ༀ འཕྲུལ་གཤུང་འཕུལ་རིག་ལས་ཁྲེ

GovTech
Bhutan

Guideline for Generative Artificial Intelligence Usage in the Civil Service
2024

Developed by
Royal Civil Service Commission
and
Government Technology
Royal Government of Bhutan

FORWARD

The Generative AI Guideline is a timely and comprehensive resource designed to introduce and establish best practices for the ethical and effective use of generative AI within the Civil Service. In this transformative digital era, the potential of generative AI to enhance our government's ability to deliver efficient, innovative, and citizen-centered services is immense. As we collectively embark on leveraging this technology to meet Bhutan's unique needs, building a strong foundation for its responsible application is essential.

Following the practices recommended in this guideline will equip us to address the opportunities and challenges presented by generative AI. Through mindful integration of AI into public service, we can enhance decision-making, optimize resource allocation, and create avenues for smarter, more responsive governance. Embracing these practices will not only enable us to improve public service delivery but also foster a culture of innovation and transparency within our civil service.
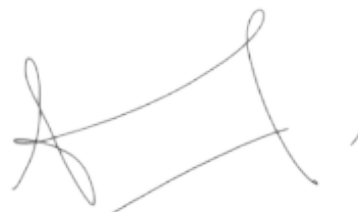
This guideline reflects the invaluable contributions of stakeholders across various agencies, and we extend our gratitude to all who have provided their expertise and insights. Their commitment has ensured that this document remains aligned with Bhutan's vision of using technology to benefit society while upholding our nation's principles of Gross National Happiness.

We are hopeful that the Generative AI Guideline will serve as a trusted and practical resource for civil servants, empowering them to responsibly and effectively harness AI while minimizing adverse impact, as we move into the digital transformation.

Tashi Pem
Chairperson
Royal Civil Service Commission

Jigme Tenzing
Secretary
Government Technology (GovTech)

Introduction

As Bhutan embraces the transformative potential of digital technologies, the adoption of Generative Artificial Intelligence stands at the forefront of this evolution. Generative AI systems are capable of creating new content ranging from text, images, video and other simulation, code and use as virtual assistant offering unprecedented opportunities to revolutionize business delivery. Using AI responsibly can enable positive disruption in government business delivery by enhancing creativity and productivity. These opportunities also raise ethical questions around authenticity and responsible use of AI.
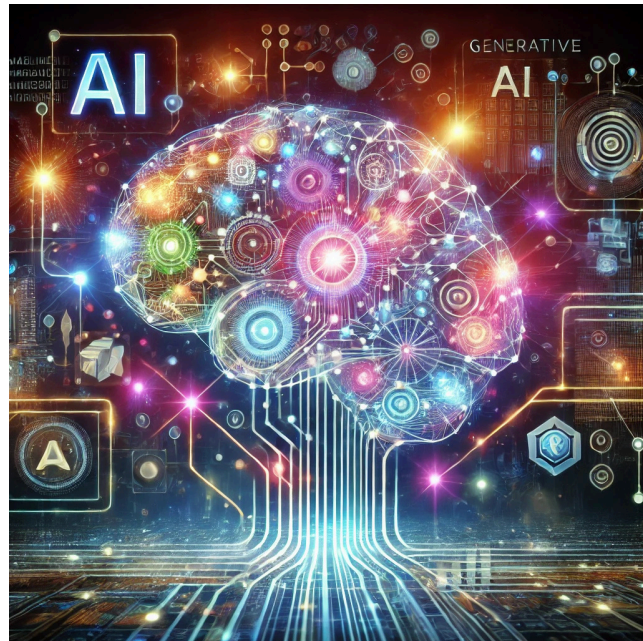
This guideline therefore will provide an overarching framework on using AI responsibly in the Civil Service. As Bhutan currently does not have a specific policy and regulations on AI, this guideline is prepared referring to the AI guidelines of Canada, UK, Switzerland, USA, United Arab Emirates, EU ACT and it will serve as an interim guidelinefor Generative AI usage within the government.

