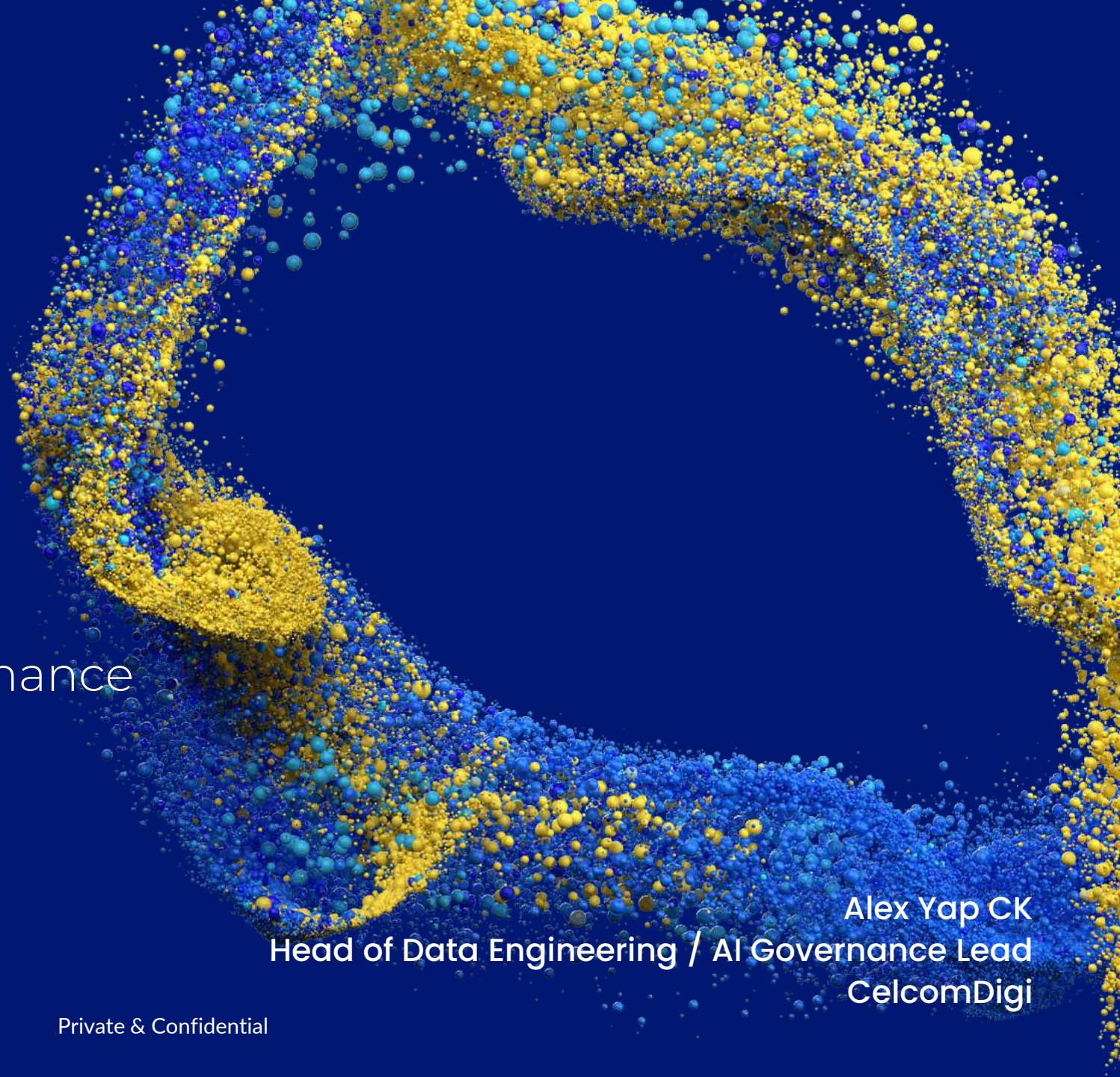
IERP®
Institute of Enterprise Risk Practitioners

