

# МУЛЬТИМОДАЛЬНИЙ RAG З ВИКОРИСТАННЯМ ТЕКСТОВИХ ТА ВІЗУАЛЬНИХ ДАНИХ

Виконав: Шевченко Михайло Григорович  
Науковий керівник: Андрощук Максим Віталійович



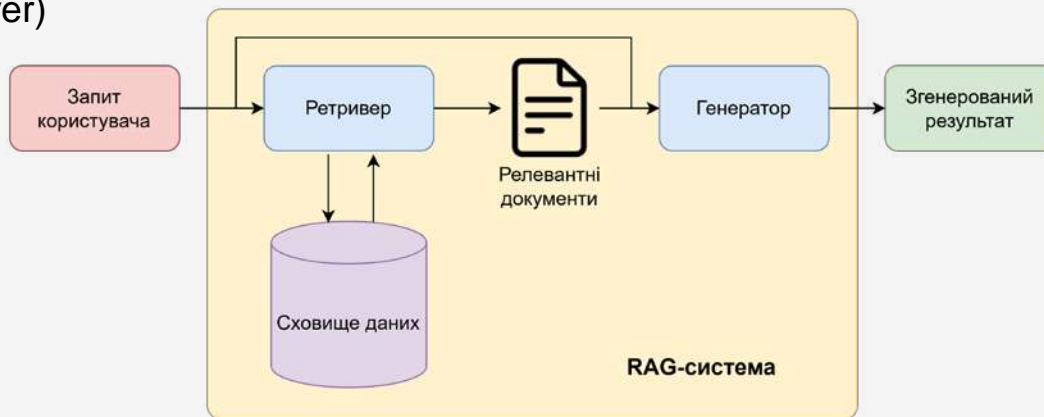
# Мета роботи

- Дослідити можливості та функції мультимодального RAG з використанням візуальних та текстових даних
- Виконати аналіз інструментів та програмних засобів для створення мультимодальних RAG-систем з використанням візуальних та текстових даних
- Розробити мультимодальну RAG-систему для роботи з рентгенівськими знімками грудної клітки та і їхніми звітами
- Протестувати розроблену систему за різними метриками



# RAG

- RAG — Retrieval-Augmented Generation, генерація з доповненою вибіркою
- RAG системи складаються з трьох основних частин:
  - генератор (англ. generator)
  - ретривер (англ. retriever)
  - сховище даних



Приклад можливої архітектури RAG-системи

# Мультиmodalний RAG

- Мультиmodalний RAG працює з двома або більше типами даних. Наприклад: текст, зображення, відео, аудіо тощо.



Приклад можливої архітектури мультиmodalної RAG-системи

# Функціональність створеної мультимодальної RAG-системи

- Працює з рентгенівськими знімками грудної клітки та відповідними звітами
- Відповідає на запитання до вхідного зображення
- Генерує звіт на вхідне зображення

