Omniverse Autonomous X-Ray Workstreams

2025

GE HealthCare

POC Planning Scope and Requirements

This project is to plan and estimate the effort to create an autonomous X-ray demo in October 2025

Scope of POC Project

Autonomous Chest X-ray for patient

Refine requirements for autonomous X-Ray POC

Decide on most cost-effective technology components to meet those requirements

- This will require creation of rapid prototypes for testing and feedback

Define solution architecture

Target scope and estimate work to create a GE-NVIDIA demo for October 2025

POC Planning Timeframe:

2 months starting May 1, 2025

- This leaves ~4 months for POC demo development

Autonomous X-Ray Requirements

Identity check and EMR communication

User-friendly environment to maximize trust

Multi-language

Guidance to Chest X-ray position

Safety mitigation, especially for moving robotic parts

Image quality check that may trigger patient re-positioning and re-acquisition of image

Assumptions:

Only one person in the room during autonomous guidance (to avoid interference with autonomous functionality)

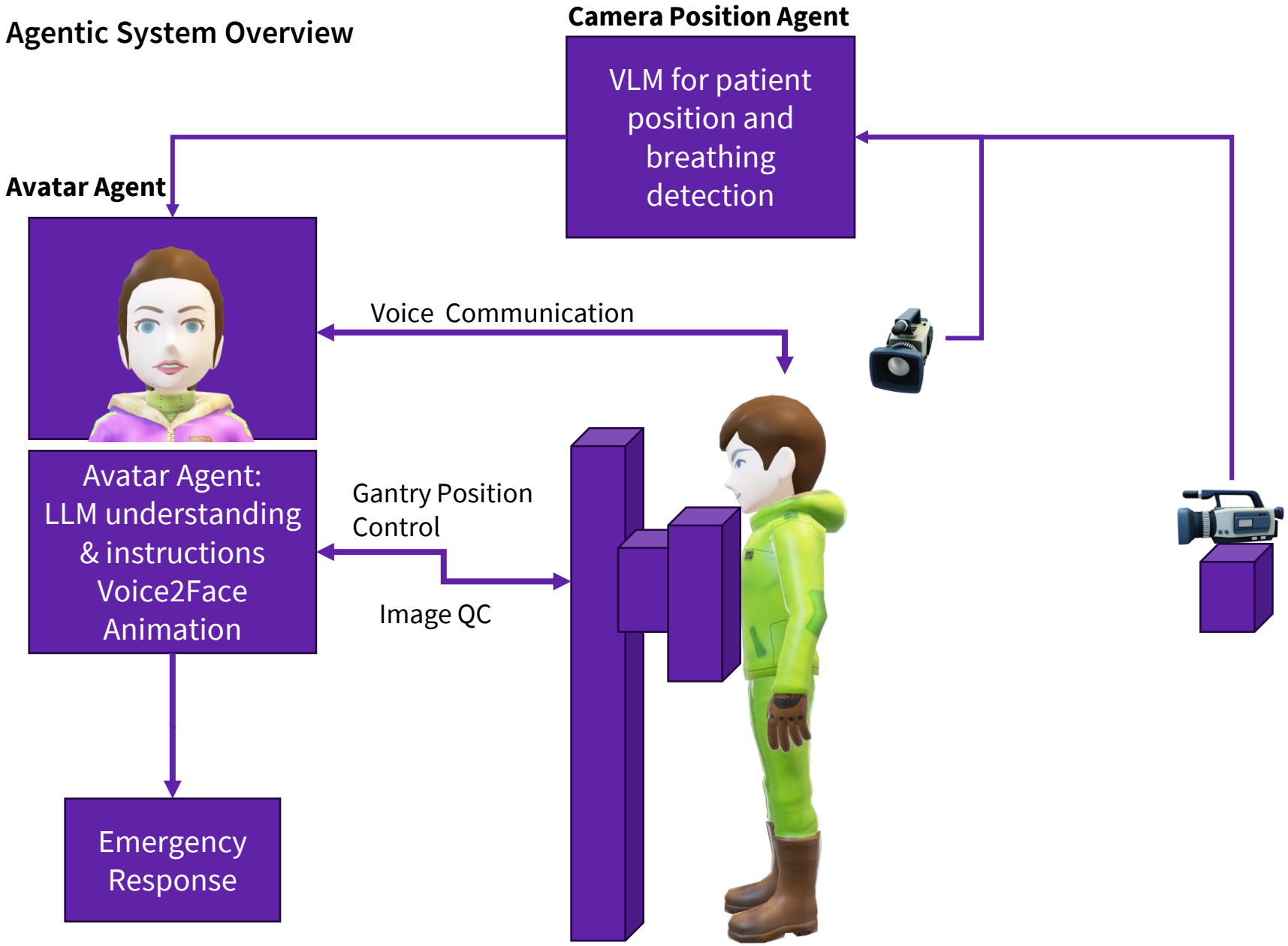
Workstreams

Overview

