

How Roche's Data Platform accelerates Feature Engineering through Generative AI

Vishakha Sharma, PhD
Senior Principal Data Scientist

Roche Diagnostics



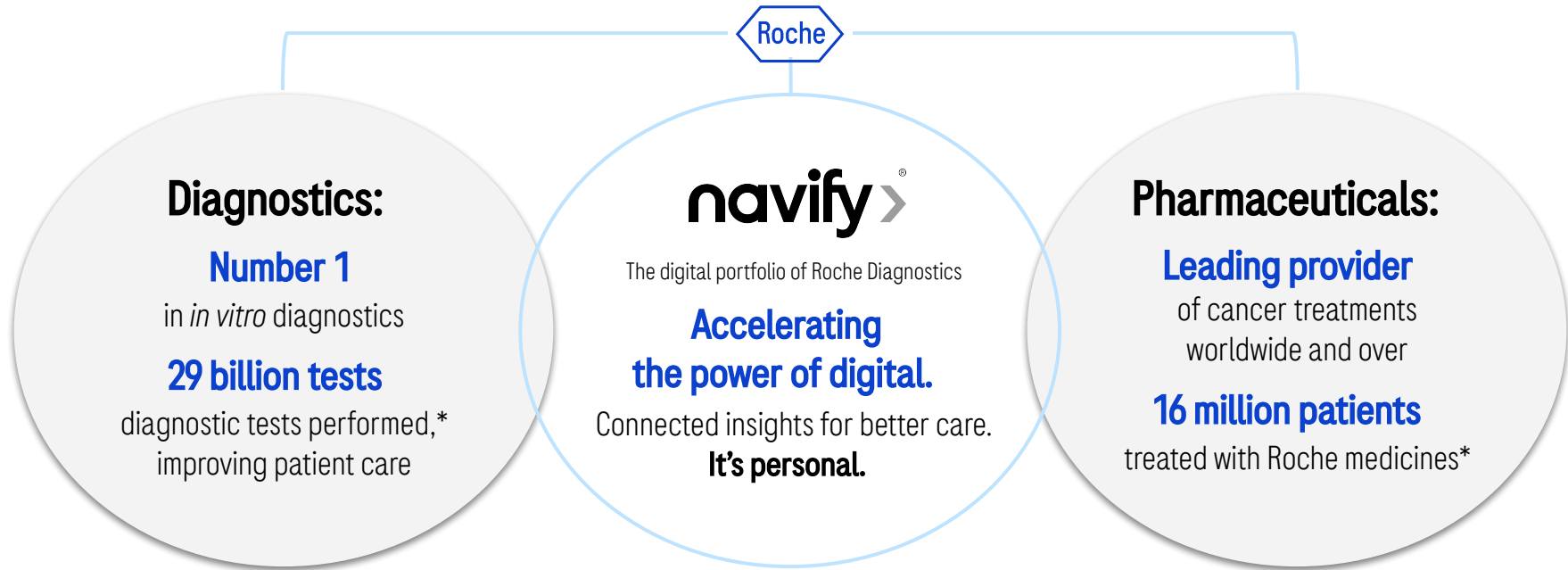
FEATURE STORE SUMMIT 2024

DATA FOR AI:
REAL-TIME, BATCH, AND LLMS

Organized by  **HOPSWORKS**

Bringing healthcare understanding to technology

Leveraging expertise from screening through diagnosis to personalized care

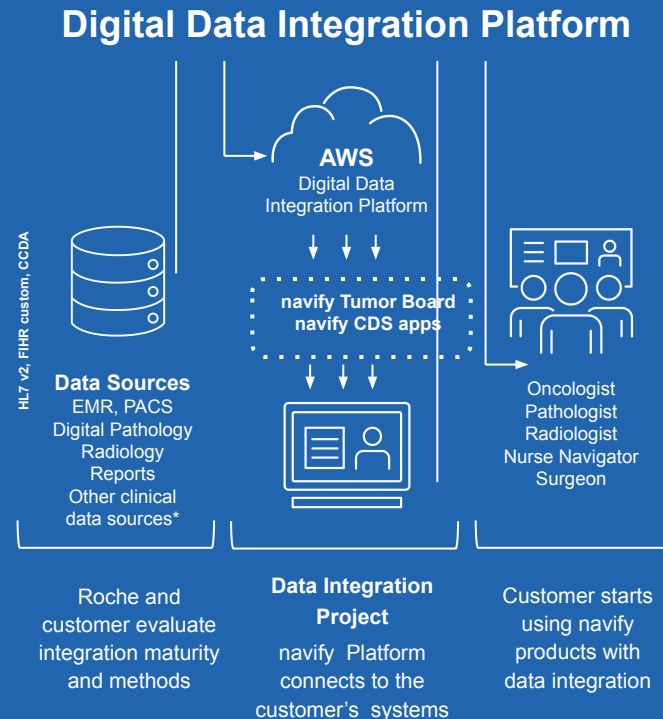


Integrating your data with navify products

Implementation builds the interface between navify portfolio and your IT infrastructure

Data integration and implementation

Phases	Added capabilities
<p>Integration Connects hospital sources to transfer data in various formats using encryption protocols</p>	Evaluates systems authorized to send data and verify data transmitted
