NCI Virtual Workshop on Medical Imaging De-Identification (MIDI)

Pixel De-Identification Using AI

May 23, 2023 12:00pm ET

George Shih, MD MS FACR

Professor and Vice-Chair for Informatics

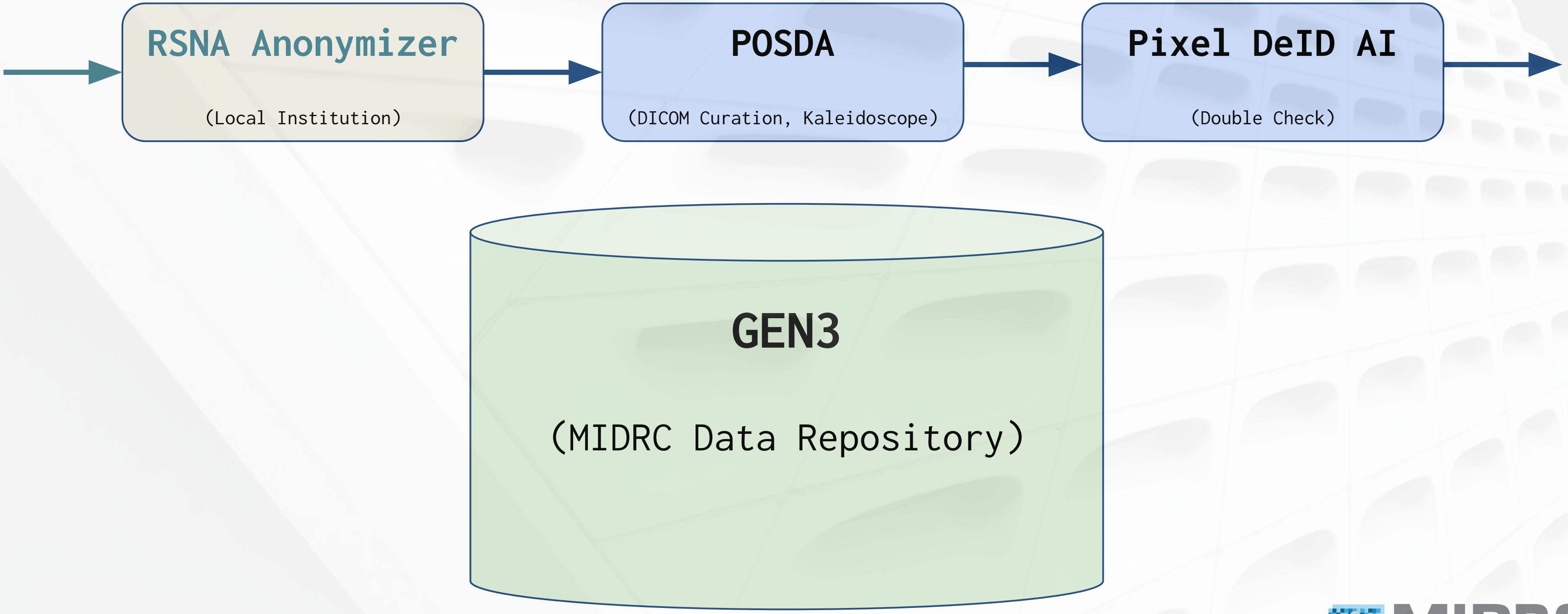
Weill Cornell Radiology

MIDRC

MEDICAL IMAGING AND DATA RESOURCE CENTER.

RSNA MIDRC

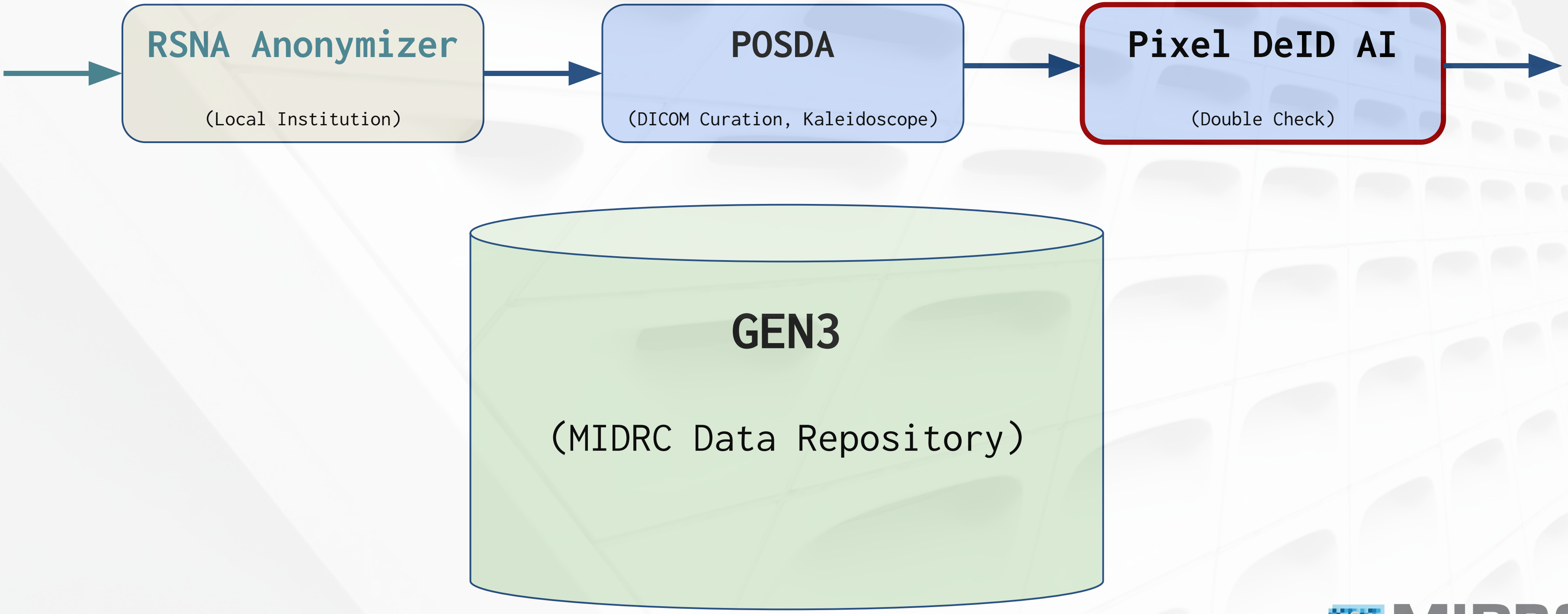
Data Workflow



MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

RSNA MIDRC

Data Workflow



MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

DiscoveryCT750HD^WithDoseSRAndScreenShot[83749123...

