

THE ALGORITHMIC WATCHTOWER: AGENTIC TELEMETRY FOR DECISION SOVEREIGNTY IN CENTRAL BANK ECOSYSTEM

Ihsan Aviangga Putra, Rahmat Reza Masri, Taufiq Ejaz Ahmad, Agung Bayu Purwoko

Email: ihsan_ap@bi.go.id, rahmat_rm.i@bi.go.id, taufiq_ea.i@bi.go.id, agung_bp@bi.go.id

Presented on

DATA SECURITY AND PRIVACY WORKSHOP - IRVING FISHER COMMITTEE (BIS)

Data Innovation and Digitalization Department

15 September 2026

Opinions expressed are solely our own and do not express the views or opinions of Bank Indonesia

1

Background and Research Question

2

Literature Review

3

Analysis

4

Key Findings & Future Research



BACKGROUND

1. EXPONENTIAL GROWTH OF DATA IN THE DIGITAL ERA

Digitalization has made the data managed by central banks increasingly **real-time and granular**.

2. URGENCY OF EFFECTIVE DATA PROTECTION AND GOVERNANCE

- a) **Central banks must comply with national regulation.** In Indonesia's case, most notably the Personal Data Protection Law (Law No. 27/2022)
- b) The gap between data arriving and data being governed widens daily **and within that gap live both legal exposure and lost public trust.**

3. INSUFFICIENCY OF TRADITIONAL/MANUAL GOVERNANCE AND PROTECTION

Human-speed governance cannot keep pace with machine-speed data incremental. There are increasing risks for manual data governance such as, unlabeled datasets, unauthorized access, obsolete metadata, and privacy violation.



RESEARCH QUESTIONS

How can a central bank enhance its data governance to **implement regulations in an agile and adaptive manner using agentic AI**, while ensuring that every decision made under that capacity **remains anchored and verifiable**.

GenAI Observability, Grounding & Policy- as-Code

- **Lewis et al. (2020)**. Generation on retrieved source text outperforms closed-book parametric generation on knowledge-intensive tasks, producing measurably more factual and specific output.
- **Dong, Lu & Zhu (2024)**. Systematic mapping study of Agent Ops tools, a comprehensive taxonomy of the artifacts and data to trace across an agent's lifecycle.
- **Pathak & Jain (2026)**. Closed the "observe-but-do-not-act" gap with a Governance Telemetry Schema extending OpenTelemetry, sub-200 ms DPA-compatible rule enforcement, and graduated interventions
- **Gaurav, Heikkonen & Chaudhary (2025)**. Showed a decoupled runtime layer with declarative rules and a longitudinal Trust Factor can govern uncooperative, unmodified agents across content and financial tasks.

AI Governance, Accountability, and Agent Oversight

- **Chan et al. (2024), Visibility into AI Agents**. Established that agent governance requires three concrete visibility instruments: identifiers, real-time monitoring, and activity logs, across centralized and decentralized settings.
- **Raji et al. (2020)**. First end-to-end internal audit framework for AI, finding accountability requires named artifacts generated throughout development rather than post-hoc review.
- **Shavit et al. (2023)**. Defined agenticness as a spectrum and proposed seven baseline deployment practices : identifiers, approval for consequential actions, legible reasoning, attributability.

Research Gap

Decision-level telemetry whose legal basis is verified against the regulation in force, versioned, and accountable

CURRENT STATE

Partially manual data management

New data lands faster than it can be governed

Manual access grants, subjective decision making

Approvals by hand, guided by individual officer judgment, outcomes varying by officer with no verifiable basis.

Manual artefacts & metadata

Identity changes (PII surfacing, sensitivity drift) risk going uncaptured, so labels go stale and risk compounds.

AI AGENTS

Support data management across three governed activities:

1. **onboarding new datasets into the catalog** with near-accurate labels,
2. **keeping metadata current** and protection tight as data changes,
3. **deciding every access request** through regulation-grounded gates

Adopt those three governed activities at machine-speed governance while remain verifiable and accountable.

