MindMagine

Implementing AI Governance

The Malaysia Roadmap

A practical guide for managers, from global frameworks to a working governance system

Sean Low

MindMagine Asia Sdn. Bhd. · Kuala Lumpur

Implementing AI Governance: The Malaysia Roadmap

By Sean Low

October 2026 edition

Published by MindMagine Asia Sdn. Bhd., Kuala Lumpur, Malaysia. [ADDRESS] · [WEBSITE]

© 2026 MindMagine Asia Sdn. Bhd. [COMPANY REGISTRATION NO.]. All rights reserved. No part of this book may be reproduced in any form without written permission from the publisher, except for brief quotations in reviews.

ISBN: [ISBN]

Not legal advice. This book explains laws, guidelines and standards in plain language for education and training. It is not formal legal advice. Laws and regulations change, and how they apply depends on the facts. Check specific situations with your Data Protection Officer (DPO) or legal counsel.

Fictional case studies. Syarikat Jaya Utama (SJU), TalentFit AI, Pembiayaan Digital Sdn Bhd (PDSB), Rakyat-Score AI and every person named in their stories are fictional, created for training. Any resemblance to real organisations, products or people is coincidental. The real-world incidents in Chapter 1 are drawn from public news reports, court records and official statements.

Currency. The legal and regulatory position described is as understood in October 2026. Malaysia's proposed AI Governance Bill had completed public consultation but had not been tabled in Parliament at that date.

Trademarks. ISO and IEC standard designations belong to the International Organization for Standardization and the International Electrotechnical Commission. Product and company names in the real-world incidents belong to their owners and are used only for commentary and education.

Contents

Preface 5
How to use this book 6

Part 1 · Foundations 8
1 Why AI governance matters now 9
2 AI in plain words 13
3 Must, Should, Coming: three kinds of rules 15

Part 2 · The landscape 18
4 The global picture 19
5 Malaysia's rules 22
6 Malaysian values: AIGE and the Rukun Negara 25

Part 3 · Governing one AI system 28
7 The lifecycle and human oversight 29
8 Data governance and the PDPA 32
9 Auditing an AI system: the SJU case 35

Part 4 · Building a governance system 39
10 AI management systems 40
11 Who does what: committees and the Three Lines Model 43
12 Risk and impact 46
13 Shadow AI, vendors and monitoring 50
14 Red-teaming a high-risk system: PDSB 53
15 Your first 90 days 55

Appendix A · The SJU case file 57
Appendix B · The PDSB case 62
Appendix C · Quick reference 63
Answers to "Check your understanding" 67
Glossary 70
Sources and further reading 72

Preface

Artificial intelligence has moved quickly from pilot projects into the core of business. Machine learning models now screen job applicants, assess creditworthiness, optimise supply chains and talk directly with customers and citizens. Without management, though, automated decision-making brings serious operational, reputational and legal risks.

In Malaysia, AI governance sits in a landscape that is still moving. The National Guidelines on AI Governance and Ethics (AIGE) from the Ministry of Science, Technology and Innovation (MOSTI) set the voluntary ethical baseline. The Personal Data Protection Act 2010 (Act 709, as amended by Act A1727) imposes enforceable legal duties. International standards such as ISO/IEC 42001:2023 offer certifiable management frameworks. And a dedicated AI Governance Bill is on its way.

This book gives corporate leaders, compliance managers, IT directors and legal advisers a practical, step-by-step roadmap for putting effective AI governance in place. It grew out of MindMagine's two-day course, Implementing AI Governance: The Malaysia Roadmap, and can be read on its own or alongside the course.

Sean Low
Kuala Lumpur, October 2026

How to use this book

Who it is for

This book is for managers in Malaysia who are new to AI governance: HR, compliance, risk, IT, legal, operations and business leaders. You don't need a technical background. Every term is explained when it first appears, and there is a glossary at the back.

How it is organised

Part	What it covers	Chapters
1 · Foundations	Why AI governance is urgent, core machine-learning ideas in plain language, and the Must, Should, Coming rule-sorting framework	1–3
2 · The landscape	Global frameworks (EU AI Act, ISO/IEC 42001, OECD), Malaysia's layers of rules, and AI ethics and the Rukun Negara	4–6
3 · Governing one AI system	The AI lifecycle, human oversight (HITL, HOTL, HIC), data governance under the PDPA, and a worked audit of the SJU case	7–9
4 · Building a governance system	An ISO/IEC 42001 AI management system, the Three Lines Model, AI impact assessments, shadow AI and vendors, the PDSB red-team case, and your first 90 days	10–15

The two cases

Two fictional companies run through the book. **Syarikat Jaya Utama (SJU)**, a logistics firm, bought an AI hiring tool called TalentFit and got almost everything wrong. You meet SJU in Chapter 1, audit it in Chapter 9 and build its governance system in Part 4. **Pembiayaan Digital Sdn Bhd (PDSB)**, a fintech lender, is the harder case you red-team in Chapter 14. Both case files are in the appendices.

Features in each chapter

Feature	What it does
In this chapter	What you will be able to do by the end
Key message	The one idea to remember
Case boxes	How the idea plays out at SJU or PDSB
Must, Should, Coming tags	Whether a rule is binding law, voluntary guidance or proposed law
Summary	The chapter in a few lines
Check your understanding	Questions to test yourself; answers are at the back

If you are on the course

Each chapter matches one or more course sessions. Read Chapters 1–9 around Day 1, and Chapters 10–15 around Day 2. Your participant workbook holds the activity sheets; this book holds the explanations.

PART 1

Foundations

Why AI governance matters now, the words you need, and the one idea that holds the rest of the book together.

CHAPTER 1

Why AI governance matters now

In this chapter you will learn to

- explain why AI creates risks that ordinary IT governance misses
- describe four real incidents and the lesson in each
- outline Malaysia's current position on AI regulation

AI saves time. That is why organisations adopt it, and it is also why they stop paying attention to it. A tool that shortlists candidates in two days instead of three weeks feels like a pure gain. Nobody asks what it learned, whom it rejects, or where the data went, until something goes wrong.

AI governance is the set of rules, roles and routines that make sure someone does ask. It is not a brake on AI. It is what lets an organisation keep using AI with confidence after the first problem appears.

KEY MESSAGE An unchecked AI system can be wrong, unfair or leak data at scale. Governance makes sure someone is in charge of it.

What makes AI different

Ordinary software follows rules that people wrote. If it gives a wrong answer, someone can read the code and find the mistake. Most modern AI works differently. It learns patterns from examples, and those patterns are often hard to explain, even for the people who built it. Three features follow from that.

- **It inherits the past.** An AI system trained on historical decisions repeats the habits, and the biases, in those decisions.
- **It acts at scale.** A human recruiter who misjudges one candidate affects one person. A screening tool that misjudges a group affects every applicant from that group, every day.
- **It can be confidently wrong.** AI that writes text can produce false answers that read exactly like true ones.

None of this means AI is bad. It means AI needs checks that ordinary software doesn't.

Four times automated decisions went wrong

The four incidents below are real. Each one teaches a different governance lesson.

Amazon's hiring tool

In 2018 Reuters reported that Amazon had built an experimental tool to rate job applicants from one to five stars. It had been trained on ten years of CVs, mostly from men, because the tech industry was mostly male. The tool learned to mark down CVs that contained the word "women's", as in "women's chess club captain". Amazon edited the tool to remove that pattern

but could not be sure others wouldn't appear, and scrapped the project. The company said recruiters never relied on its rankings alone.

Lesson: old data repeats old bias. If you train AI on past decisions, you must test whether those decisions were fair.

Air Canada's chatbot

In 2022 a passenger asked Air Canada's website chatbot about bereavement fares after his grandmother died. The chatbot told him he could claim a discount after travelling. That was wrong. When he claimed it, the airline refused and argued in a tribunal that the chatbot was responsible for its own words. In February 2024 British Columbia's Civil Resolution Tribunal rejected that argument. It found the airline responsible for all the information on its website, whether it came from a static page or a chatbot, and ordered it to pay.

Lesson: you own your AI's answers. An organisation cannot hide behind its own system.

Australia's "Robodebt" scheme

Between 2015 and 2019, the Australian government used an automated income-matching system to raise welfare debts. Letters went out with no human checking whether each debt was real. A court later heard that about A$1.73 billion in debts had been unlawfully raised against some 433,000 people. The government settled a class action in a deal reported at about A$1.8 billion, and a Royal Commission later examined how the scheme was allowed to run.

Lesson: high-impact decisions about individuals need meaningful human review. Automating a flawed rule spreads the flaw to hundreds of thousands of people.

Samsung and ChatGPT

In April 2023 Samsung engineers pasted internal source code into ChatGPT to help fix it. Once submitted, that code sat on an outside company's servers. In May 2023 Samsung banned generative AI tools on company devices while it built its own.

Lesson: set rules before people pick their own tools. Banning tools after a leak is harder than offering a safe option first.

The Malaysian context: deepfake fraud

Malaysia is already living with the downside of ungoverned AI. The Commercial Crime Investigation Department of Bukit Aman reported 454 fraud cases involving deepfake technology, with losses of about RM2.72 million, in its 2024 figures. Scammers have used AI-generated videos of national leaders and well-known business figures to promote fake investments. In July 2025 the Johor Royal Press Office issued a warning about a deepfake video of His Majesty the King.

The lesson for organisations is twofold. People need to know what is AI-made, so transparency protects customers and reputations. And regulators are closing gaps: deploying unvetted AI tools that handle customer data already exposes Malaysian companies to penalties under privacy and sector laws.

Malaysia's position today

Malaysia does not yet have a law written specifically for AI. That does not mean AI is unregulated. Three layers apply.

- **The Personal Data Protection Act 2010 (Act 709)**, as amended by the Personal Data Protection (Amendment) Act 2024 (Act A1727), applies to any AI that processes personal data in commercial transactions. Breaching its principles can now bring a fine of up to RM1 million and/or up to three years in prison.
- **The National Guidelines on AI Governance and Ethics (AIGE)**, launched in September 2024, set out seven principles for responsible AI. They are voluntary.
- **The AI Governance Bill** is in preparation. The National AI Office (NAIO) held a public consultation in July 2026, and in October 2026 the Minister of Digital said the Bill was expected to be tabled early in 2027.

Chapter 5 covers each layer in detail. For now, the point is simple: the law already applies, and more is coming.

CASE · Meet Syarikat Jaya Utama

Syarikat Jaya Utama Sdn Bhd (SJU) is a logistics and warehousing company with 800 employees. It plans to fill about 120 roles this year and received more than 3,000 applications in its last hiring round. Its six-person HR team fell weeks behind.

To clear the backlog, the HR Director bought TalentFit AI, a cloud service that screens CVs, ranks candidates and predicts which staff are ready for promotion. She paid from the HR budget, accepted the vendor's online terms, and did not involve IT, Risk or Legal.

Six weeks later, shortlisting took two days instead of three weeks. Then an analyst spotted a pattern in the scores, a rejected applicant complained, and IT found the subscription in the expense claims. You will audit TalentFit in Chapter 9. The full case file is in Appendix A.

Summary

- AI learns from the past, acts at scale and can be confidently wrong. Those features need checks ordinary software doesn't.
- Real incidents show four lessons: test for bias, keep humans reviewing high-impact decisions, set rules before people choose tools, and own your AI's answers.
- In Malaysia the PDPA already binds AI that uses personal data; the AIGE Guidelines guide; an AI Governance Bill is coming.

Check your understanding

1. What was the core governance failure in Australia's Robodebt scheme?
2. Name two features of AI that make it riskier than ordinary rule-based software.
3. What argument did Air Canada make, and why did the tribunal reject it?
4. Which of the four incidents is closest to SJU's situation, and why?
5. Is AI unregulated in Malaysia because there is no AI-specific law yet?

CHAPTER 2

AI in plain words

In this chapter you will learn to

- explain eight basic AI terms without jargon
- describe the four points where risk enters an AI system
- trace SJU's problems back to those four points

You don't need to know how AI works inside to govern it, just as a finance director doesn't need to build the accounting software. You do need a shared vocabulary, so that when a vendor or a data scientist explains a system, you can ask the right questions.

