# MOSTI's Large Language Models (LLMs) – Some Legal Concerns

25 June 2025

*For more information, please contact:*

**Chia Loong Thye**
ltchia@rdslawpartners.com

**David Lee Lai Huat**
david.lee@rdslawpartners.com

**Falisa Abu Bakar**
falisa@rdslawpartners.com

**Lim Sheh Ting**
shehting@rdslawpartners.com

**Ooi Bee Hong**
beehong@rdslawpartners.com

**Raphael Tay Choon Tien**
raphael@rdslawpartners.com

**Tan Gek Im**
gekim@rdslawpartners.com

**Teo Siang Ly**
siangly@rdslawpartners.com

**Yap Wai Ming**
waiming@rdslawpartners.com

## Personal Data Security

LLMs are trained on extremely large datasets of text, enabling them to learn intricate language patterns and nuances. LLMs are designed to continuously learn and evolve by storing data that has been fed into the system. This raises concerns regarding personal data protection, which falls under the Personal Data Protection Act 2010 (PDPA). In order to develop a more efficient and accurate LLM system, companies may unlawfully collect[1] and process personal data from millions of users without consent, in violation of the PDPA.

A recent example is LinkedIn, which was sued in January 2025 for allegedly sharing premium users' private messages with third parties to train AI models without obtaining user consent. The lawsuit claims that LinkedIn updated its privacy settings in August 2024, automatically opting users into data sharing without proper notification and later revised its privacy policy in September 2024 to include AI model training data usage.

Similarly, OpenAI and Microsoft, are facing their second class-action lawsuit in a San Francisco Federal Court for allegedly breaching multiple privacy laws in developing OpenAI's chatbot, ChatGPT[2].

As the Ministry of Science, Technology & Innovation (MOSTI) plans to develop an LLM, concerns arise regarding the use of personal data to train the model. Since the Government has access to extensive datasets, there is a genuine risk that personal data could be utilised to make the LLM more compatible with the Malaysian user interface. Additionally, the collaboration between MOSTI and local companies in developing the LLM raises further concerns about whether these companies would have access to sensitive personal data stored by MOSTI.

The storage and use of personal information create potential risks of data breaches and unauthorised access. Section 6 of the PDPA emphasises that data controllers must obtain explicit consent before processing personal data, while Section 9 of the PDPA mandates the accuracy and protection of such data. Meanwhile, Section 8 of the PDPA specifies that personal data can only be retained for as long as necessary for its intended purpose. If MOSTI fails to implement proper security measures, it could lead to serious issues regarding the misuse or leakage of personal data, ultimately violating the protections afforded under the PDPA.

## Data Confidentiality And Integrity

Another key legal concern regarding the development of LLM by MOSTI is data confidentiality of inputted data by users. While LLMs are designed to generate responses based on vast amounts of training data, they also rely on user-provided input to refine their outputs. Users often submit sensitive or confidential information to achieve more personalised and relevant responses. However, there is a risk that this data may be stored, processed or even repurposed for unintended uses beyond the original user query.

The model's ability to retain contextual knowledge means that sensitive information could be integrated into future outputs, potentially exposing confidential or proprietary data. This raises significant concerns, particularly for industries dealing with confidential business communications, legal documents or financial transactions. If an LLM system trained on such data inadvertently discloses or replicates proprietary information, it could lead to breaches of confidentiality agreements and legal liabilities. Additionally, the concern is amplified as government agencies have access to extensive datasets, including individual records, company information, and national records.

If such information is fed into the system without robust confidentiality protection, it could expose sensitive data to exploitation. There is a risk that this data may be stored, processed or even repurposed for unintended uses beyond the original user query. For example, data originally provided by one user to personalise responses could later be used to train the model for different purposes, such as improving the system's general output or targeting advertisements, without the user's explicit consent. This could lead to the inadvertent disclosure of confidential or personal information.

## Concerns About Enforcement

While MOSTI has committed to adhering to the National Guidelines on Artificial Intelligence Governance and Ethics (AIGE) in developing LLM, several legal concerns remain. One of the key challenges is that these guidelines lack the force of law, unlike legislated statutes, they do not carry direct penalties for non-compliance. This raises concerns about enforcement, particularly in ensuring transparency and confidentiality in the development and deployment of LLM.

Furthermore, Malaysia currently lacks an independent regulatory body with investigatory or enforcement powers to oversee LLM effectively. Without a comprehensive legal framework to address these issues, the reliance on ethical guidelines alone may be insufficient to safeguard against potential risks and ensure responsible LLM development.

[1] https://www.reuters.com/legal/microsofts-linkedin-sued-disclosing-customer-information-train-ai- models-2025-01-22/?utm_source=chatgpt.com

[2] https://www.reuters.com/legal/litigation/openai-microsoft-hit-with-new-us-consumer-privacy-class- action-2023-09-06/

www.rdslawpartners.com

This publication is for educational and informational purposes only and is not intended and should not be construed as legal advice.

KUALA LUMPUR
Level 16, Menara 1 Dutamas No. 1, Jalan Dutamas
1, Solaris Dutamas, 50480 Kuala Lumpur
T: +603 6209 5400
F: +603 6209 5411
enquiry@rdslawpartners.com

PENANG
Suite S-21E & F21st Floor, Menara Northam No.
55, Jalan Sultan Ahmad Shah, 10050 Penang
T: +607 585 6414
F: +607 509 7614
generalpg@rdslawpartners.com

JOHOR BAHRU
8-35, Menara Delima Satu, Jalan Forest City 1,
Pulau Satu, 81550 Gelang Patah, Johor Bahru
T: +607 585 6414
F: +607 509 7614
generaljb@rdslawpartners.com

4