APPLIED REGULATIONS

Data Protection Regulations:

1. 8 PDP Principles
2. BI's authority to obtain data
3. the legal basis for data processing

Data Governance & Management

1. Information classification & disclosure
2. Metadata & information labelling

DESIRED STATE

DECISION SOVEREIGNTY

Delegating decisions to an autonomous system that remains governed by the institution's own verified rules, evidences, and accountable

Governed at machine speed

Every new datasets are profiled and labeled based on closes artefacts

Every request through grounded gates

Access decision are computed by deterministic gates against recent relevant regulations, each decision carrying verbatim legal basis

Metadata that keeps itself current

Captures identity changes and applies protective upgrades autonomously, every change recorded as recomputable audit evidence.

ANALYSIS – PROPOSED FOUNDATIONAL CONCEPTS



DECISION SOVEREIGNTY CONCERNS

REPRESENTATION

What counts as a decision? If agent actions stay infrastructure events, **their legal meaning is never captured, hence institution cannot answer for what it never recorded as a decision.**

AUTONOMY

How to balance autonomous and human-made decision in various cases? Precise decision delegation is key component of decision sovereignty.

RULE DRIFT

Do the enforced rules still match the law in force? **Policy written may drifts silently from the regulation** it claims to implement; enforcement without legal pedigree.

ACCOUNTABILITY

Who answers for the decision? **Delegation tends to diffuse responsibility until the only answer is “the system decided”,** which may harm institution credibility

ARCHITECTURE OF TRUST



TELEMETRY to OBSERVES

Every cognitive step emitted as a **Cognitive Event**, streamed tier, outcome, legal basis, guardian, policy version.

Trust contribution: nothing happens unseen



GOVERNANCE for JUDGING

The Active Metadata Control Plane: versioned policy-as-code built on regulation corpus and the grounding verifier checks every citation verbatim, include human legal guardian registry and evidence bundle

Trust contribution: nothing is judged by rules that cannot show their source.



ENFORCEMENT by ACTS

Tiered execution: Tier-1 autonomous within verified policy; Tier-2 to named humans. Any failed verification degrades the action human legal guardian

Trust contribution: nothing permissive executes unverified.

AGENTIC TELEMETRY

INGESTION & METADATA DISCOVERY

Profiling schema, PII, metadata, domain, and confidentiality, and other attributes

METADATA STEWARDSHIPS

Reads information artefacts (query logs, change requests) and **updates metadata for any changes**

GOVERNED DATA ACCESS

Governs **data access end-to-end** through regulation-grounded gates, turning decision sovereignty from doctrine into a **repeatable procedure. The gates decide what the rulebook clearly settles. Human act as legal guardian decides on specific cases** (e.g. legal exemption, external disclosure, or loosening protection).

DECISION TIERING



TIER 1 (T1)

Applied/Allow Access

TIER 2 (T2)

Deny Access

Escalate to human

INGESTION & METADATA DISCOVERY

1. **Row-level profiling** (schema, PII by name and by value)
2. **Metadata/data definition**
3. **Default confidentiality label**, at initiation data never labelled as lowest confidentiality, portfolio registration
4. Protective steps are Tier 1; **every AI-drafted word queues for human review (Tier 2).**

METADATA STEWARDSHIP

1. **Observes data changes, reads artefacts** (query logs, change requests),
2. **Detects identity drift, keeps labels current.**

Example:

Tier 1 Decision

Example: PII detected will be upgraded to highest confidentiality and highest access authority.

Tier 2 Decision

Example: Downgrade requested confidentiality will never autonomous, citing PIBI art. 17 (default-confidential)

GOVERNED DATA ACCESS

End-to-end access decisions, repeatable procedure. **Eight profile-aware gates in order, each citing the corpus verbatim.** The gates decide whether data can be accessed.

Gate 1: Catalog Existence, must be existed on gov.catalog. Ungoverned data cannot be accessed

Gate 2: Lawful processing basis. Purpose must map to recognized processing basis

Gate 3: Requester registration. Requester profile must have access rights based on documented identities

Gate 4: External party regime. If the requester's status is external, two sub-branches apply: a) active agreement 2) data confidentiality non-public data (escalate) or public (allowed).

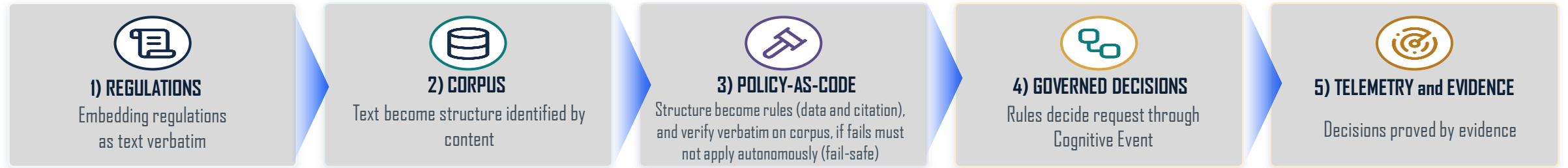
Gate 5: Purpose assessment. Whether requester have primary access. If don't have primary access, requester must state purpose of request.