This development program is divided into 3 general workstreams that will explore creation of AI Agents trained on various aspects of autonomous X-ray. The work in each workstream will involve estimation and design of both training of the AI Agents and execution/inferencing of the AI Agents that will work in tandem in the autonomous X-ray solution

1. AI agent for patient interaction
2. Camera (sensor) based applications (patient position, field of view and potentially breathing detection)
3. X-Ray image analysis (quality check etc)

Agentic System Overview



Workstream 1: Agentic-powered Avatar

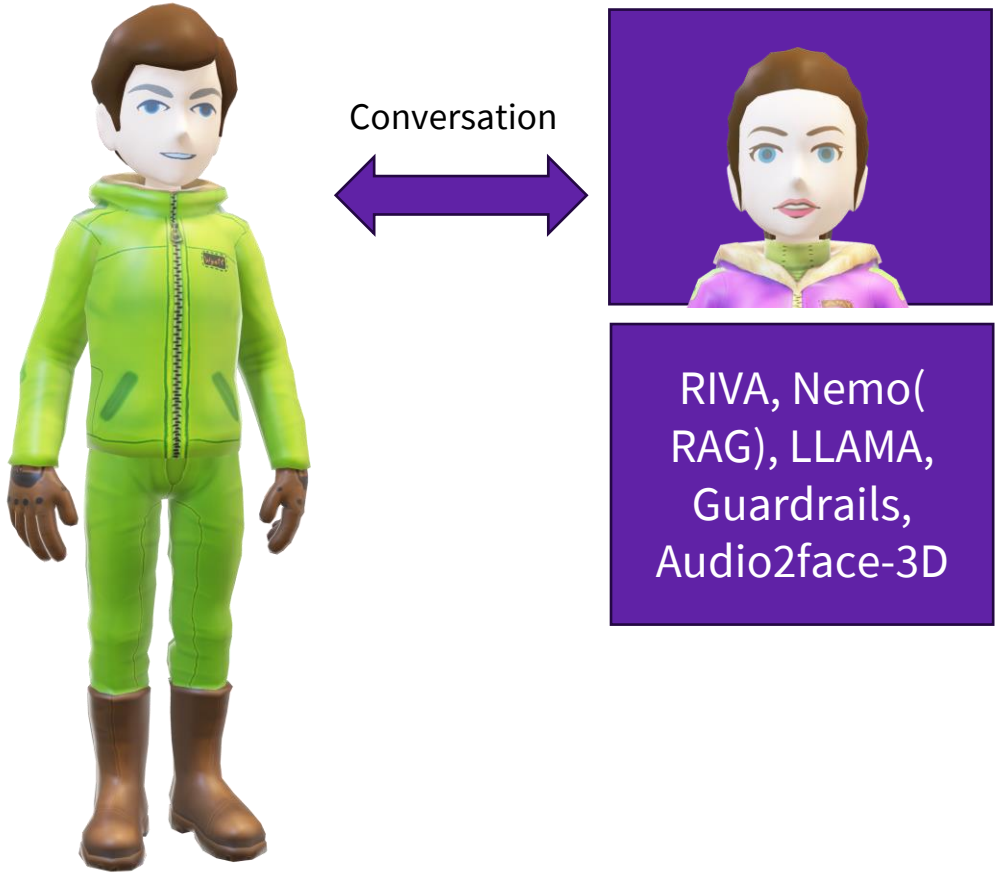
Overview

This workstream shall focus on creation of an agentic AI-powered Avatar. Avatar shall interact with patient and AI Agents for patient position and X-Ray quality and understand technologist procedures for various reasons for exam to help guide patient.

Work breakdown:

- Language model: Investigate different agentic language models to determine most cost-effective solution that meets required multi-language understanding for patient interactions.
- Avatar display: Investigate performance and cost of 3D and 2D audio2face to determine most cost-effective solution that meets user needs. Explore how avatar can be modified for gender, race/ethnicity and look (for ex photorealistic vs cartoon) to help patient trust factor.

