

Guide on

Artificial Intelligence (AI) Governance and Ethics for Brunei Darussalam





Contents

Introduction	2
Definitions & Terminology	5
Purpose & Scope	8
* Al System Lifecycle	11
Guiding Principles on Al Governance and Ethics	14
Al Governance Framework	28

Introduction



Artificial Intelligence (AI) is rapidly transforming our world, with capabilities which can influence how organisations operate and offer their products and services, enhancing business processes as well as personalising users' experiences. Al encompasses range technologies that enable machines to simulate human intelligence, including machine learning, deep learning, and natural language processing. These technologies are revolutionising multiple industries such as energy, telecommunications, healthcare. finance, education. transportation, information and communications technology (ICT) services, etc.

The trend in AI development is clear – AI is becoming increasingly sophisticated and versatile, with more powerful algorithms and ever-growing datasets, leading to AI systems that can handle complex tasks and make informed decisions, whilst offering convenience to its users.



However, it is crucial to acknowledge that alongside the immense potential of AI lies a complex threat landscape, such as the rise of misuse for disinformation, biases in AI algorithms, implications of job displacement and other associated security risks. Proper safeguards are thereby essential to prevent misuse and exploitation of these AI products and services. These considerations are paramount as we move towards the journey of a Smart Nation and the Brunei Vision 2035.

In order to shape a digitally inclusive future for Brunei Darussalam, it is more than about designing, developing, deploying and using emerging technologies, but also leveraging the power of these technologies responsibly and ethically, with appropriate cultural and religious considerations as well as alignment with the principles of Melayu Islam Beraja (MIB) and Maqasid of Shariah. Such considerations play an important role in the efforts to build a safe and trusted digital ecosystem in Brunei Darussalam.

Definitions and Terminology

This section will include relevant definitions for the purposes of the Guide on AI Governance and Ethics for Brunei Darussalam ("the Guide").



Artificial Intelligence (AI)

Refers to the discipline of making analytical machines intelligent, enabling an organisation to function appropriately and with foresight.

Source: Association of Southeast Asian Nations (ASEAN) Guide on Al Governance and Ethics

AI Systems

Refers to a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives. It uses machine and/or human-based data and inputs to:

- (i) perceive real and/or virtual environments:
- (ii) abstract these perceptions into models through analysis in an automated manner (e.g. with machine learning), or manually; and
- (iii) use model inference to formulate options for outcomes.

Al systems are designed to operate with varying levels of autonomy.

Source: ASEAN Guide on Al Governance and Ethics

Deployer

Refers to an entity that uses or implements an AI system, which could either be developed by their in-house team or via a third-party developer.

Source: ASEAN Guide on Al Governance and Ethics

Developer

Refers to an entity that designs, codes, or produces an AI system.

Source: ASEAN Guide on Al Governance and Ethics

Human-in-the-loop

Suggests that human oversight is active and involved, with the human retaining full control and the AI only providing recommendations or input. Decisions cannot be exercised without affirmative actions by the human, such as human command to proceed with a given decision.

Source: Infocomm Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC) Singapore's Model Al Governance Framework, Second Edition

Human-out-of-the-loop

Suggests that there is no human oversight over the execution of decisions. The AI system has full control without the option of human override.

Source: IMDA and PDPC Singapore's Model Al Governance Framework, Second Edition

Human-over-the-loop

Suggests that human oversight is involved to the extent that the human is in a monitoring or supervisory role, with the ability to take over control when the AI model encounters unexpected or undesirable events (such as model failure).

Source: IMDA and PDPC Singapore's Model AI Governance Framework, Second Edition

Machine Learning

Refers to a subfield in AI where algorithms learn by identifying patterns and correlations within data using statistical techniques to enhance performance, all without being explicitly programmed.

Source: ASEAN Guide on Al Governance and Ethics

User

Refers to an entity or person (internal or external) that interacts with an AI system or an AI-enabled service and can be affected by its decisions.

