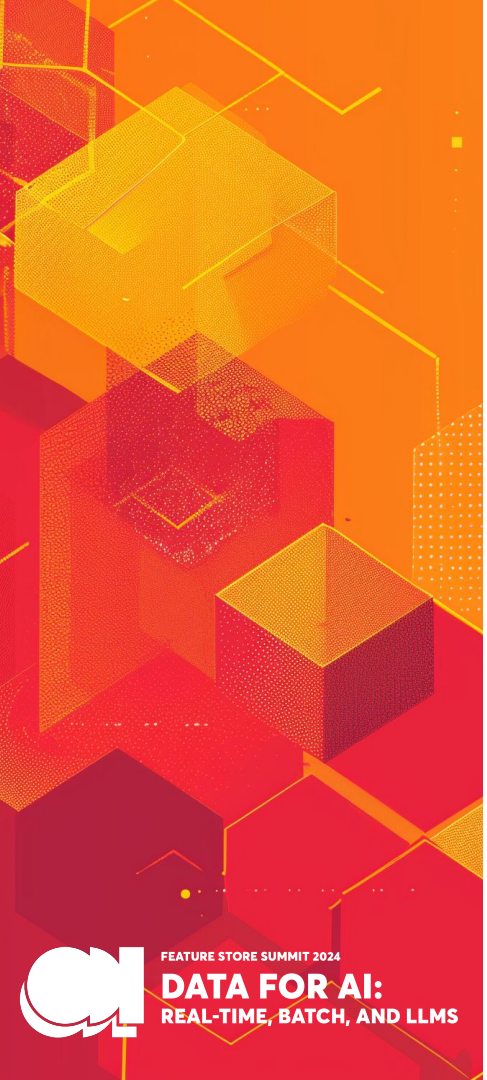


How Roche's Data Platform accelerates Feature Engineering through Generative AI

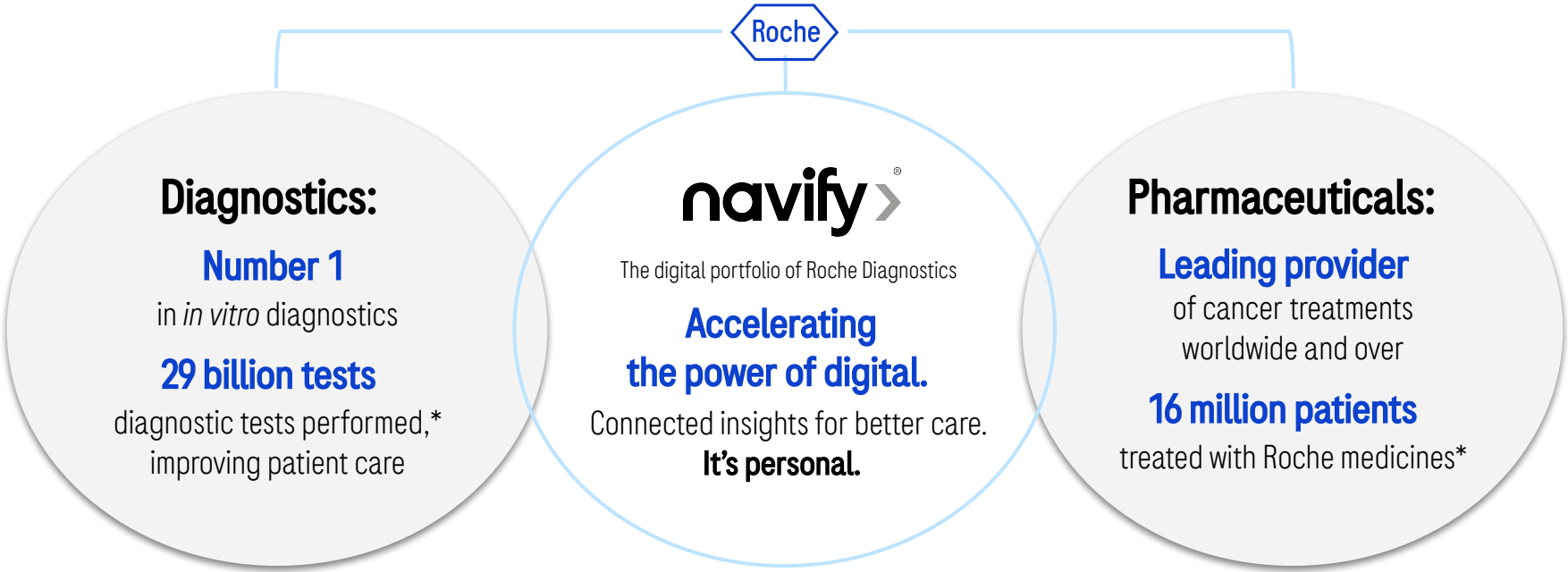
Vishakha Sharma, PhD
Senior Principal Data Scientist