Agentic AI Avatar Interaction with Patient



Workstream 1 Agentic-powered Avatar (cont'd)
- Team Composition and Required Resources -

Team Composition for LLM-powered Avatar

Clinical expert – X-Ray technologist with IT background. Will create LLM requirements for:

- Comprehension of X-Ray technologist patient positioning workflow and emergency response
- Information required from other AI agents on Patient Position and Image Quality
- Lead high level risk analysis to define mitigation requirements

LLM datascience expert – to explore most cost-effective LLM meeting above requirements

User Experience Expert- to estimate and design user experience testing

NVIDIA experts/ training for: RIVA & Whisper, NeMo (RAG, Guardrails), Nemotron LLAMA, NVIDIA Audio2face-3D&2D

Resource Requirements

Expert panel for gathering input from:

- Other X-ray technologists
- Patients

Menlo session:

- Collect feedback from above users to feed into requirements
- Clinical Partner (Radnet?)

NVIDIA/AWS Sandbox for:

- Multiple LLM models for testing functionality to requirements

Workstream 2 Camera Based Applications - Patient position and breathing detection -

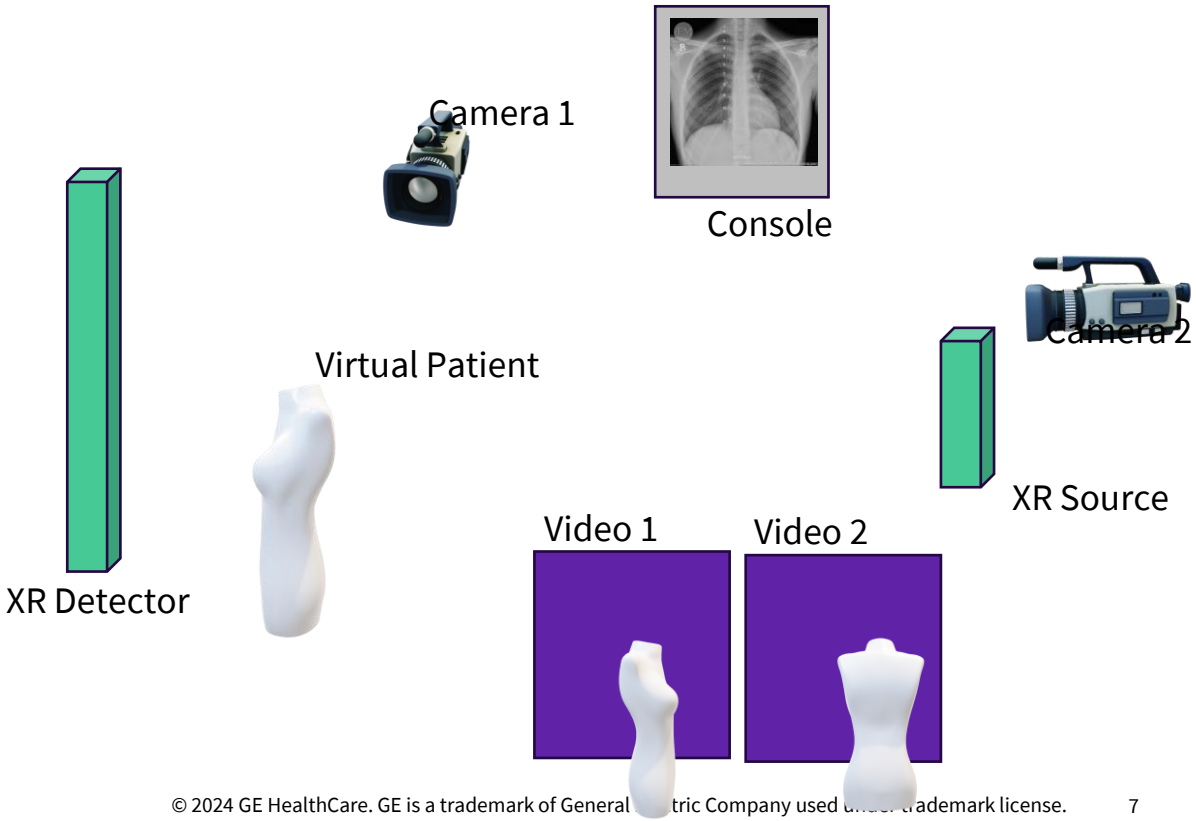
Overview

This workstream shall focus on estimating the creation of an Omniverse environment to train VLM's to track both patient position and potentially hazardous objects (ex moving X-ray device, chairs etc.) in real-time.