GLOBAL CONFERENCE 2026

FROM CHAOS TO COMMAND:
RISK LEADERSHIP IN INDUSTRY 6.0

From Risk Appetite to AI Deployment: Making AI Governance Auditable and Scalable

Alex Yap CK
Head of Data Engineering / AI Governance Lead
CelcomDigi



AI Application Risks and Governance

18th Aug 2026, Mandarin Oriental KL
IERP

Alex Yap CK
Head of Data Engineering / AI Governance Lead
CelcomDigi

IERP®
Institute of Enterprise Risk Practitioners

TABLE OF CONTENTS

What We'll Cover

AI GOVERNANCE LANDSCAPE

Three Pillars, One Outcome – Managed Shadow AI

AI Application

- Lifecycle governance (design → retire)
- Model risk management
- Fairness, transparency, explainability
- Human oversight & accountability
- Data Governance

AI Tools

- Acceptable use & guidelines
- Data privacy & confidentiality
- Risk-based usage policies
- Shadow AI visibility & control

AI Security

- Threat modelling for AI
- Data & model security
- Adversarial attack detection
- Incident response & recovery

Key Questions

What is being used? Who is using it? What data is involved? What risks exists? What happens when things go wrong?

IERP®
Institute of Enterprise Risk Practitioners

STANDARDS ALIGNMENT

Global Regulatory Frameworks

NIST AI RMF

Translates 7 voluntary characteristics into measurable dimensions.

EU AI Act

Automates classification of “High-Risk” systems for Article 14 compliance. Unacceptable risk – avoid unnecessary trouble upfront!

Risk Assessment Engine

ISO/IEC 42001

Provides the process and exact logic required for the mandatory Statement of Applicability.
8 cycles; AIIA metadata

NAIO Upcoming AI Bills

National AI Principles, AI Safety functions, Investigation and enforcement functions to exercise AI Enablement functions

The operational bridge between local execution and international compliance.

What the Engine Is — and What It Isn't

What it is

- **Consistent & deterministic** A repeatable scoring engine — same answers in, same tier out.
- **Explainable** Translates technical mechanics into business impact.
- **Traceable** Provides an audit trail for every governance decision.
- **Proportional** Multi-speed — low-risk systems move fast.
- **Agentic-AI ready** Purpose-built for autonomous and multi-agent systems.

What it isn't

An innovation blocker.

Compliance paperwork.

A generic AI principles document.

Governance that people route around isn't governance. It's an obstacle with a policy number.

The objective is not to slow AI adoption. It is to accelerate safely — by applying governance proportionate to risk.

THE LINE YOU'RE ABOUT TO WALK

Six Stages, One Line

6 core areas that is important to get right

1

Risk Appetite

Decided before any system exists. What we govern, and against which standards.

2

Universal Gate

8 questions, asked of everything. A tier out in minutes, enforced in the portal.

3

Deep Dive

Thirteen dimensions scored

4

Deployment Gates

F · R · X · B. Evidence that it ran, not an attestation that it exists.

5

Proportionate Controls

The same small toolkit, mixed to fit what the system can actually do.

6

Live + Telemetry

Monitored in production — and the numbers feed back in the loops

ONE THREAD THROUGHOUT — the customer bot enters at Stage 2, meets the emergency brake at Stage 3, and ships at Stage 6.

DEFINING THE BOUNDARY

Is It Workflow Automation, or Is It AI?

Does a human define every single path?

Yes — Workflow automation

No — AI

Does it get smarter from data, with no manual update?

Yes — AI

No — Workflow automation

Is the goal efficiency — or judgment?

Efficiency — A repetitive task, faster and without mistakes

Judgment — Mimics human decisions on complex problems

Automation fails at a line of code you can point to. AI fails as a distribution you have to measure — different failure mode, different governance.

DEFINING THE UNIT

Consistent Scoring Needs a Consistent Unit

1. CANDIDATE FILTER

Is it an AI capability?

NOT AN AI SYSTEM

- **Oven** — Ordinary software
- **Timer** — Rule-based automation
- **Data retriever** — Retrieval only (no AI decision)

Do not license

AI CAPABILITY

Generates predictions, recommendations, decisions or content

Continue to Separation Test

2. SEPARATION TEST

Does it need a separate license?

S1	**Distinct Identity?** Unique credentials or identifier.	YES
S2	**Independently Deployable?** Can be deployed, updated or retired on its own.	YES
S3	**Outside Direct Control?** Operates or acts in systems without human mediation.	YES
S4	**Different Intended Purpose?** Materially different purpose, stakeholders or impact.	YES
S5	**Executes Actions / API Calls?** Can call APIs, tools or issue instructions that cause actions.	YES

NO

LICENSE AS SEPARATE AI SYSTEM

DO NOT SEPARATE

3. OUTCOME

Two possible outcomes

LICENSE AS SEPARATE AI SYSTEM

Register and assess independently.

