**Example Artificial Intelligence Policy**

**Version Date: August 2023**

This Artificial Intelligence Policy was prepared by Next-Gen Solutions in accordance with the legal landscape known as of the date hereof.

The new EU AI Act, China AI Act presented by the Cyberspace Administration of China (CAC), the NIST AI Risk Management Framework (As directed by the US National Artificial Intelligence Initiative Act of 2020) and any other applicable rules and regulations, amongst other things, aim to impose stringent new requirements introducing a risk-based approach, which requires any entity using, developing or deploying AI tools to assess the potential risks associated with such. This draft Policy aims to set the foundation for building, adopting and implementing an AI framework within your organization.

***IMPORTANT NOTE: This template is not legal advice and any organization using this template should not rely upon it as such. This template is no more than a suggested example approach. There is no requirement to use this template or if it is used, to use it as drafted. Organizations should feel free to amend, cherry pick or ignore this template entirely.***

**Artificial Intelligence Policy**

|  |  |
| --- | --- |
| **Document Classification:** | Internal |
| **Document Ref.** | **XXX** |
| **Version:** | **1.0** |
| **Dated:** | **01 May 2023** |
| **Document Author:** |  |
| **Document Owner:** |  |

1. **Purpose**

The purpose of this policy is to establish guidelines for the use of Generative AI ("GAI") tools or systems within [*Company Name*] (the "Company"). GAI tools or systems are a type of artificial intelligence that can be used to generate text, images, and other content or to make some automated decision-making processes.

GAI tools can be a valuable tool for a variety of purposes, but they also pose some potential risks on the Company. The processes set out below are intended to ensure your development, deployment and/or use of GAI tools as part of your role in the Company are aligned with applicable legal and ethical standards, and that you use proper processes to mitigate the risks associated with such use.

For the purposes of this policy, "Generative AI" or "GAI" refers to computer systems that use machine learning, deep learning, or other AI techniques to generate outputs, such as natural language, images, videos, or recommendations.

1. **Scope**

This policy applies to all activities related to the development, testing, deployment, and operation of GAI tools and systems or tools within the Company, as well as any related activities. Examples of specific use cases that fall under the policy include the use of chatbots, voice assistants, image or video generation tools, and automated decision-making systems.

This policy does not prevail over any other provision covered by other policies, such as data privacy or security policies.

This policy applies to all business units, functions, departments, and subsidiaries of the Company that engage in activities related to GAI tools and systems. Any use of GAI tools and systems must be approved in advance by the [AI Steering Committee] [compliance team] [legal team] [other- please specify].

All business units should be involved in the development and implementation of this policy to ensure that it is consistent with their specific needs and activities.

1. **Escalation Procedure**
   1. *Start with a self-guided compliance analysis:* A requester to conduct a self-guided compliance analysis with [updated training materials] and [Q&A documents] [other- please specify] available at [insert link] for day-to-day use of GAI tools or systems to first handle any questions related to this policy.
   2. *Go to local team leaders:* In any case of an issue not resolved by the requester’s self-compliance analysis, such a case shall be brought to the local business owner and/or team leader within the relevant department. This stage shall also apply to any third-party service providers performing development activities for the Company.
   3. *Escalate:* Any issue not resolved on a local team leader level, shall be brought to the [AI Steering Committee] [compliance team] [legal team] [other- please specify] via [insert log/ticketing system] [dedicated email address] [other- please specify].
   4. *Management or Business Approval*: Highly strategic/risky matters not resolved by the [AI Steering Committee] [compliance team] [legal team] [other- please specify] shall be brought to the relevant business and/or management teams as final decision-making stage.
2. **Policy Governance and Review**

This policy will be reviewed on an annual basis by the [AI Steering Committee] [compliance team] [legal team] [other- please specify] to ensure that it remains up-to-date and effective. Any changes to the policy will be communicated to relevant stakeholders, and feedback will be solicited from employees, customers, and other stakeholders to inform future updates.

