



# **Gen AI-Powered Clinical Trials**

October 2023



# **Generative AI for Clinical Trials**

# Current Landscape of Patient Recruitment in Clinical Trials

Clinical trials play a critical role in drug development but often suffer because of **insufficient, inaccurate and expensive** patient recruitment process.



## Drug Development

\ Average cost of drug development is **\$2.5B - \$3B**

SciTranslational Medicine



## Conducting Trials

\ Median cost of conducting clinical trials is **\$19M - \$20M**

JAMA



## Delayed Recruitment

\ Loss due to delayed recruitment are **\$1M - \$5M**

PharmaVoice

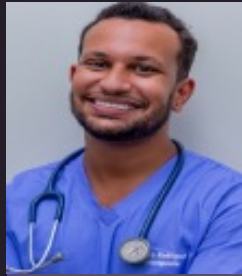


## Patient Recruitment

\ Recruitment costs adds to **40%** of clinical trial expenditure

OracleHealthSciences

# Discover / User Journey – Clinical Trial Coordinator

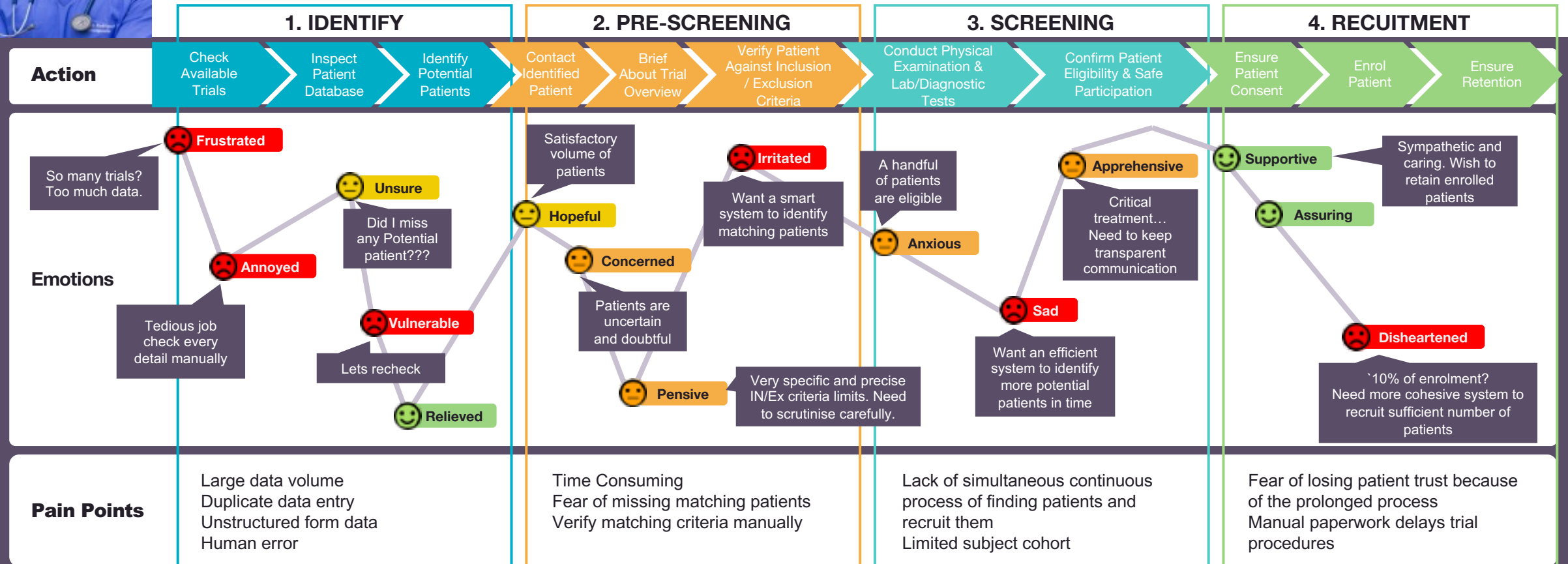


## James Morrison

Age: 39 Years	Location: New York, USA
Position: Coordinator	Experience: Over 10 Years




## Scenario

James wants to recruit adequate number of patients for the trial



# Smart Trials — Generative AI Powered Cohort Identification





## Current Landscape

	Average cost of drug development is <b>\$2.5B - \$3B</b>
	Median cost of conducting clinical trials is <b>\$19M - \$20M</b>
	Loss due to delayed recruitment range from <b>\$1M - \$5M</b>
	Recruitment costs adds to <b>40%</b> of clinical trial expenditure

## Benefits

- \ Reduce patient recruitment cost by 60%.
- \ Increase the clinical trial success rate by 30% using better cohort selection.
- \ Reduce annotation time by 50 - 60% for custom model creation to convert unstructured to structured data.