DO NOT SEPARATE

Part of the same AI system.

ALIGNED WITH LEADING FRAMEWORKS

ISO/IEC 42001

Govern based on intended purpose, context and lifecycle.

Different purposes require separate management.

EU AI ACT

Obligations and risk levels apply to the AI system for a particular purpose.

Assess by intended purpose and impact.

NIST AI RMF

Manage AI risk within its specific context and use case.

Each use case/capability is a separate AI system.

GARTNER

Inventory and govern AI at the use-case level, not just the platform level.

Use-case based governance reduces risk and complexity.

THE UNIVERSAL GATE, IN PLAIN ENGLISH

Four Questions Every AI System Answers First

Before any dimension, any score, any gate — every application/system answers these four questions first. The Blast radius screen

C — Customer Reach
Does this reach customers or Public

0 — Internal only
Internal staff only. No external visibility

1 — Trusted partners
Regulated partners, B2B Clients, Monitored public spaces

2 — Public customers
End customers, Corporate Clients, General public

D — Data Sensitivity
What data does it access, process, store?

0 — Public data
Publicly available data only

1 — Internal data
Internal Company Data – but not personal nor regulated

2 — Personal data
Customer PII/CDR /Financial/Health/Location

L — Decision Impact
How is output used in decision making?

0 — Just informs
Advise only – Human decides

1 — Helps, human signs off
Requires formal review/sign-off

2 — AI answer is the answer
AI Output stands as the decision. No human review

E — Execution Capability
What AI can do?

0 — Just content generation
Read/inform only; No system actions

1 — Bounded Actions
Executes bounded actions with human approvals

2 — Acts on its own
Execute autonomously via API, code, system writes.

GATEWAY

One rule: any single “level 2” answer, on any of the four, automatically means High tier.

STAGE 2 IN PRACTICE

The AI App Registry Portal

Universal gate assessment, enforced in code — not filled out from memory.

1

Use Case Submitted

Team logs a new AI use case directly in the portal — no spreadsheet, no email chain.

2

X Gate Questions

Reach, data, decision impact, execution, irreversibility — scored against defined criteria.

3

Escalation Logic Applied

Any level-2 answer force-escalates to High — automatically, the same way every time.

4

Tier & Scope Logged

Tier and S0/S1/S2 scope written straight to the AI registry — no manual re-entry.

5

Handoff to Deep-Dive Team

High/Medium systems route to the task force for full assessment — by design, not by tooling.

Automated by the portal

Deliberately handed to people

Six questions, one rule, applied the same way every time. Tiering is no longer a matter of who filled out the form.

THE ASSESSMENT ENGINE- DEEP DIVE

Thirteen Dimensions, Four Business Lenses

Impact

Who is affected — and what can go wrong?

C · L · I

Customer reach · Decision impact · Irreversibility

Data

What is the payload, and whose hands has it passed through?

D · T

Data sensitivity · Third-party risk

Autonomy

Is it a tool — or is it an agent?

E · P · G · A · M · O

Execution · Goal · Orchestration

Governance

Can we control it, and can we audit it?

F · R · S

Fairness · Reliability · System

Teams score the technical dimensions. Leadership gets answers to four business questions.

USE CASE PROFILE · LOW RISK

Internal Productivity Copilot

LOW RISK

SCENARIO

An AI assistant that summarises internal meetings and drafts emails. It reads and suggests — a person does the rest.

THE READOUT

1. **Impact:** Low — informs only, a human validates
2. **Data:** Low — internal, non-PII
3. **Autonomy:** Zero — read-only, no agentic behaviour

THE CONTROLS THAT FIT

Tool Checklist

Listed on the registry, so nobody loses track of what it can read.

Traffic Checkpoint

Every request through one logged gateway — cheap insurance for something already safe.

Everything else stays switched off — on purpose. Heavier controls here would only slow down something that is already safe.

GOVERNANCE OUTCOME · Minimal governance. Basic logging. Approved for rapid deployment — no committee review.

USE CASE PROFILE · HIGH RISK

Customer-Facing Generative Bot

HIGH RISK

SCENARIO

A customer service bot handling live queries and recommending products based on a customer's profile and billing data.