Roche Diagnostics

Organized by HOPSWORKS

Bringing healthcare understanding to technology

Leveraging expertise from screening through diagnosis to personalized care



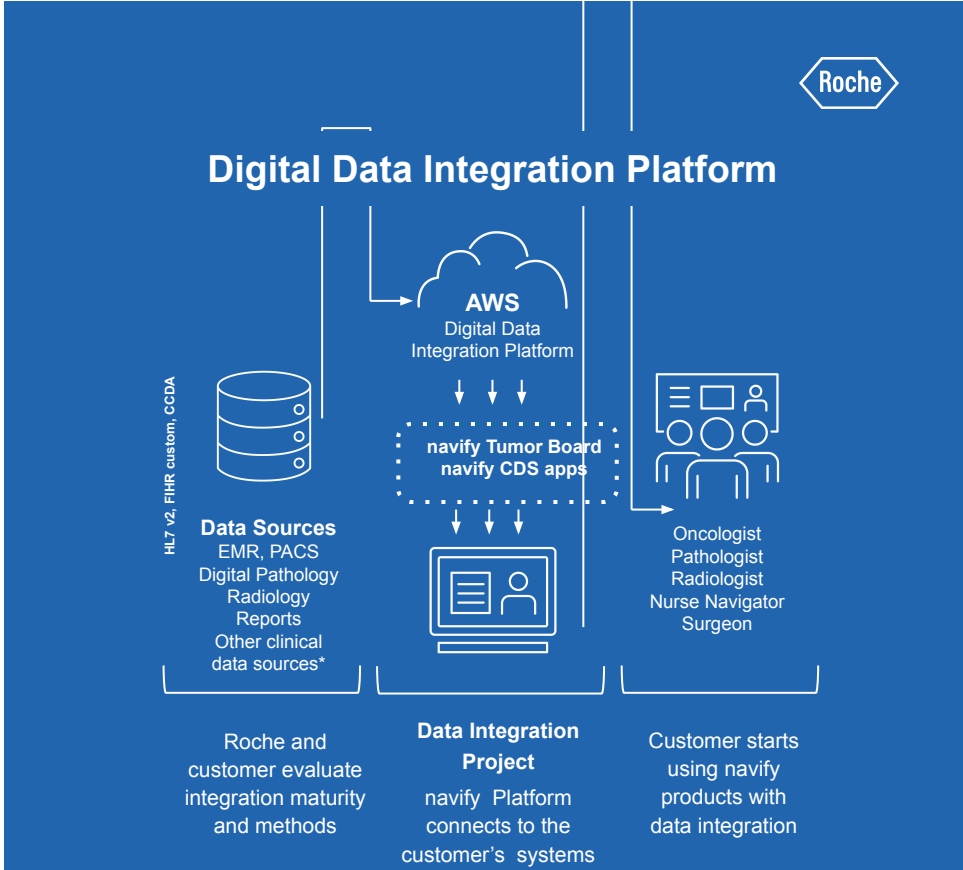
*Roche Annual Report, 2023

Integrating your data with navify products

Implementation builds the interface between navify portfolio and your IT infrastructure