The integration of Generative AI in public administration can significantly enhance efficiency, enabling civil servants to automate routine tasks, generate data-driven insights, and engage creatively with citizens. However, it is imperative to address potential risks, such as misinformation, bias, and privacy concerns, that could undermine public trust and the integrity of government processes.

This guideline serves as a primary guiding document for the civil servants in Bhutan, outlining the principles, best practices, and ethical considerations associated with the use of Generative AI within government operations.

Recognizing the unique socio-cultural context of Bhutan, particularly the emphasis on Gross National Happiness and sustainable development, the guidelines are designed to align with national priorities while promoting transparency, accountability, and inclusivity. Therefore, this document provides a strategic framework to empower civil servants with the knowledge and tools necessary to navigate the complexities of AI responsibly.

Scope

The guideline aims to equip civil servants of Bhutan with the tools and knowledge to harness the benefits of Generative AI while safeguarding national interest, ethical standards, personal information, and data privacy.

Definitions

Gen AI or Generative AI

means a type of artificial intelligence technology that can generate various types of content, including text, imagery, audio, video and synthetic data.

The term Generative means that GPT can create something new.

Personal information/data

means any information that can be used to identify an individual directly or indirectly. It could be a name, CID, address, individual details, fingerprints etc..

Generative Pretrained Transformer (GPT)

is a powerful language model based on the transformer architecture, designed to generate human-like text by predicting the next word in a sequence based on prior context. It excels in various tasks such as text generation, summarization, question answering, and conversational interactions, making it widely applicable in content creation, customer service, education, and more. However, its use also raises ethical concerns regarding bias, misinformation, and responsible AI deployment, emphasizing the need for careful oversight.

Human oversight

Human supervision refers to the involvement of human judgment and decision-making in monitoring and managing automated systems to ensure they operate correctly and ethically. It encompasses the capacity for humans to review, intervene, and challenge the decisions made by these systems, thereby maintaining accountability and aligning outcomes with societal values.

Definitions

Artificial Intelligence

Artificial Intelligence (AI) refers to the development of computer systems capable of performing tasks that typically require human intelligence, such as problem-solving, learning, and decision-making. These systems use algorithms and data to analyze information, recognize patterns, and make predictions or recommendations.

Government Data Protection Regulation (GDPR):

GDPR is a comprehensive data protection law enacted by the EU to safeguard the privacy and personal data of individuals within the EU. It sets strict rules for how organizations collect, store, and process personal data, giving individuals greater control over their data, including rights to access, rectify, and delete data.

Hallucination

Hallucination refers to instances where the AI model generates information that is factually incorrect, misleading, or entirely fabricated, even though it appears coherent or plausible. This can happen when the model confidently provides responses based on patterns in the data it was trained on, without access to real facts or a proper understanding of context

Log Data

Log data refers to the records of interactions between users and the model. This can include details such as, (Timestamps) when the interaction or query was made, (User Input) the text or prompt provided by the user, (Model Output) the response generated by the GPT model. (System Information) metadata such as session ID, model version, and processing time. (Usage Metrics) how often the model is called, response length, or latency.

Deep Learning

Deep learning is a subset of machine learning that uses neural networks with many layers to automatically learn and identify patterns in large amounts of data. It is particularly effective for tasks like image recognition, speech processing, and natural language understanding.

Definitions

Machine learning

Machine learning is a branch of artificial intelligence that enables computers to learn from and make predictions based on data without being explicitly programmed for specific tasks.

Categorization of AI based on risk Factor

Artificial intelligence can be grouped under four categories based on their risk factors. Generative AI falls under the limited risk AI where the use of such AI requires human oversight and organization guidelines to mitigate risk(EU, 2024). Since Bhutan does not have any regulation based on AI, public servants are advised not to use the AI that is categorized under unacceptable risk AI and high risk AI. If required to use high risk AI individual employees are required to get expert guidance from the additional support contact given, on the use of such AI products.