THE READOUT

1. **Impact:** High — public, customer-facing
2. **Data:** High — customer PII and billing data
3. **Autonomy:** Medium — generates novel text, bounded actions

THE CONTROLS THAT FIT

Stress Tests

Adversarial questions before launch — and continuously after.

Privacy Shield

Billing and account details masked before the model ever sees them.

Change Log

Every prompt tweak tracked, so a bad answer can be traced and undone.

Fairness Gate

Training data; Testing across different customer segments

Public-facing and handling real customer data — three of the five core controls, plus one extra for privacy.

GOVERNANCE OUTCOME · Mandatory red-teaming. Prompt-injection blocking. Continuous hallucination monitoring. Privacy shield on PII.

USE CASE PROFILE · CRITICAL RISK

Trade Autopilot

CRITICAL RISK

SCENARIO

A multi-agent orchestrator that independently rebalances client portfolios, executes trades, and moves funds between accounts.

THE READOUT

1. **Impact:** High — moves client capital directly
2. **Data:** High — account holdings and personal data
3. **Autonomy:** Maximum — independent execution, dynamic rebalancing

THE CONTROLS THAT FIT

Action Map

Every trade and transfer mapped in advance — no undocumented path.

Human Sign-Off

Above a set size, a person approves before it fires.

Emergency Stop

One tested kill-switch, with named 24/7 authority.

The checklist and stress tests from the earlier examples still apply — this is what gets added on top, because money moves on its own.

GOVERNANCE OUTCOME · Hard trade-limit containment. Human sign-off above threshold. Emergency stop. Full CIO / CRO and Board visibility.

STAGE 4 · DEPLOYMENT GATES — MECHANISMS, NOT INTENTIONS

Good Intentions Don't Work. Mechanisms Do.

A policy

"Teams should test for bias."

A mechanism

The test is a gate. Code doesn't ship until it passes and the result is logged — and every failure found in production goes back into the suite.

Defined input · concrete tool · clear owner · a feedback path that makes it stronger.

"Good intentions never work. You need good mechanisms." — Jeff Bezos

The four mechanisms in our engine

Tier at intake

Six gate questions before any pilot or purchase. Any level-2 on reach, data, impact, execution or irreversibility → High tier. Output: a tier in the registry.

Architecture enforcement

Every call via the LiteLLM proxy, virtual key per application, direct model execution blocked at the infrastructure layer — not in the prompt.

Gates that block

F · R · X · B. No AI-specific red-team → Medium and High blocked. No tested kill-switch with named 24/7 authority → blocked.

Re-assessment trigger

Contractual model-update notification: any model or version change re-fires the X gate and a full re-assessment. This is the closed loop.

The proof it scales — Waymo: flag every mile → test in simulation → verify → only then deploy. 170M+ autonomous miles → ~12× fewer serious-injury crashes (92% reduction). 'Demonstrably safe' — proven, not promised

CONTROLS IN PRACTICE

The higher the consequence, the higher the control intensity.

The same five controls, used in a different mix depending on what each system can actually do.

	Copilot LOW	Customer bot HIGH	Trade autopilot CRITICAL
Tool checklist (tool & data registry)	●	●	●
Traffic checkpoint (AI proxy gateway)	●	●	●
Stress tests (evaluation & red team)		●	●
Action map (agent action audit map)			●
Change log (version control)		●	●

Plus the extras: the customer bot adds a **privacy shield** · the trade autopilot adds **human sign-off** and an **emergency stop** — because money moves on its own.

THE FULL LINE, IN ONE PICTURE

From Risk Appetite to AI Deployment

Six stages, one line — and a return arrow, because the line is a loop, not a lecture.

SAME THREAD, ONE REAL SYSTEM — the customer bot scores H2 at the gate → proportionate controls: **stress tests · privacy shield · change log** → deploys live with red-teaming and PII masking already running.

This is the title, drawn once. Risk appetite in — a governed, monitored deployment out.