<p>Implementation Maps current and future processes to create and execute a transition plan between process states</p>	Identifies gaps/obstacles for navify portfolio use at your institution and suggests process improvements
<p>Support Trains clinicians and supports navify product users</p>	Tracks and manages issues



Technology and services that connect hospitals to navify portfolio

*In development

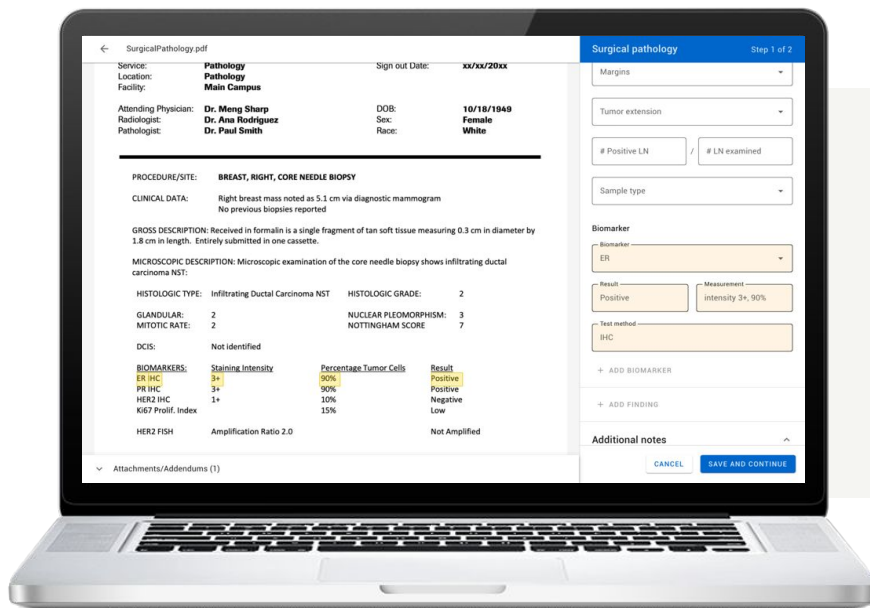
navify® Clinical Hub: AI-Assisted Data Abstraction

Leveraging Artificial Intelligence (AI) and Natural Language Processing (NLP) to structure and activate more patient data to enhance clinical decision-making

navify



navify Clinical Hub



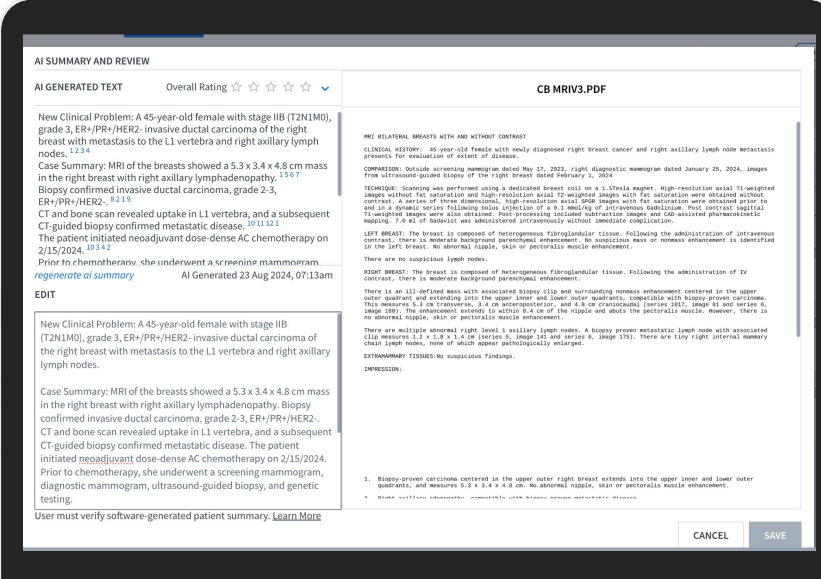
Assistive technology leverages Machine Learning and Natural Language Processing to auto-detect and extract valuable medical concepts from unstructured reports for downstream use, saving time while minimizing manual data entry

