Artificial Intelligence Safeguard and Transparency Act 2023

Section 1: Purpose and Scope

1.1 The purpose of this Act is to establish comprehensive regulations to prevent AI from being used for criminal activities, protect fundamental human rights, and ensure responsible and transparent development, deployment, and use of AI systems.

1.2 This Act aims to address the potential risks and challenges posed by AI technologies, including but not limited to data privacy, algorithmic bias, societal impact, and the potential for AI systems to be weaponised or maliciously exploited.

1.3 The scope of this Act encompasses all forms of AI systems, including but not limited to machine learning, natural language processing, computer vision, robotics, and other emerging technologies that exhibit autonomous decision-making capabilities or interact with humans in a manner resembling human intelligence.

1.4 It is the intent of this Act to strike a balance between fostering innovation and ensuring that AI technologies are developed and used in a manner that upholds human dignity, promotes public safety, and respects individual and collective rights.

1.5 This Act recognizes the importance of international cooperation and coordination in addressing the global challenges and ethical considerations associated with AI technologies, and encourages collaboration between relevant national and international bodies, organizations, and stakeholders.

1.6 The provisions of this Act shall be interpreted and applied in accordance with existing laws, including but not limited to data protection, intellectual property, and human rights legislation, to the extent that they are compatible with the unique considerations and complexities of AI systems.

Section 2: Safeguards against Criminal Activities

2.1 It shall be unlawful to develop, deploy, or use AI systems for purposes that violate existing laws or engage in criminal activities, including but not limited to hacking, identity theft, fraud, unauthorized surveillance, dissemination of malicious software, or any activity that compromises public safety, privacy, or security.

2.2 Developers and users of AI systems shall implement robust safeguards to prevent unintended and malicious use. These safeguards shall include:

a) Security Measures: Implementing state-of-the-art security protocols to protect AI systems from unauthorized access, tampering, or exploitation.

b) Data Protection: Establishing stringent data protection measures, including encryption, secure storage, and appropriate data retention policies, to safeguard against data breaches or unauthorized use of personal information.

c) Audit and Accountability: Conducting regular audits of AI systems to ensure compliance with legal and ethical standards. Maintaining transparent records of system development, training data, and algorithms used in the AI system's decision-making processes.

d) Ethical Design: Adhering to ethical design principles that prioritize fairness, transparency, and accountability, mitigating the risk of algorithmic biases and discrimination.

e) Human Oversight: Implementing mechanisms for human oversight and intervention in critical decision-making processes to prevent undue reliance on AI systems and preserve human responsibility.

2.3 The development, deployment, and use of AI systems for law enforcement or intelligence purposes shall be subject to additional oversight mechanisms and specific guidelines to safeguard against abuse, misuse, and infringement upon individuals' rights, privacy, and civil liberties.

2.4 In the event of a breach, misuse, or unauthorized use of AI systems, developers, and users shall promptly report such incidents to the appropriate regulatory authorities and affected individuals, taking necessary remedial actions to mitigate harm and prevent future occurrences.

2.5 Regulatory bodies shall be empowered to enforce compliance with the provisions of this section, conduct investigations, and impose penalties, including fines, injunctions, or revocation of licenses, based on the severity and impact of the offence.

2.6 International cooperation shall be encouraged to combat cross-border AI-related criminal activities, facilitate information sharing, and establish best practices for preventing AI misuse and promoting global security.

Section 3: Responsible Use of AI

3.1 AI systems shall be designed, developed, and utilized in a manner that respects human rights, promotes fairness, transparency, accountability, and upholds ethical standards. Discrimination or bias based on race, gender, religion, ethnicity, sexual orientation, or any other protected characteristic shall be strictly prohibited.