IERP®
Institute of Enterprise Risk Practitioners

AGENTIC DIMENSIONs(E -> PGAMO)

The Agentic Ladder — Controls by Rung

RUNG	WHAT CHANGES	CONTROLS THAT SWITCH ON	DESIGN RULE
1 · Single chatbot *Answers questions from a script.*	Talks only. Cannot touch anything.	Registry entry · gateway logging. The agentic block is not scored at all.	***Don't over-govern. Anything heavier here is theatre.***
2 · Chatbot + tools *Can now press buttons for you.*	Becomes a doer. The biggest jump on the ladder.	Task-scoped credentials, not a shared service account · tool allow-list · session monitoring of inputs, outputs and tool calls · injection testing incl. via retrieved docs and tool responses	***Least privilege, enforced at the gateway — never in the prompt.***
3 · Autonomous agent *Sent off for a week; reports back.*	Nobody watches each step. Failure is slow drift, not one bad action.	Goal-drift testing over extended runs · context management (attention pinning, memory distillation) · session budget cap · tested kill-switch with named 24/7 authority	***Bound the session, not just the action.***
4 · Orchestrator *A team with named roles and a manager.*	Several agents — but you can list every one.	Cryptographic agent-to-agent authentication — identity by signature, not by system prompt · rate limits and allow-lists on every external tool reachable	***Enumerate the team. Hardcode the routing.***
5 · Dynamic orchestrator *The team reassigns its own roles.*	Roles carry permissions — so a role switch is a permission switch.	Switching confined to a predefined state chart under supervisor oversight · role-change frequency metric per session, with an alert threshold	***Switching is fine. Unbounded switching is not.***
6 · Self-spawning swarm *The team hires people without telling you.*	You can no longer enumerate who is in the system.	Cascade testing — collusion, deadlock, viral instruction propagation · ecosystem allow-listing · per-agent ephemeral credentials so spawned agents inherit nothing	***Trust boundary must be cryptographic, not architectural.***

Every rung is delegation to a new hire: what can they touch, what can they decide, and can they bring other people in?

IERP®
Institute of Enterprise Risk Practitioners

THE STORY, END TO END

One Engine. Proportionate Governance. Trustworthy AI.

Foundations → Grounded in standards → The scoring engine → Proven in practice → Embedded in operations

TRUSTWORTHY AI OUTCOMES

Safe · Fair · Secure · Compliant · Effective · Human-Centric

Governance isn't a brake on AI adoption.
It's how we scale it safely — at the speed the business needs.

Demonstrably safe — proven, not promised.

How to Read What has been shared

Written from one context

This reflects a Malaysian telecommunications operator — our regulatory perimeter, data estate, risk appetite and organisational maturity. Yours may differ.

Patterns travel; parameters don't

The mechanisms — gates, tiers, evidence requirements — are portable. Specific thresholds, weightings and control mixes are calibrated to us and should be re-derived, not copied.

Reference, not prescription

Nothing here constitutes legal or regulatory advice. Figures and thresholds shown are illustrative. Please validate against your own regulators and advisors.

Shared in the spirit of open practice — so you can shorten your own path, on your own terms. And we could learn together

IERP®
Institute of Enterprise Risk Practitioners

GLOBAL CONFERENCE 2026

FROM CHAOS TO COMMAND:
RISK LEADERSHIP IN INDUSTRY 6.0

THANK YOU

SCAN THE QR CODE
TO KNOW MORE ABOUT US!

enquiry@insterp.com
www.insterp.com

Reach me at https://www.linkedin.com/in/alexyap/

IERP®
Institute of Enterprise Risk Practitioners

ADAPTIVE GOVERNANCE

One Number, Two Loops

Telemetry doesn't just prove a system is behaving. It feeds back into the system — and into the rules themselves.

THE TELEMETRY

Counted at the gateway, per release

- Injection attempts blocked, per 1,000 calls
- Hallucination rate vs the R-gate benchmark
- Containment time in kill-switch drills
- PII guardrail catch rate

LOOP 1 **The system gets harder to break**

Every blocked injection joins the red-team suite. Drift past threshold re-fires re-validation. A vendor model update re-fires the X gate. Month twelve is measurably harder to break than month one — and the curve is on file.

LOOP 2 **The framework gets smarter**

When the same gap recurs across assessments, the framework itself is amended. v2.3 split the R gate into three separately recorded legs — because one compressed question let "tested for accuracy, never measured for hallucination" pass as a yes.

Where the loop closes: every amendment resets the baseline the telemetry measures against — a tighter gate, a new evidence requirement, a new thing to count.

Loop one makes each system safer over time. Loop two makes every future assessment smarter. Governance that improves at the rate the threat does.