* Initially optimized for pathology reports. Support for additional report types will be added in future releases.

Automating Case Summarization with GenAI

Streamline and standardize patient case summarization for tumor board meetings

navify Clinical Hub



AI SUMMARY AND REVIEW

AI GENERATED TEXT Overall Rating ☆☆☆☆

CB MRIV3.PDF

NEW CLINICAL PROBLEM: A 45-year-old female with stage IIB (T2N1M0), grade 3, ER+/PR+/HER2- invasive ductal carcinoma of the right breast with metastasis to the L1 vertebra and right axillary lymph nodes.^{1,2,3,4}

Case Summary: MRI of the breasts showed a 5.3 x 3.4 x 4.8 cm mass in the right breast with right axillary lymphadenopathy.^{1,5,7} Biopsy confirmed invasive ductal carcinoma, grade 2-3, ER+/PR+/HER2-.^{1,1,9} CT and bone scan revealed uptake in L1 vertebra, and a subsequent CT-guided biopsy confirmed metastatic disease.^{10,11,12} The patient initiated neoadjuvant dose-dense AC chemotherapy on 2/15/2024.^{10,12} Prior to chemotherapy, she underwent a screening mammogram, regenerate ai summary AI Generated 23 Aug 2024, 07:13:am

EDIT

NEW CLINICAL PROBLEM: A 45-year-old female with stage IIB (T2N1M0), grade 3, ER+/PR+/HER2- invasive ductal carcinoma of the right breast with metastasis to the L1 vertebra and right axillary lymph nodes.

Case Summary: MRI of the breasts showed a 5.3 x 3.4 x 4.8 cm mass in the right breast with right axillary lymphadenopathy. Biopsy confirmed invasive ductal carcinoma, grade 2-3, ER+/PR+/HER2-. CT and bone scan revealed uptake in L1 vertebra, and a subsequent CT-guided biopsy confirmed metastatic disease. The patient initiated neoadjuvant dose-dense AC chemotherapy on 2/15/2024. Prior to chemotherapy, she underwent a screening mammogram, diagnostic mammogram, ultrasound-guided biopsy, and genetic testing.

User must verify software-generated patient summary. [Learn More](#)

MEI BILATERAL BREASTS WITH AND WITHOUT CONTRAST

CLINICAL HISTORY: 45-year-old female with newly diagnosed right breast cancer and right axillary lymph node metastasis pending for evaluation of extent of disease.

COMPARISON: Outside screening mammogram dated May 17, 2023, right diagnostic mammogram dated January 25, 2024. Images from ultrasound-guided biopsy of the right breast dated February 5, 2024.

TECHNIQUE: Scanning was performed using a dedicated breast coil on a 1.5Tesla magnet. High-resolution axial T1-weighted images without fat saturation and high-resolution axial T2-weighted images with fat saturation were obtained without contrast. A series of three-dimensional, high-resolution axial DWI images with fat saturation were obtained prior to and in a dynamic origin following administration of 0.1 mL/kg of gadolinium diethylenetriamine. Post-contrast sagittal T1-weighted images were also obtained. Post-processing included subtraction images and CAD-assisted pharmacokinetic mapping. T1 as of contrast was administered intravenously without immediate complication.

LEFT BREAST: The breast is composed of heterogeneous fibroglandular tissue. Following the administration of intravenous contrast, there is moderate background parenchymal enhancement. No suspicious mass or nonmass enhancement is identified in the left breast. No abnormal lymphatic, skin or pectoralis muscle enhancement.

There are no suspicious lymph nodes.

RIGHT BREAST: The breast is composed of heterogeneous fibroglandular tissue. Following the administration of IV contrast, there is moderate background parenchymal enhancement.