Data integration and implementation

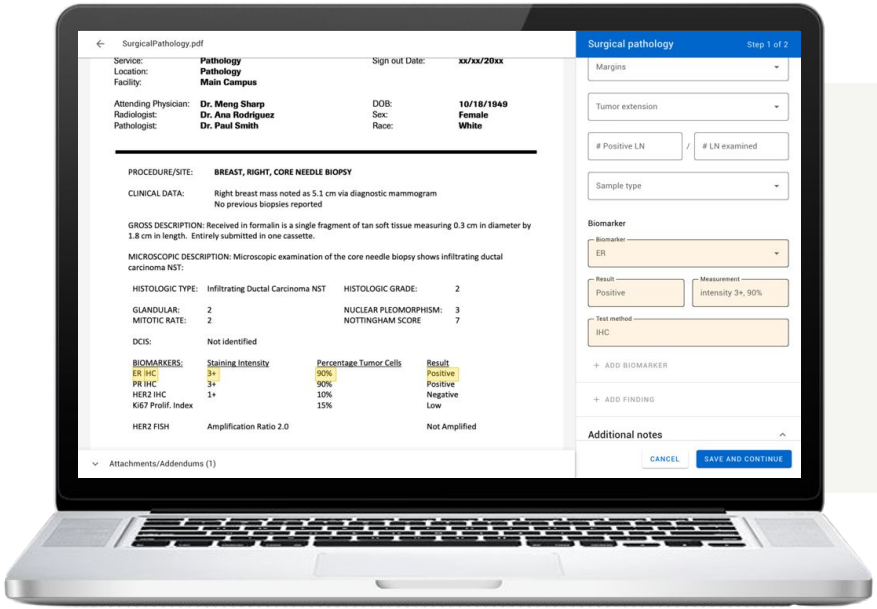
Phases	Added capabilities
Integration Connects hospital sources to transfer data in various formats using encryption protocols	Evaluates systems authorized to send data and verify data transmitted
Implementation Maps current and future processes to create and execute a transition plan between process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



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® Clinical Hub



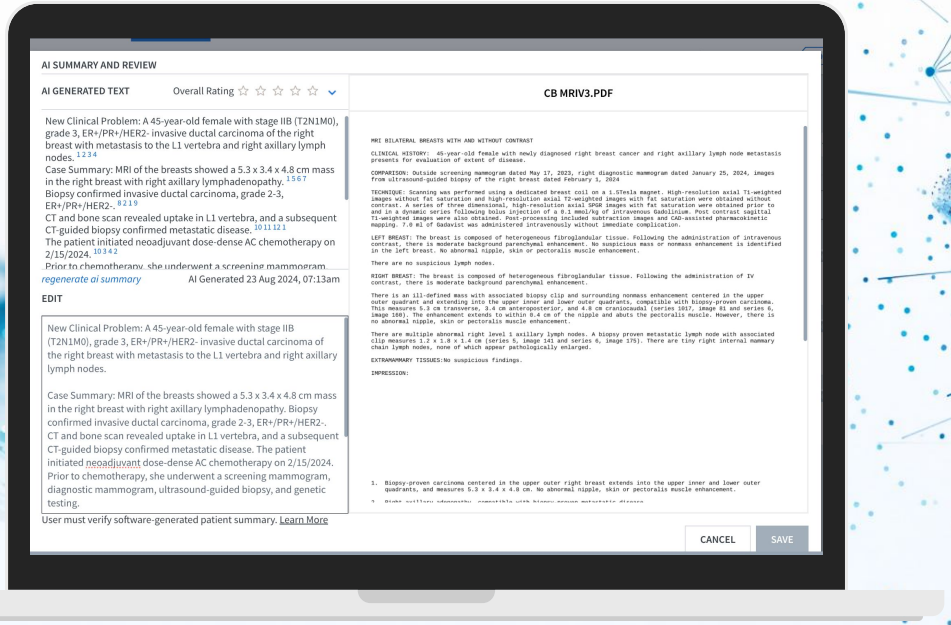
Assistive technology leverages Machine Learning and Natural Language Processing to auto-detect and extract valuable medical concepts from unstructured reports for downstream use, saving time while minimizing manual data entry