Source: ASEAN Guide on Al Governance and Ethics



Purpose & Scope



Audience

The Guide is intended for organisations which integrate or adopt AI solutions within existing or upcoming business and IT systems. It is intended for organisations that design, develop, deploy and/or use AI technologies in Brunei Darussalam. This includes organisations involved in various aspects of AI, including system designers, developers, researchers, providers and users of AI technologies as well as those hiring or outsourcing such services to third-party companies which may be located overseas.



Purpose

This document serves as a practical guide for governance and ethics in AI, through a principles-based approach suitable for various organisational contexts. It provides recommendations and guidance on the principles and safeguards which organisations should consider.

Objectives

The overall objective is to build trust among stakeholders in managing the risks related to AI technologies, through promoting governance and ethical practices as outlined in the guiding principles.

Scope

It is important to note that the Guide is developed to serve as a baseline set of considerations and safeguards in accordance with the following considerations:

Technology-agnostic

The Guide will not focus on specific systems or software and will apply regardless of any technological components within AI systems.

Sector-agnostic

The Guide will serve as a baseline for organisations operating in any sector to adopt. Specific sectors or organisations may choose to include additional considerations and safeguards to meet their needs.¹

Note

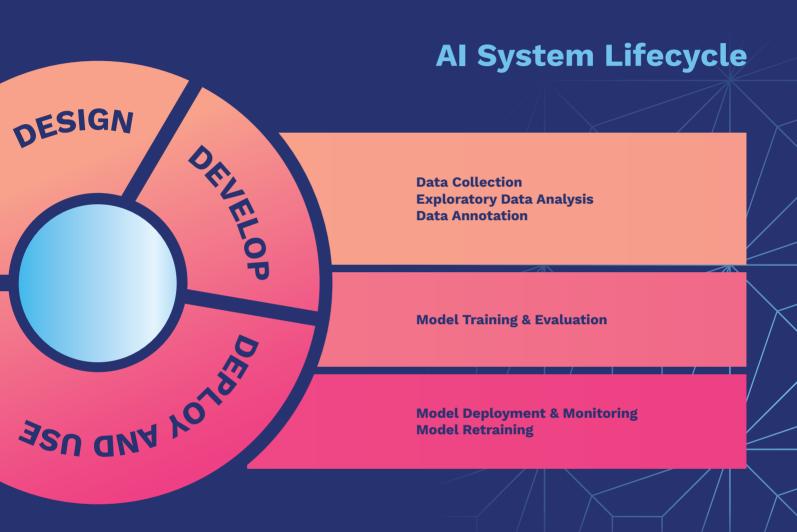
The Guide is a working document developed by Authority of Info-communications Technology Industry of Brunei Darussalam (AITI), in collaboration with the AI Governance and Ethics Working Group.² The Guide may be amended periodically with prior notice, to ensure the document remains relevant to emerging trends, practices and technologies. In essence, the Guide serves as a foundational resource for organisations in Brunei Darussalam to responsibly design, develop, deploy and/or use AI technologies whilst ensuring ethical considerations and risk management factors are adequately addressed.

- Sector-specific guides may be developed once the deployment and use of AI technologies are more prevalent and mature for organisations in Brunei Darussalam. Any existing and upcoming guides developed by respective sector agencies would act as a complement to this Guide.
- ² The members of the AI Governance and Ethics Working Group consists of representatives from organisations as indicated under page 36.

Al System Lifecycle



Understanding AI's potential and threats requires examining its development process, known as the AI system lifecycle. It is important to incorporate or align this lifecycle into organisation's system development and project management framework. This lifecycle refers to various stages that an AI system goes through and can be broadly divided into these key stages:



Design

Data Collection, Analysis and Annotation

This initial stage involves defining the system's purpose, its intended users, the amount of data, its use cases, secure data management practices³ and the ethical considerations. It encompasses the collection and processing of data to ensure data quality, data integrity, relevance, and absence of bias. At this stage, organisations should consider the sourcing of reliable data as well as engaging relevant and diverse stakeholders to ensure the system meets its main purpose.

³ This is in reference to any applicable standards, guidelines, policies and/or any relevant laws within Brunei Darussalam.