There is an ill-defined mass with associated biopsy clip and surrounding nonmass enhancement centered in the upper outer quadrant and extending into the upper inner and lower outer quadrants, compatible with biopsy-proven carcinoma. This measures 5.3 cm transverse, 3.4 cm anteroposterior, and 4.8 cm craniocaudal (series 1817, image 81 and series 6, image 182). The enhancement extends to within 5 mm of the nipple and abuts the pectoralis muscle. However, there is no definite nipple, skin or pectoralis muscle enhancement.

There are multiple abnormal right level 1 axillary lymph nodes. A biopsy proven metastatic lymph node with associated clip measures 1.1 x 1.0 x 1.4 cm (series 5, image 143 and series 6, image 175). There are tiny right internal mammary chain lymph nodes, none of which appear pathologically enlarged.

EXTRAMAMMARY TISSUES: No suspicious findings.

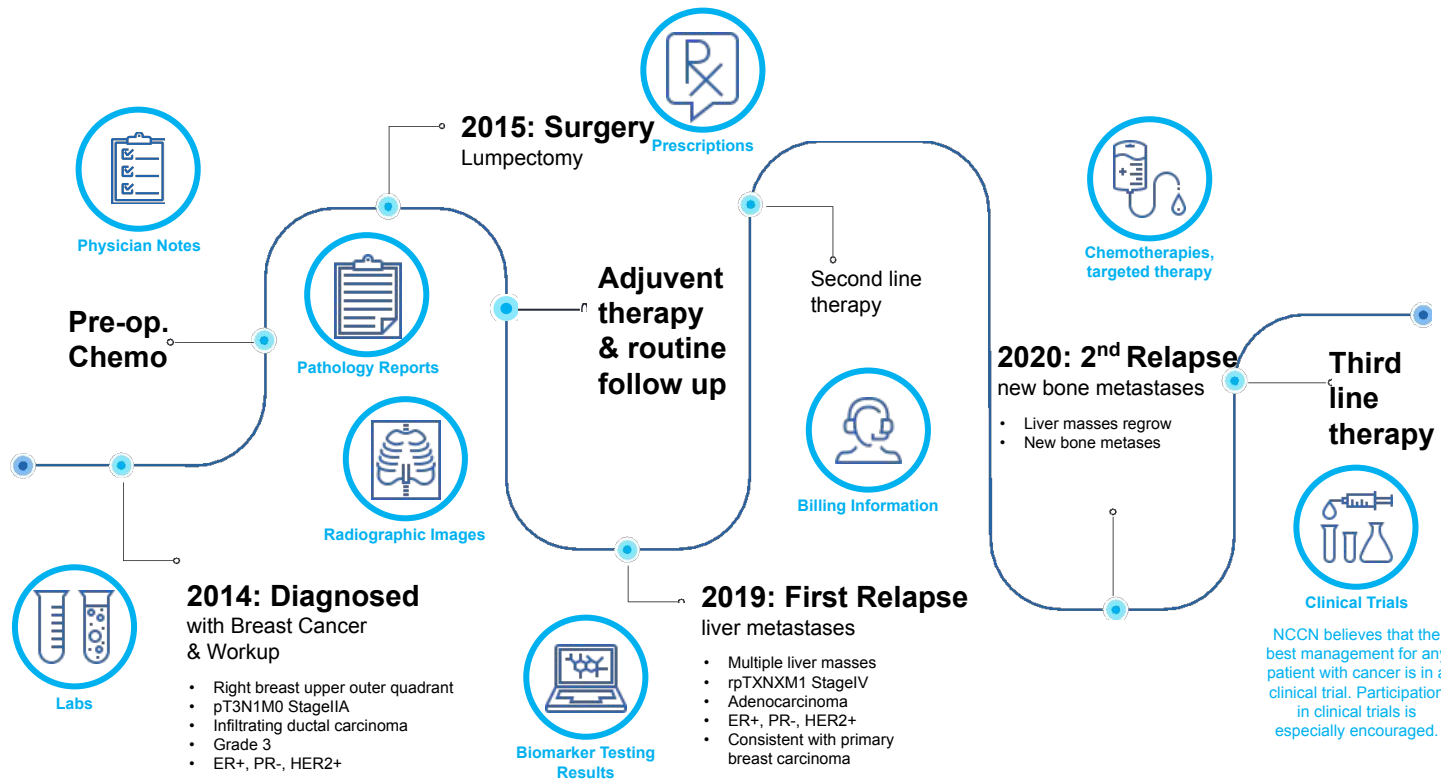
IMPRESSION:

- Biopsy-proven carcinoma centered in the upper outer right breast extends into the upper inner and lower outer quadrants, and measures 5.3 x 3.4 x 4.8 cm. No abnormal nipple, skin or pectoralis muscle enhancement.
- Mass within axilla compatible with biopsy-proven metastatic disease.