Work Breakdown:

- Omniverse training environment: Investigate creation of environment to train VLM's based on photo-realistic patient 3D avatars, equipment and other objects/artefacts in virtual X-ray room.
  - Determine best technology for virtual X-ray room simulations (Omniverse/OpenUSD vs Unreal Engine)
  - Determine best VLM technology (VILA, Isaac, VSS) and training method (for example imitation learning) & design training environment
- Omniverse execution test environment: Assuming VLMs are trained, investigate feasibility and design of real-time interaction between patient position sensing VLM/VLA's , Avatar (LLM) and virtual patient (LLM-powered to understand commands from Avatar.
- Useability testing: leverage virtual patient room and head mounted display to explore different room and avatar location configurations

Resource Requirements



Workstream 2 Camera Based Applications (cont'd)
- Team Composition and Required Resources -

Team Composition

Clinical expert – X-Ray technologist with IT background. Will create VLM requirements for:

- Optimal clinical patient position (for VLM training and testing)
- Variance in patients, X-Ray room configurations and obstacles

VLM datascience expert – to explore most cost-effective VLM methods for training and testing

3D Gaming expert to design configurable avatar creator to generate a variety of virtual patients to run under simulation and design configurable animation sequencer for training and testing

NVIDIA experts/ training for : VILA, Isaac, VSS, Ominverse/USD

Resource Requirements

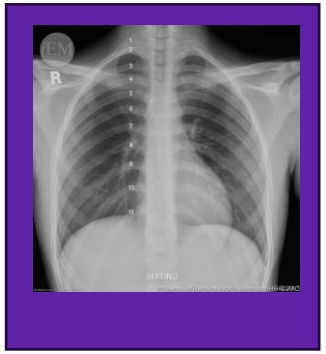
NVIDIA/AWS Sandbox for Omniverse for:

- Omniverse/OpenUSD and Unreal Engine Integration comparison and room prototyping
- Multiple VLM models for model selection and design of training environment
- Test environment for trained VLMs, Avatar and virtual patient

Head Mounted Displays to explore immersive testing with real patients in virtual X-ray room guided by Avatar

