

Feeding the AI Machine: Training Data Under Regulatory Scrutiny in Asia

Training Data Under Regulatory Scrutiny
in Korea, Vietnam, China, and Thailand



#IAPPForum26

Name the 3 main instruments your regulator reaches for first in a training-data audit and what it actually requires.

Do you think all companies are ready to demonstrate compliance? And why do you think so?



Map the Terrain

AI Training Data Landscape Across Asia (2026)

	KOREA	CHINA	THAILAND	VIETNAM
 Core Framework	PIPA + AI Framework Act + Sectoral Act Privacy law strengthened by PIPA reforms and AI Framework Act (in force 22 Jan 2026)	PIPL / DSL / CSL + AI Regulations + Other Relevant Laws Multi-layered legal and regulatory framework for data and AI governance	PDPA + AI Governance Guidelines PDPA provides the legal basis; general and sectoral guidelines operationalize AI governance expectations	PDPL + AI Law + Cybersecurity Regime Comprehensive 3-layer framework covering personal data, AI and cybersecurity
 Key AI Training Data Position	✓ Tier 1 — legitimate interests for publicly available data (PIPC 2024–25 AI guides) ✓ Tier 2 — pseudonymized data for scientific research incl. AI development (Art. 28-2) ⚠ Tier 3 pending — PIPA amendment: PIPC-approved use of original data (plenary vote awaited)	✓ Permitted subject to extensive compliance controls ✓ Legitimacy of data source ✓ IP protection ✓ Personal information protection ✓ Data quality requirements	✓ Lawful Basis for the Use of Personal Data in AI Training ✓ Governance expectations through NCSA, MDES, ETDA and sectoral guidelines ⚠ No AI-specific law in force; ETDA released the draft AI Act in July 2026 for public consultation	✓ Lawful basis required for training data ✓ Purpose compatibility (purpose-fit) assessment ✓ Accountability owner for AI systems ⚠ Cross-border transfer triggers additional obligations
 Key Watchpoint	 Amendment completes a 3-tier cascade: public data → pseudonymize → ask the PIPC	 Highly regulated and documentation-intensive environment	 Guideline-driven and PDPA compliance; monitor the enactment of the AI Act	 Cross-border AI training data flows are likely to attract heightened scrutiny from MPS

What are the first three questions a regulator in your jurisdiction asked or is likely to ask in a training-data investigation?

From an expert perspective, what is the expected evidence to demonstrate compliance underlying each of these questions?



Stress-Testing Under Scrutiny



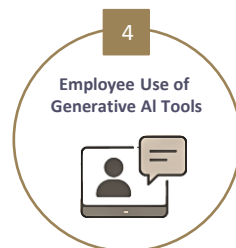
Foreseeable AI Training Data Problems

The journey of data across the AI ecosystem and the legal risks at each step

AI Developers & Model Providers



Enterprise AI Deployers & Corporate Users



Key Concerns

- Scraping secured websites or platforms for AI training
- Individuals unaware their personal data is reused for AI purposes
- Scraped content may include personal or illegal data

Where These Problems Arise by Country



Key Concerns

- Commercial datasets may have unclear origin
- Developers cannot verify lawful collection
- Hidden provenance risks become embedded into AI models

Where These Problems Arise by Country



* Including cross-border issues

Key Concerns

- Personal data related to historical customer, HR, or employee reused for internal AI training
- Original collection purpose differs from AI-development purpose
- Reuse occurs without updated governance review

Where These Problems Arise by Country



Key Concerns

- Employees upload contracts, emails, support tickets, or internal data into GenAI tools
- Enterprise data may unintentionally become training material
- Organizations lack visibility and internal controls

Where These Problems Arise by Country



#IAPPForum26

AI Data Team come and ask you that the company has years of customer-service logs, internal chats, support tickets.

Can we train in this? What should we do?



Company buys packaged consumer behavioral data. AI Data Team claimed that they are compliant with anonymisation. Purchasing Department said the company has a contract, a compliance certificate, and not much else.

Is it good to go train AI with such data package?



If the regulator opens a training-data inquiry tomorrow, the company wants to show up with the best possible response.

Please name the top 2–3 documents or governance artefacts they should have ready.



Key Documentation Regulators Expects to See

KOREA	CHINA	THAILAND	VIETNAM
 Choice of lawful basis per data source Legitimate Interest Assessment (LIA); compatible-purpose memo; consent records	 Legitimacy of Data Source Data Source Agreement, Authorization logs, Technical documentation of public data collection methods	 PDPC Security Measures Policy, DPIA, DPA, Privacy Notice, etc.	 PDPL Personal Data Processing Record - Named responsible person (statutory requirement) - Legal basis for using personal data - Data categories processed and their sensitivity classification - Retention schedule and deletion mechanism - Cross-border transfer details
 Sensitive / unique-ID handling Filtering rules; exclusion logs; PIPIA (privacy impact assessment) for public bodies / regulated sectors	 Content Security Filtering rules and logs; Safety assessment reports	 NCSA Secure coding standards, Evidence of secrets management controls, Access logs for training datasets, Documentation of data validation pipeline	 AI Law Data Governance Record - Complete list of data sources - Provenance documentation - Purpose-fit analysis: original collection purpose vs. AI training use, documented decision - Named AI system accountability owner - Date and sign-off record for the governance decision
 Data subject communication & rights Privacy notice changes; opt-out / withdrawal logs; deletion-on-request workflow	 IP Protection IP management strategies and rules; IP complaint & Reporting mechanism; disclose IP risks of the generated content in user agreement	 OIC Documentation of AI data sources and training processes, AI validation and testing record, Explainability and accuracy assessment documentation	
		 BOT AI model input/output documentation, Documentation of model parameters and assumptions	
 Regulator-facing readiness PIPC prior-adequacy review file; prior-deliberation dossier under pending amendment; sandbox approvals; monitoring reports	 Personal Information Protection Privacy notice update; consent records; PIPIA may apply; clear withdrawal path	 SEC Data provenance documentation, Data lineage and transformation records	 Cross-Border Transfer Impact Assessment - Identification of all data flows involving Vietnamese personal data leaving Vietnam - Legal basis for the transfer - Risk assessment for the transfer destination and processing arrangement - MPS notification or approval status where required - Contractual safeguards with overseas processors

How Did Things Go? (We Really Want To Know)

Did you enjoy this session? Is there any way we could make it better? Let us know by filling out a speaker evaluation.

1. Open the IAPP Events app.
2. Select **IAPP Forum 2026**.
3. Tap "Schedule" on the bottom navigation bar.
4. Find this session. Click "Rate this Session" within the description.
5. Once you've answered all three questions, tap "Done".

Thank you!