



Scientific QA System with Verifiable Answers

Nikola Milošević

Nikola.milosevic@bayer.com

Bayer A.G.

The Institute for Artificial Intelligence Research and Development of Serbia



TEAM LEAD Dr Nikola Milošević



Dr Adela Ljajić



Dr Bojana Bašaragin



Miloš Košprdić



Darija Medvecki



Lorenzo Cassano
Intern



Nataša Radmilović



Angela Pupovac



+



THE INSTITUTE FOR ARTIFICIAL INTELLIGENCE
RESEARCH AND DEVELOPMENT OF SERBIA



Petar Stevanović

OUR TEAM



LLMs and hallucinations

- Generative LLMs can produce answers that appear coherent, confident and articulate
- However, the information conveyed may not be correct or verifiable
- the limited internal knowledge of generative LLMs can hinder their ability to deliver factually accurate answers, particularly within specialized field
- notably concerning in biomedicine, where accurate and factual answers are critical
- privacy, sovereignty and security concerns in pharma and biomedicine often necessitate building systems where all components are controllable (e.g., deployed in-house), to avoid reliance on third-party APIs

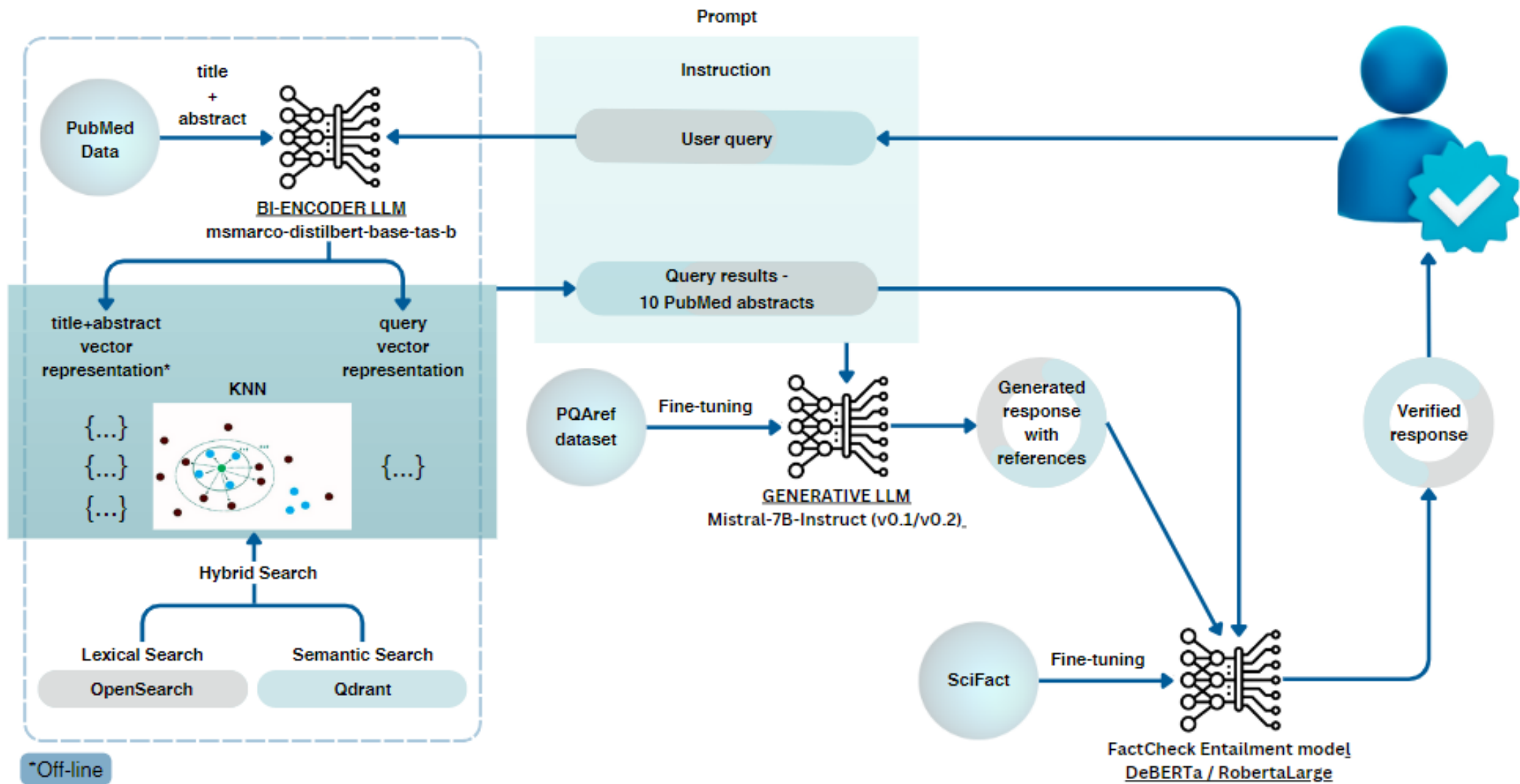
Retrieval Augmented
Generation (RAG)