Omniverse Environment

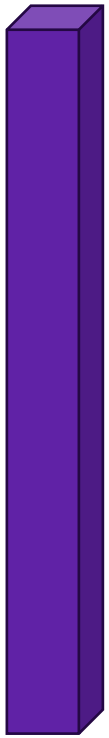
Camera 1



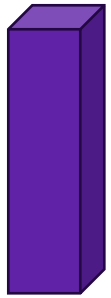
Console

Camera 2

Detector



Large patient



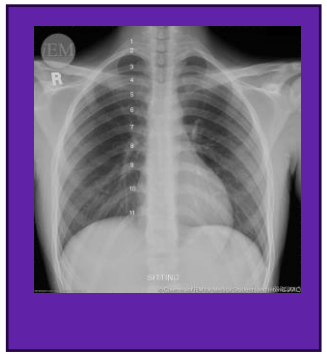
Source

+ 0 deg

Video 1

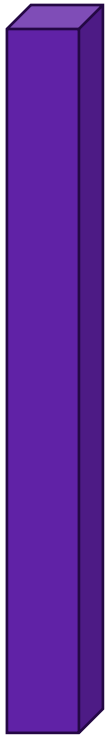
Video 2

Omniverse Training - Simulate virtual patient movement -

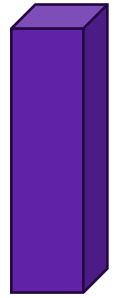


Console

Detector



Large patient



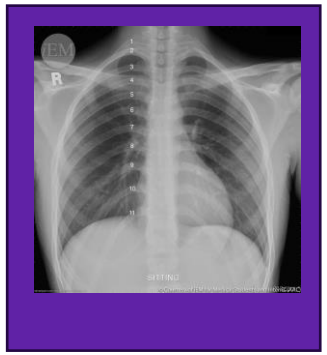
Source

+ 15 deg

Video 1

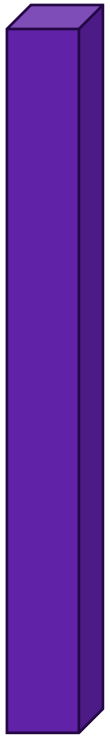
Video 2

Omniverse Training - Simulate patient size and morphology -

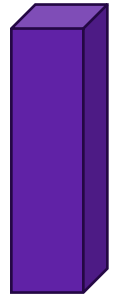


Console

Detector



Small patient

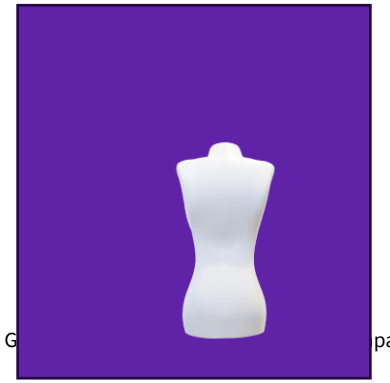


Source

+ 0 deg

Video 1

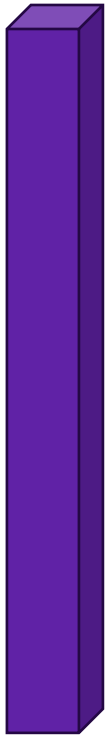
Video 2

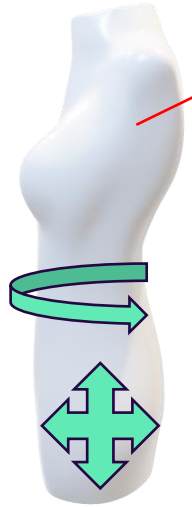


Omniverse Training - Training VLM -

For each different patient body size, vary patient position and orientation while recording video. Video-patient position information will help VLM learn patient position from video feed for future inference.

Detector

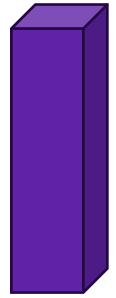




Vary patient size and body type

Vary patient position

Source



Video 1

Video 2

Training Data For X-Ray Exam Room Variability

Large room-to-room variability between exam rooms.

Differences in:

- Room dimensions
- Ceiling height
- Room layout
- Equipment and accessories
- Colors (wall, floors, equipment)
- Light color and intensity
- Possible paths for patient movement & expected target locations for patients
- Possible location of in-room cameras

It is important to consider variability in all associated parameters when training a model to monitor patient position/condition and perform patient interaction. Realistic digital simulations will help to augment training data.