## LLM based workbench tools for Clinical Trials

	<b>Criteria Extraction</b>
	<b>Annotation Assistant</b>
	<b>Chat Interface for Data &amp; Doc</b>
	<b>Patient Centric Clinical Trials</b>

**80%**

Faster Patient  
Matching

**60%**

Cost Savings

# Large Language Model based tools to accelerate clinical trials

## Annotator Assist

- \ Tool to **warm-start annotations** of protocol documents for extracting eligibility criteria
- \ Uses LLM's to **auto annotate** eligibility criteria from protocol documents
- \ Helps to **reduce 50% - 60%** of overall **time** required for annotations
- \ The LLM based annotated data is verified by SME and **can be used to train local models** for better criteria extraction

## Chat Protocol

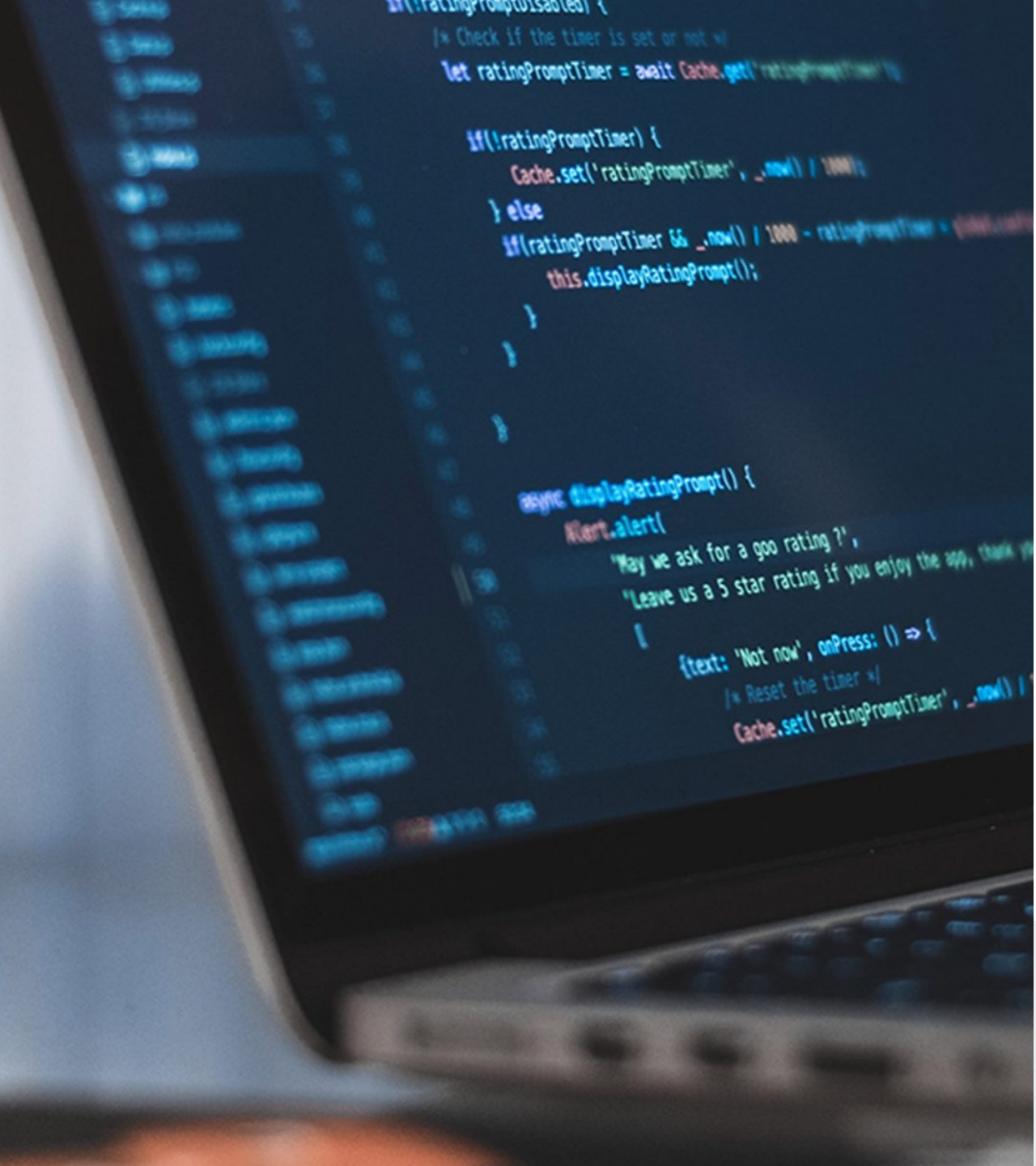
- \ An efficient BOT to “chat” with protocol document for clinical trails
- \ LLM techniques such as **chaining and embeddings** are used to make chats senseful
- \ Takes **queries in natural language**, does semantic search to provide most accurate answer while staying within boundaries of the document
- \ **Real time answers** saving time to browse hundreds of pages