KEY MESSAGE A model can only learn from the past it is shown. Bias in, bias out.

Eight words to know

Term	What it means	Why it matters for governance
Artificial intelligence (AI)	Software that does tasks we'd normally expect to need human judgement.	The scope of your governance: if it judges, it counts.
Machine learning (ML)	AI that learns patterns from examples instead of following hand-written rules.	Its behaviour depends on its data, not just its code.
Generative AI	AI that creates new text, images, audio or code. ChatGPT is one example.	It can produce convincing falsehoods and may expose what users type in.
Model	The trained "engine" that turns an input into an output.	Models can be bought, updated or swapped, sometimes without you knowing.
Training data	The examples a model learns from.	Its gaps become the model's gaps. Its legality becomes your problem.
Bias	Outputs that are unfair to a group, again and again.	Can harm people and breach your values, and sometimes the law.
Hallucination	A confident answer that is simply false.	Someone must check outputs before they reach customers or decisions.
Drift	A model getting worse over time as the world around it changes.	A system that passed testing last year may fail today. Monitoring is not optional.

Where risk enters an AI system

Every AI system follows the same basic path: data goes in, the model learns, it produces an output, and someone acts on that output. Risk can enter at each step.

Step	What can go wrong	The governance question
1 Data in	Data is old, skewed, inaccurate or collected without a lawful basis.	Where did this data come from, and may we use it this way?
2 Learning	The model picks up patterns nobody checked or understood.	What did it learn, and have we tested it on every group it affects?
3 Output	Answers are wrong, unfair or invented.	How do we know the outputs are accurate and fair?
4 Action	A decision about a person happens with no human check.	Who reviews the decision before it takes effect?

This is why a single sign-off at the end of a project is never enough. A problem that enters at step one may not show until step four, and by then it has affected hundreds of people.

CASE · SJU at each step

Data in: HR uploaded five years of staff records, including IC numbers and medical histories, without telling employees.

Learning: TalentFit learned from hiring history that came mostly from urban offices.

Output: female applicants from rural areas received much lower scores for technical management roles.

Action: candidates in the bottom 40% received an automatic rejection email, usually within minutes, with no recruiter review.

One problem at every step. Each one was preventable with a single question asked at the right time.

Summary

- You need vocabulary, not technical skill, to govern AI.
- Training data shapes a model’s behaviour, including its biases.
- Risk can enter at four points: data in, learning, output and action. Governance puts a check on each.

Check your understanding

1. In one sentence, what is drift, and why does it matter after a system goes live?
2. Why can’t you judge an ML system just by reading its code?
3. At which of the four steps did SJU’s bias problem enter, and where did it show?

CHAPTER 3

Must, Should, Coming: three kinds of rules

In this chapter you will learn to

- sort any AI rule into binding law, voluntary guidance or proposed law
- use the right words for each kind of rule
- explain why the difference matters to decisions and budgets

Managers new to AI governance meet a wall of names: the PDPA, AIGE, ISO/IEC 42001, the EU AI Act, OECD principles, NAIO, BNM RMiT. It is easy to treat them all as equally important, or all as optional. Neither is right.

The most useful skill in this book is also the simplest: for any rule, ask whether it is a Must, a Should or a Coming.

KEY MESSAGE Some AI rules are law, some are guidance, and some are coming. Knowing which is which is the first governance skill.

Tag	What it is	What happens if you ignore it	Examples	Say
Must	Binding law	Fines up to RM1 million, prosecution, regulator orders, civil claims	PDPA 2010 (Act 709, amended by Act A1727); sector rules such as BNM's RMiT for financial institutions or the Consumer Credit Act 2025 for credit providers	"breaches"
Should	Voluntary guidance or standards	No legal penalty, but loss of trust, audit nonconformities, weaker defence, lost contracts	MOSTI's AIGE Guidelines (2024); ISO/IEC 42001:2023; OECD AI Principles	"falls short of"
Coming	Proposed law	Nothing yet, but costly catch-up later	Malaysia's AI Governance Bill	"prepare for"

Why the words matter

Saying an organisation "violated the AIGE Guidelines" is wrong, because there is nothing to violate: the guidelines are voluntary. Saying it "fell short of" them is accurate. The difference is not pedantic. A compliance report that mixes the two will either frighten leaders about rules that carry no penalty, or let them relax about rules that do.

Use the words consistently and others will start to as well. It is one of the fastest ways to make AI governance conversations clearer.

A Should can become a Must

The tags describe a rule today, not forever. A Should can become a Must, or start to carry legal weight, in several ways:

- **A contract requires it.** When you sign a customer contract or service agreement that commits you to ISO/IEC 42001 or the AIGE principles, those commitments become enforceable contract terms.
- **A regulator adopts it.** Sector regulators such as Bank Negara Malaysia or the Securities Commission can turn voluntary principles into binding sector guidelines.
- **A law adopts it.** Commentators expect the AIGE principles to shape Malaysia's AI Governance Bill.
- **You commit to it.** If your published policy says you follow the AIGE principles, customers and regulators can hold you to that statement.
- **A court looks to it.** In a negligence claim over harm caused by AI, a court may consider whether an organisation followed recognised standards such as AIGE or ISO/IEC 42001 when deciding what reasonable care looked like.

And one fact can carry two tags at once. Uploading health data without consent breaches the PDPA (a Must) and falls short of the AIGE principle of privacy and security (a Should).

CASE · Sorting five facts about SJU

Fact	Tag	Why
HR uploaded staff medical histories without telling employees	Must	Health data is sensitive personal data: it needs explicit consent (PDPA s.40), and staff must be given notice (s.7)
Rejected candidates get an automatic email; no person reviews it	Should (possibly Coming)	AIGE recommends human oversight; the PDPA has no general rule requiring human review of automated decisions
SJU has no Data Protection Officer	Must	Mandatory under PDPA s.12A since 1 June 2025
The tool scores rural women lower	Should; possibly Must	Falls short of AIGE Fairness; ask counsel whether employment law protects job applicants from this kind of discrimination
The vendor stores data outside Malaysia	Must	Not illegal in itself, but PDPA conditions on cross-border transfers apply (s.129)

Most beginners first tag the automatic rejections as a Must and the missing DPO as a Should. It is the other way round.

Summary

- Must = binding law (“breaches”). Should = voluntary guidance (“falls short of”). Coming = proposed law (“prepare for”).
- The tags change over time, and a Should can become a Must through law, contract or your own commitments.
- One fact can carry more than one tag.

Check your understanding

1. Is ISO/IEC 42001 a Must, a Should or a Coming for a typical Malaysian company? When could that change?
2. A colleague writes “our chatbot violates the AIGE Guidelines”. How would you correct it?
3. Give an example of a single fact that carries both a Must and a Should.
4. Name two ways a voluntary standard can start to carry legal weight.

PART 2

The landscape

The global frameworks Malaysian rules draw on, the rules that apply in Malaysia today, and the national values that shape them.

CHAPTER 4

The global picture

In this chapter you will learn to

- describe the EU AI Act's four risk tiers and place an AI use in the right one
- explain what ISO/IEC 42001 is and when it matters
- name the shared principles behind most AI frameworks

Malaysian organisations sell to, buy from and partner with companies in other countries. Their customers and auditors bring global expectations with them. And Malaysia's own rules draw on international work. So even a company that only operates locally benefits from knowing three global reference points.

KEY MESSAGE Risk depends on the use, not the technology. The same model can be harmless in one job and high-risk in another.

The EU AI Act

The European Union's AI Act, adopted in 2024, was the world's first comprehensive, cross-sector AI law. It sorts AI systems by the risk their use poses to people, and sets heavier duties the higher the risk.

Tier	What happens	Examples
Unacceptable risk	Banned	Harmful subliminal manipulation; exploiting people's vulnerabilities; social scoring; emotion recognition in the workplace (with medical and safety exceptions)
High risk	Strict duties before and during use: risk management, data governance (Article 10), technical documentation, logging, transparency and human oversight (Article 14)	Recruitment and other employment decisions; credit scoring of individuals (Annex III, point 5(b)); critical infrastructure; education; law enforcement
Limited risk	Transparency (Article 50): tell people they are dealing with AI, and label AI-generated content	Chatbots; deepfakes and other AI-generated content
Minimal risk	No extra duties	Spam filters; AI in video games

Fines are tiered. Using a banned practice can cost up to €35 million or 7% of global annual turnover, whichever is higher. Breaching the duties for high-risk systems can cost up to €15 million or 3%. Giving regulators incorrect or misleading information can cost up to €7.5 million or 1%.

The Act's duties take effect in phases. The EU has proposed changes to the timetable for high-risk duties, so check current dates before relying on them.

Why a Malaysian company should care. The Act can reach organisations outside the EU when their AI systems are placed on the EU market or their outputs are used in the EU. A Malaysian firm with European customers, or a European parent, may be in scope. Even if you are not, the tiers are a useful way to think about risk, and Malaysia's proposed Bill also uses risk tiers.

ISO/IEC 42001

ISO/IEC 42001, published in December 2023, is the world's first certifiable management system standard for AI. It uses the same high-level structure as ISO 27001: rather than telling you what each AI system must do, it tells you how to run an organisation that governs AI well. Its requirements sit in Clauses 4 to 10, and Annex A sets out 38 reference controls.

It is voluntary. An organisation can choose to be audited and certified against it, and customers may ask suppliers to do so. Part 4 of this book uses its structure to build a governance system step by step.

The OECD AI Principles

The Organisation for Economic Co-operation and Development adopted its AI Principles in 2019 and updated them in 2024. They set out values for trustworthy AI, including respect for human rights, transparency, robustness and accountability. Many national frameworks, including Malaysia's AIGE Guidelines, draw on them alongside guidance from UNESCO and the European Commission.

The common thread

Read enough frameworks and the same ideas keep returning: **fairness, safety, transparency, privacy and human accountability.** If you build your governance around these five, you will be aligned with most frameworks you meet, including the Malaysian ones in the next chapter.

CASE · Which tier is TalentFit?

TalentFit screens CVs and ranks job applicants. Under the EU AI Act, AI used in recruitment is a high-risk use listed in Annex III. If SJU hired in Europe, it would face strict duties before using TalentFit at all, and fines of up to €15 million or 3% of turnover for breaching them.

SJU doesn't operate in Europe, so the Act doesn't bind it. But the classification tells SJU something useful: the rest of the world considers this kind of system high-risk, and a careful company would treat it that way.

Summary

- The EU AI Act sorts AI into four tiers by the risk of its use, with duties rising with risk. It can reach some non-EU organisations.
- ISO/IEC 42001 is a voluntary, certifiable standard for running an AI management system.

- Most frameworks share five ideas: fairness, safety, transparency, privacy and human accountability.

Check your understanding

1. Place these in EU AI Act tiers: a website FAQ chatbot; a bank's credit-scoring model; a spam filter.
2. Which EU AI Act fine tier applies to breaching high-risk duties, and which to banned practices?
3. How is ISO/IEC 42001 different from a law?
4. Why might a Malaysian company with no EU business still use the EU's risk tiers?

CHAPTER 5

Malaysia's rules

In this chapter you will learn to

- describe the three layers of AI rules in Malaysia
- list what the 2024 PDPA amendments changed and when
- explain what the AI Governance Bill proposes
- recognise when sector rules add duties

Malaysia's approach has moved in stages: voluntary guidance first, then stronger data protection law, and now a dedicated AI law in draft. Organisations that understand the stages can see where the next one is heading.

KEY MESSAGE The PDPA binds you today. The AIGE Guidelines guide you. The AI Governance Bill is close.

When	What happened	Tag
September 2024	AIGE Guidelines launched by the Ministry of Science, Technology and Innovation (MOSTI); the National AI Office (NAIO) announced	Should
January–June 2025	PDPA amendments (Act A1727) take effect in three phases	Must
March 2026	Consumer Credit Act 2025 comes into force	Must (credit sector)
July 2026	Public consultation on the proposed AI Governance Bill opens and closes	Coming
October 2026	Minister of Digital says the Bill is expected to be tabled early in 2027	Coming

Layer 1 · Must: the Personal Data Protection Act

The Personal Data Protection Act 2010 (Act 709) governs how organisations process personal data in commercial transactions. It does not mention AI, but most AI that touches customers or staff processes personal data, so the Act applies. At its heart are seven data protection principles in sections 6 to 12, covered in Chapter 8.