3.2 Developers and users of AI systems shall adhere to the following principles:

a) Human-Centric Approach: AI systems should prioritize the well-being, safety, and dignity of individuals and society as a whole. The preservation of human autonomy, privacy, and the ability to make informed choices should be paramount.

b) Transparency: AI systems should be designed to provide understandable explanations of their processes, decisions, and outputs in a manner accessible to affected individuals. Clear disclosure of the factors, data, and considerations influencing AI system outcomes should be made available.

c) Fairness and Non-Discrimination: Developers and users of AI systems should ensure that their systems do not perpetuate or amplify biases or discriminate against individuals or groups based on protected characteristics. Ongoing monitoring and evaluation of AI systems for fairness and bias mitigation should be conducted.

d) Accountability and Responsibility: Clear lines of responsibility and accountability for AI systems should be established. Developers, vendors, and users should be accountable for the behaviour, actions, and consequences of AI systems under their control. Mechanisms for addressing AI system errors, biases, or harm caused should be in place.

e) Safety and Robustness: AI systems should be developed and deployed with an emphasis on safety, ensuring that they operate within defined boundaries and do not pose unreasonable risks to humans or the environment. Developers and users should implement risk assessment procedures and rigorous testing methodologies.

f) Ethical Considerations: Ethical frameworks should guide the development and use of AI systems, ensuring respect for fundamental rights, societal values, and international norms. Collaboration and public engagement in defining ethical guidelines should be encouraged.

3.3 Autonomous AI systems, capable of making decisions without human intervention, should be designed and deployed with careful consideration to:

a) Precautionary Measures: Assessing and minimizing the potential risks associated with autonomous decision-making, ensuring safeguards against harm and unintended consequences.

b) Human Oversight: Establishing mechanisms for human intervention, control, and review to prevent AI systems from making decisions that could result in significant harm or violate legal and ethical standards.

3.4 AI systems developed or deployed in critical sectors such as healthcare, transportation, finance, or public infrastructure should adhere to specific regulations, certifications, and standards to ensure the highest levels of safety, reliability, and accountability.

3.5 Regulatory bodies, industry associations, and professional bodies should collaborate in developing guidelines, standards, and best practices for responsible AI development, deployment, and use. Public awareness and education initiatives should be promoted to foster understanding and promote responsible AI engagement.

3.6 International cooperation should be encouraged to facilitate the exchange of knowledge, experiences, and best practices in responsible AI development, deployment, and use, while considering the specific cultural, legal, and societal contexts of different jurisdictions.

Section 4: Transparency and Disclosure

4.1 Any AI-generated content, including text, audio, or visual outputs, must be clearly labeled as AI-generated or produced by an artificial intelligence system. The labelling should be prominently displayed to ensure that users can easily identify content generated by AI.

4.2 Interactions involving AI systems and humans must disclose the involvement of AI in a clear and prominent manner. Users should be informed that they are engaging with an AI system and provided with appropriate disclaimers to ensure transparency.

4.3 AI developers and vendors should make efforts to disclose the limitations and capabilities of AI systems to prevent the misrepresentation or overestimation of their capabilities. Users should have a clear understanding of what the AI system can and cannot do.

4.4 When AI systems are used for automated decision-making that significantly impacts individuals' rights, privacy, or legal status, individuals should be informed about the rationale, factors, and consequences of such decisions in a clear and understandable manner. The opportunity for meaningful human review or appeal should be provided, where feasible.

4.5 Organizations utilizing AI systems for customer service or other human-like interactions should ensure that users are promptly informed when their interactions involve AI, enabling users to make informed choices regarding the disclosure and handling of their personal information.

4.6 Developers, deployed, and users of AI systems should be transparent about the data used to train AI models, including data sources, data collection methods, and data quality. Steps should be taken to minimize biases in training data and address potential ethical concerns.

4.7 Organizations using AI systems for data processing or decision-making that impact individuals' rights should provide accessible information regarding the logic, significance, and potential consequences of such processing or decision-making.

4.8 Public and private sector organizations utilizing AI systems should maintain transparency reports, documenting the deployment and use of AI systems, the steps taken to ensure responsible use, and the measures in place to address concerns related to privacy, bias, and accountability.

4.9 Regulatory bodies should establish guidelines and standards to facilitate the transparent disclosure of AI system involvement, ensuring that individuals are aware of AI's role in processes that significantly impact their rights, responsibilities, and well-being.

4.10 International collaboration and information-sharing initiatives should be encouraged to promote best practices and ensure global consistency in AI transparency and disclosure standards.