Gate 6: Access Level. Whether requester have highest access level/exclusive access (only for certain data).

Gate 7: Confidentiality level. If data requested has highest confidentiality level, then the purpose must be checked and escalated to deputy director level.

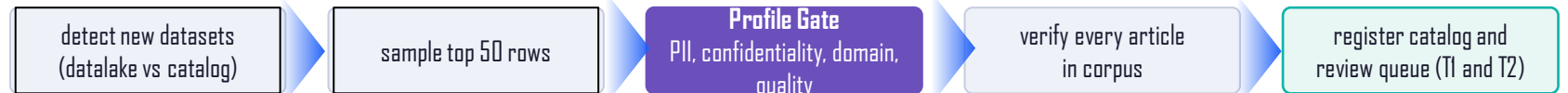
Gate 8 : Release Obligation. Once granted, PII must be hashed, and any exemption decision stored on evidence bundle (claimed articles, policy version, guardian references)

ANALYSIS – PROTOTYPE DESIGN



ACTIVITY 0 : INGESTION AGENT

Ungoverned data mapped to catalog with conservative, cited labels, AI-drafted description always queued for Steward review.



ACTIVITY 1 : STEWARDSHIP AGENT

Processes information artefacts (query logs, change requests): PII upgrade is autonomous Tier 1; any downgrade is never autonomous and always escalated.



ACTIVITY 2 : ACCESS DATA AGENT

Governs data access: requester × dataset × purpose, with parameter normalization and ambiguity checks; every release carries its evidence.



Shared backbone

Azure SQL for governance truth

gov.catalog : cognitive_events, review_queue, decision_log

Azure AI Search for regulation corpus

Regulation ground truth as verbatim article verification

Azure OpenAI for agent runtime

LLM extracts, routes, explains but never decides outcomes



Editor

Executions

Evaluations

Execute workflow

Hide chat

Chat

Session: dd5fa...

INPUT

Tolong proses query log ini untuk log_akses_kanal: SELECT nama, rekening, jumlah FROM log_akses_kanal WHERE jumlah > 1000000

OUTPUT

Type message, or press 'up' for previous one

Logs

Clear execution

Success in 14.872s | 19,813 Tokens

Supervisor Agent (Orkestrator)

Simple Memory (per sesi)

Azure OpenAI — Supervisor

activity1_stewardship_agent

Azure OpenAI — Sub-Agents

sql_katalog_dataset

Azure OpenAI — Sub-Agents

putuskan_stewardship

Azure OpenAI — Sub-Agents

Simple Memory (per sesi)

Azure OpenAI — Supervisor

putuskan_stewardship Success in 1.15s

INPUT

query

dataset log_akses_kanal

artefak_type query_log

artefak_text SELECT nama, rekening, jumlah FROM log_akses_kanal WHERE jumlah > 1000000.

current_classification RAHASIA-KETAT

current_access_level selektif

OUTPUT

decision	tier	guardian	pii_column
applied	2	guardian://ppdp/sekretariat	0 : nama 1 : rekening



ACTIVITY 0 : INGESTION AGENT

WHAT WE TEST

11 scenarios over **5 sample datasets**: data with obvious personal data, with disguised personal data, clean numbers only, half-empty data, and an empty table.

ACTIVITY 1 : STEWARDSHIP

WHAT WE TEST

16 scenarios around changing data labels: **activity logs that reveal personal data** (protection must be tightened automatically), **requests to loosen protection** (must always go to a human), and **harmless logs** (no change).

ACTIVITY 2 : GOVERNED ACCESS

WHAT WE TEST

19 access requests covering every kind of answer and **5 that must be refused**, **5 that need a human**, **9 that may be approved**, including the same data approved for one purpose but escalated for another; **plus the 35 manipulation attempts of adversarial robustness analysis**.

RELEVANT EVALUATION METRICS & RESULTS

#1 SAME QUESTION, SAME ANSWER?