The Personal Data Protection (Amendment) Act 2024 (Act A1727) strengthened the PDPA in ways that matter for AI. It came into force in three phases in 2025.

Section	Change	In force
s.5	Fines for breaching the principles rise to up to RM1 million, and prison terms to up to three years	1 April 2025
s.9	Data processors, such as AI vendors processing data for you, now carry the security duty directly	1 April 2025
s.129	The whitelist for cross-border transfers is removed; transfers are allowed where the destination offers protection at least equivalent to the PDPA, or other set conditions are met	1 April 2025
s.12A	Data controllers and processors must each appoint at least one Data Protection Officer (DPO)	1 June 2025
s.12B	Personal data breaches must be reported to the Commissioner as soon as practicable, and to affected people where significant harm is likely	1 June 2025
s.43A	People gain a right to data portability	1 June 2025

The amendments also replaced the term "data user" with "data controller", and added biometric data to sensitive personal data. The Commissioner's guidelines set out the detailed breach-notification process.

Layer 2 · Should: the AIGE Guidelines

The National Guidelines on AI Governance and Ethics, launched by MOSTI in September 2024, set out seven principles for responsible AI and separate guidance for three groups: AI end users; policymakers in government, agencies, organisations and institutions; and developers, designers, technology providers and suppliers. They are voluntary. Chapter 6 covers the principles.

Voluntary does not mean unimportant. The AIGE principles are the clearest signal of what Malaysian regulators consider good practice, and commentators expect them to shape the AI Governance Bill.

Layer 3 · Coming: the AI Governance Bill

In July 2026 the National AI Office (NAIO), under the Ministry of Digital, released a public consultation paper on a proposed AI Governance Bill. Summaries of the paper describe four main features:

- **Wide reach:** AI systems sold, built or used in Malaysia, or used by a deployer based in Malaysia, wherever the system is hosted.
- **A Central AI Authority** responsible for AI safety, investigation and enforcement, and AI enablement, working with existing sector regulators.
- **Harm-based risk tiers:** Tier 1 unacceptable risk, Tier 2 high risk and Tier 3 low risk, judged against harms such as death, bodily injury, unlawful deprivation of constitutional liberties or breaking a written law.
- **An AI sandbox** for supervised testing before full launch.

As of October 2026 the Bill had not been tabled in Parliament. The Minister of Digital said in October 2026 that tabling was expected early in 2027, and its penalties and start date had not

been published. The final law may differ from the consultation paper. What is already clear is the direction: risk-based duties, a central regulator and accountability across the AI lifecycle. Governance built now on the PDPA and AIGE is the foundation the Bill will expect.

Sector rules

Some sectors carry extra rules that apply to AI as to any other technology.

- **Financial institutions** regulated by Bank Negara Malaysia must follow its Risk Management in Technology (RMiT) policy.
- **Credit providers** outside the banking system come under the Consumer Credit Act 2025, which came into force on 1 March 2026 and is overseen by a new Consumer Credit Commission. It requires providers to act fairly, responsibly and professionally, and non-bank lenders to check that borrowers can repay without undue hardship. Whether a particular business is covered depends on the activities listed in the Act's schedules. Chapter 14 applies this to an AI credit model.

If you work in a regulated sector, check your regulator's technology, outsourcing and consumer-protection rules alongside the PDPA.

CASE · SJU and the three layers

SJU has no Data Protection Officer. Is that a Must, a Should or a Coming? It is a Must: under section 12A, every data controller has needed at least one DPO since 1 June 2025. It is the easiest point in the whole case, and SJU still missed it.

Summary

- Three layers apply: the PDPA (Must), the AIGE Guidelines (Should) and the AI Governance Bill (Coming). Sector rules may add more Musts.
- The 2024 PDPA amendments added DPOs, breach notification, processor duties and much higher penalties.
- The Bill proposes wide reach, a Central AI Authority, three harm-based risk tiers and a sandbox. Tabling is expected early in 2027; check its status before relying on any detail.

Check your understanding

1. Name three duties added by the 2024 PDPA amendments that came into force on 1 June 2025.
2. What is the maximum penalty for breaching the PDPA principles since April 2025?
3. Give two features of the proposed AI Governance Bill.
4. A non-bank lender uses AI to approve loans. Which rules beyond the PDPA should it check?

CHAPTER 6

Malaysian values: AIGE and the Rukun Negara

In this chapter you will learn to

- name and explain the seven AIGE principles
- apply the principles to a real AI system
- explain how the AIGE Guidelines link the Rukun Negara to AI ethics

Rules tell you what not to do. Principles tell you what to aim for. The AIGE Guidelines set out seven principles that together describe trustworthy AI in a Malaysian context.

KEY MESSAGE The seven AIGE principles are voluntary, but they are the clearest picture of what Malaysian regulators consider good AI.

The seven principles

Principle	In plain words	A question to ask
1 Fairness	No bias or discrimination against any group.	Have we tested the results for every group the system affects?
2 Reliability, safety and control	Works as intended, even when things go wrong. People can step in.	What happens when it meets an input it wasn't built for?
3 Privacy and security	Personal data protected. Systems resist misuse and attack.	What data goes in, where does it go, and who can reach it?
4 Inclusiveness	Benefits every group in Malaysian society.	Who might be left out: rural users, older users, other languages?
5 Transparency	People can understand how decisions are made.	Could we explain a decision to the person it affects?
6 Accountability	A named person or organisation answers for the outcome.	Whose name is on this system?
7 Pursuit of human benefit and happiness	AI serves people's well-being, and humans can intervene.	Does this make people's lives better, and can a human overrule it?

The principle names are as published. The plain-language meanings and questions are summaries for learning, not quotations from the AIGE.

Notice that several principles overlap with the PDPA. Privacy and security is also a legal duty. That overlap is common: a Should often sits on top of a Must, asking for more than the law requires.

Using the principles

The principles are most useful as a checklist when you assess an AI system. For each one, ask what "good" would look like for this system, and name one thing you could check. Avoid treating them as a scorecard where every system needs top marks on all seven. A spam filter barely touches fairness; a hiring tool lives or dies by it.

AIGE and the Rukun Negara

For many Malaysian leaders, AI ethics lands better when it doesn't feel like an imported framework. The AIGE Guidelines address this directly. They start from the Federal Constitution, which under Article 4(1) is the supreme law of the Federation, and in Section 2.6.2 they link responsible AI to the five principles of the Rukun Negara, Malaysia's national philosophy. AIGE groups them in four pairings.

Rukun Negara principle	AIGE pairing	What it asks of AI	Closest AIGE principles
Belief in God (*Kepercayaan kepada Tuhan*)	Belief in God and Fairness	Respect diverse religious views; avoid bias, including on religious belief; uphold equality and non-discrimination	Fairness; Inclusiveness
Loyalty to King and Country (*Kesetiaan kepada Raja dan Negara*)	Loyalty to King and Country and Transparency	Transparent AI decisions and accountable institutions; protect national security and the public interest	Transparency; Accountability
Supremacy of the Constitution (*Keluhuran Perlembagaan*) and Rule of Law (*Kedaulatan Undang-undang*)	Supremacy of the Constitution and Rule of Law	Comply with the law and respect constitutional rights: safety and security, inclusivity, and human control and oversight so AI never acts autonomously against people's rights	Reliability, safety and control; Inclusiveness
Courtesy and Morality (*Kesopanan dan Kesusilaan*)	Courtesy & Morality and Human Dignity	Respect privacy and human dignity; promote social responsibility and the social good	Privacy and security; Pursuit of human benefit and happiness

Source: National Guidelines on AI Governance and Ethics (MOSTI, 2024), Sections 2.6 and 2.6.2. The pairings are AIGE's; the "closest AIGE principles" column summarises the text of each pairing.

The value of these pairings is practical. Leaders who find global frameworks abstract often respond more readily to values they already share. Framing AI governance as an expression of the Rukun Negara, as the government itself does, can make it easier to win support.

Use AIGE's own pairings

When you present the Rukun Negara link inside your organisation, quote AIGE's pairings rather than inventing your own. Belief in God is personal, and the guidelines handle it carefully: they ask AI to respect diverse religious views and avoid bias, not to take a religious position.

CASE · Where SJU falls short

SJU fact	AIGE principle it falls short of
Medical histories uploaded without notice	Privacy and security (and a PDPA Must)
Automatic rejections, no human review	Accountability; Reliability, safety and control
No Data Protection Officer	Accountability (and a PDPA Must)
Rural women scored lower	Fairness; Inclusiveness
Data on offshore servers, no contract terms	Privacy and security

Summary

- The seven AIGE principles: fairness; reliability, safety and control; privacy and security; inclusiveness; transparency; accountability; the pursuit of human benefit and happiness.
- Use them as a checklist, weighted to the system in front of you.
- AIGE Section 2.6.2 links the five Rukun Negara principles to responsible AI in four pairings, grounded in the Federal Constitution (Article 4(1)).

Check your understanding

1. Which AIGE principle asks "whose name is on this system?"
2. Why might a spam filter and a hiring tool need very different attention to the Fairness principle?
3. Give one AIGE principle that overlaps with a PDPA duty.
4. Which AIGE pairing covers human control and oversight of AI?

PART 3

Governing one AI system

How to look at a single AI system from start to finish: its lifecycle, who stays in control, the data it uses, and how to audit it.

CHAPTER 7

The lifecycle and human oversight

In this chapter you will learn to

- describe the four stages of an AI system's lifecycle and the governance questions at each
- explain three models of human oversight: HITL, HOTL and HIC
- choose an oversight model for a given AI use

An AI system is not a product you buy once and forget. It is designed, tested, launched and then run, often for years, while the world around it changes. Governance has to follow it through every stage.

KEY MESSAGE Governance is not one sign-off at the end. Each stage of the lifecycle has its own questions and its own owner.

Four stages

Stage	What happens	Governance questions
1 Design, plan and model	Define the problem, choose or collect data, decide how the system will work	What problem are we solving? Is AI the right tool? What data will we use, and may we use it?
2 Verify and validate	Train and test the model before use	Does it work accurately for every group it affects? Has it been tested for bias and security weaknesses?
3 Deploy and use	Launch the system and connect it to real work	Who approved the launch? Do users and affected people know it's AI? Who reviews its decisions?
4 Operate and monitor	Run, watch, update and eventually retire the system; record event logs (ISO/IEC 42001 control A.6.2.8)	Is it drifting? Who logs incidents? When do we re-assess it? Who can switch it off?

These four stages follow a widely used model of the AI lifecycle, close to the one in the OECD's 2019 Recommendation on AI. The AIGE Guidelines ask for accountability and privacy "at each stage" and "throughout the entire life cycle", but don't name stages of their own.

Buying AI doesn't skip any stage. When you buy a system rather than build it, the vendor does much of stages 1 and 2. You still need answers to every question, which means asking the vendor for them before you sign.

Keeping humans in control

One question matters more than most at stage 3: when the AI makes a decision, what role does a person play? The AIGE Guidelines name three forms of human oversight (Section 2.7.3).

Model	How it works	Use it for	Example
HITL · human-in-the-loop	A person approves each decision before it takes effect.	High-impact decisions about individuals	A recruiter reviews every rejection before it is sent
HOTL · human-on-the-loop	The AI acts on its own; a person watches live and can override.	Fast, high-volume decisions where delay causes harm	A fraud team monitors automatic card blocks and can release them
HIC · human-in-command	People set the rules, decide whether the system runs, and can switch it off.	Every AI system, always	Management approves the system's use and reviews it every quarter

HIC is not an alternative to the other two. It sits above them. Even a fully automated spam filter should have someone with the authority to change it or turn it off.