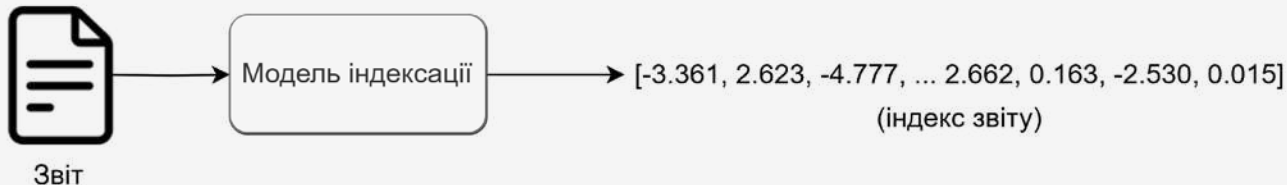


Приклад рентгенівського знімку



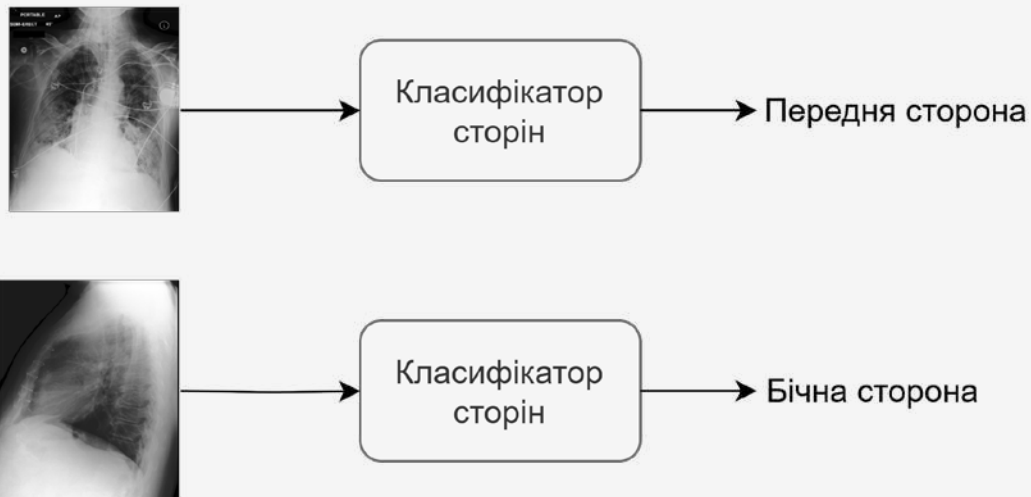
# Ретривер. Моделі індексації

- Основа — BioMedCLIP
- Створюють з зображень та звітів вектори (індекси), що знаходяться в одному просторі
- Створено 3 моделі індексації



# Ретривер. Класифікатор сторін

- Основа — DenseNet121
- Визначає «сторону» рентгенівського знімку (передня, бічна)



# Ретривер. Класифікатор хвороб

- Основа — DenseNet121
- Визначає наявність кожної з 14-ти клінічних ознак для певного зображення

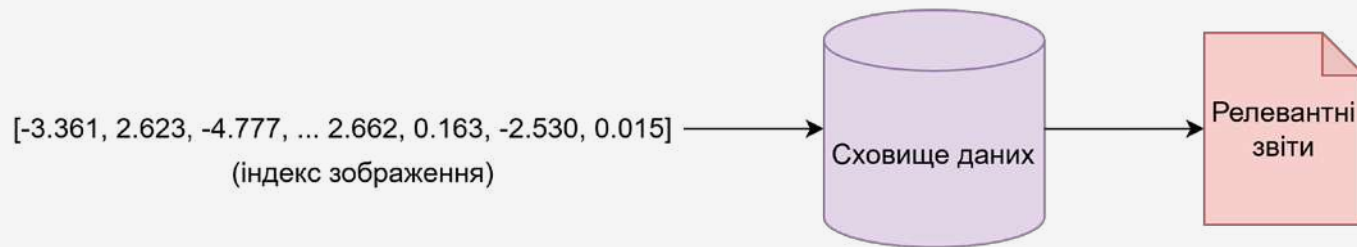


Класифікатор  
хвороб

Enlarged Cardiomeastinum: Наявне  
Cardiomegaly: Не визначено  
Lung Opacity: Наявне  
Lung Lesion: Наявне  
Edema: Не визначено  
Consolidation: Не наявне  
Pneumonia: Наявне  
Atelectasis: Наявне  
Pneumothorax: Не наявне  
Pleural Effusion: Не наявне  
Pleural Other: Не наявне  
Fracture: Не наявне  
No Finding: Не наявне  
Support Devices: Не визначено

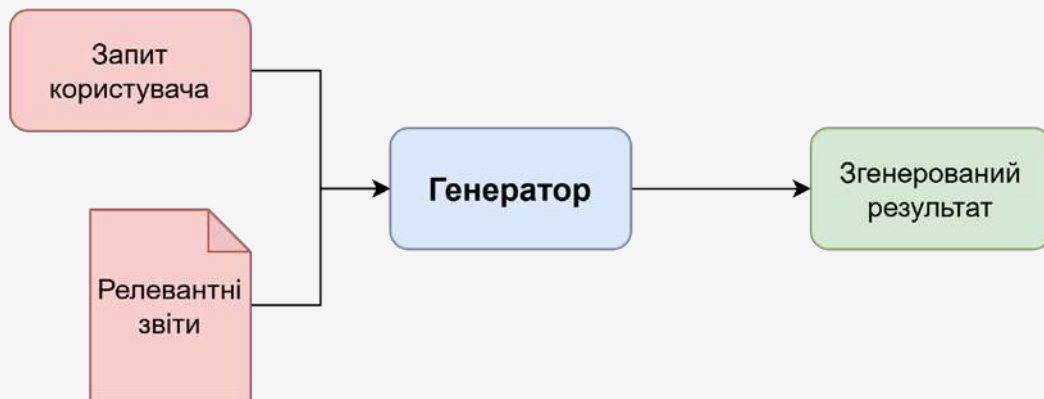
# Сховище даних

- База даних — ChromaDB
- Знаходить  $n$  найрелевантніших звітів за вхідним індексом за допомогою косинусної подібності