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

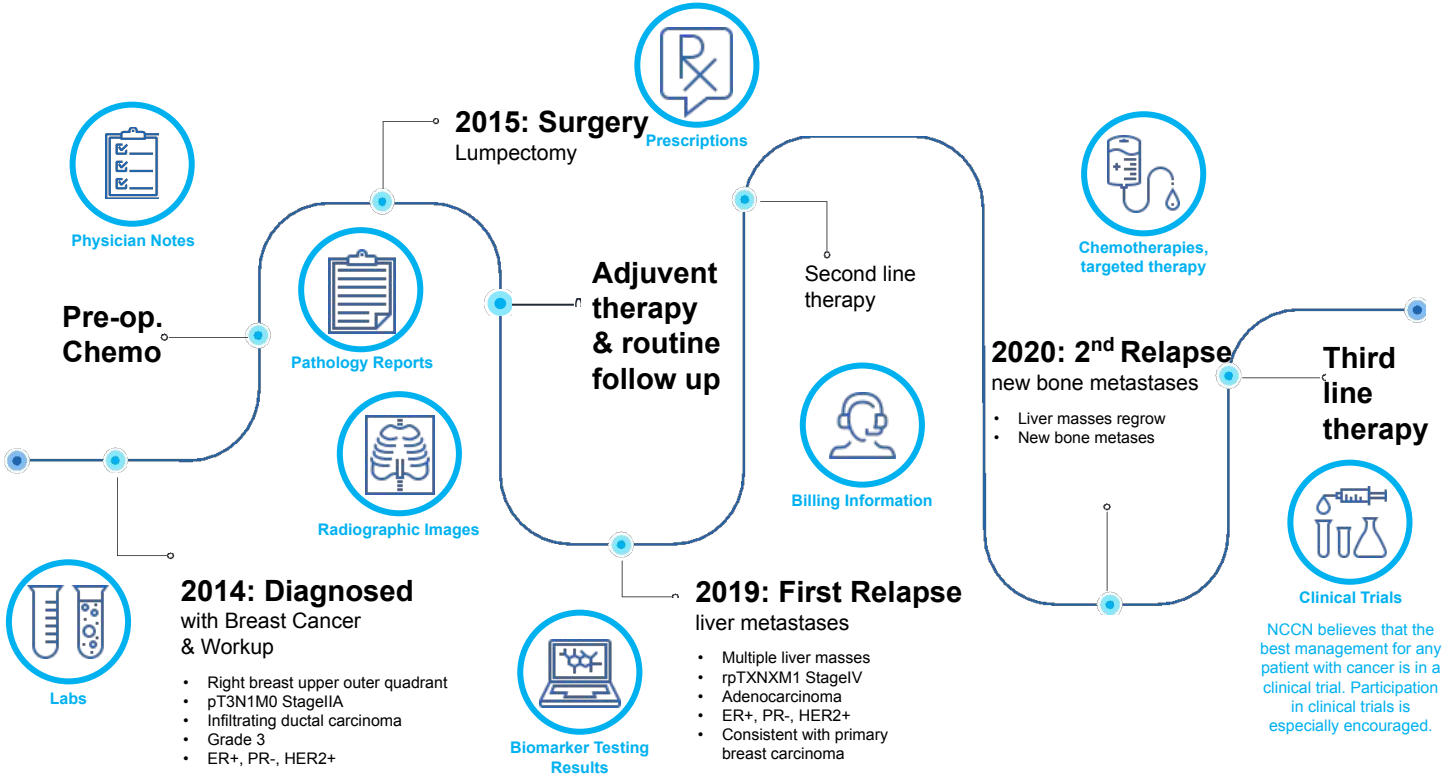
Automating Case Summarization with GenAI