Development

Model Training and Evaluation

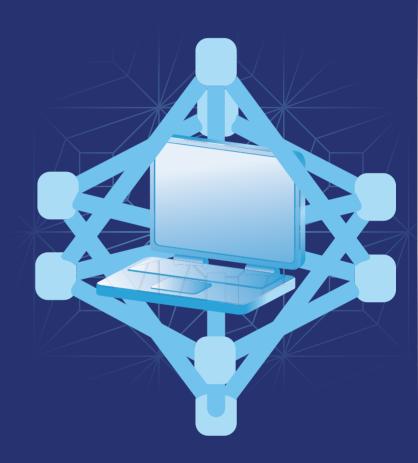
After model creation, rigorous testing is crucial. This stage ensures the AI system performs as intended, delivers accurate results and avoids unintended consequences. It also involves assessing the system's fairness⁴, robustness, and explainability which can provide humans with the understanding on how the AI system arrives at its decisions. To ensure consistency, organisations may want to continue to incorporate continuous feedback loop from the relevant and diverse stakeholders to refine and improve the AI model throughout its development.

Deployment and Use

Model Deployment and Monitoring

Once validated, the AI system is deployed in the real world. Continuous monitoring is essential to track the system's performance, identify any biases or errors that may emerge over time and ensure it adapts to changing circumstances.

At this stage, it may be essential to evaluate the safeguards which may include levels of human interventions and checks on the data inputs, algorithm and model outputs.



Model Retraining

Data gathered during monitoring is used to refine the AI system. This may involve retraining the model with new data or modifying the algorithms to address identified issues. This iterative process of feedback and improvement is crucial for maintaining the system's effectiveness and mitigating potential risks. Such periodical review is important to ensure that the system continues to be automated and adapts itself in accordance to ethical frameworks, laws and regulations.

⁴ For example, minimising bias in algorithm selection and training the AI model to use only high-quality or representative data.

Guiding Principles on Al Governance and Ethics



Guiding Principles

The seven (7) guiding principles⁵ below help to gain users' trust in AI systems and tools through guidance on how it should be designed, developed, deployed and used by organisations. The principles below take into consideration its impact to technology development, harnessing the foundation of a safe and trusted digital economy for Brunei Darussalam.

Transparency and Explainabilty

Data Protection and Data Governance

Security and Safety

Robustness and Reliability

Fairness and Equity

Human Centricity

Accountability and Integrity

⁵ The principles cite elements of the ASEAN Guide on AI Governance and Ethics.

Transparency and Explainability

Transparency involves disclosing to users that an AI system is being used and contributed in the decision-making, the types of data it uses, and its purpose. This ensures that users understand that such system is utilised, therefore enabling users to make an informed choice of whether to use the AI-enabled system.

Explainability refers to the ability of an AI system to convey the reasoning behind its decision and conclusion in a format that is simple and understandable. This allows users to understand the factors contributing to the AI system's recommendation.

It is worth noting that there may be situations where organisations may not be able to fully practice the above principles and certain exceptions may apply with justifications. For example, disclosing confidential or proprietary information at risk such as information containing organisation's Intellectual Property (IP).



Organisation A is a company that provides real-time predictive insights and offers its cost-predictor solutions to clients in the healthcare sector.

It practices the Transparency principle by disclosing to its clients the exact parameters used in developing their AI model and providing detailed explanation on all algorithms that had any foreseeable impact on its decision. Further, it practices the Explainability principle by training its AI model to explain each prediction result to its users without the need to go through Organisation A's support staff.

Data Protection and Data Governance

Al systems should have proper mechanisms in place in order to maintain and protect the quality and integrity of data throughout their entire lifecycle.

Data protection by design and privacy by design principles should be considered throughout the entire lifecycle of the AI system. The way data is collected, used, disclosed, processed, stored, generated, and deleted throughout the AI system lifecycle must comply with applicable laws and regulations, including Brunei Darussalam's upcoming Personal Data Protection law.



Organisations should be transparent about their data collection practices, including the types of data collected, where it is sourced, how it is used, who has access to it and whether such data will be disclosed and/or anonymised.