GE HealthCare

Photo-Realistic Rendering Examples
- Unreal Engine-

Camera 1 Posterior View

Photo-Realistic rendering examples
- Unreal Engine -

Camera 2 Lateral View

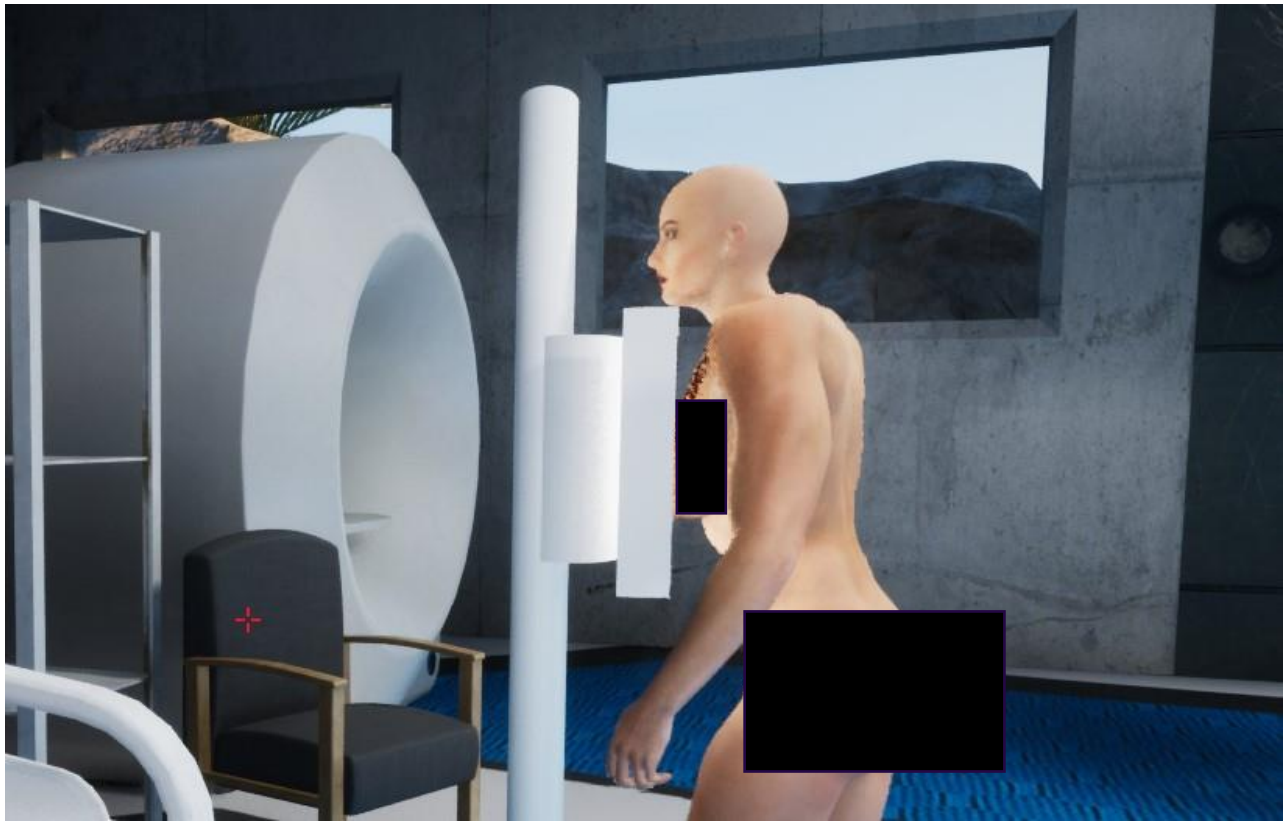
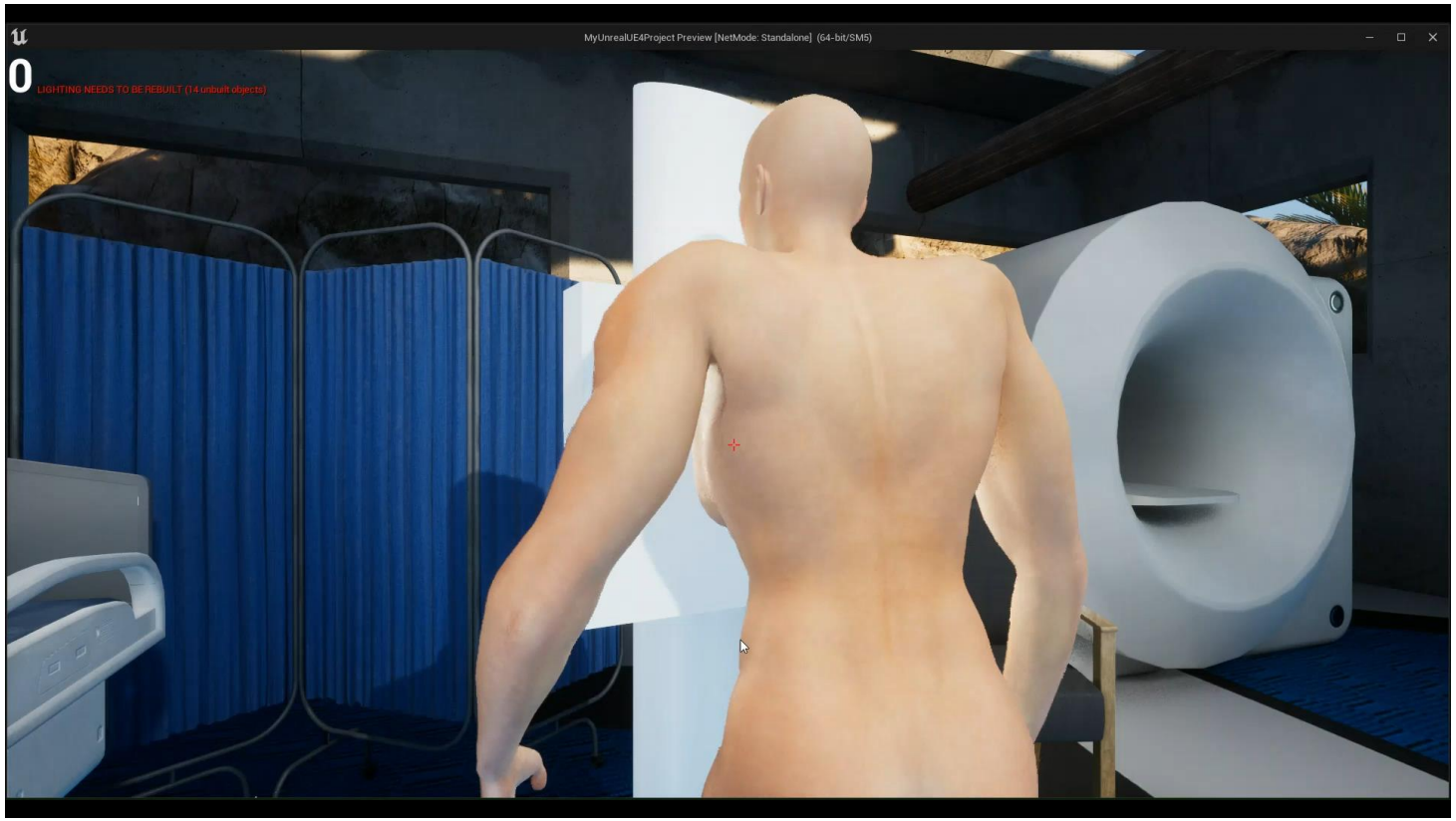