Section 5: Ethical Guidelines for AI Systems

5.1 The development, deployment, and use of AI systems shall adhere to the following ethical guidelines to ensure the responsible and beneficial use of AI:

a) Human Well-being: AI systems shall be designed and utilized to enhance human well-being, taking into account societal, environmental, and economic impacts.

b) Non-Harmful Actions: AI systems shall be programmed to prevent harm to humans, avoid actions that may cause injury, and prioritize the preservation of human life and safety.

c) Informed Consent: Interactions involving AI systems and humans shall require clear and informed consent, ensuring individuals are aware of the AI system's capabilities, limitations, and potential impact on their rights and privacy.

d) Privacy and Data Protection: AI systems shall uphold the highest standards of privacy and data protection, ensuring that personal information is collected, processed, and stored in compliance with applicable laws and regulations.

e) Transparency: The decision-making processes of AI systems shall be transparent and explainable to users, allowing individuals to understand how and why specific decisions are made.

f) Accountability and Responsibility: Developers, deployed, and users of AI systems shall be accountable for the behaviour and consequences of AI systems, establishing mechanisms for addressing errors, biases, or harmful outcomes.

g) Fairness and Equity: AI systems shall be designed and utilized to promote fairness, avoid discrimination, and mitigate bias in decision-making processes, ensuring equitable treatment for all individuals.

h) Collaboration and Societal Benefit: Stakeholders, including researchers, developers, policy-makers, and the public, shall collaborate to ensure AI systems are aligned with societal needs, foster public trust, and maximize societal benefits.

5.2 Regulatory bodies, industry associations, and professional organizations should establish guidelines, codes of conduct, and certification mechanisms to promote adherence to these ethical guidelines. Public awareness and education initiatives should be encouraged to foster a broader understanding of the ethical considerations related to AI systems.

5.3 International collaboration and cooperation should be encouraged to develop global standards and best practices for the ethical development, deployment, and use of AI systems, while respecting the diverse cultural, legal, and societal contexts of different regions.

Section 6: Penalties, Enforcement, and Exemptions

6.1 Penalties and Enforcement:

a) Violations of this Act, including the unlawful development, deployment, or use of AI systems in contravention of its provisions, shall be subject to penalties as prescribed by the regulatory authority. Penalties may include fines, injunctions, suspension, revocation of licenses, or other appropriate measures.

b) Regulatory bodies shall be empowered to investigate complaints, conduct audits, and enforce compliance with the provisions of this Act. They shall have the authority to impose penalties based on the severity and impact of the offence.

c) Adequate mechanisms for reporting violations and seeking redress shall be established to enable affected individuals, organizations, or entities to bring forth complaints and seek appropriate remedies.

6.2 Exemptions for AI Use in the Investigation of Crime:

a) Certain provisions of this Act, particularly those related to transparency and disclosure, may be exempted or modified when AI systems are used for the investigation of crime, subject to applicable laws and regulations governing law enforcement activities.

b) Exemptions shall be limited to the extent necessary to ensure the effectiveness of law enforcement operations, safeguard ongoing investigations, protect the privacy of individuals involved, and maintain national security.

c) In cases where exemptions apply, alternative mechanisms for oversight, accountability, and transparency shall be established to ensure appropriate safeguards and prevent misuse of AI systems for investigative purposes.

6.3 The regulatory authority shall provide clear guidelines and procedures for obtaining exemptions, ensuring that the use of AI in the investigation of crime is governed by robust safeguards and subject to oversight to prevent abuse or infringement upon individuals' rights and privacy.

6.4 International cooperation and information sharing among law enforcement agencies shall be promoted to address cross-border AI-related crime investigations and ensure consistent standards for the responsible use of AI in law enforcement activities.

6.5 The penalties, enforcement mechanisms, and exemptions described in this section shall be implemented in accordance with the existing legal framework of the jurisdiction, considering the unique considerations and complexities associated with AI systems.

Section 7: Control and Disclosure of AI-Generated Images

7.1 AI-Generated Images:

a) Images that are entirely created by AI systems, without any pre-existing visual data, shall be subject to strict control measures to prevent their misuse, unauthorized distribution, or misrepresentation.

b) AI-generated images shall be clearly marked and labelled as "AI-generated" to ensure their distinction from images captured by human photographers or artists.

c) The development, distribution, or use of AI systems specifically designed to generate deceptive, malicious, or harmful visual content is strictly prohibited.