Unacceptable AI Risk	Unacceptable AI risks refer to potential dangers posed by an AI system that are deemed too great to justify its deployment or use. These risks could result in significant harm to individuals, society, or the environment, and may breach legal, ethical, or societal norms. Examples include systems that could cause loss of life, serious injury (automated missile system), or widespread discrimination among religious communities, different cultural and racial groups, regional divides within a country. Civil servants are advised not to use such AI products under any circumstances.
High-Risk AI	High-risk AI refers to AI systems that have the potential to significantly impact people's rights and safety, typically in critical areas such as healthcare, law enforcement, or public services. The deployment of such systems often requires stringent regulatory oversight due to the potential for severe adverse effects, including those related to accuracy, privacy, or security. Examples might include AI used for diagnostic purposes in healthcare or court using AI for evaluating evidence reliability during criminal investigations or prosecutions, facial recognition technology. Adopting such AI required stringent regulation and if needed, as Bhutan is embracing digital transformation, individuals or organizations within Bhutan need to get expert guidance.

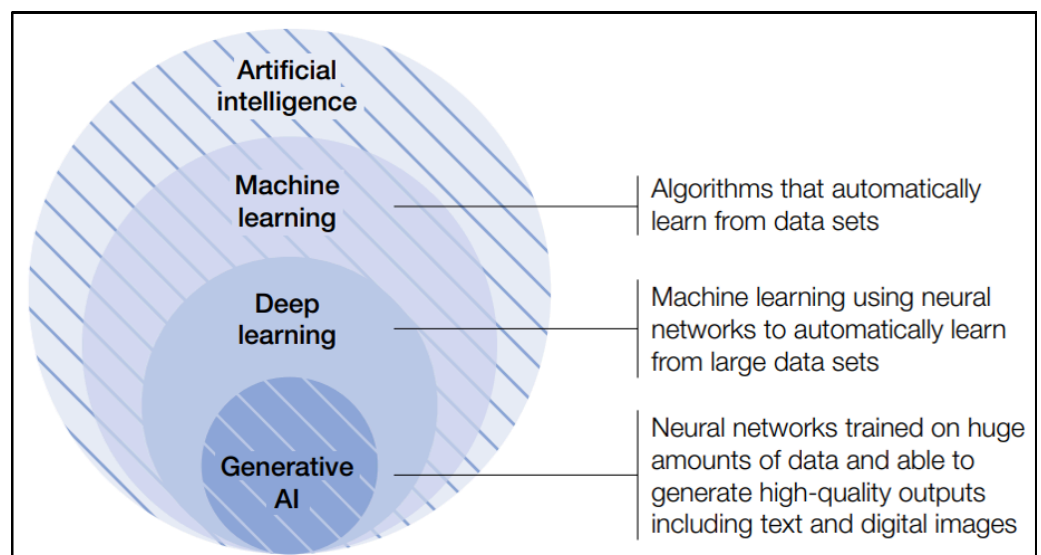# Categorization of AI based on risk Factor

Limited Risk

Limited-risk AI refers to AI applications that pose a moderate level of risk to individuals or society. These systems may require some level of oversight and regulation but are not considered as critical as high-risk AI. Limited-risk AI might include applications such as chatbots or recommendation systems that provide users with personalized content but have a lower potential for harmful impacts. Examples include the AI chatbot deployed on various government websites. Civil servants using this kind of AI need to follow the organization guidelines and human oversight.

Minimal Risk

Minimal-risk AI involves AI systems that are considered low risk, with negligible potential to harm individuals or society. These systems typically operate within well-defined boundaries and do not handle sensitive data or critical tasks. Examples of minimal risk AI might include simple automation tools or basic user interface enhancements that do not impact safety or privacy of individuals or organizations significantly. It is at the discretion of Civil servants to use such AI products.

Understanding Generative AI

Generative AI is a type of AI that produces content such as text, audio, code, videos and images. This content is produced based on information the user inputs, called a prompt, which is typically a short instructional text.