Patient Name: [redacted]

Exam no: 1744

Accession Number: [redacted]

[redacted]

Patient ID: [redacted]

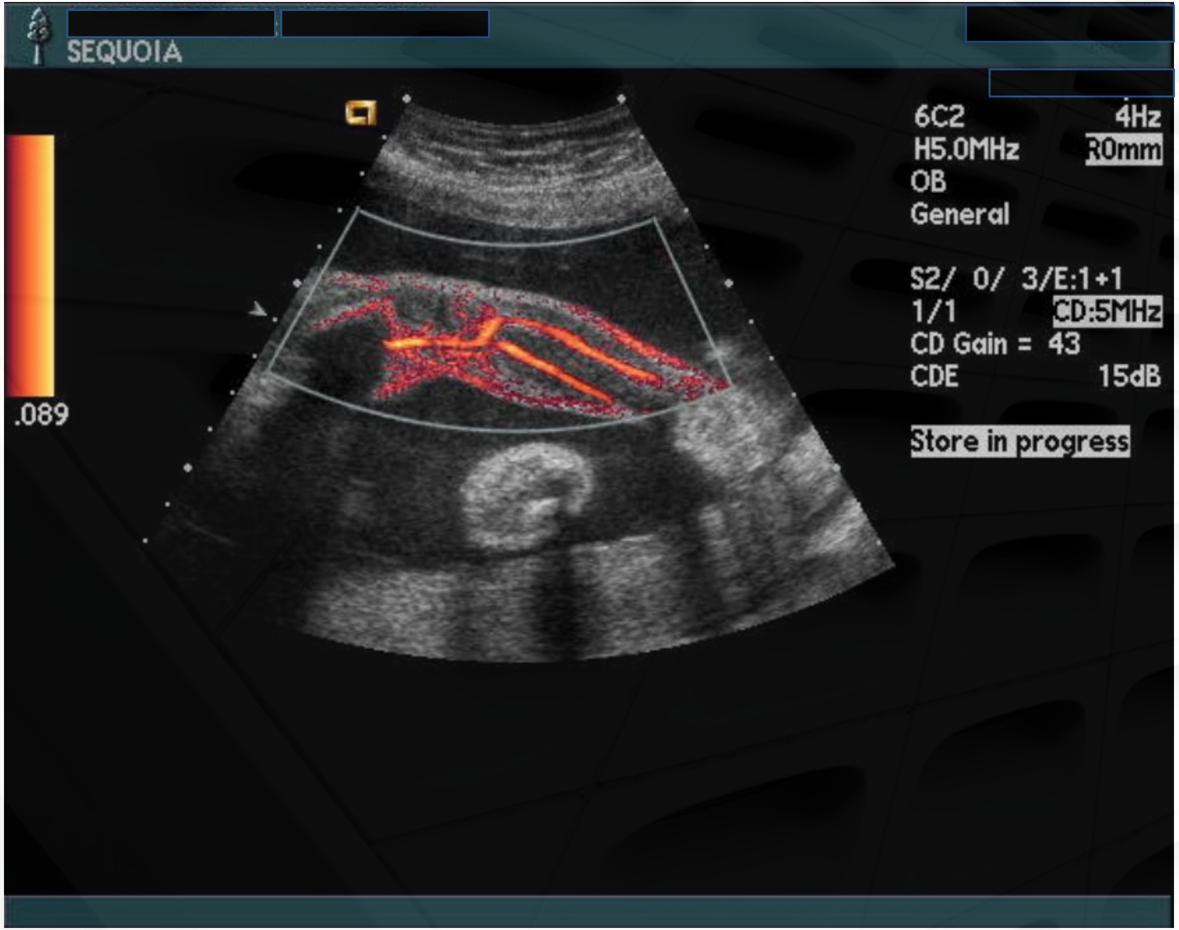
Discovery CT750 HD

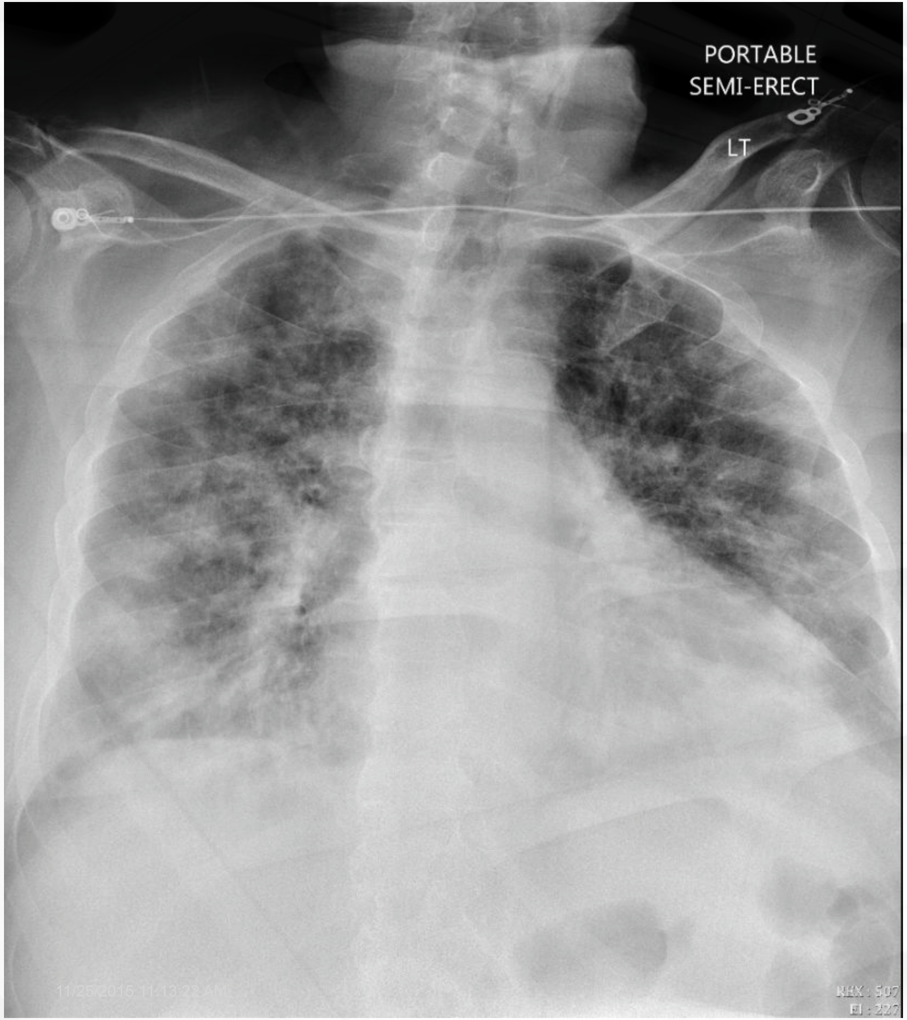
Exam Description: CT HALS/THORAX/ABDOMEN