CANCEL SAVE



Cancer care, healthcare data and decision making are more and more complex!



Choices depend on:

- Patient Demographics & Preferences
- Cancer Diagnosis
- Labs, Vitals, Biomarker/Genomic Results
- Previous Treatments & Results
- Healthcare Coverage
- Clinical Guidelines
- Eligible Clinical Trials
- Regional/Local Standards and Access

NOTE: Not a real patient case, an imaginary example for illustrative purposes



Unstructured healthcare data challenges

Transform unstructured data into actionable insights

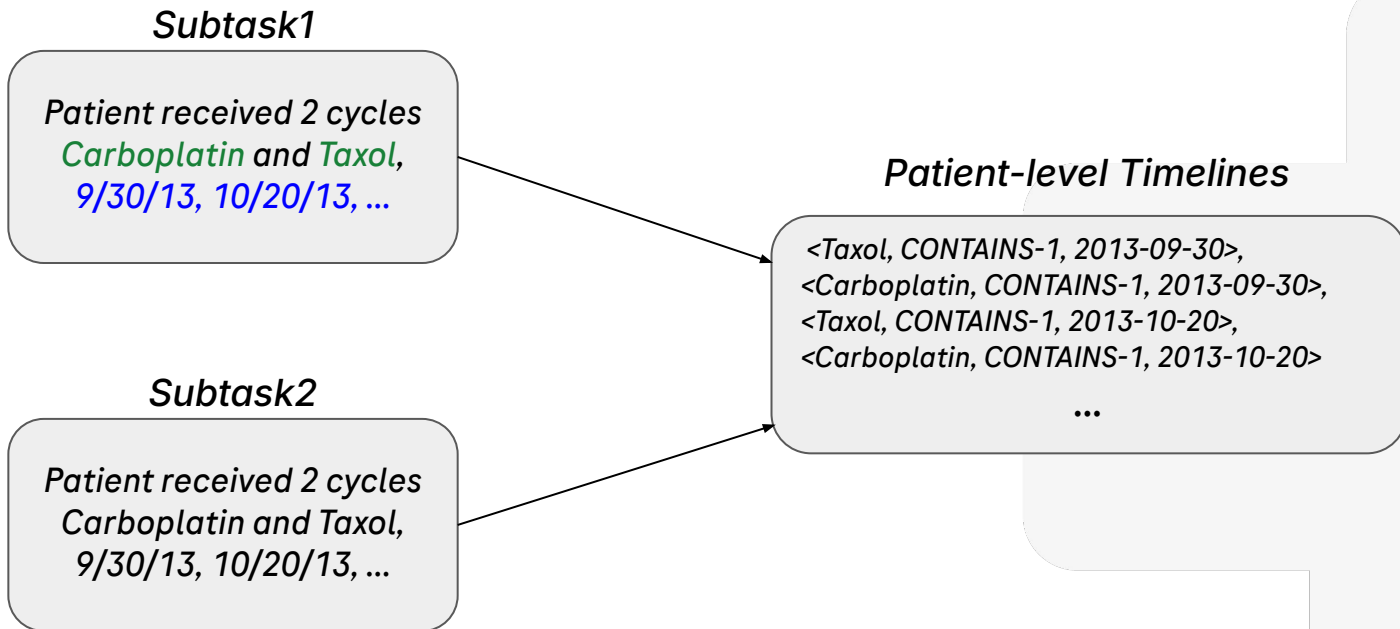
- Diverse report formats
- Multiple languages
- Semantic ambiguity
- Manual annotation can be labor-intensive and subjective

Extracting relevant information from unstructured sources requires Optical Character Recognition (OCR) and Natural Language Processing (NLP) techniques



Chemotherapy Treatment Timeline Extraction

Electronic Health Records (EHRs) of patients with breast, ovarian and skin cancers



Disclaimer: There is no real patient data being displayed here.