## Protocol Assist

- \ Generates a **baseline editable template for clinical trials**
- \ Uses NLP techniques like text clustering, summarization, natural language generation
- \ Gives **head-start in creating the protocol document** for clinical trials by providing baseline content
- \ **Reduces the time** taken to write a clinical trial protocol document

# Accelerate your clinical trials by partnering with Persistent

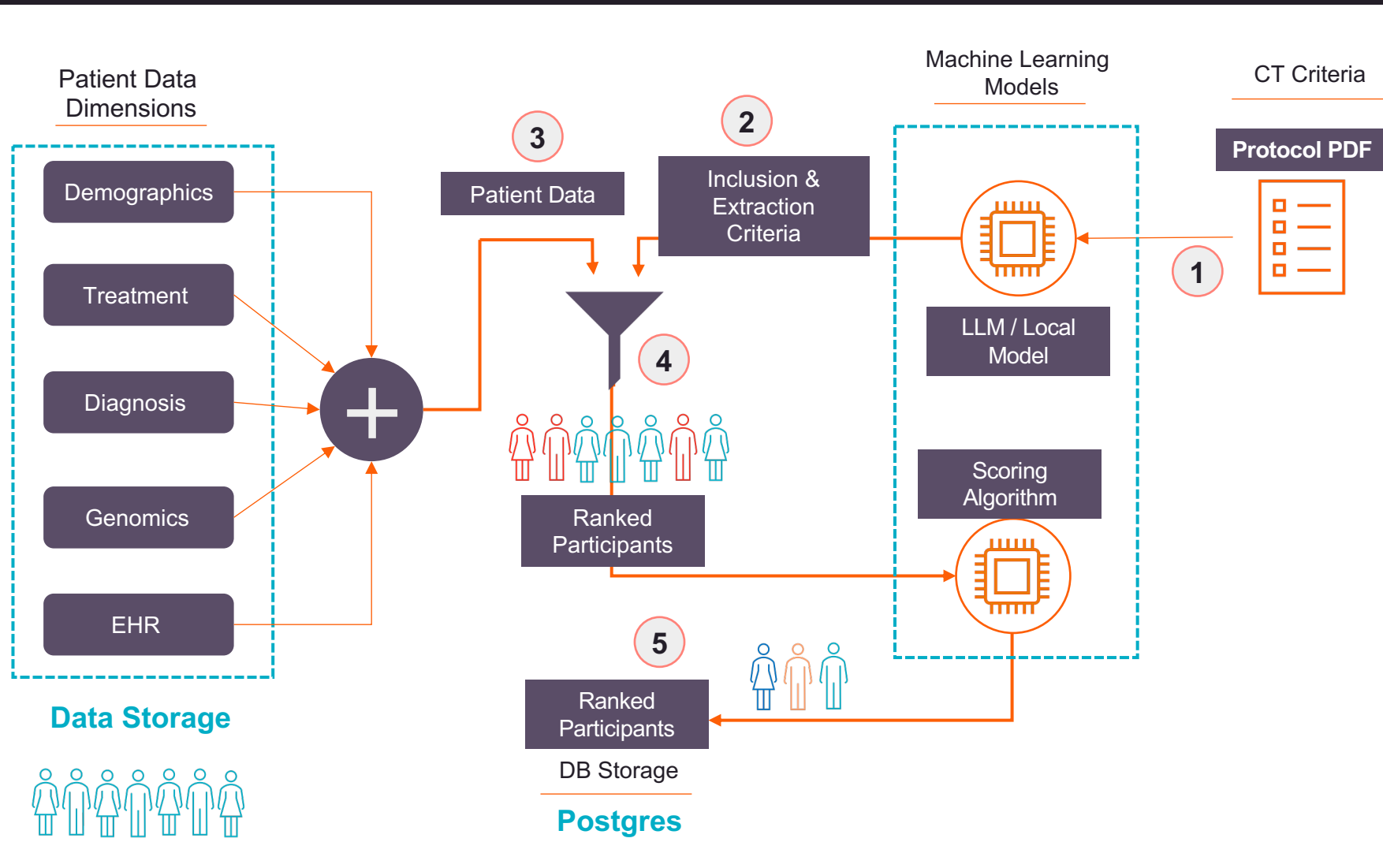




## Technical details



# Oncology CT High Level Architecture Diagram

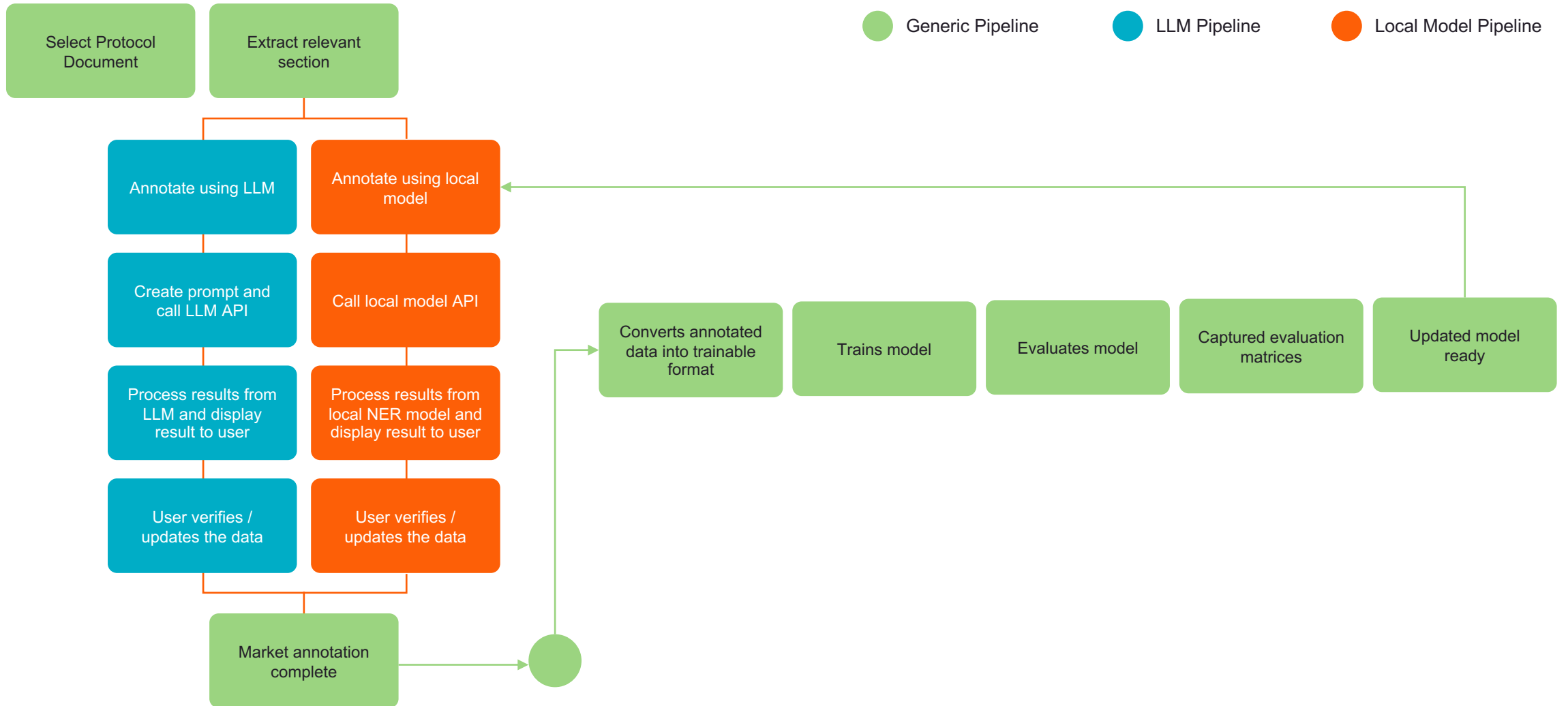


1. Inclusion and exclusion criteria are extracted from the protocol document
2. The criteria are passed to NER model (LLM / local model) to extract key entities
3. Patient data set is filtered based on the extracted key entities
4. The filter returns those patients which are eligible for the clinical trials
5. The scoring model will rank the patients based on who will benefit the most from clinical trial

# Annotator Assist: Fast-pace your annotations [1]

- \ Need for annotation: protocol documents annotations can be time consuming
- \ Varied tools for annotation
- \ Idea: use of LLM's to fast-track the annotations !
- \ Were conscious of cost factor and IP while calling API's