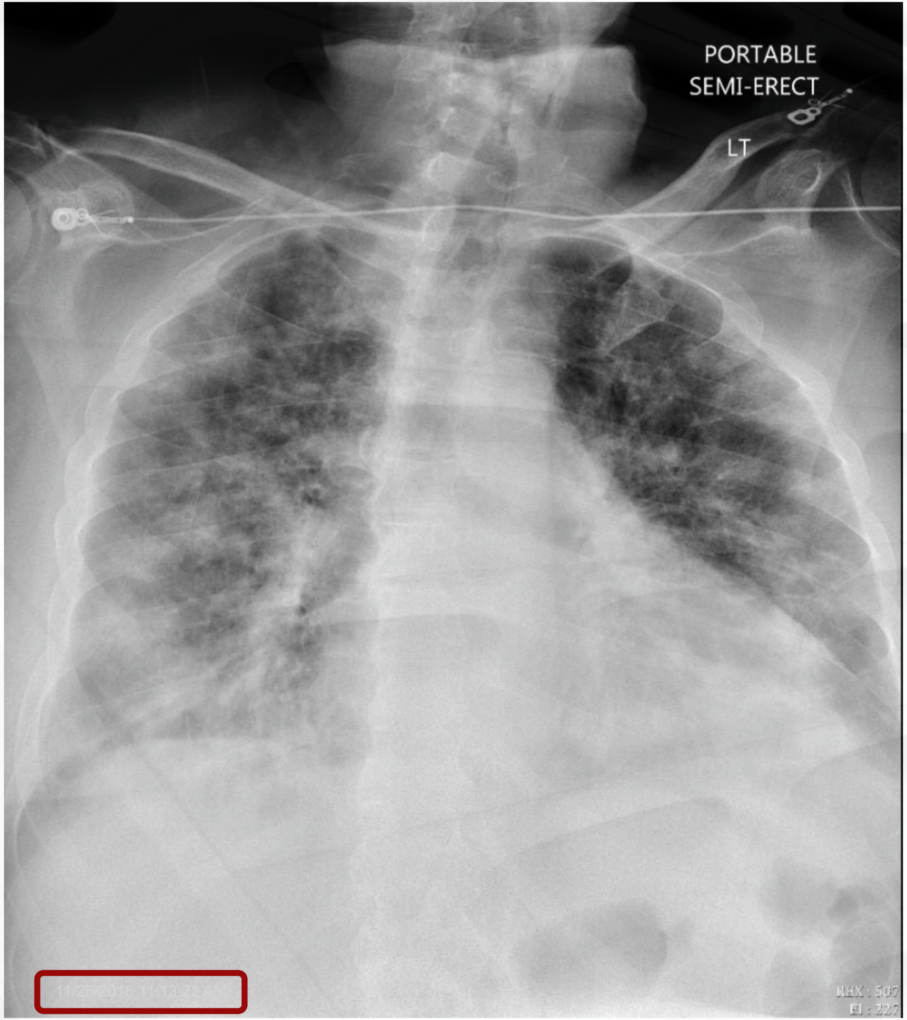
Dose Report

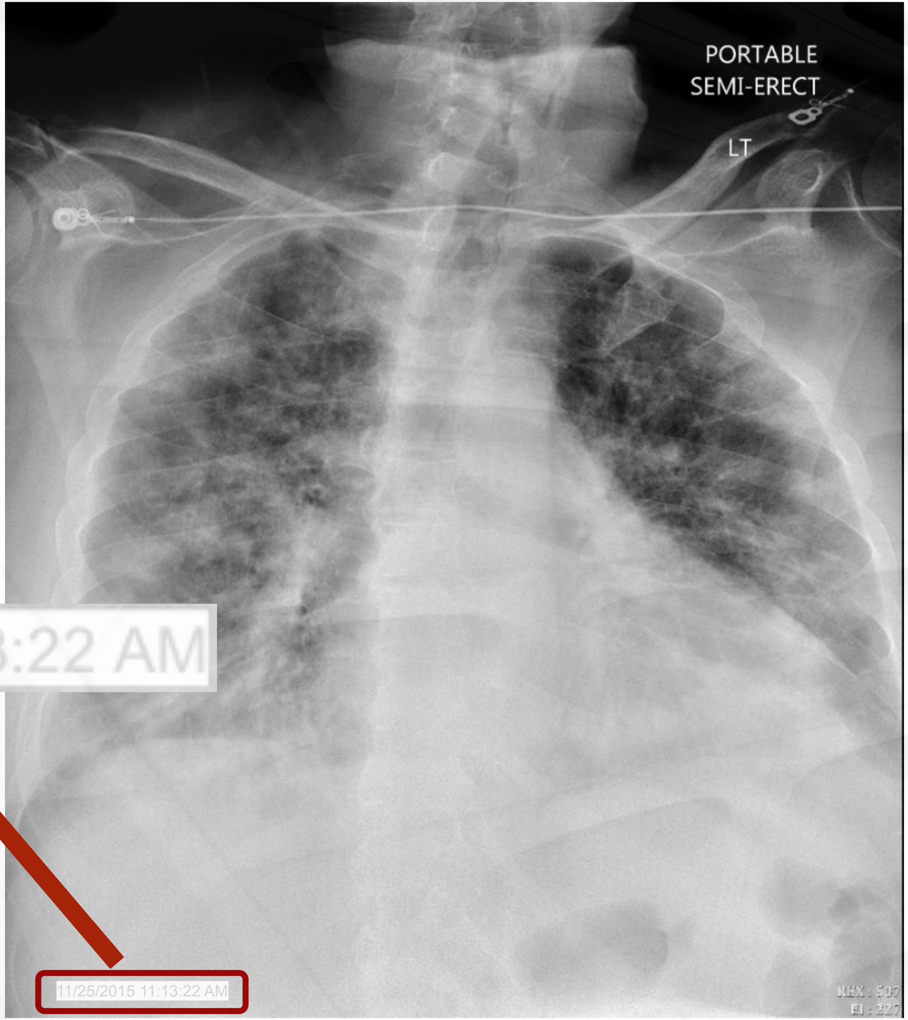
Series	Type	Scan Range (mm)	CTDIvol (mGy)	DLP (mGy-cm)	Phantom cm
1	Scout	-	-	-	-
2	Helical	S15.750-I650.250	5.10	373.00	Body 32
5	Helical	S188.000-I105.000	5.10	182.72	Body 32
			Total Exam DLP:	555.72	

1/1


[redacted]
SEQUOIA
6C2
4Hz
H5.0MHz
R0mm
OB
General
S2/ 0/ 3/E:1+1
1/1
CD:5MHz
CD Gain = 43
CDE
15dB
Store in progress
.089


PORTABLE
SEMI-ERECT
LT
11/25/2015 11:13:22 AM
RHX : 507
EI : 227


PORTABLE
SEMI-ERECT
LT
11/25/2015 11:13:22 AM
RHX : 507
EI : 227



11/25/2015 11:13:22 AM

Pixel DeID AI

Leveraging AI

- Manual de-identification is time consuming, requires significant resources and prone to human error.
- Many tools require specific configuration files or manual selection that may fail to capture PHI in all scenarios
- CNN/transformer based models scalable, flexible and adaptable across different modalities.
- Large language models (eg. GPT4) may help improve and make post processing of outputs more robust.

