۹۷۰۸	۰
تعداد ملاقات‌ها	پرسش‌های تکراری	پرسش‌های مرتبط	برچسب‌ها	شناسه پاسخ تأییدشده	نام انجمن
۷۷	۸۲۹۹۳	۸۲۹۹۳	memoir, chapters, naming	۱۶۹۷۰۸	Tex
<p>Replace the text 'Chapter' with custom text in Memoir</p> <p>How can I replace the word 'Chapter' with some other text string in Memoir class? So, instead of Chapter 1, it appears as, say, MyChapter 1. Also, the text is in Devanagari script so MyChapter will be some Devanagari text string.</p> <p>Use <code>\renewcommand{\chaptername}{MyChapter}</code> at any position before you the first chapter which should get another chapter name. If you want to rename all chapter names, put the command in the preamble.</p>					
<p>عنوان</p> <p>بدنه</p> <p>پاسخ تأییدشده</p>					
مثال ۲. برچسب: غیرمتضمن					
شناسه پرسش	شناسه پرسشگر	تاریخ ثبت	شناسه نظرات	پاسخ‌ها	تعداد علاقمندان
۸۲۹۹۳	۵۱۰	۲۰۱۲-۱۱-۱۶T۱۴:۲۴:۵۷.۰۵۰	-	۸۲۹۹۴	۵۰
تعداد ملاقات‌ها	پرسش‌های تکراری	پرسش‌های مرتبط	برچسب‌ها	شناسه پاسخ تأییدشده	نام انجمن
۱۰۳۶۱	۵۴۸۱۹	۵۴۸۱۹-۳۶۵۷۵-۶۲۲۱۳-۸۲۹۶۰	naming	۸۲۹۹۴	Tex
<p>How to change the name of document elements like Figure Contents Bibliography etc.</p> <p>I'd like to modify the names used to typeset some of my document elements. For example, the caption of figure floats should change from "Figure" to "Fig.", and my <code>\tableofcontents</code> sectioning heading shouldn't read "Contents", but "Table of Contents". How can I do that? <code>\documentclass{article} % \usepackage[english]{babel} \begin{document} \tableofcontents \section{foo} \begin{figure}[h] \centering \rule{1cm}{1cm} placeholder for graphic \caption{A figure} \end{figure} \end{document}</code></p> <p>The answer depends on whether or not you use the <code>\babel</code> [...] To change the formatting of the "Contents" heading, either [...] For captions, the <code>\caption</code> package offers a host of customization possibilities.</p>					
<p>عنوان</p> <p>بدنه</p> <p>پاسخ تأییدشده</p>					
مثال ۳. برچسب: غیرمتضمن					
شناسه پرسش	شناسه پرسشگر	تاریخ ثبت	شناسه نظرات	پاسخ‌ها	تعداد علاقمندان
۱۵۳۷۷۸	۲۴۰۷۲	۲۰۱۴-۰۱-۱۳T۰۵:۳۰:۰۱.۲۳۷	۴۲۲۵۲۵	۱۵۳۷۹۰	۲
تعداد ملاقات‌ها	پرسش‌های تکراری	پرسش‌های مرتبط	برچسب‌ها	شناسه پاسخ تأییدشده	نام انجمن
۲۹۵	-	-	fonts	-	Tex
<p>Open dyslexic font</p> <p>Is it straight forward to use the OpenDyslexic font in a latex document? What steps would be required?</p> <p>-</p>					
<p>عنوان</p> <p>بدنه</p> <p>پاسخ تأییدشده</p>					
مثال ۴. برچسب: غیرمتضمن					
شناسه پرسش	شناسه پرسشگر	تاریخ ثبت	شناسه نظرات	پاسخ‌ها	تعداد علاقمندان
۱۳۸۶۵	۴۳۲۲	۲۰۱۱-۰۳-۱۹T۰۹:۲۰:۱۷.۱۴۷	۲۵۴۷۵-۲۵۴۷۶-۲۵۴۹۰...	۱۳۸۶۸	۲۱
تعداد ملاقات‌ها	پرسش‌های تکراری	پرسش‌های مرتبط	برچسب‌ها	شناسه پاسخ تأییدشده	نام انجمن
۸۸۱۱	-	۳۷۹۷	online	۱۳۸۶۸	Tex
<p>How to use LaTeX on blogspot?</p> <p>I'm trying to use LaTeX on my blog, but I could not figure out how. Has anyone done it before? Any feedback or related information would be greatly appreciated.</p> <p>You can put a <code>\strut</code> just before <code>c</code> to enlarge the depth of your equation and if you still find the curly bracket too close, you can also define your own "strut":  <code>\p{&lt;code&gt; [...] &lt;code&gt;</code></p>					
<p>عنوان</p> <p>بدنه</p> <p>پاسخ تأییدشده</p>					