1. **Roles and Responsibilities: [AI Steering Committee] [Compliance Team], Responsible AI Officer** 
   1. The Company shall appoint [an][a][the] [AI Steering Committee] [**Compliance Team**] [**Legal Team**] [other- please specify] in order to:
      1. Review and approve this policy;
      2. Ensure and approve the Company's strategy for use, deployment and/or development of GAI tools or systems;
      3. Establish a procedure for reviewing requests to use, deployment and/or development of GAI tools or systems.
   2. The Company shall appoint a Responsible AI Officer ("RAIO"), who shall hold the following responsibilities:
      1. Oversee new tools, with a main focus on identifying risks and adopting relevant mitigations. Review responsible/ethical compliance implications of new tools and processes;
      2. Build and implement new risk assessment and mitigation processes;
      3. Review and approve new GAI vendors and engagements with third parties related to the development, deployment and/or use of GAI;
      4. Act as focal point for queries and requests from individuals affected by GAI tools and/or systems, and regulatory entities.
2. **Potential risks**

The following are some potential risks associated with the use of GAI tools. These risks should be communicated to all of the Company's employees and contractors and should be explained and emphasized as part of periodical training sessions (as elaborated under Section ‎8 below):

* 1. **Confidential information leakage:**
     1. **Risk of Company's confidential information being exposed**: Some GAI tools use user-input data to train its algorithms. In case inputs contain any information which is considered confidential to the Company or to its customers, such information could be exposed by the GAI tools to unauthorized individuals.
     2. **Potential exposure to third party's confidential information**: GAI tools can also be used to generate content that is similar to or even identical to content that is confidential to third parties. If this content is shared publicly, it could harm the third party's business or reputation.
  2. **Data privacy:**
     1. **Risk of sharing personal data**: Same as with confidential information, in case user-inputs contain any Company or customers' personal data, such data could be unlawfully disclosed by the GAI tools to third parties.
     2. **Risk of receiving personal data**: Content generated by GAI tools may contain personal data without an appropriate legal basis.
  3. **Mis/Disinformation:** GAI tools can be used to generate content that is false or misleading.
  4. **Ethical issues:**
     1. **Discrimination**: GAI tools can be used to generate content that is biased or discriminatory. This could lead to discrimination against individuals or groups of people.
     2. **Transparency and Accountability**: It can be difficult to hold GAI tools or its creators accountable for their actions.
     3. **Explainability**:GAI tools are often complex and opaque, making it difficult to understand how they work or why they generate the content that they do. This lack of explainability can make it difficult to trust generative AI tools and to hold them accountable for their actions.
     4. **Automated decision making**: GAI tools can be used to automate decision making. This can be beneficial in some cases, such as when it can help to improve efficiency or accuracy. However, it can also lead to a number of risks, such as bias, reflecting the biases that are present in the data that they are trained on, error and/or lack of transparency Automated decision making can be opaque, making it difficult to understand how decisions are made. These can lead to a lack of trust in the decision-making process or to actual harm to humans affected by such decisions.
  5. **Intellectual Property infringement:** GAI tools can be used to generate content that infringes on the copyright of others and/or not properly attributing or including an appropriate credit for its original creator.
  6. **Data security:** GAI tools can be used to generate content that contains vulnerabilities, malware or which can otherwise expose the Company, its customers or any other third parties to data security issues. For instance:
     1. GAI tools can be used to generate fake content. Such content can be used to spread phishing attacks, or even impersonate real people. For example, GAI tools could be used to create fake login pages or phishing emails, used to collect personal information, such as passwords and credit card numbers.
     2. GAI tools can be used to create malware. For example, a GAI tool could be used to create fake software programs or websites. These fake programs and websites could be used to infect computers with malware, such as viruses and ransomware.
     3. GAI tools can also be used to scan for code or systems for vulnerabilities.
     4. Code generated by GAI tools may contain malware or vulnerabilities, which could potentially contaminate the Company's code base.
  7. **Open-Source infringement and contamination:** GAI tools are often trained on large datasets of text and code. If these datasets contain open-source copyrighted material (whether permissive or non-permissive), the GAI tools may generate content that infringes on the underlaying source-code copyright attribution and/or copyleft requirements.