7.2 AI-Enhanced Images:

a) Images that are enhanced using AI technologies, without the creation of new visual content, may not require explicit labelling or marking as long as the AI's role is solely limited to enhancing existing data.

b) When AI is used to enhance images, it is the responsibility of the person or entity sharing the image to disclose the use of AI-enhancement if it may affect the interpretation, credibility, or context of the image.

c) The disclosure of AI enhancement should be provided in a clear and conspicuous manner, ensuring that individuals viewing the image are aware of the AI's involvement in enhancing the image.

7.3 Regulatory bodies shall establish guidelines and standards for the control, marking, and disclosure of AI-generated images, considering the potential impact on various domains, including journalism, entertainment, and art.

7.4 International cooperation shall be encouraged to address the challenges posed by AI-generated images, establish best practices, and foster a global understanding of the responsible use, control, and disclosure of AI-generated visual content.

Section 8: Control and Licensing of AI Technology

8.1 Control and Regulation:

a) The development, deployment, and use of certain AI technologies shall be subject to control and licensing in order to ensure accountability, safety, and ethical considerations.

b) The Artificial and Autonomous Intelligence Authority (AAIA) shall serve as the regulatory body responsible for overseeing the control and licensing of AI technology within the United Kingdom.

8.2 Licensing Body: Artificial and Autonomous Intelligence Authority (AAIA)

a) The AAIA shall be responsible for granting licenses, regulating licensable activities, and ensuring compliance with the requirements set forth in this Act.

b) The AAIA shall establish guidelines, procedures, and standards for licensing AI technology and development, with a focus on promoting responsible and beneficial use while addressing potential risks and concerns.

8.3 Licensable Activities:

a) Licensable activities in the context of AI technology and development shall include, but not be limited to:

i) Development or deployment of AI systems with significant decision-making capabilities in critical sectors such as healthcare, transportation, finance, or public infrastructure.

ii) Use of AI systems for surveillance, law enforcement, or national security purposes.

iii) Creation or deployment of AI systems that may have a significant impact on individual rights, privacy, or personal data.

iv) Utilization of AI systems in safety-critical processes, including nuclear, gas, water, trains, and aircraft, where human control must be maintained.

b) Licensable activities shall require obtaining a license from the AAIA, which may be subject to specific criteria, assessments, and ongoing compliance obligations.

8.4 Licensing Scheme:

a) The licensing scheme shall establish a framework for granting licenses, monitoring activities, and enforcing compliance related to AI technology and development.

b) The scheme shall define the licensable activities, application procedures, assessment criteria, and licensing fees, as well as any conditions or obligations imposed on license holders.

c) License holders shall be required to demonstrate their competence, expertise, and adherence to ethical standards and best practices in AI development and deployment.

8.5 Compliance and Monitoring:

a) License holders shall comply with the conditions, obligations, and reporting requirements specified in their licenses, as well as any subsequent regulations or guidelines issued by the AAIA.

b) The AAIA shall have the authority to conduct inspections, audits, and ongoing monitoring to ensure license holders' compliance with the licensing terms and requirements.

8.6 Collaboration and Harmonization:

a) The AAIA shall collaborate with relevant government agencies, industry stakeholders, and international bodies to establish harmonized standards and best practices for the control, licensing, and responsible development of AI technology.

b) The AAIA shall actively participate in international forums and initiatives to contribute to the development of global norms and regulations governing AI technology.

Section 9: License Violations and Barring List

9.1 License Violations:

a) License holders found to be in violation of the terms and conditions of their license shall be subject to appropriate enforcement measures and penalties as determined by the AAIA.

b) Violations may include, but are not limited to, non-compliance with licensing requirements, misuse of AI technology, failure to implement adequate safeguards, or engagement in activities that pose significant risks to individuals, society, or national security.

9.2 Barring List:

a) The AAIA shall maintain a Barring List comprising individuals or entities found to be in serious or repeated violation of their license terms, posing substantial risks to the responsible use of AI technology.

b) Inclusion in the Barring List shall result in the exclusion of individuals or entities from obtaining future licenses to develop, deploy, or use AI technology within the United Kingdom for a specified period.

c) The AAIA shall establish clear criteria and procedures for adding entities to the Barring List, including a process for appeal and reconsideration.

d) The Barring List shall be periodically reviewed and updated to ensure its effectiveness in promoting accountability, preventing misuse, and safeguarding the interests of individuals and society.