Artificial intelligence

Machine learning — Algorithms that automatically learn from data sets

Deep learning — Machine learning using neural networks to automatically learn from large data sets

Generative AI — Neural networks trained on huge amounts of data and able to generate high-quality outputs including text and digital images

Models which generate content are not new, and have been a subject of research for the last decade. However, the launch of ChatGPT in November 2022 increased public awareness and interest in the technology, as well as triggering an acceleration in the market for usable generative AI products. Other well known generative AI applications include Claude, Bard, Bedrock and Dall-E, microsoft copilot, which are LLMs (UK, 2024).

# What data are collected by Generative AI while using those platforms?

Civil servants need to be aware of what kind of data are collected by those Generative AI platforms, and how the data provided are being processed and used.

The Generative AI platforms like ChatGPT, and Google Gemini will collect user account information, usage data, log data, device information, AI response, and user content. Those information are then used to train those models and enhance their product, which may lead to loss of personal data to those AI platforms. To avoid processing and training user data all civil servants using AI products need to turn off chat history. If a user turns off chat history the user data will be stored for 30 days and deleted and not used for training the model.

Under certain privacy or data protection laws, such as the GDPR, a user has the right to object to the processing of their personal data by OpenAI's models. For that, the user can submit a personal data removal request. Furthermore, you can restrict your data from being used to train the AI model by filling out the OpenAI User Content Opt Out Request form.

Application of Gen AI in Government

Under certain privacy or data protection laws, such as the GDPR, a user has the right to object to the processing of their personal data by OpenAI's models. For that, the user can submit a personal data removal request. Furthermore, you can restrict your data from being used to train the AI model by filling out the OpenAI User Content Opt Out Request form.
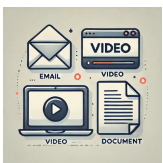
drafting emails, documents, Content generation(images, video, audio, text).
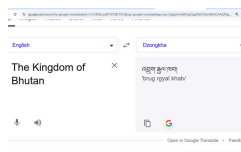
translating text between languages
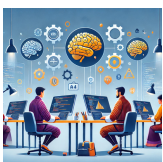
supporting coding activities by computer programmers

Data analysis and insight (AI can analyze large datasets to extract insights, identify trends, and support decision-making)

Citizen Services (AI-powered chatbots and virtual assistants can provide 24/7 support to citizens, answering queries and guiding them through)

brainstorming ideas or research information

Civil Servants to adopt a few best practices that are listed below while using Gen AI

1   Clearly indicate that you have used generative AI to develop content.

2   Don't consider generated content as authoritative. Review it for factual and contextual accuracy.

3   Learn about bias, diversity, inclusion, anti-racism, and values and ethics to improve your ability to identify biased, non-inclusive or discriminatory content.

4   Assess the impact of inaccurate outputs. Don't use generative AI when factual accuracy or data integrity is needed.

5   Consider your ability to identify inaccurate content before you use generative AI. If you can't confirm the quality of the content, don't use it.

6   Learn how to create effective prompts and provide feedback to refine outputs to minimize the generation of inaccurate content.

7   Don't use generative AI tools as search engines unless sources are provided so that you can verify the content.

Important considerations while using Generative AI

Never put sensitive and personal information into these tools and protect confidential data

Civil servants using the platform should be mindful that putting information into those tools is similar to uploading information to the public domain and can be accessed by anyone. While leveraging these technologies users need to be cautious on what information needs to be fed into those tools. Avoid inputting government data, personal information such as name of person, employee ID number, citizenship ID number, unpublished materials, proprietary or confidential data or any other data that is classified as L2, L3, and L4. There is risk of compromising data and losing proprietary rights, intellectual proprietary rights, which may bring damaging impact to individuals and organizations.

In unavoidable circumstances if civil servants need to use personal data, then civil servants need to do the De-identification of personal information as mentioned below.

De-identification of personal information

De-identification involves removing, anonymising, or masking personal information so that it can no longer be used to identify an individual. This process is crucial to protect privacy and safeguard personal information.