1. **Risk Mitigation Measures**
   1. **General Mitigations**:
      1. **Keep a human in the loop**. This means:
         1. Before any generated content is released to the public, it should be reviewed by a human to ensure that it is accurate, safe, and appropriate. This is a critical step in ensuring the accuracy, safety, and appropriateness of generated content. Humans can identify and correct errors in generated content, as well as flag content that is harmful, illegal, or discriminatory.
         2. The training and operation of generative AI models should be overseen by humans to ensure that they are not being used for malicious purposes and are not incorporating any bias, discrimination or any other wrongful basis. Humans can monitor the training and operation of these models to ensure that they are not being used to generate harmful or discriminatory content.
         3. GAI model is generating content that is harmful, illegal, or discriminatory, a human should be able to intervene and stop the process.
      2. **Risk Assessment Process**:

[AI Steering Committee] [compliance team] [legal team] [other- please specify] shall establish a risk assessment process and procedure to be used internally before embarking on any GAI development, deployment and/or use project by the Company.

Risk assessments should be based on the following principles:

* + - 1. All GAI tools and systems and tools developed, deployed and/or used by the Company should be first subject to assessment.
      2. Risk assessments should consider the potential for the GAI system to cause harm to people, property, or the environment.
      3. The Company should only use GAI tools and systems that meet the required safety and performance standards set under any applicable law.
      4. Records of the development, deployment and/or use of GAI tools and systems are to be kept by the Company. These records should include information about the risks identified as part of any assessment, the mitigation measures implemented and any other relevant information.
      5. Any major incidents, have the potential to cause harm to people, property, or the environment, involving GAI tools and systems should be reported by the Company to the relevant authorities according to any applicable law.
  1. **Mitigations - Risk to Confidential Information**:
     1. Check carefully each content you intend to share with any GAI tool prior to sharing it. Remember that confidentiality could be learnt from a context of any information, and not necessarily appear explicitly.
     2. Monitor any output generated by GAI tools for any content that may be confidential to the Company and/or third parties.
     3. If you want to share any confidential data with any GAI tool, first consult with the [AI Steering Committee] [compliance team] [legal team] [other- please specify]. If you receive approval to share such data, ensure that all personal data used with GAI tools is properly authorized and compliant with applicable data protection laws and regulations.
  2. **Mitigations - Data Privacy Risk**:
     1. Carefully check each piece of data you intend to share with any GAI tool prior to sharing it. Remember that personal data is usually defined as any information relating to an identified or identifiable natural person, not necessarily highly sensitive in nature.
     2. If you want to share any personal data with any GAI tool, first consult with the [AI Steering Committee] [compliance team] [legal team] [other- please specify]. If you receive approval to share such data, ensure that all personal data used with GAI tools is properly authorized and compliant with applicable data protection laws and regulations.
  3. **Mitigations – Mis/Disinformation**:
     1. Implement quality control measures specified in Section ‎7.1 above.
     2. Use cross-validation with multiple GAI tools to identify and address any false or misleading content generated by GAI tools.
  4. **Mitigations - Ethical Risks**:
     1. **Mitigations - Discrimination**
        1. In case the Company develops any GAI tool or incorporates such tools into its systems or products, it must ensure that the training data and algorithms used in such GAI tools systems are fair, unbiased, and inclusive of diverse perspectives. This should be done by performing regular audits and evaluations of such GAI tools and systems to identify and mitigate potential sources of bias or discrimination.
        2. If any GAI tools and systems are developed by the Company, it should involve diverse teams of experts and stakeholders, including those with lived experience of discrimination, in the design and development of its GAI tools and systems to help identify and address potential biases.
     2. **Mitigations - Transparency and Accountability**
        1. When providing any GAI tool to Company's employees, contractors, customer and/or to any third party, clearly communicate the purpose and intended outcomes of its GAI tools or systems to all stakeholders, including end-users, customers, and regulators.
        2. If any GAI tools and systems are developed by the Company, it should design such GAI tools and systems with transparency in mind, such that its decision-making processes and underlying algorithms can be easily understood by users and audited.