Fine-tune the model

Provide sources

Check the sources

Ways to
address
hallucinations



Verif.ai



INFORMATION
RETRIEVAL SYSTEM

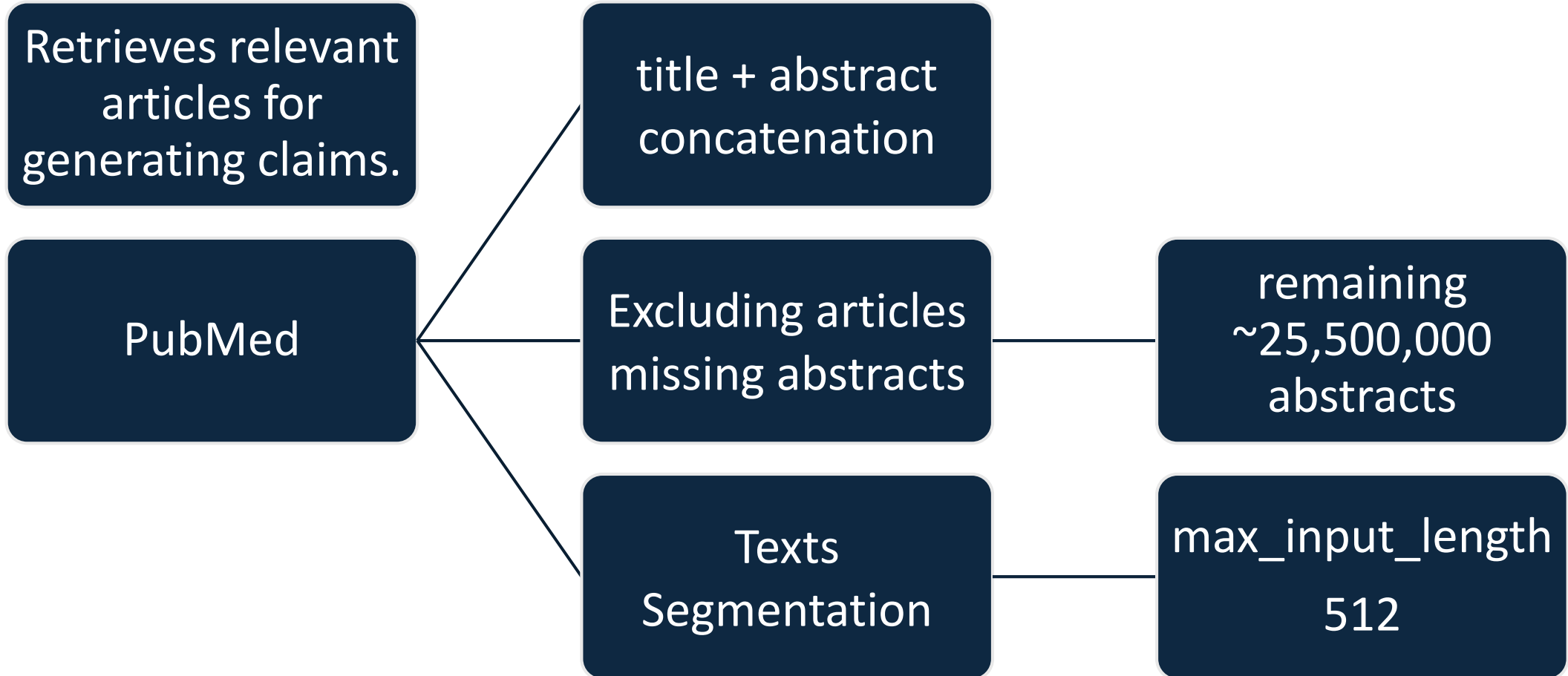


GENERATIVE
COMPONENT



VERIFICATION ENGINE

IR System



IR – Lexical Search (LS)