Organisations should ensure that necessary consent is obtained from individuals before collecting, using, or disclosing personal data for AI development and deployment, or otherwise have appropriate legal basis to collect, use or disclose personal data without consent. In addition, organisations should consider placing greater sensitivity in the design of AI systems, in particular where it collects, uses or discloses personal data of minors and vulnerable groups.

Data governance frameworks should also be considered and adhered to by designers, developers, deployers and users of AI systems and tools. Organisations should ensure a well-documented and effective audit trail. These frameworks should also be periodically reviewed and updated in accordance with relevant law and regulations.



Organisation B is a company that provides data analytics solutions to clients in the telecommunications sector.

It practices the Data Protection principles by ensuring that it anonymises its client data at source before using it for development. Furthermore, it built an AI model with the ability to automatically detect any personal data, where its algorithm was trained to minimise the use of this data in developing the AI model.

It practices the Data Governance principles by documenting the lineage of data and logging data consistently across multiple components in a secure and centralised log storage. In addition, it had also ensured proper documentation of data sources by obtaining reliable data directly from clients instead of using unreliable third-party datasets.

Organisations may refer to voluntary guides such as "ASEAN Data Management Framework (DMF) and ASEAN Model Contractual Clauses (MCC), etc.

Security and Safety

Safety involves placing emphasis on a risk-prevention approach by identifying and mitigating potential risks in order to ensure the safety of developers, deployers and users of AI systems through proper safeguards.

This can be done by conducting regular impact or risk assessments to identify suitable precautionary measures that enable human intervention when required. An example scenario is where autonomous vehicles may cause injury to pedestrians.

Safety of the public and the users of AI systems should be of utmost priority in the decision-making process of AI systems and risks should be assessed and mitigated to the best extent possible.

Before deploying AI systems, organisations should conduct risk assessments and relevant testing or certification and implement the appropriate level of human intervention to prevent harm when unsafe decisions take place. This may be verified further by engaging an internal or external agency to evaluate and carry out risk assessments from a third-party perspective. The risks, limitations, and safeguards of the use of AI should be made known to the user.



Security refers to ensuring the cybersecurity of AI systems⁷, which includes mechanisms against malicious attacks specific to AI such as data poisoning, model inversion, the tampering of datasets, as well as other attacks designed to reverse engineer personal data used to train the AI.

Organisations should implement technical security measures like robust authentication mechanisms and encryption. Organisations should also involve its cybersecurity leader throughout the AI lifecycle to maintain a strong overall security posture.

Organisation C is a technology start-up company which provides a digital care programme to tackle obesity-related chronic diseases.

It practices the Safety principles by tackling potential risks at an early stage and conducting rigorous tests on its AI models by using a build-and-deploy process that included automated code testing such as unit testing and static analysis. This allowed the company to highlight possible vulnerabilities in the code before running its AI model.

It practices the Security principles by implementing measures to strengthen their AI models by conducting annual penetration tests with an independent third-party security firm and performing post-mortem analyses to identify root causes and implementing future controls.



Organisations may refer to international best practices such as "Guidance for Secure AI System Development" developed by National Cyber Security Centre, United Kingdom.

Organisations should also implement safeguards to protect AI systems against cyberattacks, data security attacks and other digital security risks. may include ensuring These authentication mechanisms, data security safeguards, regular software updates to AI systems, incident response plans, proper access management for critical or sensitive systems and security testing such as penetration tests or vulnerability scanning.

Organisation D manufactures AI-enabled autonomous vehicles and practices the Safety principle by ensuring the designers, developers and deployers of AI systems provide for human intervention mechanisms for the human driver to easily resume manual driving, where necessary. Such precautionary safeguards allow for the autonomous vehicle system to safely disengage itself in the event that it were to make

It practices the Security principles by encrypting all data transmissions to prevent unauthorised access and ensuring real-life Geographic Information System (GIS) monitoring for timely actions to avoid any

Robustness and Reliability

Al systems should be sufficiently robust to cope with errors during execution and unexpected or erroneous input, or cope with stressful environmental conditions. It should also perform consistently. Al systems should, where possible, work reliably and have consistent results for a range of inputs and situations. Al systems may have to operate in real-world, dynamic conditions where input signals and conditions change quickly.