        3. The Company shall, after consulting with its consult with the [AI Steering Committee] [compliance team] [legal team] [other- please specify], establish clear accountability frameworks for the use of GAI tools, including defining roles and responsibilities for decision-making and ensuring appropriate oversight and governance.
     3. **Mitigations - Explainability**
        1. The Company should design its GAI tools or systems to provide explanations for its decisions and outputs to the extent technically possible.
        2. When using third party GAI tools or systems, the Company should prioritize the use of GAI tools or systems that provide interpretable and explainable outputs, even if this comes at the expense of accuracy or complexity.
     4. **Mitigations - Automated Decision Making**
        1. Conduct risk assessments process before deploying GAI tools or systems for automated decision-making, to identify and mitigate potential harms.
        2. Ensure that the use of generative AI tools for automated decision-making does not infringe on individuals' rights, freedoms, or privacy.
        3. Design the Company's GAI tools or systems to support human decision-making rather than replace it entirely, and ensure that human oversight and intervention is available when needed.
  5. **Mitigations - Intellectual Property Infringement**:
     1. **General Mitigations**
        1. In case the Company is developing any GAI tools or systems, it should conduct a thorough review of the ownership and licensing rights of any training data used in its GAI tools or systems to ensure that it have the appropriate permissions and authorizations. The Company should only use properly licensed datasets.
        2. Use GAI tools or systems only according to its license terms, and pay specific attention to any flow-down terms related to any data used to train these tools.
     2. **Mitigations - Copyright and Trademark Infringement**
        1. Ensure that any GAI tools or systems developed, deployed and/or used are not used to create content that is substantially similar to existing copyrighted works or trademarks.
        2. The Company shall, after consulting with its consult with the [AI Steering Committee] [compliance team] [legal team] [other- please specify], shall conduct regular audits and evaluations of its GAI tools or systems to identify any potential copyright or trademark infringement risks.
     3. **Mitigations - Attribution and Ownership**
        1. Ensure that appropriate permissions and licenses are obtained for the use and distribution of any GAI-generated content, including any third-party intellectual property rights that may be implicated.
        2. If any dedicated datasets were used for training any GAI tool or system developed, deployed and/or used by the Company, it should strictly comply with any attribution requirements related to the use of any such tool and/or system.
  6. **Mitigations - Data security risks:** 
     1. Any use of GAI tools or systems should be made in accordance with the Company's data security policies.
     2. The Company should conduct regular vulnerability scans of its GAI tools or systems to identify and address any potential security vulnerabilities.
     3. The Company shall regularly update GAI tools with security patches and address potential security weaknesses found.
     4. Ensure that appropriate due diligence is conducted on any third-party data sources used in our GAI tools or systems to assess the security risks associated with their use.
     5. Regularly review and update our agreements with GAI tools and/or datasets providers to ensure that they include appropriate data security protections.
  7. **Mitigations - Open-Source Infringement and Contamination**:
     1. Carefully review and validate any datasets used to train GAI tools to ensure they are free from copyrighted material or malware.
     2. If you are using GAI tools which are aimed to generate code - Don’t use too long pieces of code and make sure you edit the code generated. Don't use it as is.
     3. Use market standard scanning tools which can identify pieces of open-source code in the Company's code base, to avoid any unintentional inclusion of such code generated by GAI tools as part of the Company's proprietary code.

1. **Periodical training sessions**

The [AI Steering Committee] [compliance team] [legal team] [other- please specify] will provide periodical training sessions to all employees on the use of GAI tools or systems. These training sessions will cover the following topics:

* 1. The benefits and risks of using GAI tools or systems;
  2. How to use GAI tools or systems safely and responsibly;
  3. How to identify and avoid potential misuse of GAI tools or systems.

1. **Compliance with existing and new regulations**

The [AI Steering Committee] [compliance team] [legal team] [other- please specify] will provide periodical review of existing and new regulations and legal requirements to ensure the Company's complying with all applicable laws and regulations, including, inter-alia, the EU AI Act, the China AI regulation and FTC requirements in this field. The Company will make all necessary changes to this policy and its procedures as needed to ensure compliance.