Sparse vectors

BM25

Metadata for filtering:

- authors' names
- publication dates
- journal names

OpenSearch

IR – Semantic Search (SS)



Dense vectors



Embedding generation – bi-encoder for Assymmetric Semantic Search

sentence-transformers/msmarco-distilbert-base-tas-b
MS MARCO Dataset



Vector comparison:

Hierarchical Navigable Small World (HNSW) indexing technique
Approximate Nearest Neighbours clustering
Dot product metrics



Qdrant Vector Database

8-bit quantized embeddings

IR – Hybrid Search



Combination of LS and SS

identification of direct matches
ability to discover semantically related phrases



Prerequisite – normalizing LS and SS scores



Importance weights

Evaluation: BioASQ Dataset
Best combination: 0.7 LS and 0.3 SS

Generative Component

– PQAref Dataset

- Randomly selected ~ 9000 questions from PubMedQA dataset (PMIDs)
- For each question, 10 relevant abstracts from the PubMed repository
- GPT-4 Turbo for creating the answers based on the retrieved abstracts
 - currently the number one model on the Chatbot Arena leaderboard
- The prompt to instruct GPT-4 Turbo to use references (PMIDs)

Answer the question using relevant abstracts provided, up to 300 words. Reference the statements with the provided abstract_id in brackets next to the statement.

Generative Component – LLM training

Mistral 7B Instruct (v0.1 and v0.2)

Fine-tuned for the Referenced QA Task, using:

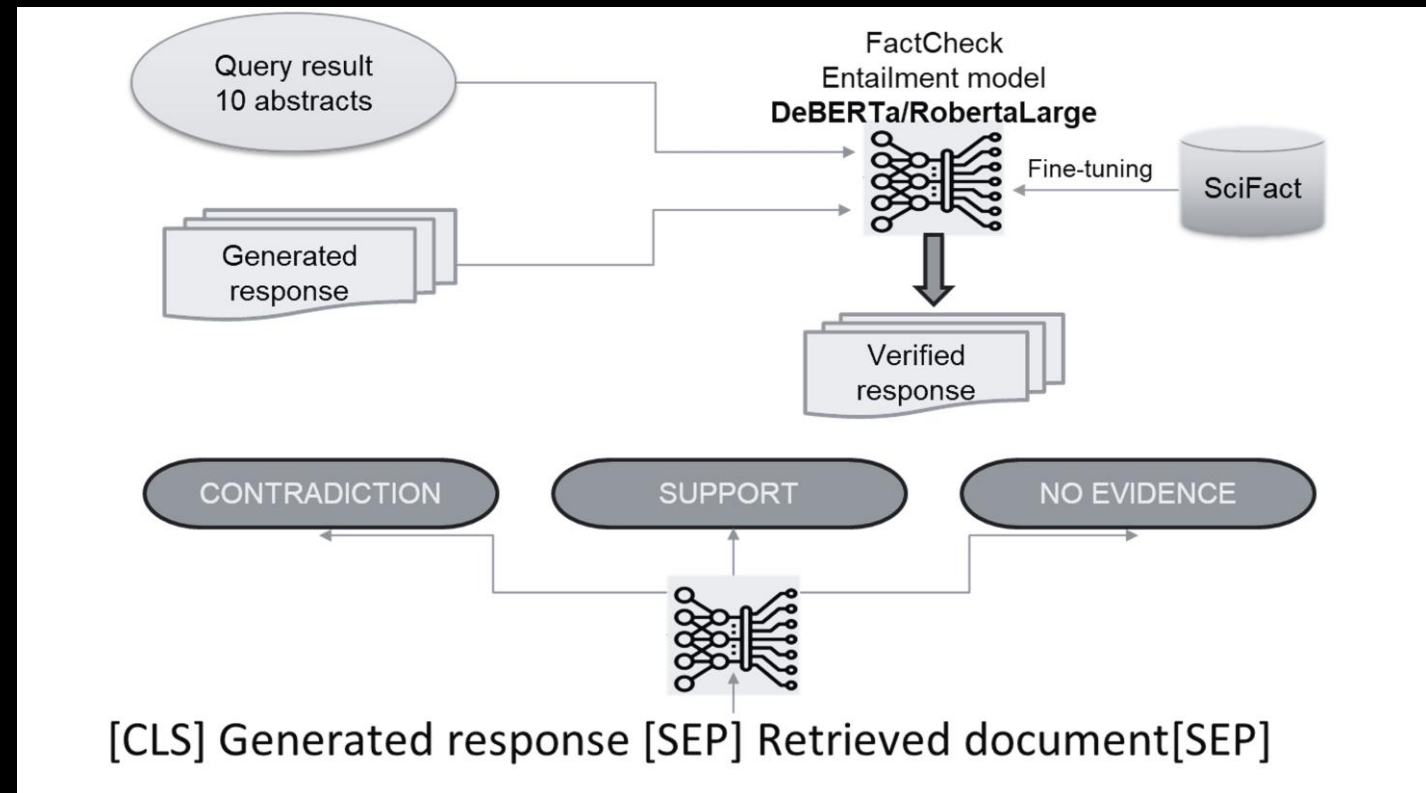
- PQAref Dataset (80 : 10 : 10)
- QLoRA methodology
 - Rank = 64
 - Alpha = 16
 - LoRA dropout = 0.1