Pixel DeID Workflow

- Three step process

Detect text ---------> OCR --------> Classification

DAVIDSON DOUGLAS M 01092012

DOB 06161976

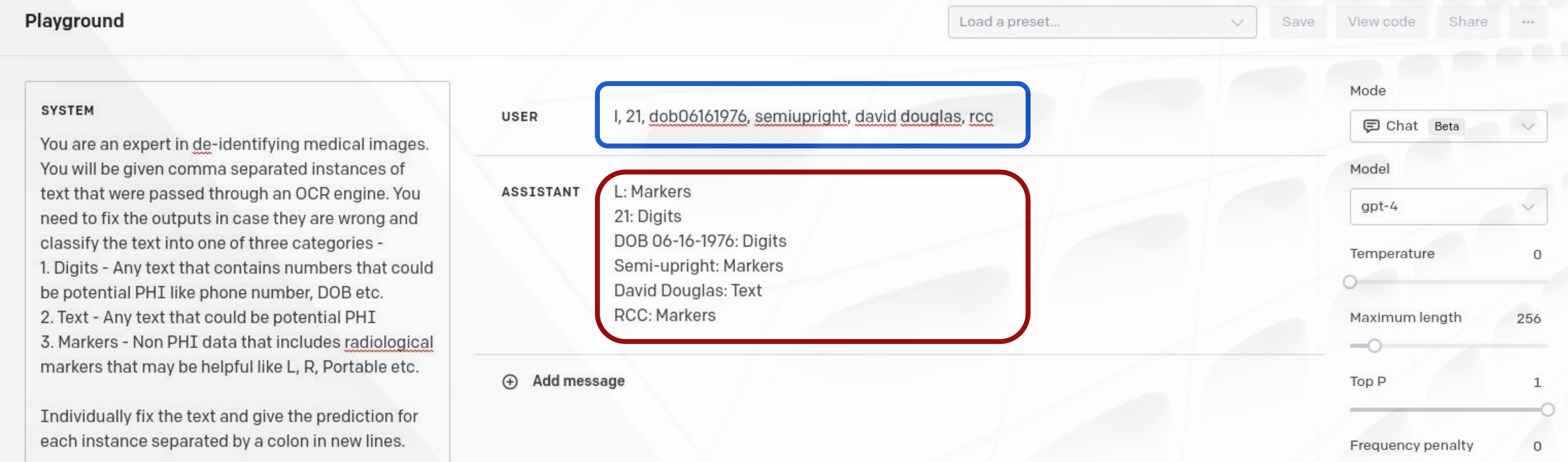
PHI Detection and Classification

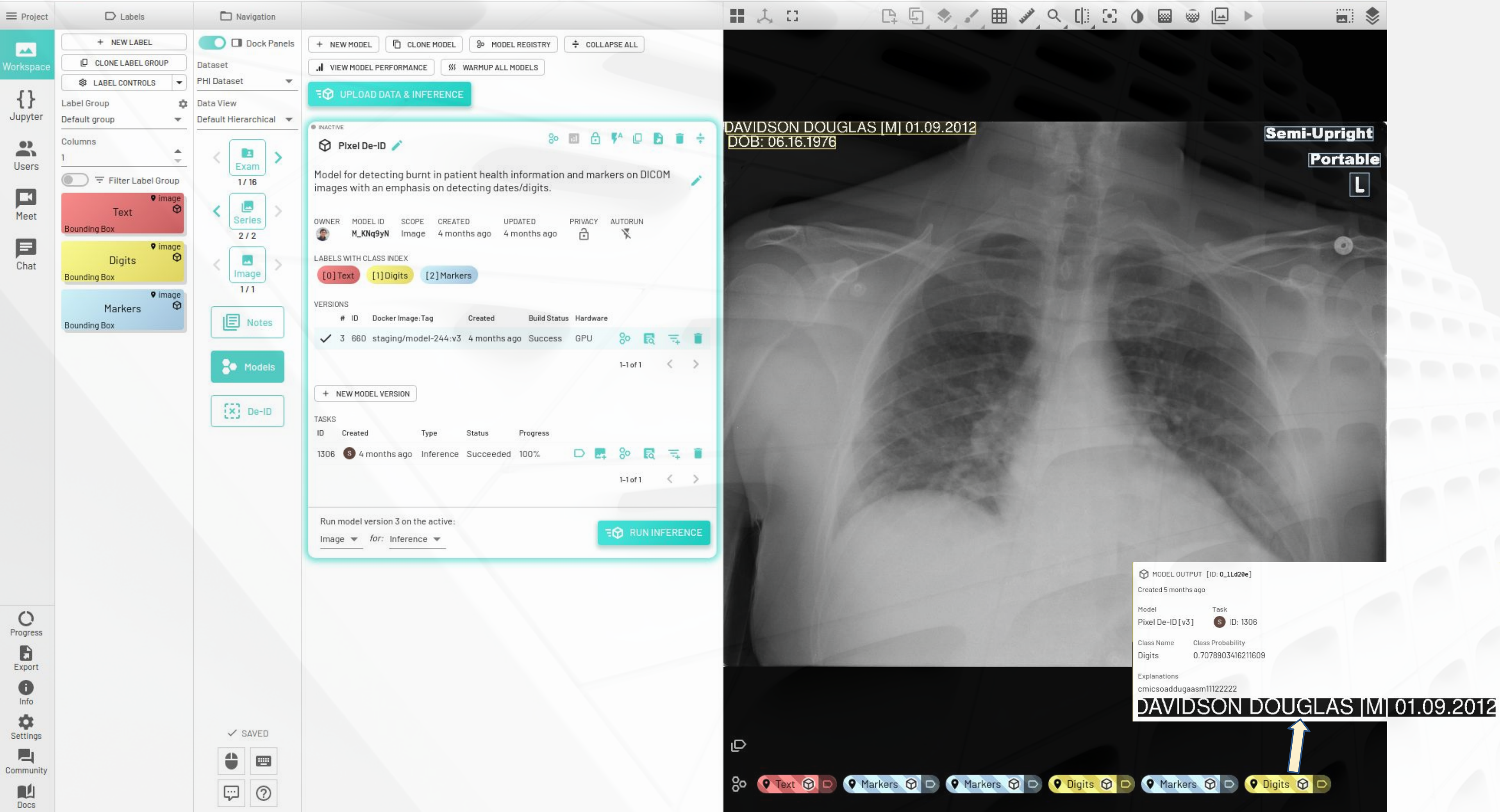
- Deep learning model that purely looks at pixel data and detects text on images. Classifies outputs into three categories

    - **Text** (potential PHI)
    - **Digits** (text with any digits, potential PHI)
    - **Markers** (radiological markers, may not want to redact)

