Scenario: Hazardous Event Automated Recognition System

You'll fill out the AIA according to the following scenario that has been provided for you. Use this scenario to answer the questions in the AIA to the best of your ability.

You work for a company that builds AI solutions for a diverse array of clients. A client in the manufacturing industry has requested your services. This client works with dangerous equipment and substances on a daily basis, and one of their main concerns is the safety of personnel and property. They want to install cameras and sensors in their manufacturing plants that can detect a potentially dangerous event quickly and more accurately than any human. Those dangerous events include fires, chemical spills, equipment misuse causing injury, and many more.

Your company has agreed to take on the job. You're the project manner for what's being called the Hazardous Event Automated Recognition System, or HEARS. HEARS will use deep learning techniques to create a model that can process several different types of input that come from environmental sensors. Particularly, it will be able to analyze an environment for still images, video, and audio. The model will be able to classify an environment as being in one of several different states, including: no hazards detected, hazardous chemical detected, fire detected, etc.

So, it might "see" that a particular chemical is about to start leaking out of a container using both image recognition and text analysis (i.e., it "reads" the container's label to determine what chemical is inside). In addition, the system also processes speech to detect distress among employees. It is essentially listening in on all conversations, including conversations that involve private information. Therefore, employees must sign a consent form stating that they understand and agree to be recorded for safety purposes.

All of the data used to train the model was collected by your client. The data comes from plants that the client has all over the world. Much of the client's workforce are considered part of minority protected groups, so your organization and your client will work together to identify any potential for biases in the system.

As far as the AI system itself, it will use openly available deep learning algorithms, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs). You plan on developing everything from open source libraries like TensorFlow.

Many of the resulting models will be black box in nature. Though, you do plan on keeping a human in the loop at every stage in the machine learning process. You also plan on keeping audit trails for each stage of the process, as well as continually monitor systems as they get pushed into production.

You're also concerned about the security of HEARS as it's deployed into production. HEARS must interface with a variety of other systems, including Internet of Things (IoT) sensors, the client's networks, backend databases, administrative interfaces, etc. All of these may be

targeted by bad actors looking to perform reconnaissance on the client for an attack. In addition, the client is responsible for collecting and storing all data used as input to HEARS, and the terms of that use are not known to your organization.

The client has agreed to conduct basic risk analysis procedures, but hasn't committed to specific defense measures like threat modeling and penetration testing. What's more, the client doesn't seem to have prepared an incident response team in the event that HEARS or its supporting systems suffer an adverse event.

Regardless, development on HEARS will begin soon.