single DGX NVIDIA A100-40GB GPU

- ~ 27M trainable parameters
- ~ 32 hours of training over 2 epochs, using a batch size of 1

Verification Engine

- Cross-checks the generated claim and the abstract from which the claim was derived
- Textual Entailment Task
 - Each sequence pair is categorized into 3 classes
- SciFact Dataset Transformation
 - 80 : 10 : 10 train : valid : test split



Results – IR System

	P@10	MAP@10	time [ms]
1. Semantic without rescore	14.0%	25.7%	245
2. Semantic with rescore	14.4%	26.0%	297
3. Hybrid with rescore (lex. 0.1 sem. 0.9)	24.7%	32.5%	442
4. Hybrid with rescore (lex. 0.2 sem. 0.8)	24.7%	32.5%	442
5. Hybrid with rescore (lex. 0.3 sem. 0.7)	24.7%	32.5%	442
6. Hybrid with rescore (lex. 0.4 sem. 0.6)	24.7%	32.6%	442
7. Hybrid with rescore (lex. 0.5 sem. 0.5)	25.2%	41.0%	442
8. Hybrid with rescore (lex. 0.6 sem. 0.4)	30.7%	42.0%	442
9. Hybrid with rescore (lex. 0.7 sem. 0.3)	30.8%	42.5%	442
10. Hybrid with rescore (lex. 0.8 sem. 0.2)	30.8%	42.5%	442
11. Hybrid with rescore (lex. 0.9 sem. 0.1)	30.8%	42.6%	442
12. Lexical with stopwords removal	28.7%	41.1%	189
13. Lexical without stopwords removal	28.3%	40.1%	189
PubMed without Mesh Terms	9.2%	15.3%	698
PubMed with Mesh Terms	12.0%	19.1%	742

Results – Generative Component

N	GPT-4 Turbo	M1	M2
0	2	8	5
1	241	86	105
2	76	138	112
3	128	185	178
4	126	172	169
5	119	117	124
6	87	72	75
7	45	66	34
8	29	27	34
9	31	22	23
10	24	15	49
Sum of referenced abstracts	3,464	3,648	3,816
% of 9080 abstracts	38.15	40.18	42.03
Avg no of references per answer	3.81	4.01	4.20

	GPT-4 Turbo	M1	M2
Relevant missed	1 (0.1%)	29 (3.5%)	10 (1.2%)
Relevant referenced	822 (99.9%)	794 (96.5%)	813 (98.8%)

Table 2: Evaluation of the generative component on a subset of the BioASQ dataset.