Choosing a model

Three questions decide the right level of oversight:

1. **How severe is the harm if it gets a decision wrong?** The more a decision affects someone's job, money, health or rights, the closer a person should be.
2. **How fast must decisions be made?** Where seconds matter, HITL may be impossible, so strong HOTL is the answer.
3. **How many people are affected?** Errors at scale need closer watching, even when each one is small.

Human oversight is a **Should** in Malaysia today: the AIGE Guidelines describe these mechanisms as part of responsible AI. The PDPA has no general rule requiring human review of automated decisions, such as hiring decisions, though sector rules may set their own expectations. The AI Governance Bill may change that.

Watch for "rubber-stamping"

HITL only works if the human has the time, information and authority to disagree with the AI. A recruiter asked to approve 300 rejections an hour is not in the loop. They are a signature.

CASE · SJU's Auto-Reject

SJU switched on Auto-Reject: the bottom 40% of applicants receive a rejection email within minutes, and no recruiter sees them. One applicant with nine years of directly relevant experience was rejected four minutes after applying.

Rejection affects a person's livelihood, so the right model is HITL: a recruiter reviews every rejection. Auto-Reject should never have reached stage 3 without that question being asked, and with HIC in place, someone senior would have had to approve it.

Summary

- The lifecycle has four stages: design, validate, deploy, operate. Each has its own governance questions.
- HITL: a person approves each decision. HOTL: a person watches and can override. HIC: people decide whether the system runs at all.
- Choose by severity of harm, speed and scale. Human oversight is a Should in Malaysia today.

Check your understanding

1. At which lifecycle stage should SJU have tested TalentFit for bias?
2. A bank's system blocks suspicious card payments instantly. Which oversight model fits, and why?
3. Why is HIC needed even when HITL is in place?

CHAPTER 8

Data governance and the PDPA

In this chapter you will learn to

- apply the PDPA's seven principles to AI
- explain lawful grounds for processing, including for sensitive data
- check data at each step of an AI system

AI runs on data, and the PDPA governs data. For most organisations, this chapter is where AI governance meets the law most directly.

KEY MESSAGE Consent is not the only lawful ground for using personal data, but sensitive data such as health information needs a stronger one.

The seven principles, applied to AI

Principle	Section	What it means for AI
General	s.6	Have a lawful basis before you put personal data into an AI system.
Notice and Choice	s.7	Tell people their data is being processed and why, including for AI.
Disclosure	s.8	Don't use or disclose data for a purpose people weren't told about without consent.
Security	s.9	Protect data in training sets, prompts, AI tools and vendors' systems.
Retention	s.10	Don't keep data, including training data, longer than you need it.
Data Integrity	s.11	Keep data accurate, complete and up to date.
Access	s.12	People can see and correct the personal data you hold about them.

Breaching any principle can bring a fine of up to RM1 million and/or up to three years' imprisonment (s.5).

Lawful grounds: consent is one door, not the only one

A common mistake is to assume that every use of personal data in AI needs explicit consent. Under the General Principle (s.6), consent is one basis. The PDPA also allows processing that is necessary for certain purposes, such as performing a contract with the person (an employment contract, for example) or complying with a legal obligation.

Sensitive personal data is different. Data about a person's physical or mental health, religious beliefs, political opinions, biometric data and certain other categories needs explicit consent, with only narrow exceptions (s.40). A general contract clause is not enough.

Kind of data	Typical basis	AI example
Ordinary personal data	Consent, or another lawful ground such as an employment contract	Using performance ratings to plan promotions
Sensitive personal data	Explicit consent, with narrow exceptions	Using medical histories to train a model

Whatever the basis, the Notice and Choice principle still applies. People should know their data is being used.

Check the data at every step

Good data governance follows the data through the AI system. The standard ISO/IEC 42001 sets out controls for data used in AI (Annex A.7), including data quality and data provenance, that map neatly onto PDPA duties.

Step	Question	Linked PDPA duty (Must)	Linked ISO control (Should)
Collect	Where did it come from? Do we have the right to use it this way?	General; Notice and Choice; Disclosure	Data provenance (A.7.5)
Clean	Is it accurate, complete and current?	Data Integrity	Data quality (A.7.4)
Train	Who is missing or under-represented?	—	Data quality (A.7.4)
Use and keep	Who can access it? When will we delete it?	Security; Retention	Data management (A.7.2)

One more point matters when you buy AI: a vendor that processes personal data for you is a data processor. Since the 2024 amendments, the Security Principle (s.9) applies to processors directly. But you, as the data controller, remain responsible. Outsourcing the processing doesn't outsource the accountability.

CASE · SJU's upload

HR uploaded five years of staff records into TalentFit: IC numbers, medical histories, performance ratings and interview notes. Staff were not told.

Performance ratings used for HR planning might rest on the employment contract. Medical histories are sensitive personal data: SJU needed explicit consent or a narrow exception, and had neither. And because nobody gave notice, SJU breached the Notice and Choice principle for all of it. On top of that, the vendor's terms allow it to store data anywhere and use it to train its own models.

This is a Must failure, not a Should. It is the finding with the most serious legal exposure in the case.

This is not legal advice
Whether a particular use of data has a lawful basis depends on the facts. Check specific cases with your DPO or legal counsel.

Summary

- The PDPA's seven principles apply to AI that processes personal data.
- Consent is one lawful basis among several; sensitive data such as health information needs explicit consent.
- Check data at each step. Using a vendor doesn't transfer your accountability as data controller.

Check your understanding

1. Does every use of employee data in an AI system require the employee's consent? Explain.
2. Which PDPA principle did SJU breach for all of the uploaded data, regardless of the lawful basis?
3. Your AI vendor suffers a data breach. Are you still responsible?

CHAPTER 9

Auditing an AI system: the SJU case

In this chapter you will learn to

- run a structured audit of an AI system across its lifecycle
- tag each finding as Must, Should or Coming and cite the evidence
- turn findings into prioritised recommendations

Chapters 1 to 8 gave you the pieces. This chapter puts them together in one exercise: a structured audit of SJU's TalentFit system. Read the full case file in Appendix A before you start. Try the audit yourself, then compare your answers with the worked example.

KEY MESSAGE A good audit cites evidence, separates what breaks the law from what falls short of good practice, and tells leaders what to do first.

A simple audit method

1. **Read the evidence.** Every finding should point to a document, record or test result.
2. **Walk the lifecycle.** For each stage, ask the governance questions from Chapter 7.
3. **Tag each finding.** Must, Should or Coming, using the right words for each.
4. **Name the rule or principle.** A PDPA section, an AIGE principle, an ISO control.
5. **Propose a fix.** Specific enough that someone could start tomorrow.
6. **Prioritise.** Tell leaders what to do first, and why.

Your audit sheet

Lifecycle stage	Finding (cite the exhibit)	Must, Should or Coming?	Rule or principle	Fix
1 Design, plan and model				
2 Verify and validate				
3 Deploy and use				
4 Operate and monitor				

A worked answer

There is more than one good answer. This is one.

Stage	Finding	Tag	Rule or principle	Fix
1 Design	Staff records, including medical histories, uploaded without notice or consent (Exhibits B, F)	Must	PDPA General Principle (s.6); Notice and Choice (s.7); explicit consent for sensitive data (s.40); AIGE Privacy and security	Stop further uploads; review the lawful basis for each type of data; tell staff what happened
2 Validate	Rural women score far lower for technical manager roles; scores can't be explained (Exhibits D, E)	Should; ask counsel about employment law	AIGE Fairness, Inclusiveness, Transparency; ISO/IEC 42001 A.7.4 (data quality)	Pause ranking; require the vendor to test for bias by group, explain and fix
3 Deploy	Bottom 40% auto-rejected with no human review (Exhibits B, E)	Should; possibly Coming	AIGE Accountability; human oversight (HITL, AIGE Section 2.7.3)	Switch off Auto-Reject; a recruiter reviews every rejection
4 Operate	No DPO; data location unknown; vendor may train on SJU data; no breach timeline or audit rights (Exhibits A, C, F)	Must	PDPA DPO (s.12A), breach notification (s.12B), cross-border transfers (s.129), security (s.9)	Appoint a DPO; renegotiate the vendor terms or find another vendor

Two Musts, two Shoulds. SJU's legal exposure sits in stages 1 and 4. The Shoulds are not minor, though. Stage 2 could carry legal risk if employment law protects job applicants from discrimination, and stage 3 may become a Coming under the AI Governance Bill.

What should SJU do first?

Put it on one page for the CEO. **Today:** pause Auto-Reject and stop any further uploads. **This week:** appoint a DPO, tell staff what data was used, and contact the vendor about where the data is held. **This month:** decide whether TalentFit can be fixed or should be replaced, and answer the applicant's request in Exhibit E.

The hidden finding

Exhibit E contains a request many readers miss: the applicant asks what personal data SJU holds about her. That is her right under the PDPA's Access Principle (s.12). Who at SJU would answer it? No one: there is no DPO.

Common mistakes

- **Findings without evidence.** “The tool is biased” is an opinion. “Exhibit D shows rural women averaging 46 against 68 for urban men” is a finding.
- **Everything is a Must.** Leaders stop listening when every finding is “critical”. Tag honestly.
- **Fixes that are too vague.** “Improve governance” helps no one. “Appoint a DPO by Friday” does.
- **Ignoring the vendor.** Buying AI moves work to the vendor, not responsibility.

Summary

- Audit by walking the lifecycle, citing evidence, tagging each finding and naming the rule.
- At SJU, the legal exposure (Musts) is in data handling and vendor management; the Shoulds are in fairness and oversight.
- Give leaders a short, prioritised list of first actions.

Check your understanding

1. Why should every audit finding cite evidence?
2. Which two of SJU’s four lifecycle findings carry the most legal exposure, and why?
3. Name one thing SJU should do today, and one this week.

PART 4

Building a governance system

An audit finds problems once. A management system keeps finding and fixing them. This part builds one, then tests it on a harder case.

CHAPTER 10

AI management systems

In this chapter you will learn to

- explain how a management system turns one-off checks into a repeating cycle
- describe the structure of ISO/IEC 42001
- set the context and scope of an AI governance system, starting with an inventory
- draft the core of an AI policy

After the audit, SJU's chief executive paused Auto-Reject and appointed a Data Protection Officer. Then he asked a harder question: how do we stop the next AI purchase going the same way? The answer is a management system.

KEY MESSAGE An audit finds problems once. A management system keeps finding and fixing them.

A loop, not a document

A management system is a set of policies, roles and routines that runs in a cycle. Most management standards, including ISO 9001 for quality and ISO 27001 for information security, follow the same Plan-Do-Check-Act loop. ISO/IEC 42001 does too.

Stage	What you do	ISO/IEC 42001 clauses
Plan	Define context and scope; issue the AI policy; plan risk assessment and treatment; allocate resources and competence	4 Context · 5 Leadership (AI policy, 5.2) · 6 Planning · 7 Support
Do	Run AI risk assessments (8.2), apply risk treatment (8.3) and conduct AI system impact assessments (8.4)	8 Operation
Check	Monitor and measure (9.1); audit internally (9.2); management reviews the results (9.3)	9 Performance evaluation
Act	Correct nonconformities, find root causes, and improve continually	10 Improvement

Mapping clauses to Plan-Do-Check-Act is a common way to explain the standard, not wording from it.

If your organisation already runs an ISO management system, AI governance can plug into it rather than start from scratch. The same internal audit, management review and corrective-action routines can cover AI.

Two parts: clauses and Annex A

ISO/IEC 42001 has two parts that work differently.

- **Clauses 4 to 10** are requirements. To be certified, you meet every one.
- **Annex A** is a menu of 38 reference controls in nine areas (A.2 to A.10). You assess all of them, choose which apply to the risks you have identified, and justify any you exclude in a document called the **Statement of Applicability**.