Optical Character Recognition (OCR) + GPT-4

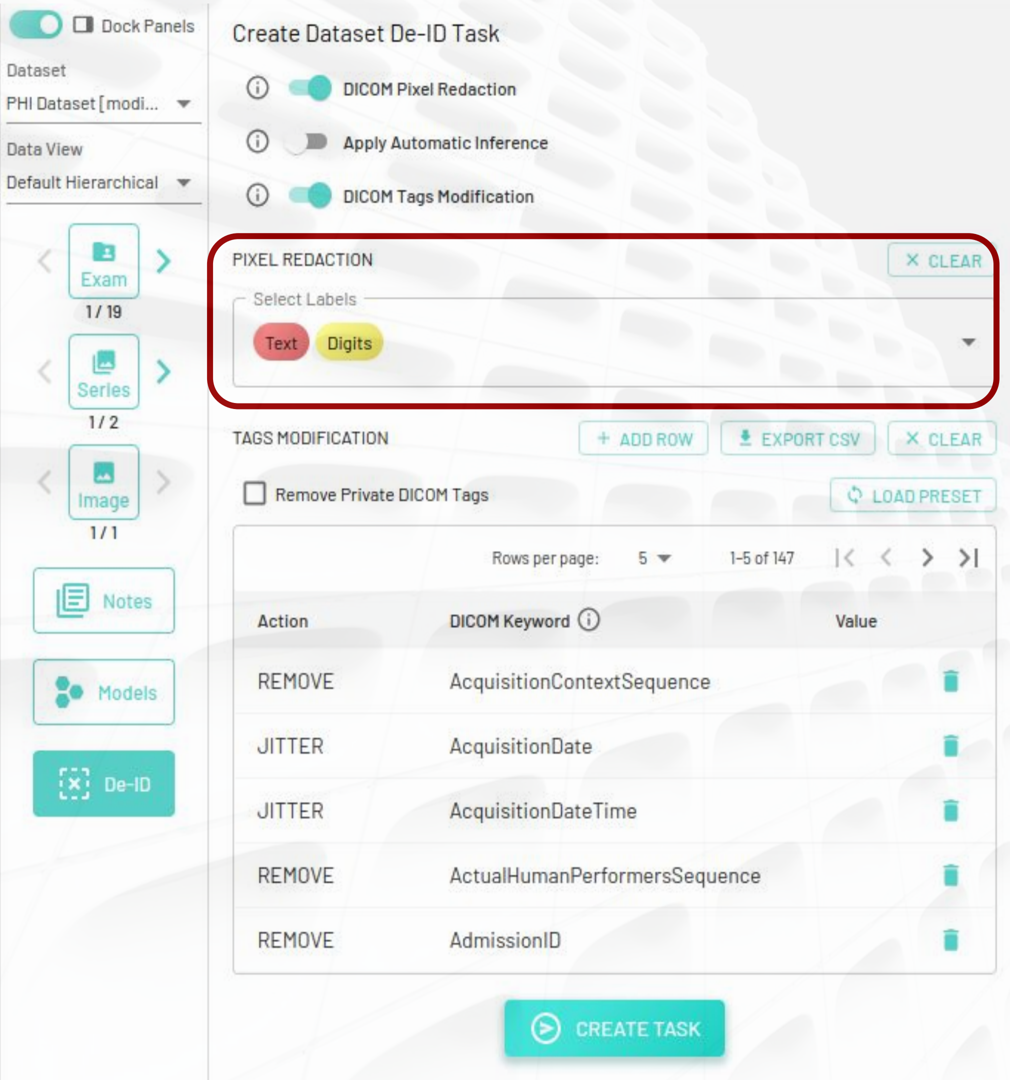
- Detected bounding boxes are passed through a robust OCR model to extract accurate textual information.
- Currently a deep learning + NER approach classifies predictions into Text, Digits or Markers. Can also utilize GPT-4 for zero shot or one shot classification -




Project
Labels
Navigation
Workspace
Jupyter
Users
Meet
Chat
NEW LABEL
CLONE LABEL GROUP
LABEL CONTROLS
Label Group
Default group
Columns
1
Filter Label Group
image
Text
Bounding Box
image
Digits
Bounding Box
image
Markers
Bounding Box
Dock Panels
Dataset
PHI Dataset
Data View
Default Hierarchical
Exam
1 / 16
Series
2 / 2
Image
1 / 1
Notes
Models
De-ID
NEW MODEL
CLONE MODEL
MODEL REGISTRY
COLLAPSE ALL
VIEW MODEL PERFORMANCE
WARMUP ALL MODELS
UPLOAD DATA & INFERENCE
INACTIVE
Pixel De-ID
Model for detecting burnt in patient health information and markers on DICOM images with an emphasis on detecting dates/digits.
OWNER
MODEL ID M_KNq9yN
SCOPE Image
CREATED 4 months ago
UPDATED 4 months ago
PRIVACY
AUTORUN
LABELS WITH CLASS INDEX
[0] Text
[1] Digits
[2] Markers
VERSIONS
ID Docker Image:Tag Created Build Status Hardware
3 660 staging/model-244:v3 4 months ago Success GPU
1-1 of 1
NEW MODEL VERSION
TASKS
ID Created Type Status Progress
1306 4 months ago Inference Succeeded 100%
1-1 of 1
Run model version 3 on the active:
Image for: Inference
RUN INFERENCE
Progress
Export
Info
Settings
Community
Docs
SAVED
DAVIDSON DOUGLAS [M] 01.09.2012
DOB: 06.16.1976
Semi-Upright
Portable
L
MODEL OUTPUT [ID: 0_1Ld20e]
Created 5 months ago
Model
Pixel De-ID [v3]
Task
ID: 1306
Class Name
Digits
Class Probability
0.7078903416211609
Explanations
cmicsoaddugaasm11122222
DAVIDSON DOUGLAS [M] 01.09.2012
Text
Markers
Markers
Digits
Markers
Digits