Never put sensitive and personal information into these tools and protect confidential data

Data anonymization	Data masking	Data Control
Replace or alter personal information in the text using information that cannot be linked to the individual.	Replace personal information in the text, such as email addresses or phone numbers, with fictional data that retains the same format. (eg: abc@test.xyz.xy)	To disallow ChatGPT to use your content for learning and improvement of the model for everyone, disable the 'the improve the model for everyone' under settings (data control).

Data Classification Levels

De-identification involves removing, anonymising, or masking personal information so that it can no longer be used to identify an individual. This process is crucial to protect privacy and safeguard personal information.

L1 Information intended and released for public use	L2 Low-Risk Confidential Information that may be shared only within the respective community
This includes information intentionally provided to the public such as content on websites, general agencies, ministries, organizations, Dzongkhags contact details, published papers, press releases, general news and announcements intended for the public, code contributed to Open Source, etc.	This includes information kept private but its disclosure would not cause material harm, such as department policies and procedures, employee web portals, research papers, work papers, non-public project plans or layouts, non-sensitive administrative survey data, etc.

Data Classification Levels

L3 Medium-Risk Confidential Information intended only for those with a business need to know.	L4 High-Risk Confidential Information that requires strict controls
Disclosure of this information beyond intended recipients might cause material harm to individuals or the community. This includes personal records, non-published staff information, sensitive administrative survey data, non-public legal work, non-public financial statements, etc.	Disclosure of this information beyond specified recipients would likely cause serious harm to individuals or the community. This includes passwords and PINS, system credentials, private encryption keys, government-issued identifiers (e.g., CID Number, Passport Number, phone number, driver's license), individually identifiable financial account information, individually identifiable health or medical information, security system procedures, and architectures, etc.

Do Verify and Fact-Check Information

While Generative AI can provide detailed responses to questions, it may not always be accurate . Therefore, civil servants are required to do fact-check responses first. Then, go further and verify them with other authoritative sources. Any kind of Generative AI should not be considered as the primary source of information. With this practice, we can ensure that the information provided is accurate and reliable. This can go a long way in preventing the spread of misinformation and inaccuracies online. In case due to the user's carelessness if there is spread of misinformation, the individual is subjected to legal implications as mentioned in section 8.

Civil servants need to adhere to academic integrity while using Generative AI

Civil servants should be aware of the policies on permitted uses of generative AI in official tasks, academic coursework, or research areas. Employees should use Generative AI as a tool to aid learning and spark ideas rather than a substitute for your own effort.

Civil servants are responsible for any content that you produce that includes AI-generated material (accountability)

AI generated content might lack accuracy, mislead, be entirely fabricated known as hallucinations(confabulation), or even contain copyrighted material. Employees should verify and review AI-generated content before official use. Generative AI can be also used to create content for advertising and dissemination of information using deepfake, but it will have legal implications and adverse impact if individuals use such technology to create deep fake avatars and spread misinformation. If use of Generative AI has intentionally or unintentionally posed threat or caused damage to any individual, the user is liable for his action as per CIVIL LIABILITY ACT OF BHUTAN 2023 as mentioned in section 8.

Legal implication of using the Generative AI by the civil servants of Bhutan

Generative AI, with its ability to craft compelling text, translate languages, and even generate art, holds immense potential for civil servants in Bhutan. However, navigating this exciting frontier requires mindful consideration of the existing laws, acts and regulations like the Information Communications and Media Act 2018 which stands as a cornerstone, safeguarding against the dissemination of misinformation and upholding responsible online conduct. Additionally, the National Digital Identity ACT Of Bhutan 2023, and Data Management Guide 2023 outlines best practices for handling sensitive information, ensuring respect for individual privacy and data security and The Copyright ACT of Kingdom of Bhutan 2001 addresses the issue related to copyright. On top of that, the Civil Liability Act of Bhutan 2023 outlines the accountability and liability of individuals be it service providers or users. Furthermore, civil servants' use of Gen AI shall be guided by the civil service core values and code of conduct as prescribed in the Civil Service Act 2010. These legal frameworks serve as crucial safeguards, ensuring your AI endeavors adhere to Bhutan's values and regulations. Civil servants using any kind of Generative AI need to comply with those existing regulations.