- \ Tool to warm-start annotations of protocol documents for extracting eligibility criteria
- \ Uses LLM's to auto annotate eligibility criteria from protocol documents
- \ Helps to reduce 50% - 60% of overall time required for annotations
- \ The LLM based annotated data is verified by SME and can be used to train local models for better criteria extraction

# Annotator Assist: Fast-pace your annotations [2]



# Fast Tracking Annotations using Large Language Models (LLMs)

## Example Pathology Text Reports

### Challenges

- Need for manual annotation by specialists
- Pathology report annotations are tedious & time consuming
- Data privacy and Security concerns

### Solutions

- Building reliable models for faster & accurate annotations of huge datasets
- Developing proprietary solution that can toggle between LLM API's and locally trained NLP Model

### Sample Pathology Report

PATIENT HISTORY: [REDACTED]

PRE-OP DIAGNOSIS: Lung cancer (right lung).  
POST-OP DIAGNOSIS: Same  
PROCEDURE: Bronchoscopy, mediastinoscopy and thoracotomy.

ADDENDA:  
Addendum  
Molecular Anatomic Pathology Microdissection Genotyping Testing.

A. KRAS Exon 1 mutation NOT identified.  
B. EGFR Exon 19 mutation NOT identified.  
C. EGFR Exon 21 mutation NOT identified.

Note:  
KRAS codon 12/13 mutations are found in approximately 10-30% lung adenocarcinomas and associated with history of smoking (1). EGFR exon 19 and 21 mutations are present in about 10% of non-Asians and up to 40% in the Asian population, common in non-smokers and associated with the tumor response to treatment with EGFR inhibitors (2). These mutations are mutually exclusive and their presence may have prognostic and/or therapeutic implications (3-4).

Addendum  
PROBE: Tris LSI EGFR SpectrumOrange CFP 7 SpectrumGreen™ Probes [REDACTED]

Fluorescence in-situ hybridization studies performed on the adenocarcinoma shows a ratio of EGFR gene to the centromere of chromosome 7 of 1.07 indicating no EGFR amplification in the targeted region. The rate of hyperploidy for the centromere of chromosome 7 was 30% of the analyzed cells. The signal to nucleus ratio (SNR) for the EGFR gene was 2.5. 60 cells were analyzed in the targeted region.

Interpretation guidelines for EGFR Gene Copy Number by FISH

### User review for extracted entities

Patient Ehr [REDACTED]

PATIENT HISTORY:  
PRE-OP DIAGNOSIS: Lung c.ancer (right lung).  
Same.POST-OP DIAGNOSIS:  
PROCEDURE: Bronchoscopy, mediastinoscopy and thoracotomy.