Photo-Realistic rendering examples (cont'd)

Patient breathing (video)



Workstream 3: X-Ray image analysis

Overview

This workstream shall focus on improving image analysis for X-ray Quality Check (QC)

A CT scan will be inserted into the virtual X-Ray room and registered with the virtual patient.

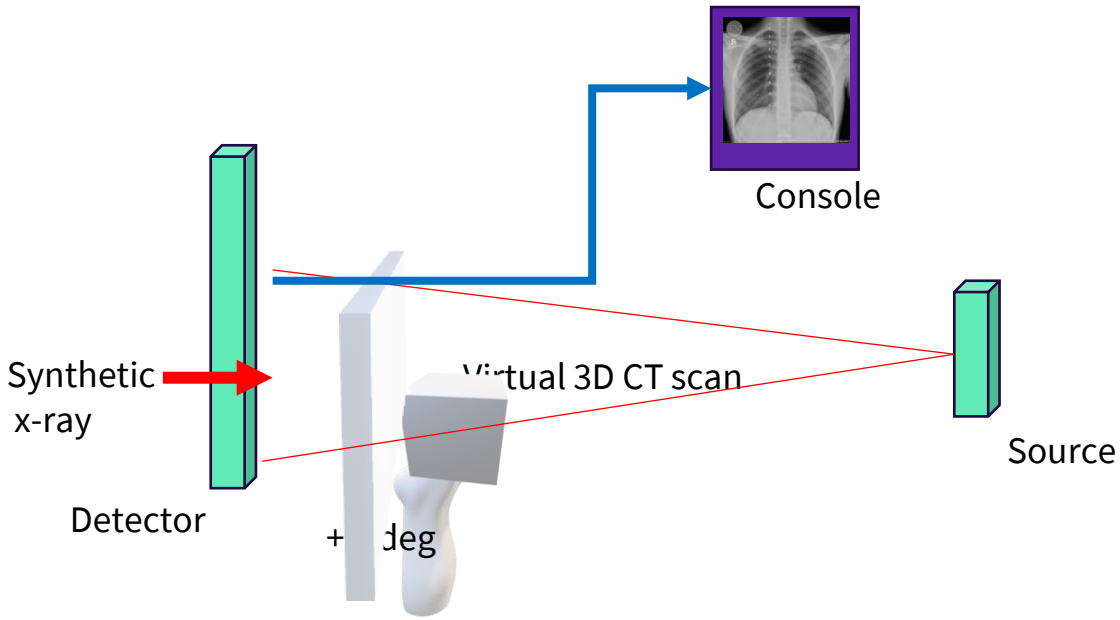
As the virtual patient moves around, a Maximum Intensity Projection of the CT scan will generate simulated X-rays

Given that the rotation of the virtual patient are known, an AI algorithm can be trained to recognize image characteristics such as image position.

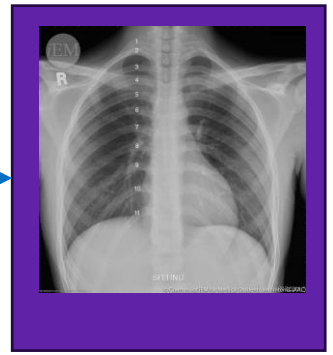
For testing, the X-ray QC algorithms can be combined with the Avatar-Camera positioning system to see how well patients are positioned for the scan.

Images that don't pass QC will have LLM communication to the LLM Avatar so that the Avatar can take corrective actions with the patient and or gantry position.