To prevent harm, AI systems need to be resilient to unexpected data inputs, not exhibit dangerous behaviour and continue to perform according to the intended purpose. Notably, AI systems are not infallible and organisations should ensure proper access control and protection of critical or sensitive systems and take actions to prevent or mitigate negative outcomes that occur due to unreliable performances. Measures such as proper documentation of data sources, tracking of data processing steps, and data lineage can help with troubleshooting AI systems.



Organisation E is a company that leverages deep learning techniques and combines biometric, geo-location and behavioural analytics with multi-factor to authenticate user identities.

It practices Robustness and Reliability principles by involving engineering, product, sales and research teams in the design and development stages. It conducts rigorous testing before the deployment of AI systems to ensure consistent results across a range of situations and environments. In addition, it ensured proper access controls on protection via a multi-level assurance framework through relevant department's Head of Unit.

Fairness and Equity

Organisations should have safeguards in place to ensure that algorithmic decisions do not further exacerbate or amplify existing discriminatory or unjust impacts across different demographics. The design, development, and deployment of AI systems should not result in unfair biasness or discrimination. An example of such safeguards would include human interventions and checks on the algorithms and its outputs.

Organisations should conduct regular testing of such systems to confirm if there is bias and where bias is confirmed, make the necessary adjustments to rectify imbalances to ensure equity. With the rapid developments in the AI space, AI systems are increasingly used to aid decision-making.

Organisation F is an Institute of Higher Learning which offers diploma courses to students. Every year, the institution conducts an early admissions exercise which is an aptitude-based admission process that allows students to apply for admissions prior to receiving their final grades.

If not properly managed, decisions made using an AI system's outputs could perpetuate existing bias against specific demographics. To mitigate discrimination, it is important that the design, development, and deployment of AI systems align with fairness and equity principles. In addition, the datasets used to train the AI systems should be diverse and representative. Appropriate safeguards should be taken to mitigate potential biases during data collection and pre-processing, training, and inference.

The institution practices Fairness and Equity principles by adopting a human-over-the-loop approach to invite the unsuccessful applicants for a face-to-face interview to determine the AI system's output and validity. It also ensures that a second review would be conducted by relevant staff (such as admissions committee or academic members of staff) to ensure reliability on the shortlisting of candidates. Excellence School Early
Admissions ENTER

Human Centricity

Al systems should respect human-centred values and pursue benefits for society, including human well-being. It is key to ensure that people benefit from AI design, development, and deployment while being protected from potential harms and not take advantage of individuals, in particular minors and vulnerable individuals. It is imperative for AI systems to be accessible, especially in instances where AI systems are used to make decisions about humans or aid them.

Human-centricity should be incorporated throughout the AI system lifecycle, starting from the design to development and deployment. Actions must be taken to understand the way users interact with the AI system, how it is perceived, and if there are any negative outcomes arising from its outputs.

Organisation G is an organisation in the education sector which has developed an AI-enabled Adaptive Learning System (ALS) for deployment within an online learning portal for primary students.

It practices the Human Centricity principles by engaging user research through involving its stakeholders during the design and development such as policymakers, academia, technical experts and its users. It had considered students individual learning readiness and personalised effective support needs for each student.

Accountability and Integrity

There needs to be human accountability and control in the design, development, and deployment of AI systems. Organisations should be accountable for decisions made by AI systems and should act with integrity throughout the AI system lifecycle.

Organisations should ensure the proper functioning of AI systems and its compliance with applicable laws, internal AI governance policies and ethical principles. In the event of a malfunction or misuse of the AI system that results in negative outcomes, responsible individuals should act with integrity and implement mitigating actions to prevent similar incidents from happening in the future.



