Algorithmic Impact Assessment

Project Owner: _____

Project Name: HEARS

Why do you want to use AI for this project? (Check all that apply.)

- [x] Improve quality of decisions
- [x] Manage backlog/improve efficiency
- [ ] Lower costs
- [x] Innovation
- [x] Augment human processes

What capabilities does the project include? (Check all that apply.)

- [x] Image and object recognition
- [x] Text and speech analysis
- [x] Risk assessment
- [ ] Content generation
- [ ] Process optimization and workflow automation
- [ ] Other

Will the AI system use personal information as input data?

- (●) Yes
- ( ) No

How does this affect your risk level?

Using personal information increases privacy risks, particularly if the system misinterprets speech or records sensitive conversations without proper safeguards. To mitigate this, strict access controls, anonymization techniques, and encryption should be implemented. Moreover, employees must be informed about how their data will be used and given an option to raise concerns.

How was the data used by the system collected?

- (●) Collected by your organization
- ( ) Purchased from or shared by a third party
- ( ) Open dataset with known owner
- ( ) Open dataset with unknown origin
- ( ) Unclear

How does this affect your risk level?

Data collected directly by the client introduces moderate risk, as the quality and legality of the collection process may vary across regions. The client must ensure compliance with global data privacy laws, such as GDPR. An audit of the data collection process is necessary to identify and address gaps in privacy protection.

Will the system use data from multiple different sources?

- (●) Yes
- ( ) No

How does this affect your risk level?

Combining data from diverse sources increases risks of data inconsistency, integration issues, and potential exposure to cyber threats. Standardizing input formats and using secure APIs for data transfer can reduce these risks. Comprehensive end-to-end testing should also be conducted to ensure data integrity across systems.

How was user consent managed?

- ( ) No consent obtained
- (●) Opt in
- ( ) Opt out

How does this affect your risk level?

While opt-in consent mitigates some privacy risks, the process must be clear, accessible, and transparent to all employees. Failure to adequately explain the implications of the consent form could lead to disputes or legal challenges. Regularly revisiting and updating the consent forms ensures alignment with evolving privacy standards.

Will you maintain a log detailing all the changes made to model and the system?

- (●) Yes
- ( ) No

How does this affect your risk level?

Maintaining logs reduces risks by ensuring traceability for all decisions and changes made to HEARS. This is particularly important for regulatory compliance and resolving any disputes or errors in the system. Audit logs also help detect potential tampering or unauthorized access.

What type of third-party components will you use? (Check all that apply.)

- [ ] Off-the-shelf commercial products
- [x] Open source libraries
- [ ] None

How does this affect your risk level?

Open-source components can introduce vulnerabilities if not properly maintained or vetted. Regularly updating libraries and monitoring for known vulnerabilities can mitigate risks. Additionally, conducting thorough code reviews and testing open-source components will ensure compatibility and security.

Will the system provide an audit trail that records all decision points made by the system?

- (●) Yes
- ( ) No

How does this affect your risk level?

An audit trail improves transparency, making it easier to identify and correct errors in the decision-making process. For HEARS, this is particularly crucial in high-risk scenarios where incorrect hazard classification could have severe consequences. Regularly reviewing the audit trail ensures ongoing accountability.

Is there ongoing monitoring of the system to ensure that the system is operating as intended?

- (●) Yes
- ( ) No

How does this affect your risk level?

Continuous monitoring helps detect deviations in system behavior before they cause harm. For example, if the system misclassifies chemical spills, real-time monitoring could alert human operators for intervention. Implementing automated alerts based on performance thresholds further reduces risk.

Will the algorithm be a trade secret?

- ○ Yes
- ◉ No

How does this affect your risk level?

Not treating the algorithm as a trade secret lowers risks related to transparency and regulatory scrutiny. Open documentation ensures that external auditors can evaluate the system for fairness and bias. However, it is still important to protect intellectual property where appropriate

Is it possible to discover how your system renders a decision or performs a function?

- ○ System is transparent.
- ◉ System is opaque, but it is possible to analyze processes to draw an accurate conclusion about how the decision was made.
- ○ System is a black box, but detailed records of design and operations have been kept.
- ○ System is a black box.

How does this affect your risk level?