Agentic AI Avatar Interaction with Patient

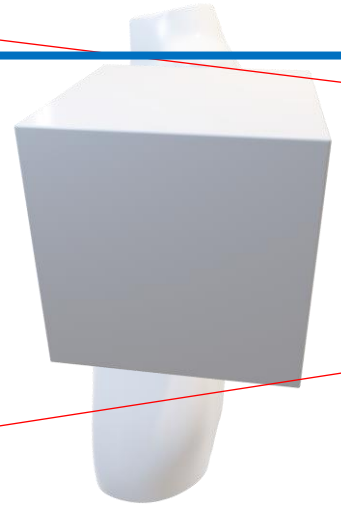


Omniverse Training Environment



Console

Detector



Virtual 3D CT scan

Synthetic x-ray

Source

+ 0 deg

Workstream 3 X-Ray image analysis(cont'd)
- Team Composition and Required Resources -

Team Composition

Clinical experts – X-Ray technologist/ application specialist:

- Criterial for proper X-ray image quality

Imaging AI datascience expert – to explore most cost-effective various MONAI-supported algorithsm methods for training and testing

NVIDIA experts/ training for : MONAI, VLM-Radiology-Agent-Framework, MONAI multi-modal, VLM-Radiology-Agent-Framework

Resource Requirements

Annotated X-ray images including breath hold images

NVIDIA/AWS Sandbox for MONAI VLM-Radiology-Agent-Framework, MONAI multi-modal, VLM-Radiology-Agent-Framework

System Level Testing

Overview

Once the Avatar Agent, Camera-based patient position Agent, Breathing Agent, X-ray quality Agent and Virtual Patient Agent are working, the system can be used for testing.

Testing can be done in a virtual environment or with a combination of the system interacting with a real patient.

- Demo would ideally be performed with a person

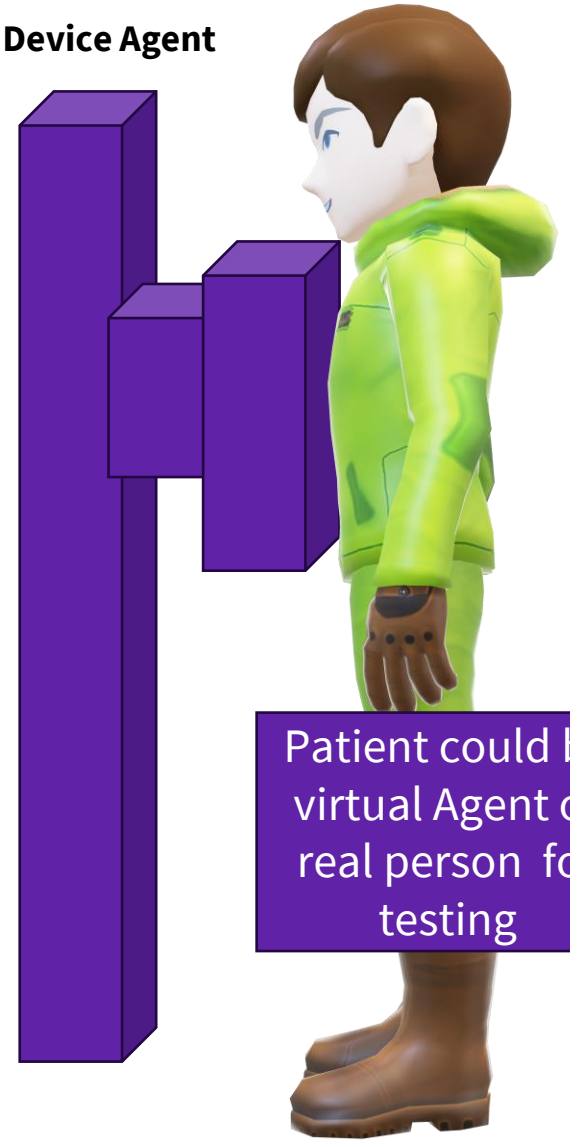
Futuristic vision: Agents could be tasked to work together to detect failures and potentially make corrective actions.

One of the outputs of this POC testing will be to understand the hardware required during real-time inferencing to understand product cost of goods and performance/cost trade-offs.

Agentic System Inference Real-time Testing Overview

Patient Agent

XR Device Agent



Patient could be virtual Agent or real person for testing

Voice Comm.

Avatar Agent

Avatar Agent

3D Position Agent

Video patient position to text

Visual camera position detection of patient, XR device and obstacles

X-Ray Quality Agent

X-ray quality to text

X-ray quality assessment Agent

Emergency Response Agent

Emergency Response


GE HealthCare