Streamline and standardize patient case summarization for tumor board meetings

navify Clinical Hub



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

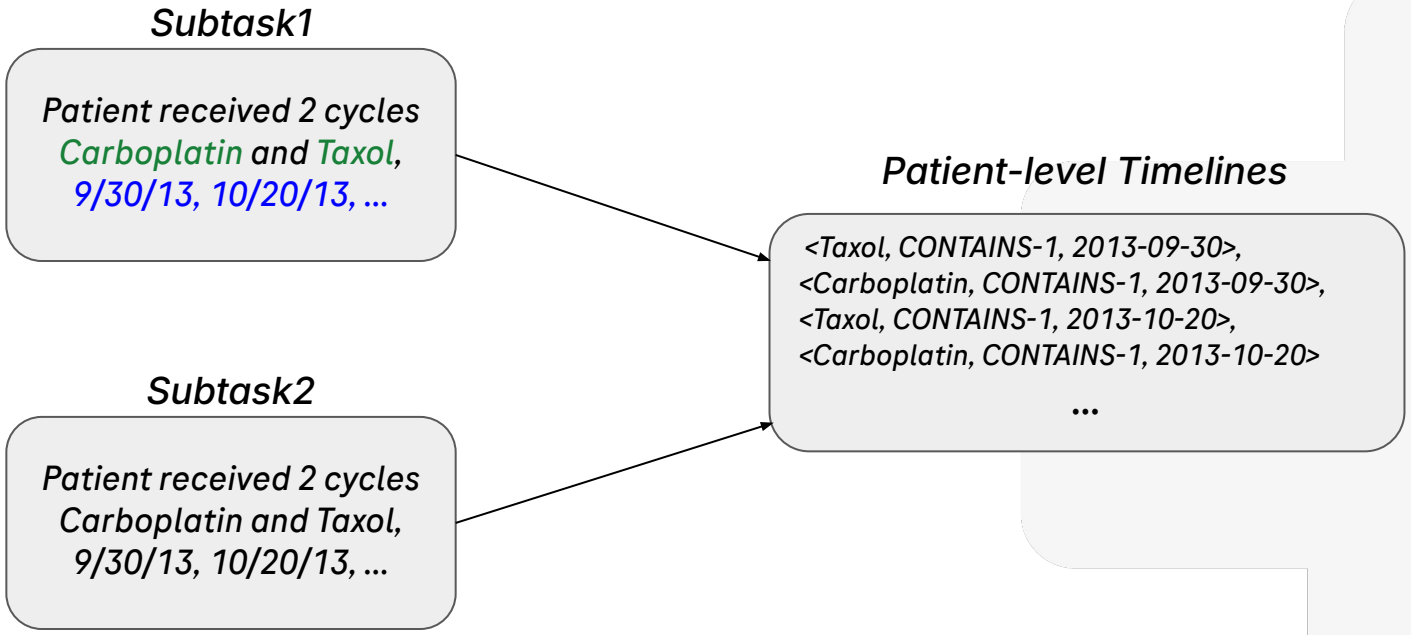
Transform unstructured data into actionable insights

- 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.