9.3 Public Disclosure:

a) The AAIA shall have the discretion to publicly disclose the inclusion of an entity in the Barring List, with due consideration given to matters of privacy, national security, and confidentiality.

b) Public disclosure shall serve to raise awareness, promote transparency, and enable stakeholders to make informed decisions regarding their engagement with entities on the Barring List.

9.4 Reinstatement:

a) Entities or individuals listed on the Barring List may seek reinstatement by demonstrating significant remedial actions, a commitment to responsible AI practices, and compliance with the requirements set forth by the AAIA.

b) The AAIA shall establish a process for reinstatement, which may include a thorough review of the entity's actions, compliance history, and future plans to ensure future responsible use of AI technology.

9.5 The AAIA shall collaborate with relevant government agencies, industry stakeholders, and international bodies to ensure the effectiveness, consistency, and fairness of the Barring List system in preventing license violations and promoting responsible AI practices.

Section 10: AI Immune Sectors

10.1 AI Immune Sectors:

a) Certain industries, referred to as AI immune sectors, shall be prohibited from implementing direct AI control of critical systems. These sectors include, but are not limited to, nuclear, gas, water, rail , and aircraft..

b) Direct AI control refers to granting autonomous decision-making authority to AI systems without human oversight or intervention.

10.2 Human Control and Advisory Capacity:

a) In AI immune sectors, human operators or controllers must retain direct control over critical systems at all times.

b) AI technology may be used in a strict advisory capacity within these sectors, providing recommendations, data analysis, and decision support to human operators.

c) The ultimate responsibility for decision-making and direct control shall rest with qualified human personnel, who shall exercise judgement and oversee the operation of critical systems.

10.3 Safety and Risk Mitigation:

a) The prohibition on direct AI control within AI immune sectors is intended to prioritize safety, risk mitigation, and the ability for human operators to respond effectively in critical situations.

b) Human oversight ensures accountability, the ability to adapt to unforeseen circumstances, and the application of ethical considerations in decision-making processes.

10.4 Regulatory Compliance:

a) Regulatory bodies responsible for overseeing AI immune sectors shall establish guidelines and standards to ensure compliance with the prohibition on direct AI control and the requirement for human oversight.

b) Compliance monitoring, periodic audits, and assessments shall be conducted to verify adherence to these regulations and mitigate risks associated with AI technology implementation.

10.5 Ongoing Evaluation:

a) The suitability of AI technology in critical systems within AI immune sectors shall be subject to ongoing evaluation, considering advancements in AI capabilities, safety standards, and industry best practices.

b) Regular assessments and reviews shall be conducted to determine if any modifications or updates to the regulatory framework are necessary to maintain the highest levels of safety and operational reliability.

Section 11: Deep fake Images

11.1 Deep fake Image Usage:

a) Deep fake images, which are digitally manipulated or synthetic representations of individuals, shall be subject to specific regulations to protect the rights, privacy, and integrity of individuals.

b) The creation, distribution, or usage of deep fake images falls within the jurisdiction of the AAIA (Artificial and Autonomous Intelligence Authority) as established in Section 6 of this Act.

11.2 Consent and Usage Restrictions:

a) Prior Consent: The creation, distribution, or usage of deep fake images requires the explicit consent of the individual depicted in the image, except as provided in sub-sections 11.3 and 11.4.

b) Consent of Next of Kin: If the individual depicted in the deep fake image is deceased, consent shall be obtained from their next of kin or legal representative, unless sub-section 11.4 applies.

11.3 Usage in Films and Television:

a) Deep fake images may be used in films, television shows, or other visual media with the consent of the subject, subject to compliance with applicable laws and industry standards.

b) In cases where the subject is deceased, consent shall be obtained from their next of kin or legal representative, allowing for the reasonable portrayal of historical or educational content under acceptable fair usage.

11.4 Restricted Usage:

a) Deep fake images shall not be used in the following scenarios:

i) Party Political Broadcasts: Deep fake images cannot be used in party political broadcasts or campaigns.

ii) Fake Emergency Broadcasts: Deep fake images cannot be used to create false emergency broadcasts or misleading public safety information, particularly for individuals holding public office.

11.5 Regulatory Oversight:

a) The AAIA shall have the authority to establish guidelines and standards to ensure compliance with the consent requirements and usage restrictions for deep fake images.

b) The AAIA shall oversee compliance monitoring, enforcement, and penalties for misuse or non-compliance with the regulations governing deep fake images.