# Генератор

- Основа — LLaVA-Med 1.5
- Генерує текстовий результат відповідно до запиту користувача та релевантних звітів, знайдених ретривером



# Приклади роботи розробленої системи

## Генерування звіту



### Згенерований звіт:

«The chest X-ray image shows a large right pleural effusion, which is an abnormal accumulation of fluid in the pleural space surrounding the lungs. This finding is concerning for pneumonia, which is an infection that causes inflammation in the air sacs of the lungs. It is important to consult a healthcare professional for a thorough evaluation and proper diagnosis of the underlying cause of these findings.»

## Відповідь на питання до зображення



### Питання:

«Is the cardiomeastinal silhouette within normal limits?»

### Згенерована відповідь:

«Yes, the cardiomeastinal silhouette appears to be within normal limits in the image.»

# Тестування системи. Конфігурація 1

- З класифікатором сторін, класифікатором хвороб, НЕ точне співпадіння хвороб



# Тестування системи. Конфігурація 2

- 3 класифікатором сторін, з класифікатором хвороб, точне співпадіння хвороб



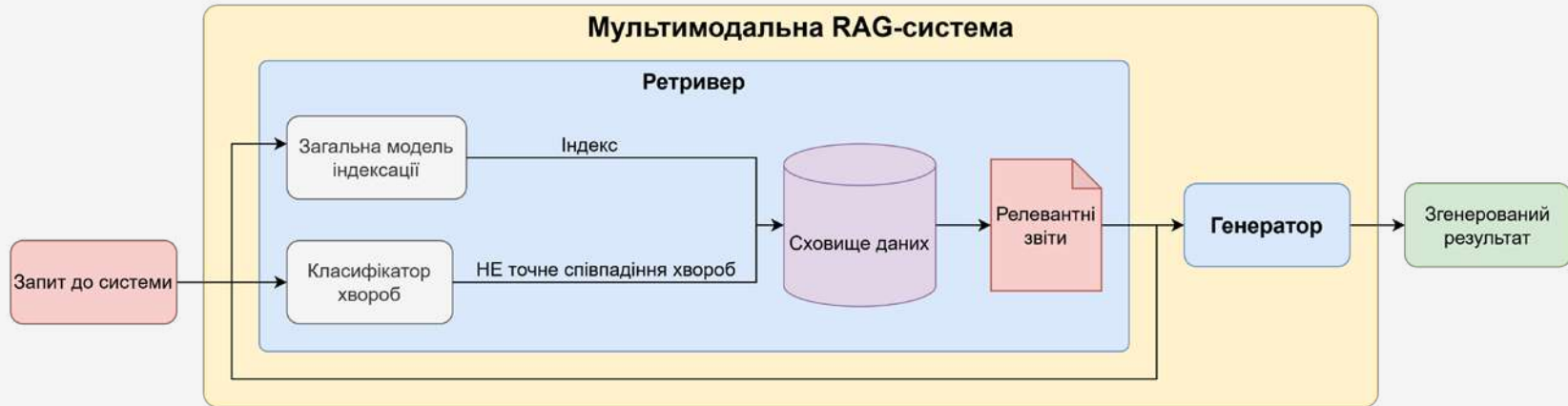
# Тестування системи. Конфігурація 3

- 3 класифікатором сторін, БЕЗ класифікатора хвороб



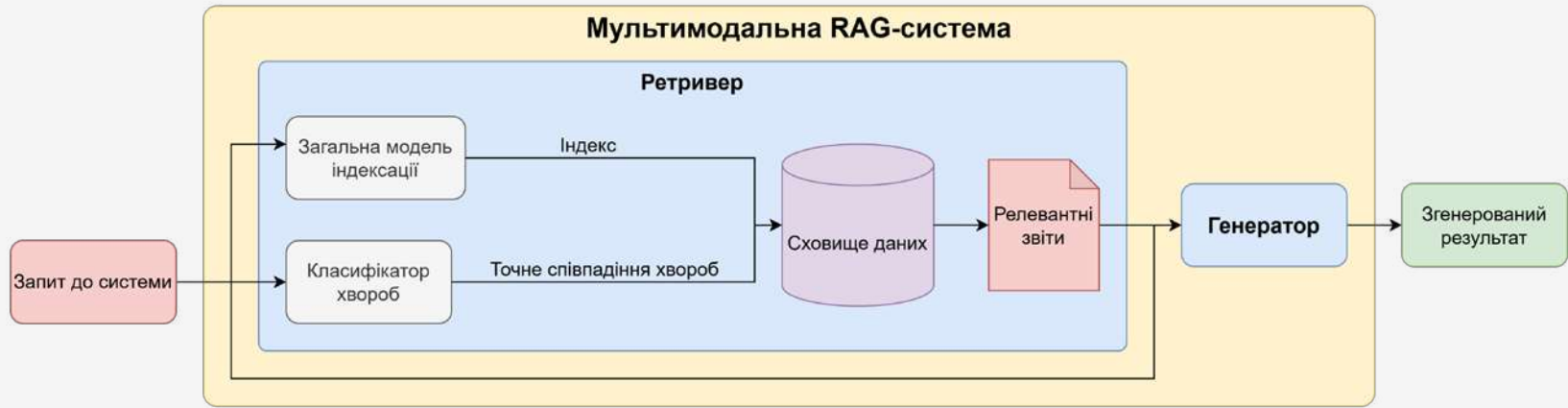
# Тестування системи. Конфігурація 4

- БЕЗ класифікатора сторін, з класифікатором хвороб, НЕ точне співпадіння хвороб



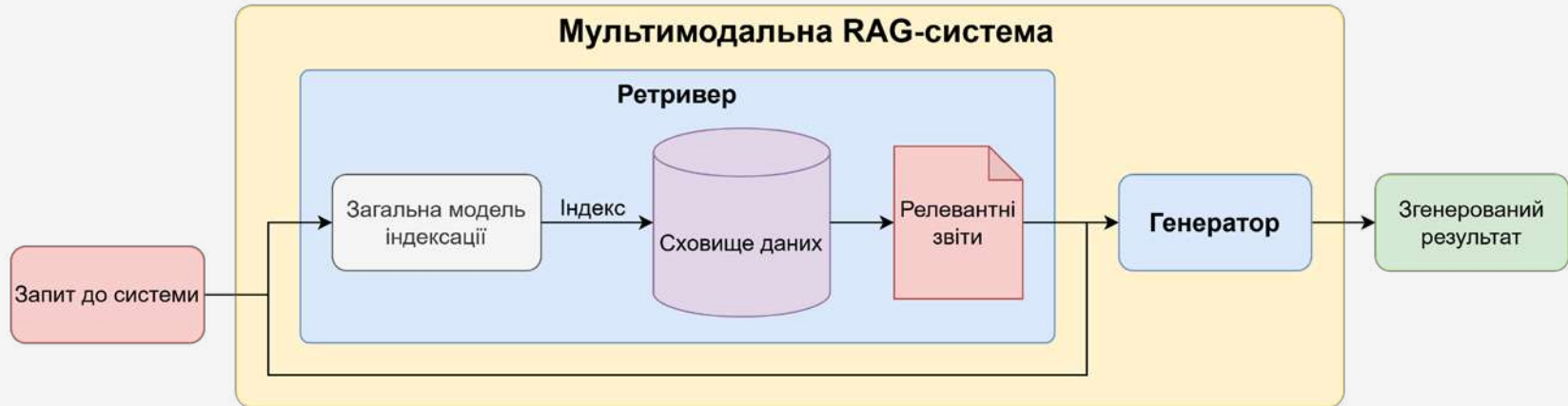
# Тестування системи. Конфігурація 5

- БЕЗ класифікатора сторін, з класифікатором хвороб, точне співпадіння хвороб



# Тестування системи. Конфігурація 6

- БЕЗ класифікатора сторін, БЕЗ класифікатора хвороб



# Тестування системи. Відповідь на питання до вхідного зображення. Метрика точності та F1

	MIMIC-CXR		IU-Xray	