Moderate risk. Although the system is not entirely transparent, detailed records and supplementary tools (e.g., LIME for explainability) can help clarify decision-making. This is especially critical when resolving disputes or addressing errors in hazard detection.

Is the user documentation easy to understand by the intended audience?

- (●) Yes
- ( ) Not sure
- ( ) No

How does this affect your risk level?

Accessible documentation decreases risk by ensuring operators and employees understand how to interact with the system. Clear guidance on troubleshooting and system limitations minimizes misuse.

Was there a human involved at every stage of the machine learning process?

- (●) Yes
- ( ) No

How does this affect your risk level?

Including humans at each stage significantly reduces risks related to unchecked biases or errors in training data. For HEARS, this ensures that sensitive decisions, such as hazard classification, are validated before deployment.

Does the AI system pertain to any of these categories? (Check all that apply.)

- [ ] Health-related services
- [ ] Economic (loans, tax benefits, debt collection)
- [ ] Social assistance (welfare, disability claims)
- [ ] Access and mobility (security clearances, border crossings)
- [ ] Licensing or permits
- [x] None of the above

How does this affect your risk level?

Lower risk since HEARS is focused on safety, not social or economic systems where fairness is critical. However, fairness is still important to ensure the system treats diverse environments and employee groups equitably.

Does your user base include protected groups of people?

- ( ) Yes—most
- (•) Yes—some
- ( ) No

How does this affect your risk level?

Including protected groups increases the risk of bias if the system underperforms for certain demographics. Regular bias testing and diverse data collection are essential to mitigate this risk.

Are the impacts of the system reversible?

- (●) Yes
- ( ) No

How does this affect your risk level?

Reversibility reduces risk by allowing corrections or retraining if HEARS generates incorrect classifications. For example, errors in fire detection can be addressed through system updates without long-term harm.

Who was consulted about design and development decisions? (Check all that apply.)

- [x] AI team
- [x] Department that will use the system
- [ ] Users that represent the entire diverse potential user base
- [ ] Business partners
- [x] Ethics board

How does this affect your risk level?

Consulting a broad range of stakeholders reduces risks by ensuring all perspectives are considered. For HEARS, input from safety experts and ethics boards ensures alignment with industry best practices.Safety and SecurityWill the system interface with other IT systems?YesHow does this affect your risk level?Higher risk due to the expanded attack surface. For HEARS, interfacing with IoT sensors and administrative systems requires stringent access controls and regular vulnerability scans to prevent

Will the system interface with other IT systems?

- (●) Yes
- ( ) No

How does this affect your risk level?

Higher risk due to the expanded attack surface. For HEARS, interfacing with IoT sensors and administrative systems requires stringent access controls and regular vulnerability scans to prevent unauthorized access.

What risk management strategies have been employed on the networks and systems that support the AI system? (Check all that apply.)

- [✔] Risk analysis
- [✔] Threat intelligence
- [ ] Threat modeling
- [ ] Penetration testing

How does this affect your risk level?

While basic risk management reduces vulnerabilities, the absence of threat modeling and penetration testing leaves gaps. For example, IoT sensors connected to HEARS could be exploited without thorough testing. Adding these strategies is critical.

How is data collection, storage, and use being managed?

- (●) Internally
- ( ) Through a third-party service (terms and conditions unknown)
- ( ) Through a third-party service (terms and conditions verified meet strict standards)

How does this affect your risk level?

Managing data internally lowers third-party risks but requires robust internal policies. For HEARS, encrypting stored data and limiting access to authorized personnel mitigates these risks.

Does the organization have a response team ready to address any safety and security incidents?

- ( ) Yes
- (●) No

How does this affect your risk level?

A lack of an incident response team significantly increases risk. Without this team, HEARS could face prolonged outages or data breaches. Immediate investment in a dedicated response team and disaster recovery plans is necessary

Describe any additional risks to consider.

Errors in hazard classification could lead to financial and reputational damage.Cyberattacks on IoT sensors could compromise safety data integrity.Misuse of employee data could result in legal challenges.

Describe any additional mitigation strategies to consider.

Implement multi-layered security protocols, including firewalls, intrusion detection systems, and encryption. Conduct regular retraining and testing to improve model accuracy and fairness. Develop a comprehensive incident response plan to address data breaches or system failures promptly.