Yao, Jiarui, et al. (2024). "Overview of the 2024 shared task on chemotherapy treatment timeline extraction." *Proceedings of the 6th Clinical Natural Language Processing Workshop*. 2024.



Dataset Details

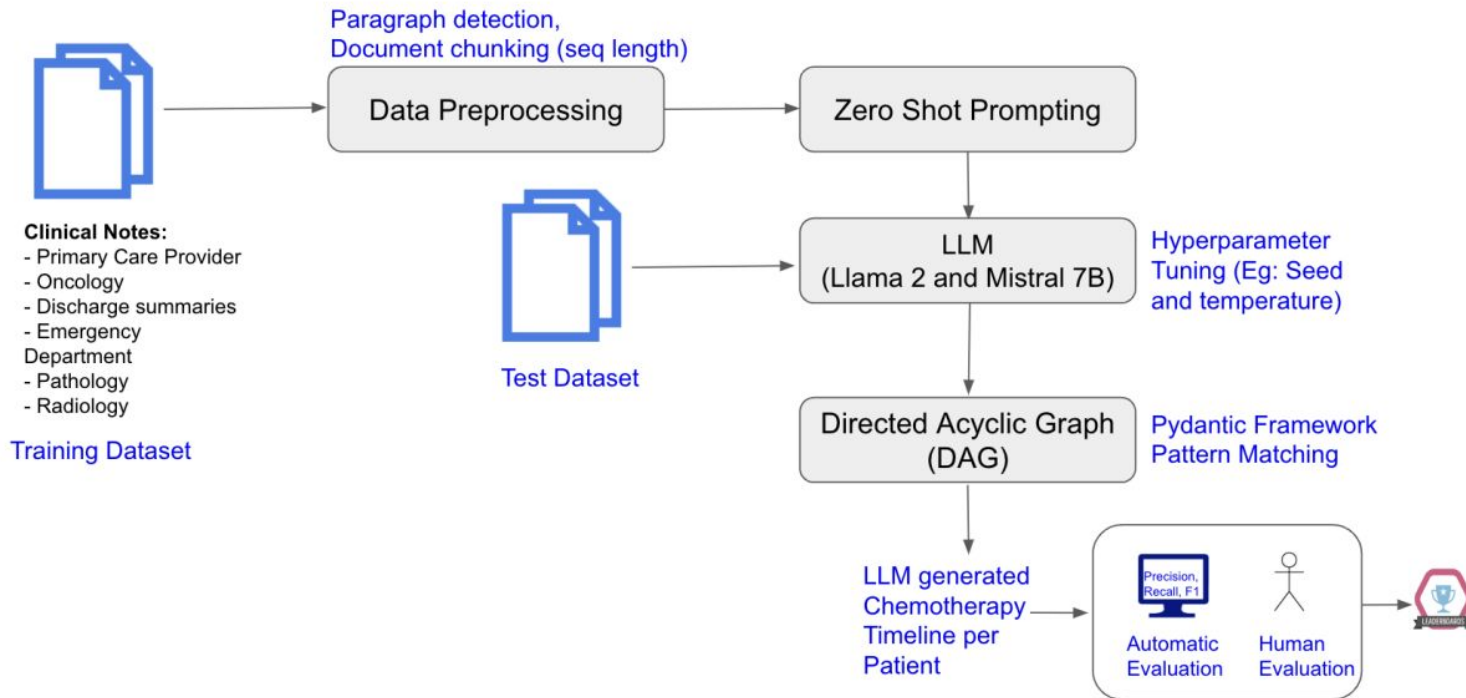
Gold labeled dataset: number of patients, notes, and words across train/dev/test sets. "Words" denotes the tokens delimited by white spaces

	Train			Dev			Test		
	Patients	Notes	Words	Patients	Notes	Words	Patients	Notes	Words
Ovary	26	1,675	1,183,632	8	562	308,814	8	559	257,116
Breast	33	1,002	465,644	16	499	225,588	35	1,333	786,896
Melanoma	10	233	124,924	3	211	178,308	10	229	156,083

Disclaimer: There is no real patient data being displayed here.