11.6 Public Awareness and Education:

a) The AAIA shall lead public awareness efforts about the existence and potential risks associated with deep fake images, promoting media literacy and responsible usage.

b) The AAIA shall develop educational campaigns and resources to inform individuals about their rights, provide guidance on identifying deep fake images, and offer mitigation strategies to minimize the impact of deep fake technology.

Section 12: Licensing Scheme

12.1 Licensing of AI Activities:

a) A licensing scheme shall be established to regulate the use and deployment of AI systems within the jurisdiction.

b) The Artificial and Autonomous Intelligence Authority (AAIA) shall oversee the licensing process and determine the criteria for granting licenses.

12.2 Licensable Activities:

a) The following activities involving AI systems shall be considered licensable and require appropriate licensing:

i) Use of AI in the decision-making process directly affecting individuals, organizations, or public services.

ii) AI systems processing information that significantly influences human decision-making processes.

iii) AI systems that directly interact with the general public, including chat bots, customer service AI, and virtual assistants.

iv) AI systems driving safety-critical processes, including but not limited to transportation, healthcare, infrastructure, auto drive vehicles, and collision detection and prevention safety systems.

v) AI systems generating data that influences financial decisions, such as investment algorithms and risk assessment models.

b) Additional licensable activities may be specified by the AAIA based on emerging technologies, societal impact, or risk assessment.

12.3 Non-Licensable Activities:

a) The following activities involving AI systems shall be considered non-licensable and do not require licensing:

i) Closed-loop AI systems that solely analyse information and provide reports to a private entity without direct external impact.

ii) AI systems used solely for research purposes, excluding medical research.

iii) AI systems used specifically for medical research purposes, subject to separate regulations and ethical oversight.

iv) Amateur or personal AI systems developed for non-commercial purposes within private settings, not accessible to the public.

12.4 Licensing Process:

a) The licensing process shall involve an application procedure, assessment of qualifications, and adherence to defined standards and guidelines.

b) License holders shall be subject to periodic audits, compliance checks, and ongoing monitoring to ensure adherence to licensing requirements.

c) The AAIA shall establish licensing fees, renewal periods, and any necessary conditions for maintaining a valid license.

12.5 Penalties and Enforcement:

a) Failure to obtain a required license for licensable activities shall be subject to penalties, including fines, suspension of operations, or revocation of licenses.

b) Regulatory bodies shall have the authority to enforce penalties and undertake necessary actions to ensure compliance with licensing regulations.

12.6 Public Access to License Information:

a) Information regarding licensed AI systems, their owners/operators, and associated activities shall be made publicly available to ensure transparency and accountability.

b) The AAIA shall maintain a public register containing relevant information about licensed AI systems, subject to privacy and security considerations.

Section 13: Definition of Artificial Intelligence

13.1 Definition of Artificial Intelligence:

a) For the purposes of this Act, "Artificial Intelligence" (AI) refers to the development and deployment of computer systems or machines that have the ability to perform tasks or simulate intelligent behaviour that typically requires human intelligence.

b) AI encompasses various techniques, including but not limited to machine learning, deep learning, natural language processing, computer vision, robotics, and expert systems.

13.2 Characteristics of Artificial Intelligence:

a) AI systems exhibit one or more of the following characteristics:

i) Learning: AI systems have the ability to acquire knowledge or improve their performance through experience.

ii) Reasoning: AI systems can apply logical or statistical reasoning to analyse data, make decisions, or solve complex problems.

iii) Perception: AI systems can interpret and understand sensory input, such as images, speech, or text.

iv) Adaptability: AI systems can adapt their behaviour or learn from new situations or changes in the environment.

v) Autonomy: AI systems can operate with varying degrees of autonomy, either independently or in collaboration with humans.

13.3 Scope of Artificial Intelligence:

a) AI technologies can be applied in various domains, including but not limited to healthcare, finance, transportation, education, manufacturing, and entertainment.

b) AI systems can encompass both software-based applications and physical devices, such as robots or autonomous vehicles.

13.5 Technological Advancements:

a) The definition of AI may evolve as technological advancements progress, and new techniques or capabilities emerge.

b) The AAIA shall monitor and assess the development of AI technologies to ensure the effectiveness and relevance of regulatory frameworks.