**Introduction of an Audit Trail for Generative Artificial Intelligence**

**Terms**

“Generative Artificial Intelligence (GenAI) Systems” – refers to the use of Artificial Intelligence to create new content, like text, images, music, audio and videos.

“VeriHash” – an on-chain solution that provides a platform for a decentralized, easily assessable database for the general public and government entities.

“Hash Records” – an alphanumeric string is generated based on content generated from GenAI. Examples of Hash records generators are the SHA-256.

“Log file” – a compilation of a single session of interaction with GenAI systems, inclusive of inputs (“prompts”) and outputs ("generated content”). The output can be in all forms such as images or text.

“General public” – refers to the citizens and residents of the Republic of Singapore.

“Government entities” -

**“Users” –**

**“Deployers” -**

**Background**

The EU AI act aims to foster trustworthy Artificial Intelligence (AI) in Europe. This is by ensuring that the AI systems respect fundamental rights, safety, and ethical principles. It uses a risk-based approach for powerful and impactful AI models. The EU AI act was rolled out on 1 Aug 2024.

[**AI Act | Shaping Europe’s digital future (europa.eu)**](https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai#:~:text=prohibit%20AI%20practices%20that%20pose,of%20high%2Drisk%20AI%20applications)

AI, and by extension, GenAI, are at the forefront of digital innovation and technology today. They can manifest in a wide variety of uses and their development is currently unpredictable. Regulation of uncertainty, hence requires a dynamic framework.

["Governance and Uncertainty" by Justin R. Pidot](https://digitalcommons.du.edu/law_facpub/14/)

The risk-based approach to regulation of AI is hence one framework aimed at standardizing and regulating AI into the potential impact and the potential risk they impose on society. This allows resources to be allocated in the enforcement and regulation of the AI that could have the largest devastating impact on society.

The risk-based approach also recognises the opportunities that GenAI provides in society, hence less regulatory restrictions are imposed to support continuous innovation and development of the technology.

[Compliance vs. Risk Management: Meaning, Difference & Tools | GEP Blog](https://www.gep.com/blog/strategy/differences-between-compliance-and-risk-management#:~:text=Tactical%20vs.,potential%20damage%20of%20a%20crisis.)

[**Microsoft PowerPoint - YordankaIVANOVA.Generative AI and the AI Act.pptx (europa.eu)**](https://www.europarl.europa.eu/cmsdata/277774/02.YordankaIvanova.pdf)

**Regulatory Purpose**

AI centric\*\*

The future of AI is largely becoming??? They manifest into XYZ and is a form of new media.

This significantly reduces the time spent by the general public in the content creation of XYZ.

This convenience is complicated by

High adoption of GenAI has made it XYZ and there is pressure on regulatory bodies around the world to ensure that these “black boxes” are safe and trustworthy for deployment and for society at large.

Risk has increased, and new risks have emerged; namely:

**Hence we need to look at GenAI have a duty to provide correct information to a third party? Are developers/deployers liable for LLM’s defamatory output? What about racist and offensive content?**

[**Liability for AI-generated Content - The Singapore Law Gazette**](https://lawgazette.com.sg/feature/liability-for-ai-generated-content/)

1. ***Goal of Regulation***

For GenAI, we need to recognise that the generated output produced in part depends on the user’s prompt, on the developers, on the training dataset, and the usage of the output.

Our policy direction is in line with our proposed solution and mirrors the high cost of non-compliance with the EU AI Act.

**Aim**

To increase accountability and transparency of currently deployed Generative Artificial Intelligence (GenAI) Systems via mandatory submission of Hash records generated from Log files of all sessions every 10 minutes in production into VeriHash.

**VeriHash Solution**

Hash records are unique to every Log file as serve as an auditable trail to the Log files. Once submitted via Verihash, these hashes are not editable, hence traceable to each session of Log file outputs by the GenAI system.

The original Log file is stored at the servers of the original GenAI and must be reproducible to the authorities when requested. Any edits to this original Log file will result in a different Hash record.

Any criminal activity that is conducted, in whole or in part, through GenAI, will be traceable to the Log file. This will both hold the owners of GenAI outputs and the users who prompt the GenAI via targeted inputs accountable for the generated content.

Such records promote traceability, allowing users to have an avenue for feedback to generated content to either the GenAI deployed or the government entities. At the same time, the GenAI developers will also have a traceable database to defend themselves against user prompts or actions created to attack or trick their GenAI into producing potential harmful, defamatory or criminal content.

**Policy**

Non-compliance to the submission of Hash records via VeriHash will result in the cessation of the deployment of the GenAI for use in the general public and/or a fine of up to 1,000,000 SGD.

GenAI systems must keep reproducible records of the original Log file for access by government entities. For the supply of incorrect, incomplete or misleading information to government entities in reply to a request, the deployers are subject to a fine of up to 10,000,000 SGD.

[Penalties of the EU AI Act: The High Cost of Non-Compliance (holisticai.com)](https://www.holisticai.com/blog/penalties-of-the-eu-ai-act)

**Proposed Regulatory Framework**

1. Audit Trail for Generative Artificial Intelligence
2. Regulating GenAI to reduce risks through engagement
3. *Public Consultations*
4. *Information Campaigns*
5. Regulating GenAI through the deployers and developers

*1) Self-regulation*

*2) Enhancing Security Measures*

1. Protecting the Vulnerable
2. Setting up a GenAI Regulatory Authority in IMDA

***Challenges to the regulation of GenAI and how the risks can be minimized***

For GenAI deployers, there is a trade-off between increasing the adoption rate of their technology (and hence subscription rate) and their willingness and ability to bear the costs of the risks.

Reputation exists as the companies and operators’ most valuable currency in that it provides a gauge of trust in the operator and company. For existing customers, reputation is what induces loyalty to the company and operator, while the trust displayed by existing customers could entice new sign ups too.

Users may be blinded to the risks of using GenAI as they focus on convenience in relying on GenAI in their day-to-day use. Regulations that impede the convenience of the consumer and innovation of the GenAI systems will hence meet strong resistance.

**Enforcement and Monitoring**

The main enforcement authority of deployed GenAI systems will be IMDA.

If VeriHash detects that a currently deployed GenAI system does not submit any hash records within the last 24 hours, they will alert IMDA to initiate further checks.

Owners of GenAI systems will need to submit an application to IMDA if the GenAI system is unable to be deployed for a period of time. During such time, the GenAI will be suspended from onboarding onto VeriHash.

Other scenarios

1. **Analysis and Review of Regulatory Strategy**