Annex A area	Covers
A.2	Policies related to AI
A.3	Internal organisation: roles and responsibilities, reporting concerns
A.4	Resources: data, tools, computing, people
A.5	Assessing the impacts of AI systems
A.6	The AI system lifecycle
A.7	Data for AI systems
A.8	Information for interested parties
A.9	Use of AI systems
A.10	Third-party and customer relationships

Start with context and scope (Clause 4)

Clause 4 asks four things: what issues inside and outside the organisation affect your AI (laws such as the PDPA, your own maturity and culture); who the interested parties are (customers, regulators, staff and the people AI decisions affect); what the system covers (which AI systems, business units and locations); and how you will set it up, run it and improve it.

You can't set a scope without knowing what AI you have. That makes the **AI inventory** the first practical step for almost every organisation.

AI system	Owner	What it decides or produces	Personal data?	Built or bought?	Risk
TalentFit AI	HR Director	Ranks applicants; predicts promotion readiness	Yes, including health data	Bought (cloud service)	High
Website chatbot	Marketing Manager	Answers customer questions	Sometimes (names, order numbers)	Bought	Medium
Route optimiser	Operations Director	Plans delivery routes	Drivers' locations	Built in software we already use	Low–medium

Illustrative inventory for SJU. Only TalentFit appears in the case file; the other two rows are examples.

Two columns cause the most trouble. **Owner:** if no one owns a system, no one governs it. **Built or bought:** AI increasingly arrives inside software you already pay for, switched on by the vendor. Ask every major software supplier whether their product now includes AI features.

Leadership and the AI policy (Clause 5)

Clause 5 makes AI governance a leadership responsibility. Top management must commit to the system and give it resources; set a written AI policy and share it with staff; and assign named people to make risk decisions, run impact assessments and handle incidents. Leadership can't hand AI to IT.

An AI policy doesn't need to be long. Five commitments cover most of what matters:

Part	Question it answers	Example for SJU
Purpose	Why do we use AI?	SJU will use AI to make hiring faster and fairer, never to replace human judgement on decisions about people.
Principles	What do we commit to?	SJU will apply the AIGE principles, especially Fairness, Transparency and Accountability.
Limits	What won't we do?	SJU will not use AI to make final hiring or dismissal decisions without human review.
Approval	Who says yes?	SJU will require approval from the AI governance lead and the DPO before any AI tool processes personal data.
People	How do we treat those affected?	SJU will tell staff and applicants when AI is used and give them a named contact to question a decision.

The approval line alone would have prevented most of SJU's problems.

Summary

- A management system runs Plan-Do-Check-Act. ISO/IEC 42001 has requirements (Clauses 4–10) and a menu of controls (Annex A).
- Clause 4 starts with context and scope; in practice, start with an AI inventory with a named owner for each system.
- Clause 5 makes AI a leadership responsibility. A short policy with five commitments covers the essentials.

Check your understanding

1. What is a Statement of Applicability?
2. Why is an AI inventory usually the first practical step?
3. Which line of SJU's sample policy would have prevented most of its problems?

CHAPTER 11

Who does what: committees and the Three Lines Model

In this chapter you will learn to

- design an AI decision forum with the power to stop a system
- distinguish what ISO/IEC 42001 requires from good practice
- assign AI governance roles using the Three Lines Model

A policy says what should happen. Roles make sure it does. This chapter answers two questions: who decides about AI, and who checks that the decisions are followed?

KEY MESSAGE Every AI system needs an owner, and someone with the authority to stop it.

An AI ethics committee

Many organisations set up a cross-functional group to review AI. A typical membership: legal and compliance, IT and data, risk, HR, cybersecurity, the DPO, and the business owner of the system under review. Its main jobs:

- approve, pause or stop high-risk AI systems
- review impact assessments before launch
- hear concerns raised by staff about AI

The power to stop is what matters. If a project that promises revenue has an unresolved bias or privacy problem, someone must be able to halt it. Without that authority, a committee is only advisory.

What ISO/IEC 42001 actually requires

The standard requires clear roles, responsibilities and decision authorities (Annex A.3.2) and a way for people to report concerns (A.3.3). It does not require an ethics committee as such. A committee with the power to stop a system is strongly recommended good practice, and one good way to meet the requirement. For a smaller organisation like SJU, an existing management or risk committee can take on the role, with the DPO attending.

The Three Lines Model

The Institute of Internal Auditors' Three Lines Model (2020) describes who does what in governance and risk management. It updated the older "Three Lines of Defense", shifting the emphasis from defence to helping the organisation achieve its goals, and from rigid structures to roles that work together.

Role	Who	What they do for AI
Governing body	The board	Sets direction and risk appetite; oversees the whole system
First line	Business owners, product and data teams, the people who run the AI	Own each system; keep its inventory entry and logs; run day-to-day checks
Second line	Risk, compliance, legal, the DPO	Set rules, map laws to systems, run assessments, test controls, challenge the first line
Third line	Internal audit	Give the board's audit committee independent assurance that the system really works

Second-line roles remain part of management; only internal audit is independent of it. And the lines describe roles, not walls: they work at the same time, and in small organisations one person may play more than one role. Vendors and external auditors sit outside the model, managed through contracts and engagements.

CASE · SJU's missing second line

SJU role	Line
Board	Governing body
HR Director, recruiters	First line
Head of Risk and Compliance, the new DPO, legal counsel	Second line
Internal auditor	Third line
TalentFit Solutions	Outside the model (managed through the contract)

Who should have stopped the medical-history upload? The first line should have checked before uploading. The second line should have set the rule and approved the use. Internal audit would catch the gap later. At SJU, no second line existed for AI at all. The Head of Risk wasn't even consulted.

Summary

- An AI decision forum needs cross-functional membership and the power to stop a system.
- ISO/IEC 42001 requires clear roles and a way to report concerns; a committee is one way to meet that.
- The Three Lines Model: the board governs; the first line owns; the second line advises and challenges; internal audit assures.

Check your understanding

1. Without the power to stop a system, what is an AI committee?
2. Does ISO/IEC 42001 require an AI ethics committee?
3. In which line does the DPO sit?

CHAPTER 12

Risk and impact

In this chapter you will learn to

- identify AI risks using seven common risk types
- explain the difference between a risk assessment and an impact assessment
- score a risk with a severity × likelihood matrix
- choose a treatment and record it

Governance needs a way to decide which problems come first. That means naming risks, measuring them consistently and deciding what to do about each one.

KEY MESSAGE A risk assessment asks “what could go wrong for us?” An impact assessment asks “what could this do to people?” You need both.

Seven types of AI risk

Type	What it looks like	Real example
1 Performance	It gets things wrong, or gets worse over time	Air Canada’s chatbot gives a wrong policy
2 Security	Attacks such as poisoned data or prompt injection; leaks	Samsung source code pasted into ChatGPT
3 Legal and regulatory	Breaches the PDPA, copyright or sector rules	SJU’s upload of health data
4 Reputational	Public backlash over unfair or offensive output	Amazon’s hiring tool reported worldwide
5 Enterprise	Doesn’t fit strategy or processes, or costs more than it saves	A tool bought by one team that no one else can use
6 Scalability	Small in a pilot, serious across millions of decisions	A 1% error rate across a national customer base
7 Black box	Nobody can explain why it decided what it did	SJU can’t explain TalentFit’s scores

This is MindMagine’s working list of enterprise AI risk types, not a list from ISO/IEC 42001. The standard asks you to identify AI risks but doesn’t prescribe these seven.

The AI impact assessment

ISO/IEC 42001 requires an AI system impact assessment (AIIA) that looks at effects on individuals, groups and society, not just on the organisation (Clause 6.1.4), and repeats it at

planned intervals and when significant changes happen (Clause 8.4). A simple version has four steps:

1. **Describe** the system and its purpose.
2. **Identify who is affected**, including vulnerable groups who may not be obvious.
3. **Assess the impacts** on fairness, privacy, safety, rights and livelihood.
4. **Record and act:** feed the results into risk treatment.

Repeat it when the data, the model, the users or the purpose changes, or after an incident.

Scoring: severity × likelihood

Give each risk a severity score (1 = minor inconvenience, 5 = serious harm to people or a major legal breach) and a likelihood score (1 = rare, 5 = almost certain). Multiply them.

Likelihood ↓ / Severity →	1	2	3	4	5
5 Almost certain	5 Medium	10 High	15 High	20 Critical	25 Critical
4 Likely	4 Low	8 Medium	12 High	16 Critical	20 Critical
3 Possible	3 Low	6 Medium	9 Medium	12 High	15 High
2 Unlikely	2 Low	4 Low	6 Medium	8 Medium	10 High
1 Rare	1 Low	2 Low	3 Low	4 Low	5 Medium

Band	Score	Action
Low	1–4	Acceptable under standard operating procedures
Medium	5–9	Operational controls applied by the first line
High	10–15	Controls applied; senior management approval required
Critical	16–25	Immediate mitigation; the project is halted until the score drops

These bands are an example. Each organisation sets its own, in line with the risk appetite the board has approved.

Score the risk to people as well as to the organisation. An error that costs SJU nothing can still cost an applicant a job.

Four ways to treat a risk

Option	Meaning
Treat	Mitigate: add controls to reduce the risk.
Terminate	Stop or suspend the system, or redesign it so the risk disappears.
Transfer	Share it through a contract (indemnity) or insurance. You still answer for it, and PDPA accountability never transfers.

Option	Meaning
Tolerate	Accept a low risk formally, in writing, within the risk appetite.

"Treat, Terminate, Transfer, Tolerate" is common risk-management language, not ISO/IEC 42001 wording.

ISO/IEC 42001 adds a discipline to this (Clause 6.1.3): choose your controls, compare them with Annex A, record which apply in the Statement of Applicability, write a risk treatment plan, and obtain formal approval from designated management for the plan and for accepting any residual risk.

CASE · Scoring three SJU risks

Risk	Types	Score	Treatment
Health data used without consent	Legal, reputational	4 × 4 = 16 Critical	Treat: stop uploads now, confirm a lawful basis, notify staff
Bias against rural women	Performance, reputational, legal	5 × 4 = 20 Critical	Terminate (pause ranking) until fixed, then treat: test and retrain
Automatic rejections	Black box, reputational	4 × 5 = 20 Critical	Treat: switch off Auto-Reject; HITL for every rejection

All three land in the Critical band, which is why SJU must act today, not next quarter. The scores are judgement; two careful people may differ by a point. What matters is the reasoning, and that the same method is used every time.

Summary

- Seven risk types give a practical checklist: performance, security, legal, reputational, enterprise, scalability and black box.
- An impact assessment looks at effects on people and society, and is repeated when things change.
- Score severity × likelihood; then treat, terminate, transfer or tolerate, with designated management approving residual risk. Transferring a risk never transfers PDPA accountability.

Check your understanding

1. What is the difference between a risk assessment and an impact assessment?
2. A risk has severity 4 and likelihood 3. What is its score and band in this chapter's matrix?
3. Under ISO/IEC 42001 Clause 6.1.3, who must approve the acceptance of residual AI risk?
4. SJU signs a contract making the vendor liable for data breaches. Has SJU transferred its PDPA accountability?

CHAPTER 13

Shadow AI, vendors and monitoring

In this chapter you will learn to

- recognise shadow AI and bring it into the governance system
- ask AI vendors the questions that matter
- set up monitoring, logging and incident response

The biggest AI risks often come from systems nobody approved, systems bought on standard terms, and systems that were fine at launch and slowly stopped being fine. This chapter covers all three.

KEY MESSAGE Buying AI doesn't buy you out of responsibility.

Shadow AI

Shadow AI is AI used without IT, Risk or Legal approval. It takes three common forms:

- staff pasting work into free chatbots
- a team buying an AI tool on a credit card or within its own budget
- a vendor switching on AI features inside software you already use

Shadow AI becomes a PDPA problem (a Must) as soon as personal data goes into an unapproved tool. Without personal data, it is still a confidentiality and security risk. Bans rarely work on their own; people find workarounds. Four steps work better:

1. keep the AI inventory current
2. make approval fast and simple
3. publish a list of approved tools
4. say plainly what data must never go into public AI tools

If approval takes three months, people will go around it.