Type	Samples	SAME_CONCLUSION	ALL_INFO	PERC_IDEAL
yesno	51	82.35%	52.94%	74.43%
factoid	49	85.71%	59.18%	74.72%
summary	42	73.81%	40.48%	69.59%
list	36	83.33%	44.44%	80.36%
all	178	81.46%	50%	74.57%

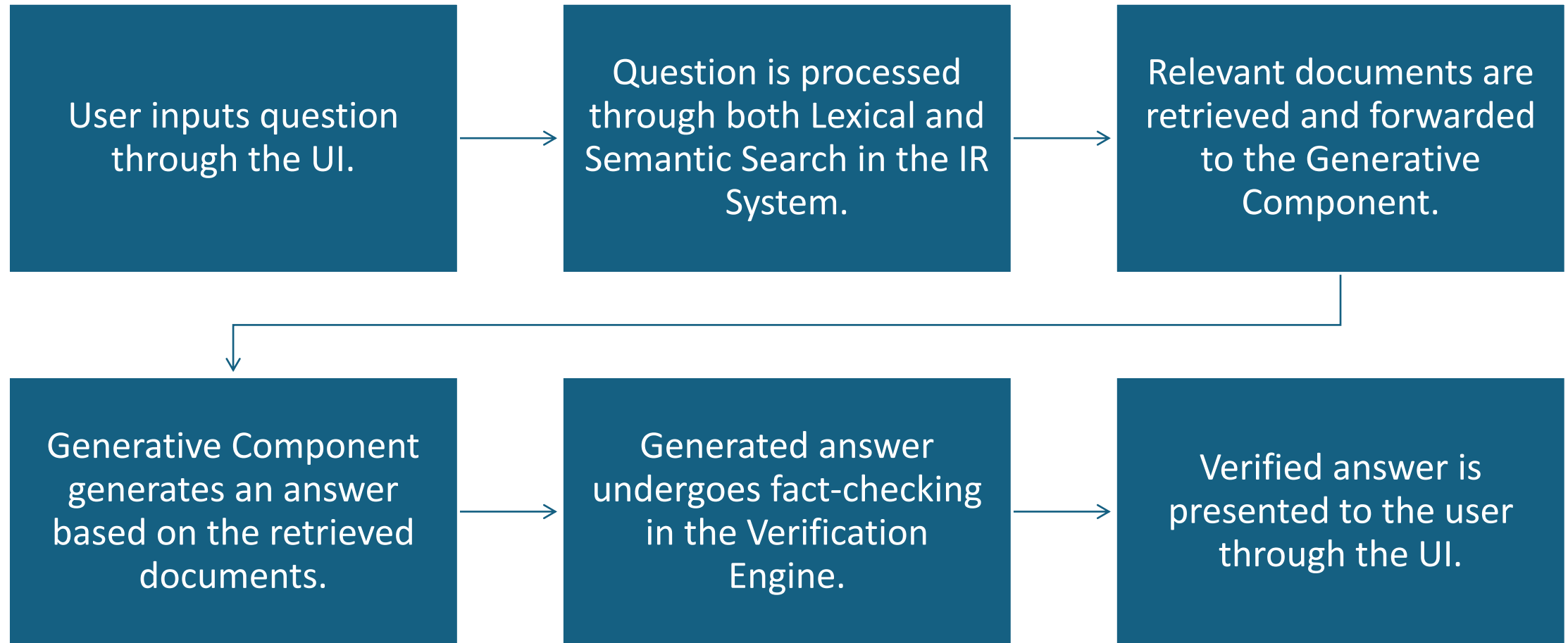
Results – Verification Engine

	RoBERTa L _{SF}				XLM RoBERTa L _{SF}				DeBERTa _{SF}				DeBERTa SQuAD _{SF}			
	NE*	S	C	wa	NE	S	C	wa	NE	S	C	wa	NE	S	C	wa
3 P	0.71	0.55	0.00	0.48	0.83	0.69	0.54	0.71	0.83	0.86	0.85	0.84	0.86	0.90	0.82	0.87
3 R	0.73	0.82	0.00	0.61	0.89	0.67	0.52	0.71	0.86	0.84	0.81	0.84	0.86	0.88	0.85	0.87
F1	0.72	0.66	0.00	0.53	0.86	0.68	0.53	0.71	0.84	0.85	0.83	0.84	0.86	0.89	0.84	0.87
Acc	0.61				0.71				0.84				0.87			
4 P	0.85	0.75	0.67	0.77	0.75	0.76	0.71	0.74	0.88	0.90	0.88	0.89	0.82	0.91	0.88	0.87
4 R	0.89	0.76	0.59	0.77	0.91	0.67	0.63	0.75	0.95	0.88	0.78	0.89	0.93	0.84	0.81	0.87
F1	0.87	0.76	0.63	0.77	0.82	0.71	0.67	0.74	0.91	0.89	0.82	0.88	0.87	0.88	0.85	0.87
Acc	0.77				0.75				0.89				0.87			