ADDENDA:  
A. KRAS Exon 1 mutation NOT identified.  
B. EGFR Exon 19 mutation NOT identified.  
C. EGFR Exon 21 mutation NOT identified.

----- codon 12/13 mutations are found in approximately 10-30% lung ad.enocarcinomas a-d. associat-d with his-ory of smokin,  
(1) EGFR exon 19 and 21 mutations are present in about 10% of non-ASians and up to 40 y, in the ASian population, common in non-smokers and associated with the tumor response to treatment with EGFR inhibitors (2). These mutations are mutually

EXTRACT ENTITIES

### Final Extracted entities

Label	Entities	
Assay type	FISH	✕
Cancer subtype	Invasive moderately differentiated adenocarcinoma Typical carcinoid	✕
Cancer type	Non small cell lung cancer	✕
Genomic Biomarkers	KRAS Exon 1 mutation EGFR Exon 19 mutation EGFR Exon 21 mutation EGFR amplification	✕
Pathology stage	T2N0Mx	✕
Tumor size	2.5cm 0.6cm	✕

SAVE TO DB

A woman with dark hair tied back, wearing a white lab coat, is looking at a computer monitor. The monitor displays a software interface with various data points and graphs. In the background, another person is partially visible, and the setting appears to be a laboratory or control room with blue lighting and equipment.

## Demo Screenshots

## SmartTrials



**Matching patients to clinical trials  
with AI precision**



**Email Id**

**Password**










# Demo - Chat with your data

The screenshot shows a chat application window titled "Analyze Patient Data". In the top right corner, there is a user profile for "JAMES" and a share icon. A dark blue message bubble at the top contains the text "Your database consists of 1846 lung cancer patients". At the bottom, there is a text input field with the placeholder "Type Something ..." and a blue send button with a white arrow. A context menu is open over the input field, listing five options: "Table", "Text", "Pie Chart", "Bar Chart", and "Line Chart". The "Text" option is currently selected and highlighted.

# Demo – Convert Unstructured data – Structured data using GenAI

## List of available documents (45):



- |  |             |
|--|-------------|
|  A Study of Atezolizumab in Combination W...<br>NCT02657434   | ANNOTATE >> |
|  Alectinib Versus Pemetrexed or Docetaxel...<br>NCT02604342   | ANNOTATE >> |
|  A Study of Atezolizumab Compared With Ge...<br>NCT02409355   | ANNOTATE >> |
|  A Study Comparing Alectinib With Crizoti...<br>NCT02075840   | ANNOTATE >> |
|  A Study of Erlotinib in Participants Wit...<br>NCT01667562   | ANNOTATE >> |
|  A Study of Atezolizumab (MPDL3280A) Comp...<br>NCT02409342   | ANNOTATE >> |
|  A Study of Carboplatin Plus Etoposide Wi...<br>NCT02763579  | ANNOTATE >> |
|  A Study of Atezolizumab in Combination W...<br>NCT02367794 | ANNOTATE >> |
|  A Study of Erlotinib in Participants Wit...<br>NCT01260181 | ANNOTATE >> |

## Upload Document:



OR

ANNOTATE USING LLM

OR

ANNOTATE USING OUR MODEL



LLM annotation: **80 mins**  
Manual annotation: **280 mins**























Model: Davinci

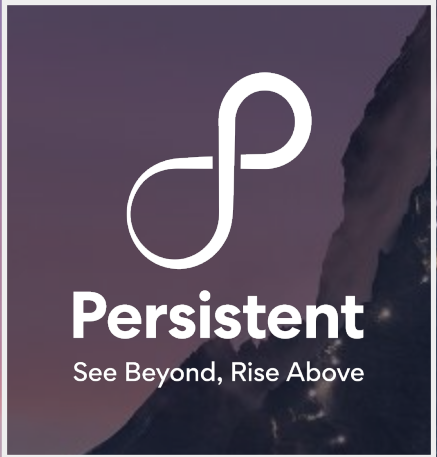


See model accuracy

## Annotated (40):

-  NCT02352948 
-  NCT03043872 
-  NCT02542293 
-  NCT03164616 
-  NCT02511106 
-  NCT02125461 
-  NCT02453282 
-  NCT03693300 
-  NCT03775486 
-  NCT03853551 





Thank You!