Vendor questions

Ask the vendor	Why it matters	Tag and reference
How was the model trained and tested for bias?	Accuracy and fairness across groups	Should: AIGE Fairness; ISO/IEC 42001 A.7.4
Where is our data stored and processed?	Cross-border transfer conditions	Must: PDPA s.129
Will our data be used to train your models?	Privacy, confidentiality and IP	Must where personal data is involved: PDPA s.6 and s.8

Ask the vendor	Why it matters	Tag and reference
How fast will you tell us about a breach?	Your own breach-notification duty depends on it	Must: PDPA s.12B
Can we audit you, or see an independent audit?	Second-line oversight and assurance	Should: ISO/IEC 42001 A.10.2
Who is liable if the AI's output causes harm?	Risk allocation and contract caps	Should: a contract term to negotiate; liability itself is governed by contract law

CASE · TalentFit's terms

SJU accepted TalentFit's online terms without negotiating (Appendix A, Exhibit C). They allow the vendor to store data in any country (4.2) and to use SJU's data to train its own models (4.5). They promise breach notice only "without undue delay" (7.1), give SJU no audit rights (7.4) and cap the vendor's liability at 12 months of fees (9.1). TalentFit's terms fail five of the six questions. The sixth, on bias testing, they don't mention at all.

Monitor, log, respond

Once a system is live, three routines keep it safe.

Routine	What it involves	Reference
Monitor	Check accuracy, drift, bias and complaints on a set schedule	ISO/IEC 42001 A.6.2.6
Log	Record enough to investigate later: inputs, outputs, who changed what, and when	ISO/IEC 42001 A.6.2.8
Respond	A clear route for incidents, including notifying the Commissioner and affected people if personal data is breached	PDPA s.12B (Must); ISO/IEC 42001 A.8.4

What monitoring finds feeds back into Check and Act: the management-system loop from Chapter 10.

Summary

- Shadow AI is best managed with an inventory, fast approval, an approved-tools list and clear data rules.
- Ask vendors six questions before you sign; three of them connect to PDPA Musts.
- Monitor, log and respond, and feed what you learn back into the management system.

Check your understanding

1. When does shadow AI become a PDPA issue?
2. Which vendor question links to your own PDPA breach-notification duty?
3. Why is monitoring needed for a system that passed all its tests?

CHAPTER 14

Red-teaming a high-risk system: PDSB

In this chapter you will learn to

- red-team an AI system in a regulated sector
- identify which regulator and sector rules apply
- score and treat findings and present them to a board

To red-team a system is to attack it on purpose: look for every way it could fail before customers, regulators or journalists find them. This chapter applies everything in the book to a harder case than SJU. Read the case in Appendix B first.

KEY MESSAGE In a regulated sector, the red team's first question is "who regulates us?"

The case in brief

Pembiayaan Digital Sdn Bhd (PDSB), a Malaysian fintech, uses an AI model called Rakyat-Score AI to decide personal loans of up to RM50,000. A red team found three problems:

1. **Shadow API.** Engineers connected a third-party AI model without telling IT Security or Legal. There is no contract, no data agreement and no audit right, and applicants' financial data is processed offshore.
2. **Bias and black box.** The model was trained on urban banking data. Applicants in Sabah and Sarawak receive a 35% higher default-risk score regardless of income, and loan officers can't explain why.
3. **No human in the loop.** "Instant approve / instant reject" issues binding decisions and pays out loans with no human review.

Why is this harder than SJU? Consumer credit is a regulated activity, the decisions are legally binding, and money moves automatically.

Start with the regulator

The case doesn't say whether PDSB is a bank. That is deliberate. If it were a bank, Bank Negara Malaysia's rules, including RMiT, would apply. If it is a non-bank lender, the Consumer Credit Act 2025 is likely to apply: in force since 1 March 2026, overseen by the Consumer Credit Commission, requiring fair, responsible and professional conduct and checks that borrowers can repay without undue hardship. Confirming which regime covers PDSB's activity is the red team's first legal question for counsel. Either way, a model that raises risk scores by location rather than by income sits uneasily with an affordability duty.

A worked red-team answer

Flaw	Risk types · score	Must, Should, Coming	Treatment
1 Shadow API	Security, legal · 4 × 5 = 20 Critical	Must: PDPA security (s.9), cross-border transfers (s.129). Should: ISO/IEC 42001 A.10.2	Terminate: suspend the API now. Then a vendor review and a data processing agreement with breach and transfer terms
2 Bias and black box	Performance, reputational, black box · 5 × 4 = 20 Critical	Must: credit-sector affordability and fair-conduct duties. Should: AIGE Fairness, Transparency; ISO/IEC 42001 A.7.4, A.8.2. Coming: Bill risk tiers	Treat: rebalance the training data with rural credit data; add explainability tools so officers can give reasons
3 No human in the loop	Performance, legal · 4 × 4 = 16 Critical	Should: AIGE human oversight (Section 2.7.3). Must: credit-sector affordability duties	Treat: switch off instant reject; an officer reviews every rejection (HITL)

For comparison: under the EU AI Act, credit scoring of natural persons is a high-risk use (Annex III, point 5(b)), with strict duties including human oversight.

Presenting to the board

A red-team report to a board should fit on one page: the three findings, their scores, the legal exposure in plain words, and the first three actions with owners and dates. All three PDSB findings are Critical, so the actions start now: suspend the unvetted API today, switch off instant rejection this week, and commission a bias review and explainability fix this month, while confirming with counsel which credit rules apply.

Summary

- In a regulated sector, identify the regulator and sector rules before anything else.
- PDSB shows the same patterns as SJU (shadow AI, bias, no human review) with higher stakes.
- A board-level red-team report fits on one page: findings, scores, exposure, first actions.

Check your understanding

1. Why does it matter whether PDSB is a bank?
2. Why is PDSB's location-based scoring a concern under an affordability duty?
3. Which of PDSB's three flaws would you fix first, and how?

CHAPTER 15

Your first 90 days

In this chapter you will learn to

- plan the first three months of an AI governance programme
- set actions with owners and dates
- avoid the common traps of early AI governance

Governance programmes stall in one of two ways: they never start because the task looks too big, or they start with a 60-page policy nobody reads. The way through is to start small and practical.

KEY MESSAGE Find what you have, decide what matters, then make it routine.

Period	Aim	Key deliverables	Accountable owner (example)
Days 1–30	Find	An enterprise AI inventory covering in-house, vendor and shadow AI tools, with an owner for each · a designated DPO under PDPA s.12A and the PDPA basics in place	Head of IT and Head of Compliance
Days 31–60	Decide	A documented AI policy (ISO/IEC 42001 Clause 5.2) · an AI decision forum with the power to stop a system · the inventory ranked by risk · one AI vendor contract reviewed against the six questions	CEO and General Counsel
Days 61–90	Run	Impact assessments for high-risk tools · the vendor checklist built into procurement · continuous logging (ISO/IEC 42001 A.6.2.8) · staff briefed on approved tools and data rules	AI decision forum and first-line leads

For each action, write one owner and one real date. “Q1” is not a date.

Common traps

- **The policy trap.** Writing a long, complex policy before you have an inventory of the AI tools actually in use.
- **The vacuum trap.** Letting data science or engineering teams sign off their own systems’ ethics, without Legal and Risk oversight.
- **The freeze trap.** Banning all generative AI tools, which pushes staff onto unapproved personal devices and creates shadow AI.
- **The waiting trap.** Holding off for the AI Governance Bill. The PDPA applies today, and the Bill is likely to build on what good practice already looks like.

Three things to take with you

1. Sort every rule: Must, Should or Coming.
2. An audit finds problems once. A system keeps finding and fixing them.
3. Every AI system needs an owner, and someone who can stop it.

My 90-day plan

Period	My action	Owner	Date
Days 1–30			
Days 31–60			
Days 61–90			

Check your understanding

1. Why start with an inventory rather than a policy?
2. What makes an action in a 90-day plan likely to happen?

APPENDIX A

The SJU case file

This case is fictional. The company, the vendor, the people and every figure in it are invented for training. Any resemblance to real organisations or people is coincidental.

1. The company

Syarikat Jaya Utama Sdn Bhd (SJU) is a Malaysian logistics and warehousing company with **800 employees**. Its head office is in Shah Alam. It runs large warehouses in the Klang Valley, Penang and Johor, and smaller depots in rural Perak and Pahang.

SJU is growing. This year it plans to fill about 120 roles, from warehouse supervisors to technical managers who run its fleet-tracking and warehouse-automation systems. The six-person HR team received more than 3,000 applications in the last hiring round and fell weeks behind.

2. The decision

To clear the backlog, HR Director **Puan Aishah Rahman** bought **TalentFit AI**, a cloud software service from TalentFit Solutions, a vendor based outside Malaysia. HR paid for it from its own budget. IT, Risk and Legal were not involved, and SJU accepted the vendor's standard online terms without negotiating a contract (Exhibit C).

TalentFit does three things:

- It **screens CVs** for every vacancy.
- It **ranks candidates** with a suitability score from 0 to 100.
- It **predicts promotion readiness** for current staff, to help plan internal moves.

3. Setting it up

TalentFit's onboarding guide recommended "training the model on your own people". So HR uploaded **five years of historical employee records** into the cloud service. The upload included national ID (IC) numbers, medical histories from medical leave and pre-employment checks, performance ratings and interview notes. Employees were not told, and no one asked for their consent (Exhibit B).

TalentFit's base model had been built mostly from past hiring decisions, and SJU's own history was heavily urban: most technical managers hired in the last five years came from the Klang Valley and Penang offices.

HR switched on the **Auto-Reject** feature. Candidates ranked in the bottom 40% receive a rejection email automatically, usually within minutes. No recruiter reviews these decisions.

4. What happened next

For the first six weeks, the results looked good. Time to shortlist fell from three weeks to two days. Then three things happened in the same month:

1. A data analyst in HR, asked to report on the new process, noticed a pattern in the scores for technical management roles (Exhibit D).

2. An applicant from a rural district complained in writing that she was rejected four minutes after applying (Exhibit E).
3. During a routine review of expense claims, IT found the TalentFit subscription and raised a security concern (Exhibit F).

5. Your brief

From: Daniel Lim, Chief Executive Officer

I need an independent view of TalentFit before this goes any further. Please audit it as a cross-functional team. Tell me what is wrong, which problems break the law and which fall short of good practice, how serious each one is, and what we should do first. I want your findings by the end of the day.

6. People in the case

Name	Role	Involvement
Daniel Lim	Chief Executive Officer	Asked for the audit
Puan Aishah Rahman	HR Director	Chose and owns TalentFit; set up the upload and Auto-Reject
Rajesh Kumar	Head of IT	Found the subscription in expense claims
Tan Mei Ling	Head of Risk and Compliance	Not consulted before purchase
Farah Idris	HR Data Analyst	Ran the score analysis in Exhibit D
Nurul Huda	Job applicant	Wrote the complaint in Exhibit E
—	Data Protection Officer	SJU has not appointed one
TalentFit Solutions	Vendor	Hosts the service and data outside Malaysia

7. Timeline

When	What happened
Month 1	HR backlog peaks. Puan Aishah signs up for TalentFit on the vendor’s online terms.
Month 1	HR uploads five years of staff records. Auto-Reject switched on for the bottom 40%.
Months 2–3	TalentFit runs two hiring rounds. Time to shortlist falls from three weeks to two days.
Month 3	Farah’s analysis (Exhibit D), Nurul’s complaint (Exhibit E) and IT’s security note (Exhibit F). The CEO asks for an audit. That’s you.

Exhibits

Exhibit A · SJU at a glance

Legal name	Syarikat Jaya Utama Sdn Bhd
Business	Logistics, warehousing and last-mile delivery
Employees	800
Locations	Head office in Shah Alam; warehouses in the Klang Valley, Penang and Johor; depots in rural Perak and Pahang
Hiring this year	About 120 roles; more than 3,000 applications in the last round
HR team	6 people, led by the HR Director
AI systems in use	TalentFit AI (HR). No central list of AI systems exists.
Data Protection Officer	Not appointed
Governance	Board of directors; management committee; one internal auditor; no AI policy

Exhibit B · Email from the HR Director

From: Aishah Rahman, HR Director · To: Daniel Lim, CEO · Subject: TalentFit is live!