To facilitate the allocation of responsibilities, organisations should adopt clear reporting structures for internal governance and procedures for complaints handling, setting out clearly the different kinds of roles and responsibilities for those involved in the AI system lifecycle.8

Al systems should also be designed, developed, and deployed with integrity – any errors or unethical outcomes should at minimum be documented and corrected to prevent harm to users upon deployment.

Organisation H is a tech company in the financial sector which provides an AI Anti-Money laundering Filter Model that identifies predictive indicators of suspicious transactions to reduce the number of false positives generated by the non-AI system, thereby reducing the number of alerts that require manual review.

It practices the Accountability and Integrity principles by ensuring that it sets up a 'Responsible Data Use' framework where a Committee from various departments was appointed to oversee and govern it, which ensured that the AI model complies with legal, compliance, security and any data quality issues.







⁸ This includes outsourced service providers or third-party vendors which have been engaged throughout the AI system lifecycle.

Al Governance Framework

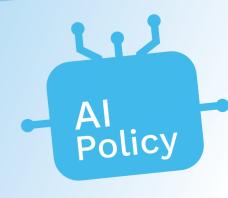
This section of the Guide includes guidance on measures promoting the responsible use of AI that organisations should consider to adopt and is designed around the established methodology of people, process and technology.





Organisations should detail a set of ethical and fair usage principles when developing or deploying AI products, services or systems. When establishing ethical principles for AI, should review their existing latest light of the organisations developments on AI technologies and the corporate values Guiding Principles mentioned in Section 5. When establishing usage policies for AI, organisations transparency. It should also ensure that such systems have proper authorisation over the use of any copyrighted materials. 1

Organisations should take steps to promote adaptability to technological advancements among the employees through reskilling and upskilling programmes. The initiative should consider the opportunities arising from digital transformation which extends beyond enhancing technical skills such as AI literacy, AI governance and AI ethics.



1

1

1

10

Level of human involvement in Al-augmented decision-making

Organisations should determine the level of human involvement in AI-augmented decision-making. This includes:

- Conducting relevant risk impact assessments to determine level of risk.
- Mitigating risks in order to build trust towards the acceptance and greater use of Al technologies.

Human involvement based on level of risk

Human-in-the-loop

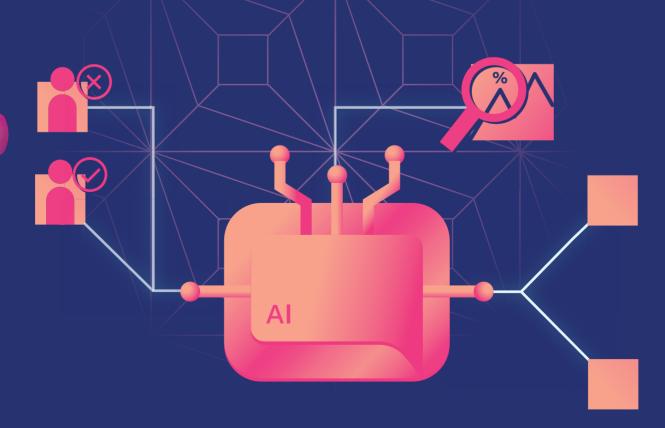
Organisation I uses AI for decision-making on its recruitment process. The AI system assists the organisation in identifying the suitability of potential candidates that meet the expected criteria of minimum entry requirements. In this approach, the AI system provides baseline recommendation or information for humans to weigh their decision.

Human-out-of-the-loop

Organisation J uses AI to enhance its digital marketing activities by learning about customers' purchases. The AI system creates customer profiles to understand their preferences and automatically recommends the products or services within similar categories or interests. In this approach, the AI system does not require any human oversight and has full control over its recommendations.

Human-over-the-loop

Organisation K uses AI to identify possible diagnosis of its patients and provides recommendations on potential medical treatments. Where it results in an undesirable outcome, humans would be able to alter the parameters to reach an acceptable outcome. In this approach, the AI system allows humans to take a supervisory role in its decision when it encounters challenges.



Process

Organisations should have in place appropriate internal governance frameworks that allow them to have oversight over AI technologies and tools.