* NE: no.evidence, S: support, C: contradict, P: precision, R: recall, F1: F1 score,
Acc - accuracy

	DeBERTa_80				DeBERTa_90			
	NE	S	C	wa	NE	S	C	wa
P	0.46	0.70	0.66	0.60	0.47	0.67	0.69	0.59
R	0.94	0.25	0.15	0.50	0.88	0.29	0.27	0.52
F1	0.62	0.37	0.24	0.44	0.61	0.40	0.39	0.48
Acc	0.50				0.52			

User Interface





SIGN IN

LOG IN

New to VerifAi app? [Sign up](#) to get instant access.

SEARCH CONFIGURATION

Type of Search:

Hybrid

Lexical Weights: 0.700, Semantic Weights: 0.300



SEMANTIC

NEUTRAL

LEXICAL

Number of Documents:

Normal - 10 documents

From:

01.01.1940

To:

01.01.2030

Temperature: 0:

The higher the temperature, the less accurate answers will be.



PRECISE

NEUTRAL

CREATIVE

Stream:

Stream



Input your question in natural language:

What genes play role in breast cancer?

WHAT GENES PLAY ROLE IN BREAST CANCER?

SOURCES:

Breast cancer gene discovery. Many important advances have...	Genetic susceptibility to breast and ovarian cancer... 5-10% of women with breast c...	Migratory gene expression signature predicts poor ... Breast cancer metastasis acc...	Associations between cytokine gene variations and... UNLABELLED Persistent pain f...	Polymorphism of the homologous recombination repa... The RAD51 protein and its pa...	View 7 more
--	---	--	---	---	-------------

ANSWER:

The genes that play a role in breast cancer include BRCA1 and BRCA2, which are high-penetrance breast cancer susceptibility genes. Mutations in these genes are responsible for approximately 5% of all breast carcinomas. However, the genes responsible for the majority of breast cancer cases remain to be identified, and there is evidence that there are additional high-to-moderate-penetrance breast cancer susceptibility genes.

There are also many candidate low-penetrance breast cancer genes that are likely to be identified, and in addition to germline and somatic sequence alterations, epigenetic changes in many genes are likely to play an important role in the pathobiology of breast cancer ([PUBMED:14987381](#)).

BRCA1 and BRCA2 genes are large, of essentially unknown function, and play a very small role in the development of tumors in non-predisposed individuals. These genes have been cloned, and a highly automated genetic test has been developed to screen high-risk populations to establish the utility of the test ([PUBMED:9296070](#)).

Other genes that have been studied for their role in breast cancer include cytokine genes, such as IL1A, IL1B, IL1RN, IL4R, IL6, and IL10, but none of these polymorphisms showed any overall predisposition to breast cancer susceptibility, severity, or time to death or occurrence of distant metastases ([PUBMED:3684207](#)).

Additionally, polymorphisms in genes involved in the metabolism of xenobiotics have been associated with breast cancer risk, particularly in premenopausal breast cancer, and there is an interaction between variants in these genes and tobacco smoking ([PUBMED:30303577](#)).

Furthermore, genetic variation in the homologous recombination repair genes RAD51 and XRCC3 may play a role in breast cancer occurrence and progression, although the role may be underlined by a mutual interaction between these genes ([PUBMED:19620772](#)).

Lastly, the expression of certain genes, such as APOA2, DLX5, APOC3, CAMK2B, and PAK6, may play an immense role in breast cancer tumorigenesis and progression in response to anti-cancer drug Paclitaxel ([PUBMED:38933604](#)).

In summary, a variety of genes, including high-penetrance genes like BRCA1 and BRCA2, as well as low-penetrance genes and genes involved in xenobiotic metabolism and homologous recombination repair, are implicated in the development and progression of breast cancer. The specific genes and their roles may vary depending on the context, such as the presence of other genetic or environmental factors.



What genes play role in breast cancer?



WHAT GENES PLAY ROLE IN BREAST CANCER?