Redaction

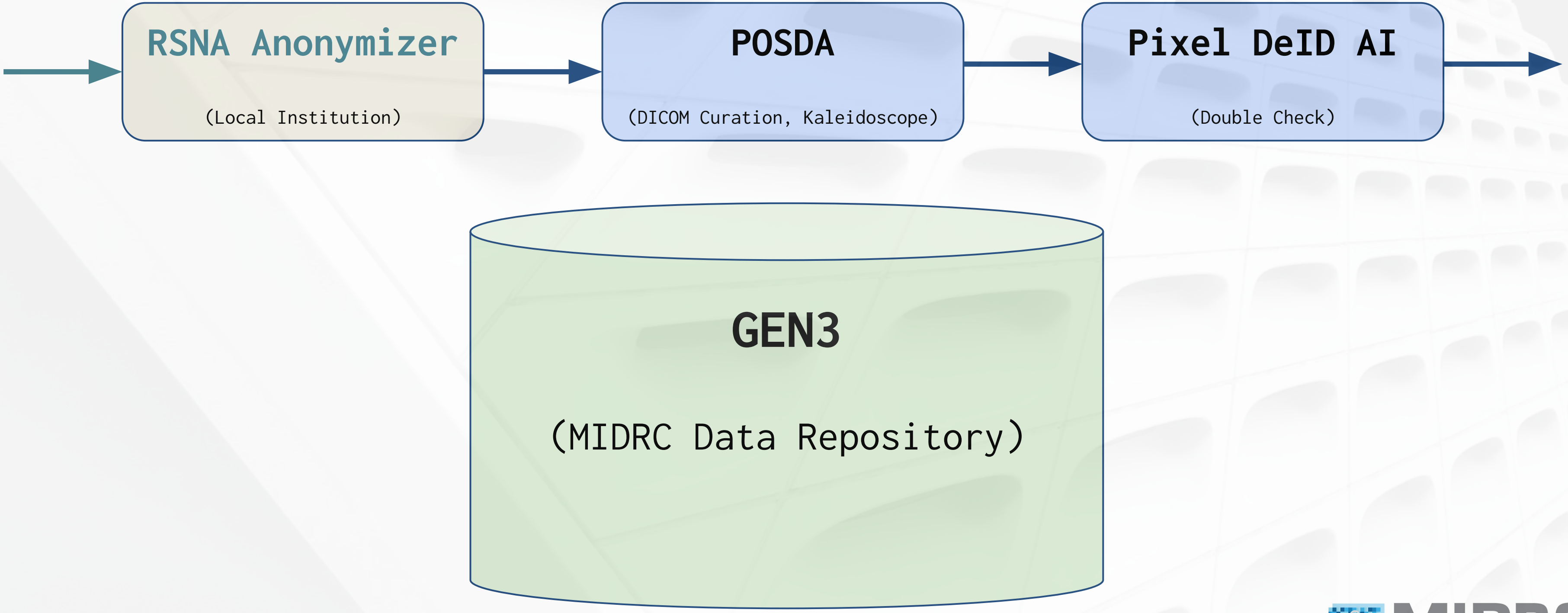
- Pixel redaction replaces text with black box.
- Can redact all predictions or choose a subset to redact (human in the loop)



DeID Workflow

RSNA MIDRC

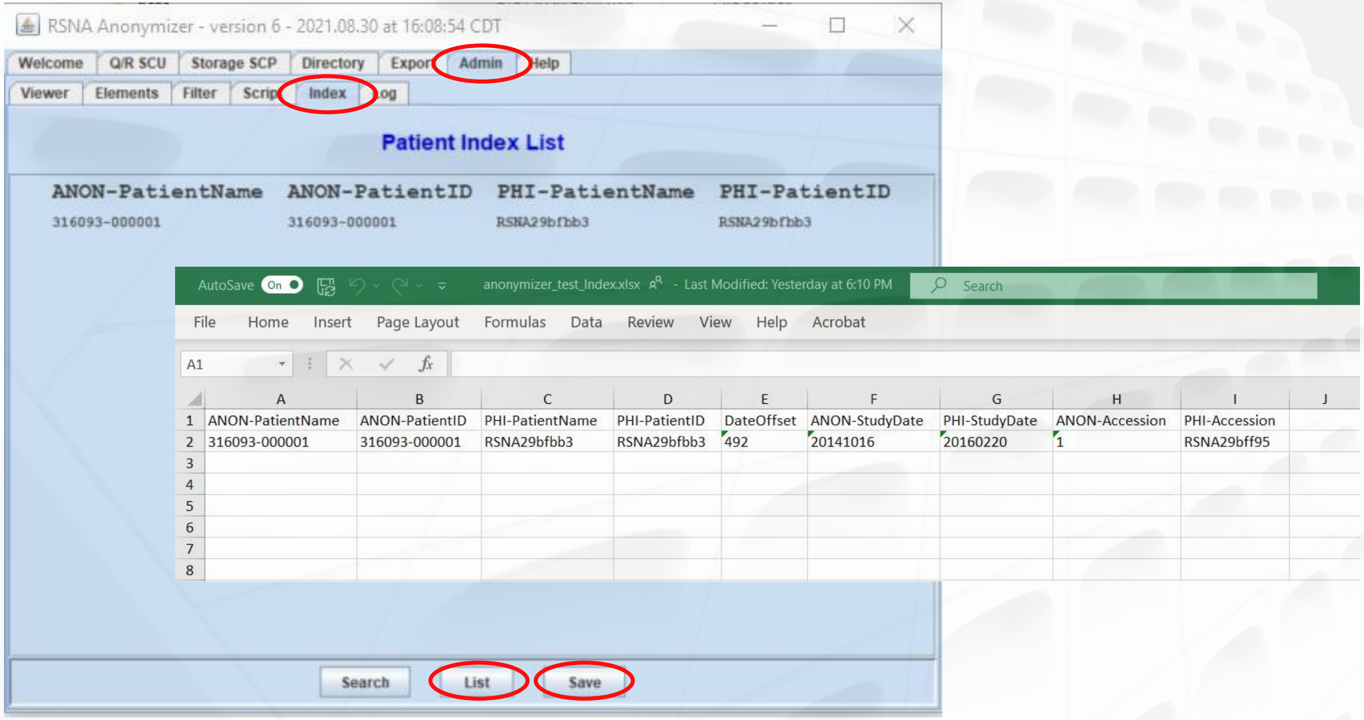
Data Workflow



MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER

1. RSNA Anonymizer For DICOM Tags

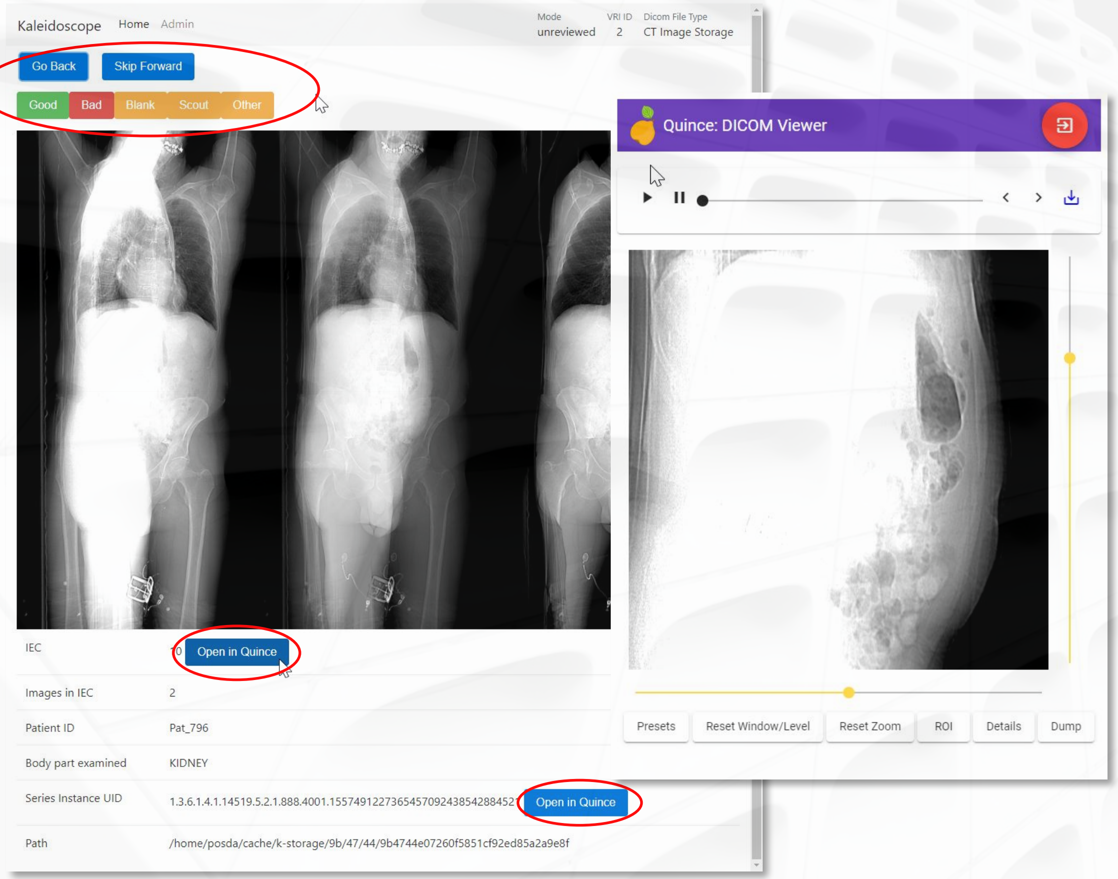
- The Admin tab of Anonymizer provides access to the script used in de-identification.
- You can also generate a table (in MS-Excel) that shows the mapping of original to pseudonymous values for patient name and ID and study date, as well as the integer representation of the date offset.
- To generate the mapping table, click the Admin tab, then Index, then the List and Save buttons.
- Keep the table secure, as you would any document containing PHI.