Internal governance framework

- Develop standards, guidelines, tools and templates to help organisations design, develop, and deploy AI responsibly and ethically.
- Clearly lay out the roles and responsibilities of personnel involved in the responsible design, development and/or deployment of AI.
- Developing procedures for continuous monitoring and feedback to regularly assess and improve AI governance practices.

AI Governance General principles of risk management that are integrated, structured and comprehensive should be considered.

Risk management should consider the whole system, with all its technologies and functionalities and its impact on the environment and relevant stakeholders.

Risk management framework should be integrated as part of the organisation's AI System framework. Organisations should conduct risk-based assessments before starting any data collection and processing or modelling.

Organisations also should be ready to deploy measures to mitigate risks of unjust bias due to insufficiently representative training, testing and validation datasets. In addition, organisations should consider to ensure that AI systems are adaptable and allow for re-evaluative assessments due to the complexities of risk scenarios and ever-changing environments in the era of technological advancements.



Technology

Building a resilient system is the cornerstone of building a robust AI system and thereby contributes to building trust in an AI system.

Testing, validation and verification considering the complexity and criticality of the system is crucial. This should not only include the core components such as the AI model, data sets and business logic but also the underlying systems and infrastructure and more importantly to test the system in entirety. Organisations should employ techniques to test applications and create dedicated environments for testing. Creating a digital twin might be a good idea for AI systems used in critical functions.

Organisations may consider using internationally recognised AI governance standards,⁹ testing framework¹⁰ and software toolkit that validates the performance of AI systems against a set of internationally recognised principles through standardised tests and is consistent with international AI governance frameworks. Organisations should implement safeguards to help employees adapt to an AI-augmented work environment.



¹⁰ An example of regional testing framework is Singapore's AI Verify – which is a single integrated toolkit that operates within enterprise environment that can perform technical tests on common supervised learning classification and regression models for most tabular and image datasets.

Al Governance and Ethics Working Group

Members of AI Governance and Ethics Working Group are representatives from the following organisations:

- 1 Ministry of Transport and Infocommunications (MTIC)
- 2 Ministry of Health (MOH)
- 3 Ministry of Finance and Economy (MOFE)
- 4 Ministry of Education (MOE)
- 5 E-Government National Centre (EGNC)
- 6 Cyber Security Brunei (CSB)
- 7 Authority for Info-communications Technology Industry of Brunei Darussalam (AITI)
- 8 Brunei Darussalam Central Bank (BDCB)
- 9 Universiti Brunei Darussalam (UBD)
- 10 Universiti Teknologi Brunei (UTB)
- 11 Universiti Islam Sultan Sharif Ali (UNISSA)
- 12 Baiduri Bank
- 13 Bank Islam Brunei Darussalam (BIBD)
- 14 Brunei LNG Sendirian Berhad (BLNG)
- 15 Datastream Digital Sdn Bhd (DST)
- 16 Dynamik Technologies Sdn Bhd (Dynamik)
- 17 EVYD Technology Sdn Bhd (EVYD)
- 18 Huawei Technologies (B) Sdn Bhd (Huawei)
- 19 Imagine Sdn Bhd (Imagine)
- 20 I.T. Protective Security Services Sdn Bhd (ITPSS)
- 21 Microsoft
- 22 Progresif Sdn Bhd (Progresif)
- 23 Royal Brunei Airlines Sdn Bhd (RB)
- 24 Unified National Networks Sdn Bhd (UNN)

COPYRIGHT NOTICE

© AITI, 2024. This document is the property of the Authority for Info-communications Technology Industry of Brunei Darussalam ("AITI"), a body corporate with perpetual succession with its address at B13 and B14, Simpang 32-5, Jalan Berakas, Kampung Anggerek Desa, Brunei Darussalam. It must not be copied, used or reproduced for any other purpose other than for which it is supplied, without the expressed written consent of AITI.

DISCLAIMER

The information contained in this document does not constitute legal advice and should not be treated as such. AITI disclaims any responsibility or liability for any use or misuse of this document by any person and makes no representation or warranty, express or implied, as to the accuracy or sustainability of the information to any third party.