Copyright Infringement

Generative AI models are trained on a variety of sources, many of which are protected by copyright. As a result, it may reproduce copyrighted content in its output. This is not only an ethical issue but also a potential legal issue. OpenAI states that users are responsible for the content of outputs, meaning that users may be liable for copyright issues that arise from the use of Generative AI outputs. This is problematic as Generative AI is unable to provide citations for the sources it was trained on, so it can be difficult for the user to know when copyright has been infringed. Therefore civil servants while using AI generated content, it is essential to cite the rightful sources to avoid copyright infringement.

Plagiarism and Deceitful Use

Generative AI can be used unethically in ways such as cheating, impersonation or spreading misinformation due to its humanlike capabilities. In academic contexts, Generative AI may be used to cheat, intentionally or unintentionally. It includes passing off AI-generated or paraphrased plagiarized content as original work and fabricating data to support your research. Because the Generative AI model can write code, it also presents a problem for cybersecurity. Threat actors can use it to help create malware. It can also be used to impersonate a person by training it to copy someone's writing and language style. The chatbot could then impersonate a trusted person to collect sensitive information or spread disinformation. As stated in Annexure 1 plagiarism related to AI generated can be detected. Therefore employees are to be cautious while leveraging Generative AI, if found practicing any of the mentioned activities, the individual is liable for his/her own action.

Civil servants who use the Generative AI to understand concepts better or learn new things. That is not unethical. However, the problem arises when individuals use it to write essays, solve problems, and do research and academic writing and claims as one's own work. Hence, having AI-based technology do your research and work comes under plagiarism, which violates the code of academic conduct. Civil servants need to be mindful of such practices and refrain from using AI generated work as individual work.

Personal Liability

Generative AI models such ChatGPT, Google Gemini, and Microsoft copilot use a click-through agreement. Click-through agreements, including terms of use. Civil servants who accept click-through agreements without delegated signature authority may face personal consequences, including responsibility for compliance with terms and conditions. It is important that civil servants are well aware of the terms and conditions of Generative AI while using such platforms to prevent unnecessary obligations.

Don't Rely Solely on Generative AI

While Generative AI is useful, never solely rely on it especially for making big decisions. The system is limited by the data it's been trained on and cannot understand context or emotions. Therefore, consider its responses alongside other sources of information. Then, you can make informed decisions based on all available data. Civil servants are requested to use human oversight when using Generative AI on making decisions related to HR recruitment, HR planning like promotion, financial planning, evaluating students performances in schools and colleges, which may otherwise lead to bias or inappropriate results if it depends only on automated AI decisions.

Non-Discrimination and Inclusivity

Avoid Bias: Civil servants need to be cautious and mindful of the language and instructions provided to Generative AI to prevent the generation of biased or discriminatory content. Promote inclusivity and avoid content that may contribute to harm or discrimination among people and organizations within Bhutan.

Cultural Sensitivity: Bhutan has strong religious and cultural heritage , and civil servants are mandated to respect cultural differences and sensitivities when engaging with Generative AI. Avoid generating content that may be offensive or disrespectful to different cultural and religious communities.