	Точність	Зважена F1	Точність	Зважена F1
LLaVA-Med 1.5	0.717	0.668	0.411	0.418
Конфігурація 1 (з класифікатором сторін, з класифікатором хвороб, НЕ точне співпадіння хвороб)	0.840	0.838	0.921	0.924
Конфігурація 2 (з класифікатором сторін, з класифікатором хвороб, точне співпадіння хвороб)	0.844	0.842	0.909	0.913
Конфігурація 3 (з класифікатором сторін, БЕЗ класифікатора хвороб)	0.831	0.830	0.929	0.931
Конфігурація 4 (БЕЗ класифікатора сторін, з класифікатором хвороб, НЕ точне співпадіння хвороб)	0.766	0.766	0.917	0.919
Конфігурація 5 (БЕЗ класифікатора сторін, з класифікатором хвороб, точне співпадіння хвороб)	0.786	0.784	0.918	0.921
Конфігурація 6 (БЕЗ класифікатора сторін, БЕЗ класифікатора хвороб)	0.773	0.773	0.921	0.923

# Тестування системи. Генерування звіту до вхідного зображення. Метрика BLEU

	<b>MIMIC-CXR</b>	<b>IU-Xray</b>
LLaVA-Med 1.5	0.006204	0.014087
Конфігурація 1 (з класифікатором сторін, з класифікатором хвороб, НЕ точне співпадіння хвороб)	0.025932	0.035426
Конфігурація 2 (з класифікатором сторін, з класифікатором хвороб, точне співпадіння хвороб)	0.042499	0.036582
Конфігурація 3 (з класифікатором сторін, БЕЗ класифікатора хвороб)	0.027740	0.034322
Конфігурація 4 (БЕЗ класифікатора сторін, з класифікатором хвороб, НЕ точне співпадіння хвороб)	0.018719	0.035311
Конфігурація 5 (БЕЗ класифікатора сторін, з класифікатором хвороб, точне співпадіння хвороб)	0.022738	0.037315
Конфігурація 6 (БЕЗ класифікатора сторін, БЕЗ класифікатора хвороб)	0.018848	0.035051

# Тестування системи. Генерування звіту до вхідного зображення. Метрика F1-CheXbert

	MIMIC-CXR	IU-Xray
LLaVA-Med 1.5	0.011978	0.120694
Конфігурація 1 (з класифікатором сторін, з класифікатором хвороб, НЕ точне співпадіння хвороб)	0.266570	0.217418