Large Language Models (LLMs) Pipeline





Healthcare-Specific Large Language Models (LLMs)

JSL-MedLlama-3-8B-v1.0

```
# Install necessary packages
!pip install -qU transformers accelerate

# Import libraries
from transformers import AutoTokenizer
import transformers
import torch

# Model and input prompt
model = "johnsnowlabs/JSL-MedLlama-3-8B-v1.0"
messages = [{"role": "user", "content": "What is a large language model?"}]

# Load tokenizer and prepare the prompt
tokenizer = AutoTokenizer.from_pretrained(model)
prompt = tokenizer.apply_chat_template(messages, tokenize=False, add_generation_prompt=True)

# Initialize text generation pipeline
pipeline = transformers.pipeline(
    "text-generation",
    model=model,
    torch_dtype=torch.float16,
    device_map="auto",
)

# Generate response
outputs = pipeline(prompt, max_new_tokens=256, do_sample=True, temperature=0.7, top_p=0.5)

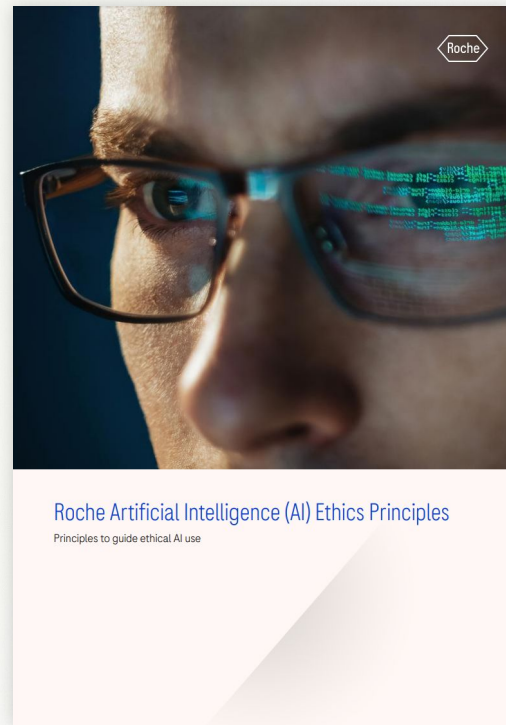
# Print the generated text
print(outputs[0]["generated_text"])
```



Roche foundation for using Artificial Intelligence (AI)

Using AI in a responsible and ethical way to benefit health care

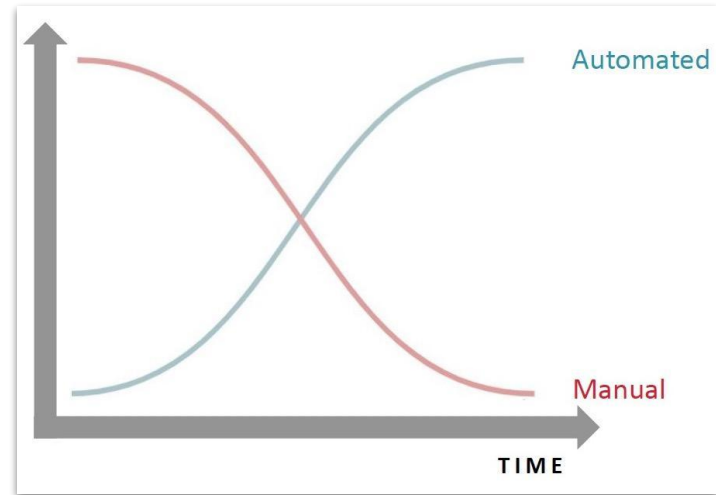
- | | |
|---------------------|-------------------------------------|
| 1 Ethical use | 6 Fairness and minimization of bias |
| 2 Transparency | 7 Accountability |
| 3 Explainability | 8 Privacy |
| 4 Human Control | 9 Security |
| 5 Empowering people | 10 Safety by design |





LLMs Platform will be a journey

- LLMs are more expensive at scale
- Inputs may not align with LLMs intended purpose
- LLMs training with personal data risks privacy and legal issues in outputs
- LLMs may “hallucinate” generating incorrect information
- LLMs can summarize vast amounts of medical literature
- Most medical generative AI solutions, the LLM is only a user interface
- Chatbots and virtual assistants powered by LLMs can improve patient engagement and education



Thank You!

We are hiring!

www.navify.com/careers/



FEATURE STORE SUMMIT 2024

DATA FOR AI:

REAL-TIME, BATCH, AND LLMS

Q&A



FEATURE STORE SUMMIT 2024

DATA FOR AI:
REAL-TIME, BATCH, AND LLMS