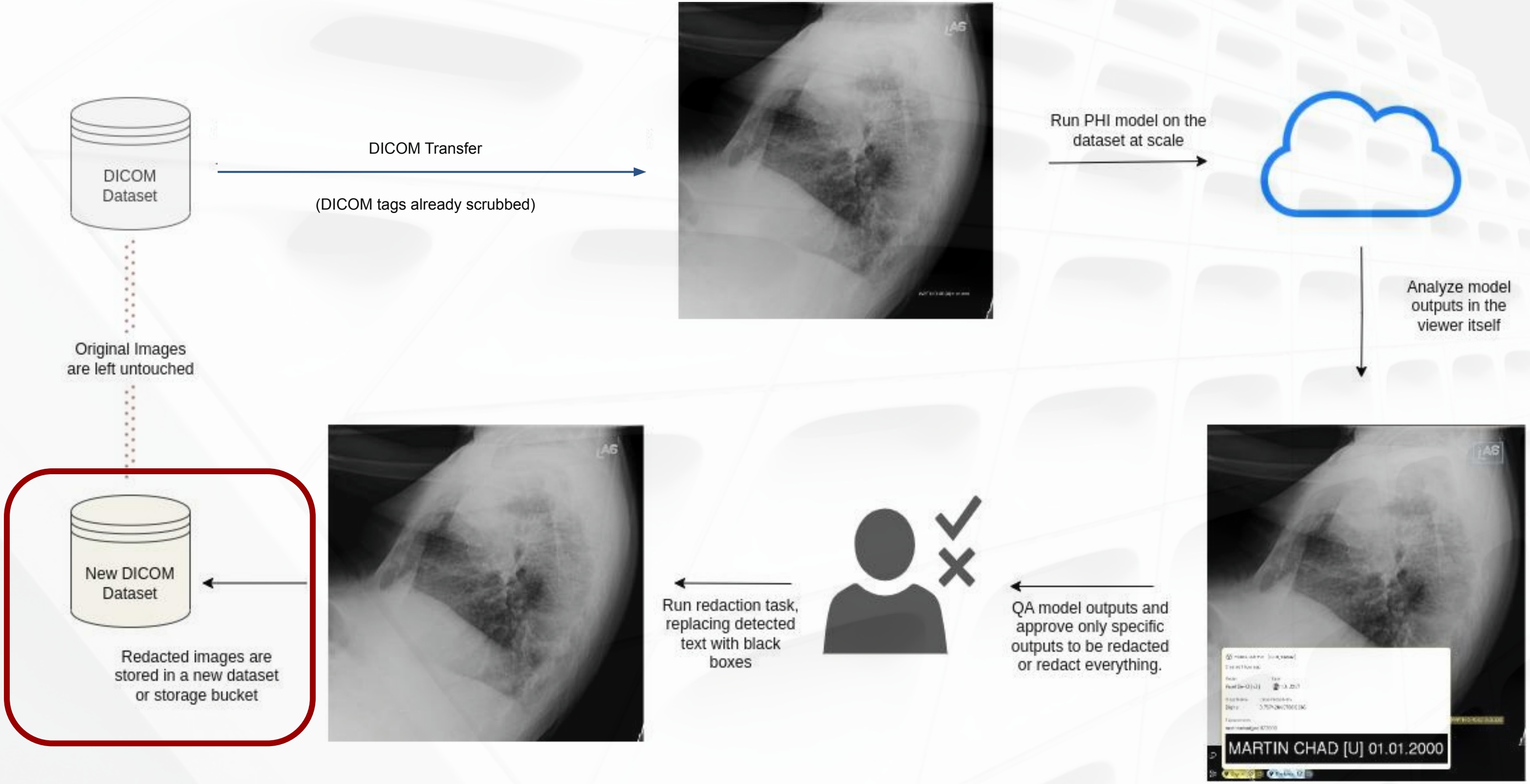
2. Human Visual Inspection Using Posda Tools

- Once Kaleidoscope opens you will have the option to review all the images, mark images as Good or Bad, etc. and to perform further review with the Quince viewer.
- See the <u>Using Kaleidoscope and Quince</u> section in the POSDA user Guide for more information on the viewers.