SOURCES:

Breast cancer gene discovery. Many important advances have...	Genetic susceptibility to breast and ovarian cancer... 5-10% of women with breast c...	Migratory gene expression signature predicts poor ... Breast cancer metastasis acc...	Associations between cytokine gene variations and ... UNLABELLED Persistent pain f...	Polymorphism of the homologous recombination repa... The RAD51 protein and its pa...	View 7 more
--	---	--	--	---	-------------

ANSWER:

The genes that play a role in breast cancer include BRCA1 and BRCA2, which are high-penetrance breast cancer susceptibility genes. Mutations in these genes are responsible for approximately 5% of all breast carcinomas. However, the genes responsible for the majority of breast cancer cases remain to be identified, and there is evidence that there are additional high-to-moderate-penetrance breast cancer susceptibility genes.

There are also many candidate low-penetrance breast cancer genes that are likely to be identified, and in addition to germline and somatic sequence alterations, epigenetic changes in many genes are likely to play an important role in the pathobiology of breast cancer ([PUBMED:14987381](#)).

BRCA1 and BRCA2 genes are large, of essentially unknown function, and play a very small role in the development of tumors in non-predisposed individuals. These genes have been cloned, and a highly automated genetic test has been developed to screen high-risk populations to establish the utility of the test ([PUBMED:9296070](#)).

Other genes that have been studied for their role in breast cancer include cytokine genes, such as IL1A, IL1B, IL1RN, IL4R, IL6, and IL10, but none of these polymorphisms showed any overall predisposition to breast cancer susceptibility, severity, or time to death or occurrence of distant metastases ([PUBMED:3684207](#)).

Additionally, polymorphisms in genes involved in the metabolism of xenobiotics have been associated with breast cancer risk, particularly in premenopausal breast cancer, and there is an interaction between variants in these genes and tobacco smoking ([PUBMED:30303577](#)).

Furthermore, genetic variation in the homologous recombination repair genes RAD51 and XRCC3 may play a role in breast cancer occurrence and progression, although the role may be underlined by a mutual interaction between these genes ([PUBMED:19620772](#)).

Lastly, the expression of certain genes, such as APOA2, DLX5, APOC3, CAMK2B, and PAK6, may play an immense role in breast cancer tumorigenesis and progression in response to anti-cancer drug Paclitaxel ([PUBMED:38933604](#)).

In summary, a variety of genes, including high-penetrance genes like BRCA1 and BRCA2, as well as low-penetrance genes and genes involved in xenobiotic metabolism and homologous recombination repair, are implicated in the development and progression of breast cancer. The specific genes and their roles may vary depending on the context, such as the presence of other genetic or environmental factors.

The claim for document [PUBMED:14987381](#) is SUPPORT ● Closest

Sentence on the abstract: In addition to germline and somatic sequence alterations, epigenetic changes in many genes are likely to play an important role in the pathobiology of breast cancer.

Future Work

- Broaden the project's scope beyond biomedicine
- Long term hosting solution
- Enterprise solution
- Incorporate feedback from the scientific community, and continuously improve generative search engine
- foster trust in generative AI across various scientific domains



PUBLICATIONS

- Adela Ljajić, Miloš Košprdić, Bojana Bašaragin, Darija Medvecki, Lorenzo Cassano, Nikola Milošević, "Scientific QA System with Verifiable Answers", The 6th International Open Search Symposium 2024
- Košprdić, M., Ljajić, A., Bašaragin, B., Medvecki, D., & Milošević, N. "Verif. ai: Towards an Open-Source Scientific Generative Question-Answering System with Referenced and Verifiable Answers." The Sixteenth International Conference on Evolving Internet INTERNET 2024 (2024).
- Bojana Bašaragin, Adela Ljajić, Darija Medvecki, Lorenzo Cassano, Miloš Košprdić, Nikola Milošević "How do you know that? Teaching Generative Language Models to Reference Answers to Biomedical Questions", Accepted at BioNLP 2024, Colocated with ACL 2024
- Adela Ljajić, Lorenzo Cassano, Miloš Košprdić, Bojana Bašaragin, Darija Medvecki, Nikola Milošević "Enhancing Biomedical Information Retrieval with Semantic Search: A Comparative Analysis Using PubMed Data", Belgrade Bioinformatics Conference 2024, BelBi2024

THANK YOU FOR YOUR ATTENTION!



www.verifai-project.com



verif.ai.project@gmail.com



<https://github.com/nikolamilosevic86/verifAI>

LinkedIn

Verif.ai Project

Instagram

verif.ai_project

X / Twitter

Verif.ai Project

Facebook

Verif.ai Project

TikTok

verif.ai