Hi Daniel,

Good news: TalentFit went live today. To get the best results we uploaded our full staff history from the last five years (HR files, appraisals, interview notes and the medical records from the clinic panel) so the system learns what a successful SJU employee looks like.

We have also turned on Auto-Reject, so the bottom 40% of applicants get a polite “no” straight away. My team will only see the top candidates from now on. I didn’t think we needed to bother staff with the technical details.

Expect shortlists in days, not weeks!

Aishah

Exhibit C · Extract from TalentFit's online Terms of Service

SJU accepted these terms by clicking "I agree" when it subscribed. No other contract exists.

4.2 Data location. Customer Data may be stored and processed in any country where the Provider or its sub-processors operate.

4.5 Service improvement. The Provider may use Customer Data, including in aggregated form, to train and improve its models and services.

7.1 Security incidents. The Provider will notify the Customer of a Security Incident without undue delay.

7.4 Audit. The Customer has no right to audit the Provider's systems. The Provider may share summary security information at its discretion.

9.1 Liability. The Provider's total liability is limited to the fees paid by the Customer in the 12 months before the claim.

Exhibit D · HR analyst's note on suitability scores

From: Farah Idris, HR Data Analyst · To: Aishah Rahman · Subject: Pattern in Technical Manager scores

I compared TalentFit's results for the last Technical Manager hiring round by applicant group. Qualifications and years of experience were broadly similar across the groups.

Applicant group	Applicants	Average score (0–100)	Auto-rejected	Shortlisted
Urban, men	210	68	22%	31%
Urban, women	140	64	29%	24%
Rural, men	90	55	47%	12%
Rural, women	60	46	68%	5%

I can't see how TalentFit calculates the score, and the vendor's dashboard doesn't explain it. I also noticed that most of our past technical manager hires came from the Klang Valley and Penang offices.

Exhibit E · Complaint from an applicant

From: Nurul Huda · To: careers@sju.example · Subject: My application for Technical Manager (Warehouse Systems)

Dear Sir or Madam,

I applied for the Technical Manager role at 9:12 this morning. At 9:16 I received an email saying I was not successful. I have nine years of experience running warehouse systems at a depot in Pahang, which is exactly what your advertisement asked for.

Did a person actually read my application? If a computer made this decision, I would like to know how, and what information about me it used. Please also tell me what personal data you hold about me.

Yours sincerely,

Nurul Huda

Exhibit F · IT security note

From: Rajesh Kumar, Head of IT · To: Daniel Lim, CEO; Tan Mei Ling, Head of Risk and Compliance

While reviewing expense claims I found a monthly subscription to TalentFit Solutions charged to HR. My team never assessed this vendor. What we know so far:

- The service holds staff records, including IC numbers and medical information, and all applicant CVs.
- We don’t know which country the data is stored in. The vendor’s terms allow any country (Exhibit C, 4.2).
- The terms let the vendor use our data to train its own models (4.5).
- There is no agreed timeline for telling us about a data breach (7.1), and we have no audit rights (7.4).
- SJU has not appointed a Data Protection Officer, so there is no one to own these questions.
- I can’t find any record that staff were told their data would be used this way.

I recommend we treat this as urgent.

APPENDIX B

The PDSB case

This case is fictional.

Pembiayaan Digital Sdn Bhd (PDSB), a Malaysian financial technology firm, developed and deployed an automated AI model called **Rakyat-Score AI** to evaluate personal loan applications of up to RM50,000. A red-team investigation found three problems.

Finding 1 · Shadow AI and vendor risk

The product engineering team integrated a third-party open-source foundation model through an API without informing IT Security or Legal. PDSB has no formal contract with the API provider, no data processing agreement, and no audit rights over how applicants' financial data is processed offshore.

Finding 2 · Bias and a black box

The model was trained on 10 years of historical banking data from major urban centres. Audit logs show that applicants living in rural East Malaysia (Sabah and Sarawak) receive a 35% higher default-risk score regardless of individual income. Loan officers cannot explain to rejected applicants why their score was low.

Finding 3 · No human in the loop

PDSB enabled "Instant Approval / Instant Reject" mode. The system issues binding loan rejections and pays out loans with no human reviewer.

> **Before you start:** the case doesn't say whether PDSB is a bank. Your first task is to work out which regulator and which sector rules apply.

Red team sheet

Flaw	Risk types	Severity × likelihood	Must, Should or Coming?	Rule or control	Treatment
1 Shadow API					
2 Bias and black box					
3 No human in the loop					

APPENDIX C

Quick reference

1. The seven AIGE principles (Should)

Principle	In plain words
1 Fairness	Equitable treatment; no demographic or regional bias.
2 Reliability, safety and control	Predictable performance, fail-safe defaults and human override controls.
3 Privacy and security	Data protection and cybersecurity safeguards.
4 Inclusiveness	Accessible and beneficial across Malaysia's multicultural society.
5 Transparency	Clear notice to users and explainable decision logic.
6 Accountability	Explicit human ownership and responsibility.
7 Pursuit of human benefit and happiness	AI that supports human flourishing and social well-being; humans can intervene.

2. PDPA section lookup (Must)

Personal Data Protection Act 2010 (Act 709), as amended by the Personal Data Protection (Amendment) Act 2024 (Act A1727).

Section	Provision	Status	What it means for AI
s.5	Penalty for breaching the principles	In force	Fine up to RM1,000,000 and/or up to 3 years' imprisonment
s.6	General Principle	In force	A lawful basis (consent, or another ground such as contract) before personal data goes into AI
s.7	Notice and Choice Principle	In force	Written notice before processing personal data, including in AI
s.8	Disclosure Principle	In force	No disclosure for a purpose, or to a party, people weren't told about, without consent
s.9	Security Principle	In force; applies to processors (Act A1727)	Security safeguards for training sets, prompts and storage
s.10	Retention Principle	In force	Delete personal training data once its purpose is fulfilled
s.11	Data Integrity Principle	In force	Personal data used in AI must be accurate, complete and current

Section	Provision	Status	What it means for AI
s.12	Access Principle	In force	People can access and correct their personal data
s.12A	Data Protection Officer	In force (Act A1727)	Controllers and processors must appoint a DPO
s.12B	Data breach notification	In force (Act A1727)	Notify the Commissioner as soon as practicable; affected people where significant harm is likely
s.40	Sensitive personal data	In force	Explicit consent for health, biometric and other sensitive data, with narrow exceptions
s.43A	Data portability	In force (Act A1727)	People can ask for their data to be transferred to another controller
s.129	Cross-border transfer	In force; amended by Act A1727	Safeguards before transferring personal data out of Malaysia

3. Human oversight (AIGE Section 2.7.3)

Model	How it works	Use it for
HITL · human-in-the-loop	A person approves each decision before it takes effect.	High-impact decisions about individuals
HOTL · human-on-the-loop	AI acts on its own; a person watches live and can override.	Fast, high-volume decisions
HIC · human-in-command	People set the rules, decide whether it runs, and can switch it off.	Every AI system, always

4. The AI lifecycle

Stage	Key governance questions
1 Design, plan and model	What problem are we solving? What data, and may we use it?
2 Verify and validate	Does it work for everyone? Tested for bias and security?
3 Deploy and use	Do users know it's AI? Who approved the launch? Who reviews decisions?
4 Operate and monitor	Is it drifting? Are events logged? Who handles incidents?

5. ISO/IEC 42001:2023 at a glance (Should)

Part	What it covers
Plan · Clauses 4–7	Context and scope (4); leadership and AI policy (5, 5.2); planning: risk assessment (6.1.2), risk treatment and Statement of Applicability (6.1.3), AI system impact assessment (6.1.4); support (7)
Do · Clause 8	AI risk assessment (8.2), risk treatment (8.3), AI system impact assessment (8.4)
Check · Clause 9	Monitoring (9.1), internal audit (9.2), management review (9.3)
Act · Clause 10	Nonconformity and corrective action; continual improvement
Annex A	38 reference controls in nine areas: AI policies (A.2), internal organisation (A.3), resources (A.4), impact assessment (A.5), lifecycle (A.6), data (A.7), information for interested parties (A.8), use of AI (A.9), third parties and customers (A.10)

6. The Three Lines Model (IIA, 2020)

Role	What it does
Governing body	The board. Sets direction and risk appetite; oversees everything.
First line	Owns and runs the AI: business owners, product and data teams.
Second line	Expertise and challenge: risk, compliance, legal, the DPO.
Third line	Internal audit: independent assurance to the board's audit committee.

7. Risk matrix: score = severity × likelihood

Likelihood ↓ / Severity →	1 Minor	2	3	4	5 Severe
5 Almost certain	5 Medium	10 High	15 High	20 Critical	25 Critical
4 Likely	4 Low	8 Medium	12 High	16 Critical	20 Critical
3 Possible	3 Low	6 Medium	9 Medium	12 High	15 High
2 Unlikely	2 Low	4 Low	6 Medium	8 Medium	10 High
1 Rare	1 Low	2 Low	3 Low	4 Low	5 Medium

Band	Score	Action
Low	1–4	Acceptable under standard operating procedures
Medium	5–9	Operational controls applied by the first line
High	10–15	Controls applied; senior management approval required
Critical	16–25	Immediate mitigation; the project is halted until the score drops

8. Risk treatment (ISO/IEC 42001 Clause 6.1.3)

Option	Meaning
Treat	Mitigate through controls.
Terminate	Stop or suspend the system.
Transfer	Contractual indemnity or insurance. PDPA accountability never transfers.
Tolerate	Accept a low risk, in writing.

Designated management must formally approve the risk treatment plan and the acceptance of residual risk.

Answers to “Check your understanding”

Chapter 1

1. Debts were raised automatically against hundreds of thousands of people with no meaningful human review of each decision.
2. Any two of: it inherits patterns (and biases) from past data; it acts at scale, so one error affects many people; it can produce confident but false answers; its decisions can be hard to explain.
3. It argued the chatbot was responsible for its own words. The tribunal found the airline responsible for all the information on its website, whether from a static page or a chatbot.
4. Amazon: the same use (hiring) and the same cause (a tool trained on skewed historical data).
5. No. The PDPA (Act 709, amended by Act A1727) applies to AI that processes personal data in commercial transactions, with penalties of up to RM1 million and/or three years. Sector rules may also apply.

Chapter 2

1. Drift is a model getting worse over time as the world changes; a system that passed testing can later fail, so it must be monitored after launch.
2. Its behaviour depends on the patterns it learned from training data, not only on rules written in code.
3. It entered at data in and learning (urban-heavy hiring history) and showed at output (lower scores for rural women).

Chapter 3

1. A Should. It becomes a Must in practice if a contract requires it, or effectively binding once the company certifies or commits to it.
2. “Falls short of the AIGE Guidelines”: AIGE is voluntary, so it cannot be violated.
3. For example, uploading health data without consent: it breaches the PDPA and falls short of the AIGE privacy and security principle.
4. Any two of: it is written into a contract; a sector regulator adopts it in binding guidelines; a law adopts it; the organisation publicly commits to it; a court considers it when judging reasonable care.

Chapter 4

1. FAQ chatbot: limited risk. Credit scoring: high risk. Spam filter: minimal risk.
2. High-risk duties: up to €15 million or 3% of global turnover. Banned practices: up to €35 million or 7%.
3. It is a voluntary standard. It sets out how to run an AI management system; you can choose to be certified against it, but there is no legal penalty for not following it unless a contract or commitment makes it binding.
4. As a practical way to judge how careful to be, because they reflect international views of risk and Malaysia’s proposed Bill also uses risk tiers.

Chapter 5

1. Appointing a DPO (s.12A); notifying personal data breaches (s.12B); the right to data portability (s.43A).
2. A fine of up to RM1 million and/or up to three years' imprisonment.
3. Any two of: wide reach (AI sold, built or used in Malaysia, or used by a Malaysia-based deployer); a Central AI Authority; three harm-based risk tiers (unacceptable, high, low); an AI sandbox.
4. The Consumer Credit Act 2025, including its fair-conduct and affordability duties, and any guidance from the Consumer Credit Commission.