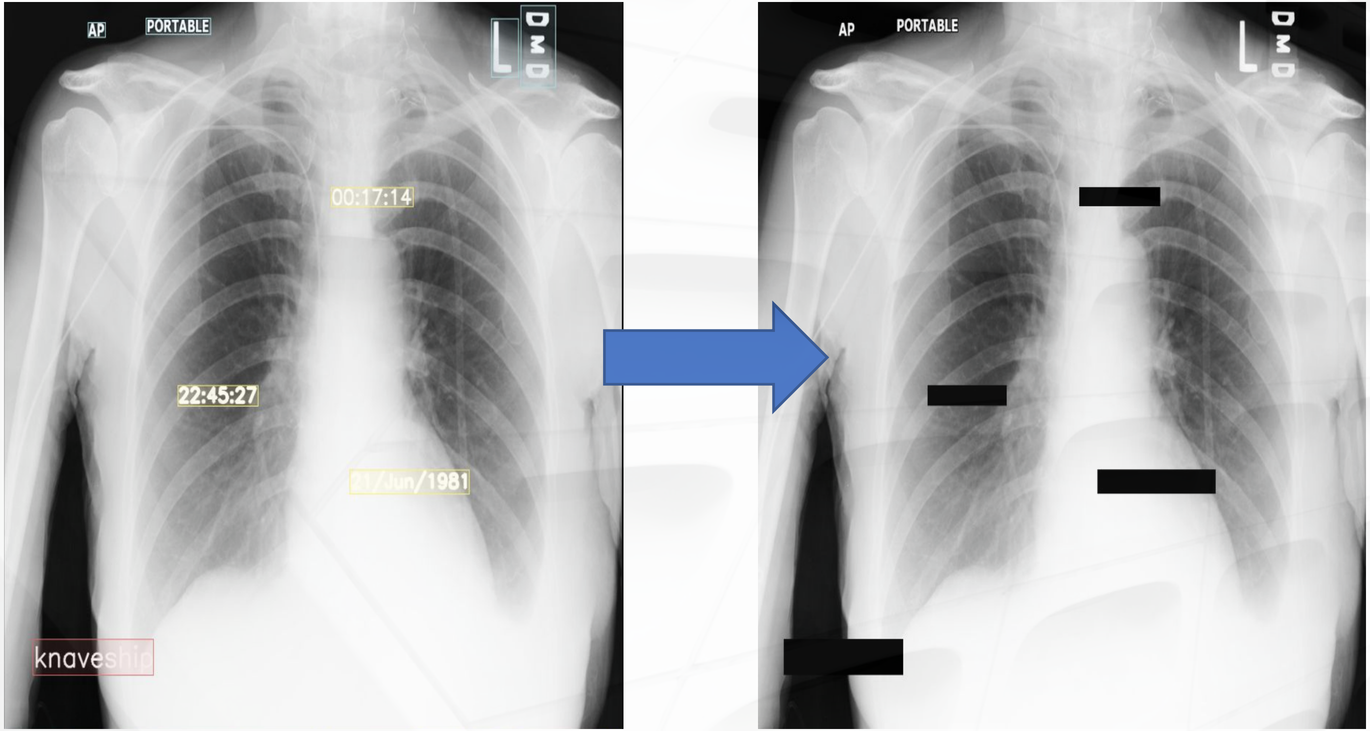
https://github.com/UAMS-DBMI/PosdaTools



3. Pixel DeID AI Workflow

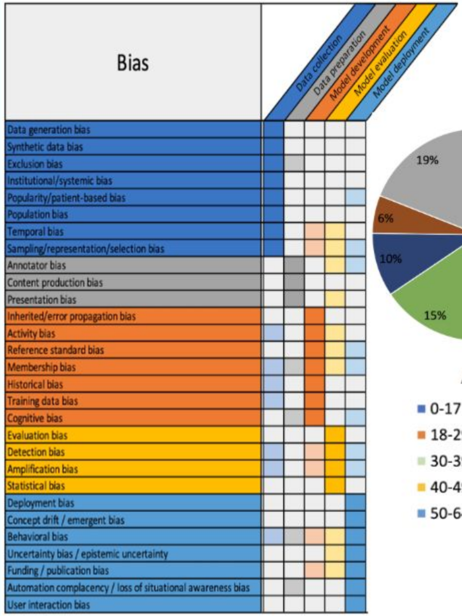


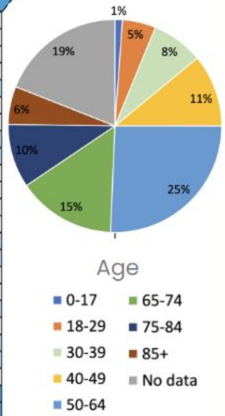
3. Pixel DeID AI Workflow



Bias & Diversity Working Group

A diverse data collection and curation strategy, as well as the mitigation of bias in data analysis within the MIDRC commons, are critically important to yield ethical AI algorithms that produce trustworthy results for all groups. MIDRC strives to mitigate bias in its study population, data collection, curation and analysis.





BDWG Members:
AAPM - Weijie Chen, Karen Drukker, Kyle Myers, Berkman Sahiner, Emily Townley **ACR** - Jayashree Kalpathy-Cramer, Judy Wawira-Gichoya **NIH** - Rui de Sá **RSNA** - Sanmi Koyejo, Zi Jill Zhang **UChicago-MIDRC Central** - Maryellen Giger, Nick Gruszauskas, Heather Whitney

Grand Challenges Working

- Performs oversight of all MIDRC Grand Challenge processes, protocols, and guidelines
- Maintains MIDRC Portfolio of Grand Challenges
- Planning several upcoming Challenges using pre-published MIDRC data with expert annotations, including:

- COVID-19 severity
- Patient outcome prediction using both clinical and imaging data
- Longitudinal assessments, disease progression
- Risk assessment for long-term sequelae (e.g., neuro, cardiovascular)
- COVID segmentation on CT data

MIDRC COVIDx CHALLENGE

A COVID classification Grand Challenge on pre-published portable chest radiographs from MIDRC

Top-ranked finishers will be acknowledged during our session at the **Innovation Theater, Booth 3316, South Hall Tuesday, November 29, 4:00 PM - 5:00 PM**

Cash awards generously sponsored by the International Society for Photonics and Optics (SPIE)

MIDRC-GCWG Members:
AAPM - Sam Armato, Karen Drukker, Lubomir Hadjiiski, Emily Townley **ACR** – Jayashree Kalpathy-Cramer, Chris Treml **NIH** - Rui de Sá **RSNA** - Robyn Ball, Adam Flanders, Tim Stearns, Carol Wu **UChicago-MIDRC Central** - Maryellen Giger, Ravi Madduri

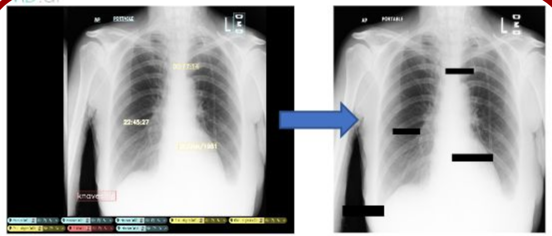
Annotations Working Group

Crowd-Sourced Annotations

- To create standard methods for labeling and annotation of COVID-19 radiographic and CT images.
- To develop best-practices for image labeling by crowds of imaging experts, including:
  - Recruitment and training of experts
  - Image label adjudication
  - Methods to assess accuracy of experts
  - Data cleaning methods for labels.
- To deploy these tools and processes to conduct 3 data science challenges to develop computer vision algorithms for COVID-19.

Helper AI & the Annotation Process:

- High quality, well-curated annotations are an essential supplement to the MIDRC effort and AI research
- Annotations can be created by human experts or automatically created by tools like AI models.



- MD.ai's recent application of Helper AI: Using an AI model to help identify pixel data in regions containing PHI. The burned-in PHI is then redacted using the AI model's annotations.

Annotation Members:
RSNA - Adam Flanders, Bhavik Patel, Carol Wu, Chris Carr, George Shih, Maryam Vazirabad, Jason Sho, Thomas OSullivan
ACR - Brian Bialecki **AAPM** - Andrey Federov, Paul Kinahan, Sam Armato
Gen3 - Bob Grossman, **UChicago-MIDRC Central** - Nick Gruszauskas
MD.ai team: Chinmay Singhal, Zhihao Wang