Annexure 1 : AI Tools for different use cases

Sl. No	Name of AI tools	Application
1	Tome, Gamma, Slide and Simplified.	These AI tools can help in preparing presentations efficiently and effectively.
2	Generative AI chatbots (ChatGPT, google gemini, microsoft copilot, claud AI, perplexity, Brave Leo AI, Phind, Jasper chat, Chatsonic)	These Generative AI Chatbots are designed to assist users in various tasks by generating contextually relevant responses from input prompts. They are widely utilized for applications such as content creation, customer support, research assistance, and productivity enhancement, leveraging advanced natural language processing (NLP) capabilities.
3	DALL-E 3, Midjourney, Firefly, DeepAI Org	These AI tools are used for designing visual content with the help of design platforms like Canva, Adobe Firefly, and Figma.
4	QuillBot	QuillBot is a versatile AI-powered writing tool designed to enhance the writing process by offering features such as paraphrasing, grammar checking, summarization, plagiarism detection, translation, and citation generation. It allows users to rephrase content while maintaining its original meaning and provides customizable writing modes to suit different contexts like academic or professional writing
5	WordTune, Scholarcy, and Research Rabbit.	Help in drafting and summarizing complex reports.
6	Surfer SEO, SEMrush, AdCreative, and Text Cortex	Marketing tools.

Annexure 1 : AI Tools for different use cases

Sl. No	Name of AI tools	Application
7	Synthesia, sounddraw, wordtune, Replit, Flicki, Remini, Pictory.	Generate Videos, Generate music, summarize notes, generate code, generate Tiktok ,edit pictures, edit videos respectively.
8	Clickup, Otter.ai, Laxis, Doodle, FireFlies.ai, Timz.Flowers	These are AI tools for note taking and minutes keeping.

Annexure 2 : Cybersecurity threat and AI solution

Sl. No	AI solution to Detect threat
1	Darktrace uses AI to detect and respond to anomalies in network behavior, addressing Advanced Persistent Threats (APTs), zero-day exploits, and insider threats
2	Vectra AI automates the detection of ransomware, insider threats, and APTs through AI-driven network analysis.
3	IBM QRadar is a security information and event management (SIEM) platform that detects Distributed Denial of Service (DDoS), SQL injection, and malware (Trojans, viruses).
4	CrowdStrike Falcon offers AI-driven endpoint security to detect malware, fileless attacks, and phishing in real-time.
5	Cynet provides automated incident response for malware, credential stuffing, and phishing.
6	FireEye monitors network traffic to detect APTs, zero-day exploits, and supply chain attacks, while Splunk uses AI to prevent DDoS, insider threats, and Man-in-the-Middle (MitM) attacks.
7	Microsoft Azure Sentinel identifies and mitigates ransomware, DDoS, phishing, and credential stuffing, while HackerOne crowdsources vulnerability detection for SQL injection, cross-site scripting (XSS), and zero-day exploits.
8	FortiAI offers AI-driven threat detection for malware, ransomware, and phishing, providing automated responses to emerging threats.

References (other international ACT and regulation related to AI)

1. European Union. (2024). EU act on AI. [EU ACT on AI](#)
2. Federal Administration of Switzerland.(2021). Use of artificial intelligence within federal administration. [Use of artificial Intelligence within Federal administration](#)
3. Federal Government of the United States. (n.d). Introduction to the AI guide for government. [Introduction to the AI Guide for Government](#)
4. Government of Canada. (2023). Guide on the use of generative artificial intelligence. [Guide on the use of generative artificial intelligence](#)
5. Government of the United Kingdom. (2024). Guidance to civil servants on the use of generative AI. [Guidance to civil servants on use of generative AI](#)
6. United Arab Emirates. (2020). AI guideline for the United Arab Emirates. [AI guideline for united Arab](#)

ACT related to ICT services in Bhutan

1. GovTech Agency. (2023). Data Management Guide 2023. [management guide 2023](#)
2. Government Of Bhutan(2023). NATIONAL DIGITAL IDENTITY ACT OF BHUTAN 2023. [NDI Act.pdf](#)
3. Ministry Of Information and Communication.(2018). Information, Communications and Media. [ICM Act.pdf](#)
4. Government Of Bhutan.(2023). CIVIL LIABILITY ACT OF BHUTAN 2023 [Civil Liability Act.pdf](#)
5. Government Of Bhutan.(2001). Copyright ACT of Kingdom Of Bhutan, 2021 [Copyright Act.pdf](#)

GovTech Support Team

✉ etd@tech.gov.bt