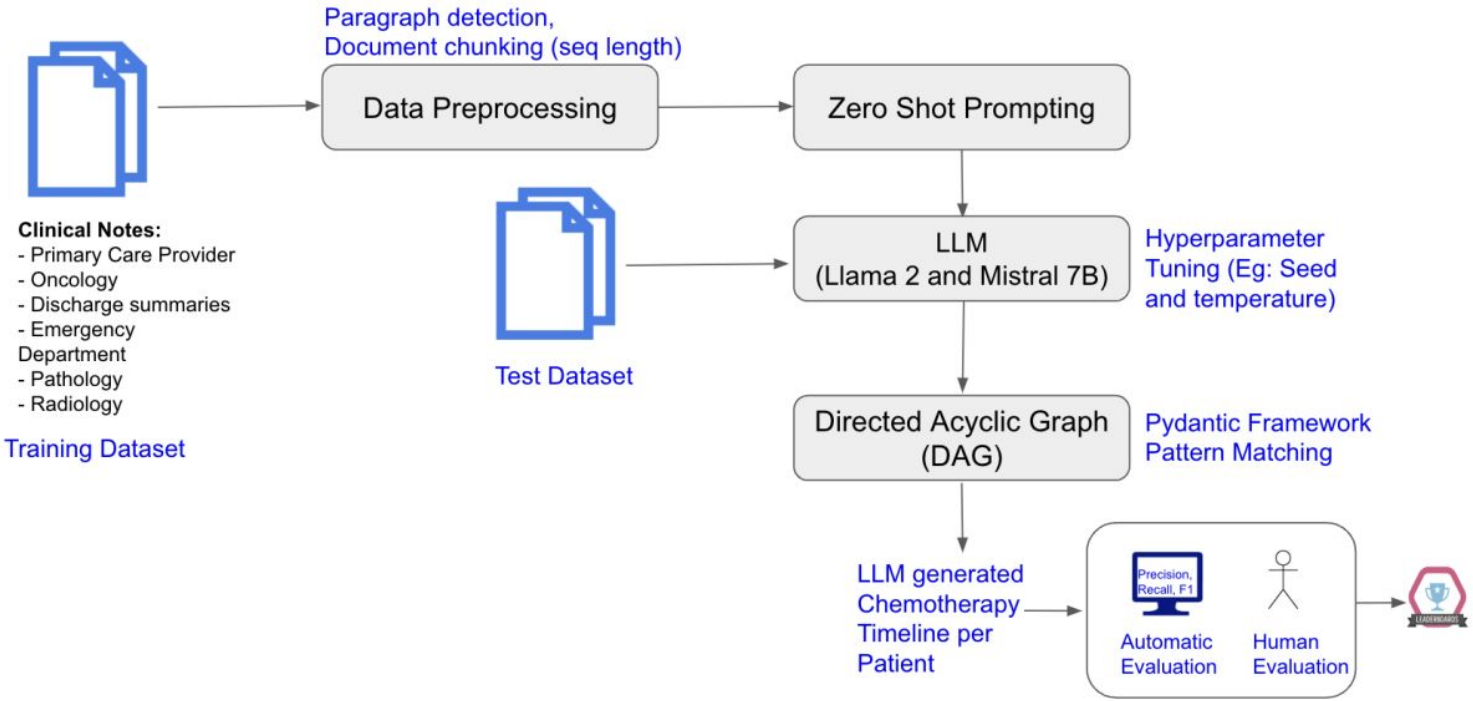
Dataset Details

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



Sharma, Vishakha, et al. "Lexicans at Chemotimelines 2024: Chemotimeline Chronicles-Leveraging Large Language Models (LLMs) for Temporal Relations Extraction in Oncological Electronic Health Records." *Proceedings of the 6th Clinical Natural Language Processing Workshop*. 2024.

Healthcare-Specific Large Language Models (LLMs)

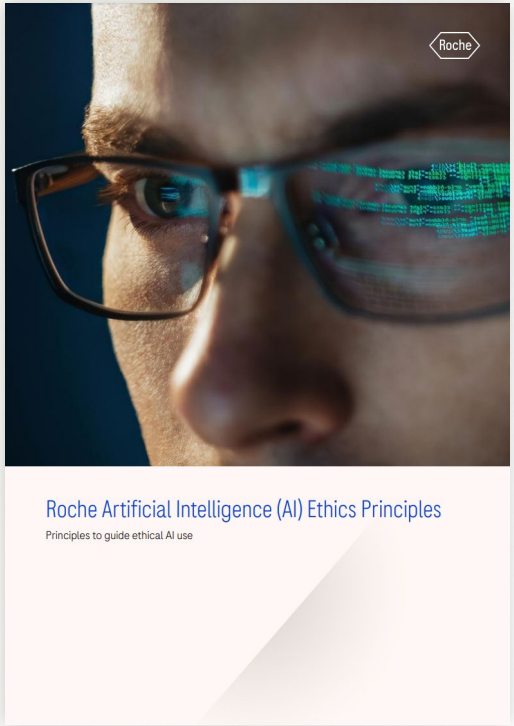
JSL-MedLlama-3-8B-v1.0

```
# 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_pro

# 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

# 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

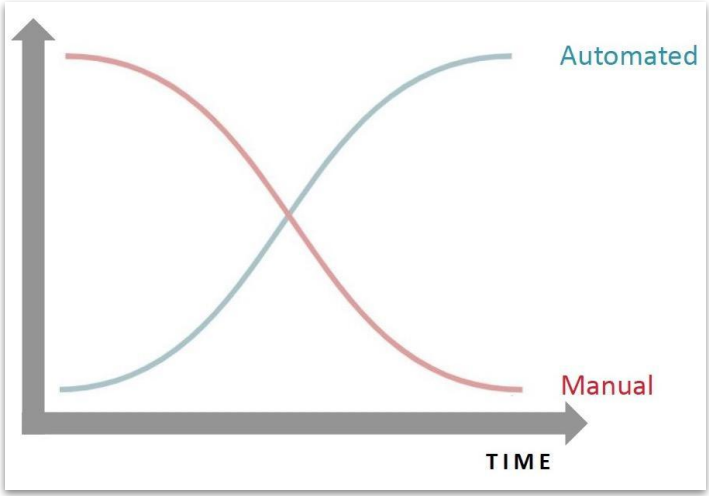
1 Ethical use
2 Transparency
3 Explainability
4 Human Control
5 Empowering people
6 Fairness and minimization of bias
7 Accountability
8 Privacy
9 Security
10 Safety by design



F. Hoffmann-La Roche Ltd. Roche Artificial Intelligence (AI) Ethics Principles [PDF guide; Dec, 2023]. https://assets.roche.com/f/176343/x/401c28049f/roche-ai-ethics-principles.pdf, accessed April 6,2024

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/

Q&A