Конфігурація 2 (з класифікатором сторін, з класифікатором хвороб, точне співпадіння хвороб)	0.289483	0.219341
Конфігурація 3 (з класифікатором сторін, БЕЗ класифікатора хвороб)	0.263669	0.226831
Конфігурація 4 (БЕЗ класифікатора сторін, з класифікатором хвороб, НЕ точне співпадіння хвороб)	0.221617	0.182259
Конфігурація 5 (БЕЗ класифікатора сторін, з класифікатором хвороб, точне співпадіння хвороб)	0.235491	0.180509
Конфігурація 6 (БЕЗ класифікатора сторін, БЕЗ класифікатора хвороб)	0.214062	0.179557

# Тестування системи. Генерування звіту до вхідного зображення. Метрика F1-RadGraph

	<b>MIMIC-CXR</b>	<b>IU-Xray</b>
LLaVA-Med 1.5	0.020411	0.059309
Конфігурація 1 (з класифікатором сторін, з класифікатором хвороб, НЕ точне співпадіння хвороб)	0.095536	0.120658
Конфігурація 2 (з класифікатором сторін, з класифікатором хвороб, точне співпадіння хвороб)	0.101452	0.119638
Конфігурація 3 (з класифікатором сторін, БЕЗ класифікатора хвороб)	0.097419	0.122248
Конфігурація 4 (БЕЗ класифікатора сторін, з класифікатором хвороб, НЕ точне співпадіння хвороб)	0.076431	0.118745
Конфігурація 5 (БЕЗ класифікатора сторін, з класифікатором хвороб, точне співпадіння хвороб)	0.079204	0.122624
Конфігурація 6 (БЕЗ класифікатора сторін, БЕЗ класифікатора хвороб)	0.078109	0.112216

# Висновки

- Досліджено можливості та функції мультимодального RAG і систем, які його використовують
- Розглянуто інструменти й програмні засоби для створення мультимодальних RAG-систем
- Створено мультимодальну RAG-систему для роботи з рентгенівськими знімками грудних кліток і їхніми звітами
- Протестовано різні конфігурації створеної системи за різними метриками: точність, F1, BLEU, ROUGE, F1-CheXbert, F1-RadGraph
- Визначено найкращу конфігурацію створеної системи