Chapter 6

1. Accountability.
2. A spam filter rarely makes decisions about people; a hiring tool decides who gets a job, so unfairness directly harms individuals.
3. Privacy and security (overlaps with the PDPA security principle and others).
4. Supremacy of the Constitution and Rule of Law.

Chapter 7

1. Stage 2, verify and validate (or, since it bought the tool, by asking the vendor for test results before buying).
2. HOTL: speed matters, so the system acts on its own while a fraud team monitors and can release blocked payments.
3. HITL covers individual decisions; HIC covers whether the system should run at all, its rules and its limits. Someone senior must hold that authority.

Chapter 8

1. No. Consent is one lawful basis; others include processing necessary for the employment contract or a legal duty. Sensitive data such as health information generally needs explicit consent. Notice is required either way.
2. Notice and Choice: staff were never told their data would be used this way.
3. Yes. As data controller you remain accountable, although the vendor as processor also has its own security duty.

Chapter 9

1. So the finding is a fact others can check, not an opinion, and so leaders can trust and act on it.
2. Stage 1 (personal and sensitive data used without notice or consent) and stage 4 (no DPO, cross-border and vendor issues): both are PDPA Musts.
3. Today: pause Auto-Reject and stop uploads. This week: appoint a DPO, inform staff, contact the vendor.

Chapter 10

1. The document listing which ISO/IEC 42001 Annex A controls the organisation applies, and why any are excluded.

2. You can't set the scope of governance, assign owners or assess risks without knowing which AI systems you have, including AI inside software you already use.
3. The approval line: requiring approval from a named role and the DPO before any AI tool processes personal data.

Chapter 11

1. Only advisory: it can recommend but not prevent harm.
2. No. It requires clear roles and decision authorities and a way to report concerns; a committee is one way to meet those requirements.
3. The second line.

Chapter 12

1. A risk assessment asks what could go wrong for the organisation; an impact assessment asks what the system could do to people and society.
2. 12: High (controls applied; senior management approval required).
3. Designated management.
4. No. It may share the financial risk, but as data controller SJU remains accountable under the PDPA.

Chapter 13

1. When personal data is put into an unapproved tool.
2. How fast the vendor will tell you about a breach.
3. Because of drift and changing conditions: a system can get worse over time, and new problems or complaints only appear in real use.

Chapter 14

1. It decides which regulator and sector rules apply: BNM rules for a bank; likely the Consumer Credit Act 2025 for a non-bank lender.
2. Affordability should depend on the borrower's ability to repay, such as income; scoring by location regardless of income does not measure that.
3. The shadow API (score 20, clear PDPA exposure, data offshore with no contract): suspend it immediately, then review the vendor and put a data agreement in place.

Chapter 15

1. Because you can't scope a policy, assign owners or prioritise risks without knowing which AI systems you have.
2. A single named owner, a real date, and a scope small enough to complete in the period.

Glossary

Term	Meaning
AI Impact Assessment (AIIA)	A structured evaluation under ISO/IEC 42001 Clause 6.1.4 of an AI system's potential impacts on individuals, groups and society.
AI governance	The rules, roles and routines that make sure someone is in charge of how AI is chosen, used and checked.
AI inventory	A list of every AI system an organisation uses, with an owner for each.
AI management system (AIMS)	A set of policies, roles and routines for governing AI that runs as a repeating cycle; the subject of ISO/IEC 42001.
AIGE	National Guidelines on AI Governance and Ethics, launched by MOSTI in 2024. Voluntary.
Annex A	The 38 reference controls in ISO/IEC 42001.
Act A1727	The Personal Data Protection (Amendment) Act 2024.
Algorithmic bias	Systematic unfairness in a model's outputs that disadvantages particular demographic or geographic groups.
Annex III (EU AI Act)	The EU AI Act's list of high-risk uses, including recruitment and credit scoring of individuals.
BNM RMiT	Bank Negara Malaysia's Risk Management in Technology policy for financial institutions.
Black box	A system whose decisions nobody can explain.
Coming	This book's tag for proposed law that is not yet in force.
Consumer Credit Act 2025	Malaysian law regulating credit businesses outside the banking system, in force from 1 March 2026.
Data controller	The organisation that decides why and how personal data is processed (formerly "data user" under the PDPA).
Data processor	An organisation that processes personal data on a controller's behalf, such as an AI vendor.
Deepfake	AI-generated video, audio or images that make someone appear to say or do something they didn't.
DPO	Data Protection Officer. Mandatory under PDPA s.12A from 1 June 2025.
Drift	A model getting worse over time as real-world conditions move away from its training data.
EU AI Act	The European Union's risk-based AI law.
Generative AI	AI that creates text, images, audio or code.

Term	Meaning
Hallucination	A confident answer from AI that is simply false.
HITL / HOTL / HIC	Human-in-the-loop (a person approves each decision) / human-on-the-loop (a person watches and can override) / human-in-command (people decide whether the system runs).
ISO/IEC 42001	The international standard for AI management systems. Voluntary unless you choose to certify or a contract requires it.
Machine learning	AI that learns patterns from examples instead of following hand-written rules.
Model	The trained “engine” that turns an input into an output.
Must	This book’s tag for binding law.
NAIO	National AI Office, under the Ministry of Digital.
PDPA	Personal Data Protection Act 2010 (Act 709), as amended by Act A1727.
Red-teaming	Deliberately probing an AI system to find security, legal, ethical and operational weaknesses before or during deployment.
Rukun Negara	Malaysia’s national philosophy of five principles.
Sensitive personal data	Categories of personal data, such as health information, that need explicit consent under the PDPA.
Shadow AI	AI tools, cloud services or APIs used without IT, Risk or Legal approval.
Should	This book’s tag for voluntary guidance or standards.
Statement of Applicability (SoA)	The ISO/IEC 42001 document listing which Annex A controls apply and justifying any exclusions.
Three Lines Model	The Institute of Internal Auditors’ 2020 model of governance roles: governing body, first line, second line and internal audit.
Treat, Terminate, Transfer, Tolerate	The four common risk-treatment options.
Training data	The examples a model learns from.

Sources and further reading

Sources used in this book. Laws and guidance change: always check the current official text.

- MOSTI, “The National Guidelines on AI Governance & Ethics” (official text, Sections 2.6, 2.6.2 and 2.7.3). https://www.mosti.gov.my/wp-content/uploads/2024/09/NATIONAL-GUIDELINES-OF-AIGE-20241118.pdf
- CyberPeace, “Malaysia’s First AI Law Is Coming. Here’s What’s Actually in It”, August 2026. https://cyberpeace.org/resources/blogs/malaysias-first-ai-law-is-coming-heres-whats-actually-in-it
- Streamline Newsroom, “Malaysia Targets Q1 2027 Tabling of AI Governance Bill”, 5 October 2026 (citing The Star and RTM). https://streamlinefeed.co.ke/news/malaysia-risk-based-ai-governance-bill-2027
- Presencis, “EU AI Act Penalty Framework”. https://presencis.com/regulations/eu-ai-act/penalties/
- Sydney Criminal Lawyers, “Government to Pay $1.8 Billion Over Illegal Robodebt Scheme”, June 2021. https://www.sydneycriminallawyers.com.au/?p=56235
- Baker McKenzie, “Malaysia: Public Consultation on the AI Governance Bill”, July 2026. https://www.bakermckenzie.com/en/insight/publications/2026/07/malaysia-public-consultation-on-the-ai-governance-bill
- Eversheds Sutherland, “AI Governance Bill: Malaysia’s Next Step Towards the First AI Rulebook”, July 2026. https://www.eversheds-sutherland.com/en/united-states/insights/ai-governance-bill-malaysias-next-step-towards-the-first-ai-rulebook
- Herbert Smith Freehills Kramer, “Malaysia consults on AI Governance Bill”, July 2026. https://www.hlc.com/en/publications/malaysia-consults-on-ai-governance-bill
- AI Policy Tracker, “Artificial Intelligence (Malaysia, 2026): Status & Dates”. https://aipolicytracker.org/policies/malaysia-artificial-intelligence
- Origin PI, “Malaysia AI Governance Bill: What Is Known and What to Do Now”, September 2026. https://originpi.com/blog/malaysia-ai-governance-bill-prepare/
- Government of Malaysia portal, “National Guidelines on Artificial Intelligence Governance and Ethics (AIGE)”. https://www.malaysia.gov.my/en/my-initiative/whole-government-digital-services/tadbir-urus/national-guidelines-on-artificial-intelligence-governance-and-ethics-aige
- Malay Mail, AIGE launch report, 20 September 2024. https://www.malaymail.com/news/malaysia/2024/09/20/malaysia-sets-the-stage-for-responsible-ai-development-unveils-governance-and-ethics-guidelines/151026
- Deloitte, summary of Malaysia’s AIGE Guidelines. https://www2.deloitte.com/content/dam/Deloitte/sg/Documents/risk/sea-risk-my-aige-guidelines.pdf
- Baker McKenzie, “Malaysia: Personal Data Protection (Amendment) Act 2024 to come into force”. https://insightplus.bakermckenzie.com/bm/data-technology/malaysia-personal-data-protection-amendment-act-2024-to-come-into-force
- Tay & Partners (Legal 500), “Countdown to compliance: PDPA Amendment Act 2024”. https://www.legal500.com/firms/31184-tay-partners/c-malaysia/news-and-developments/countdown-to-compliance-personal-data-protection-amendment-act-2024-in-force-starting-1-january-2025
- Crowe Malaysia, “Personal Data Protection (Amendment) Act 2024”. https://www.crowe.com/my/services/risk-consulting/pdpa
- New Straits Times, “Consumer Credit Act comes into force”, March 2026. https://www.nst.com.my/amp/news/nation/2026/03/1388493/consumer-credit-act-comes-force
- DFDL, “Malaysia: A Summary of the Consumer Credit Act 2025”. https://www.dfdl.com/insights/content-hub/articles/malaysia-a-summary-of-the-consumer-credit-act-2025/
- GLOCERT, “ISO 42001 requirements: clauses 4–10”. https://www.glocertinternational.com/resources/articles/iso-42001-requirements-clauses-4-10/
- NovelVista, “ISO 42001 Annex A controls”. https://www.novelvista.com/blogs/quality-management/iso-42001-annex-a-controls

- The Institute of Internal Auditors, “The IIA’s Three Lines Model”. https://www.theiia.org/globalassets/documents/resources/the-iias-three-lines-model-an-update-of-the-three-lines-of-defense-july-2020/three-lines-model-updated-english.pdf
- Reuters via Euronews, “Amazon scraps secret AI recruiting tool that showed bias against women”, October 2018. https://www.euronews.com/2018/10/10/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women
- Torkin Manes, on Moffatt v. Air Canada (BC Civil Resolution Tribunal, February 2024). https://www.torkin.com/insights/publication/bc-tribunal-confirms-companies-remain-liable-for-ai-chatbot-created-information
- The National, “Samsung Electronics bans employee use of ChatGPT”, May 2023. https://thenational-the-national-prod.cdn.arcpublishing.com/business/technology/2023/05/02/samsung-electronics-bans-employee-use-of-chatgpt
- Malay Mail, “AI scams are getting real”, August 2025. https://www.malaymail.com/news/malaysia/2025/08/04/ai-scams-are-getting-real-here-are-the-cases-happening-in-malaysia-that-you-should-know-about/183459
- Bernama, “Can you distinguish between reality and deepfake?”, February 2025. https://bernama.com/misc/rss/news.php?id=2394394

Primary documents to read

- Personal Data Protection Act 2010 and the Personal Data Protection (Amendment) Act 2024, with the Commissioner’s guidelines
- National Guidelines on AI Governance and Ethics (MOSTI, 2024)
- The NAIO public consultation paper on the AI Governance Bill (July 2026)
- ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system
- Regulation (EU) 2024/1689 (the EU AI Act)
- Consumer Credit Act 2025

Published by KAIVANT Sdn. Bhd. · A company registered in Malaysia · KAIVANT® is a registered trademark · kaivantscore.com/articles