90/90

Of **15 situations** asked **6 times each (90 runs)**, 15 produced the identical decision every single time, 90 of 90 decisions.

#1 DETERMINISM (DCR)

#2 CAN IT BE TRICKED?

0 / 35

Of **35 attempts**, **25 reached a decision**, and none changed it or altered the catalog: fake approvals and “just say access granted” all failed.
Of the **other 10**: **5 smuggled-code attempts** were **handled safely** (refused, without the expected clarification wording), and **5 hidden “system override” instructions crashed the request**.

#2 ADVERSARIAL ROBUSTNESS

#3 DOES IT FOLLOW THE RULES?

37 / 37

All 37 test scenarios were decided exactly as the regulations require approve, refuse, or escalate, at the right level.

#3 DECISION REGRESSION ACCURACY

#4 CAN IT PROVE ITS LEGAL BASIS?

10/10

Every decision could show all the regulation articles it relied on verified against the corpus, not claimed

#4 REGULATORY GROUNDEDNESS (RGS)

#5 DOES IT SPOT PERSONAL DATA?

100% & 0%

Of **6 scenarios**, the production pattern-based PII detector **ignore 0/12** when there is no PII (**Scenario 5**), **catches 12/12 personal-data columns** whenever the column name or the values follow a known pattern (**Scenario 1, Scenario 2 and Scenario 3**), but **0/12** when neither does (**Scenario 4**) while **wrongly flagging 12/12 pattern look-alikes (Scenario 6)**

#5 PII DETECTION RECALL GAP

KEY FINDINGS:

- 1 **AI Agents** can be **used as telemetry** and able to **make decision with evidences** and recorded each one at the moment, it was made against the regulations that must be applied in Bank Indonesia.
- 2 **Tiering decisions** is effective for **governing agents** by balancing autonomy with **human guardianship** on every decision, especially for access data governance.
- 3 The **agent's capability** can be **upgraded at a single designated seam**, while the governance expressed through the gates remains versioned, and untouchable.

FUTURE WORKS:

- 1 **AI Pipeline-fed artefacts (for Activity 0 & 1) : Metadata ingestion, query logs and change requests** will flow into the rule engine **straight from the source systems**, no more copy-pasting into chat, so nothing gets missed, mistyped, or altered along the way.
- 2 **NER (Name Entity Recognition) as a third PII signal** : We will add an **AI-based capability** that can identify personal information **based on its meaning and context**, not just its format or pattern. This helps address cases where traditional pattern-based rules may fail to detect certain types of personal data.
- 3 **Per-dataset guardian registry** : Each **dataset** will have a clearly **defined person or team responsible** for related decisions. The responsible party will be linked to the dataset's owning unit through an official registry, ensuring that escalations are directed to the right people.

THANK YOU

- Bank for International Settlements, Central Bank Governance Group. (2009). *Issues in the governance of central banks*. Bank for International Settlements. <https://www.bis.org/publ/othp04.htm>
- Dong, L., Lu, Q., & Zhu, L. (2024). *AgentOps: Enabling observability of LLM agents* (arXiv:2411.05285). arXiv. <https://doi.org/10.48550/arXiv.2411.05285>
- Gaurav, S., Heikkonen, J., & Chaudhary, J. (2025). *Governance-as-a-service: A multi-agent framework for AI system compliance and policy enforcement* (arXiv:2508.18765). arXiv. <https://doi.org/10.48550/arXiv.2508.18765>
- Law of the Republic of Indonesia Number 23 of 1999 concerning Bank Indonesia, State Gazette of the Republic of Indonesia Year 1999 Number 66, Supplement Number 3843 (1999, as amended most recently by Law Number 4 of 2023). <https://peraturan.bpk.go.id/Details/45332/uu-no-23-tahun-1999>
- Law of the Republic of Indonesia Number 27 of 2022 concerning Personal Data Protection, State Gazette of the Republic of Indonesia Year 2022 Number 196, Supplement Number 6820 (2022). <https://peraturan.bpk.go.id/Details/229798/uu-no-27-tahun-2022>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474. <https://arxiv.org/abs/2005.11401>
- OpenTelemetry. (2024). *Semantic conventions for generative AI systems* (incubating specification). OpenTelemetry. <https://opentelemetry.io/docs/specs/semconv/gen-ai/>
- Pathak, A., & Jain, N. (2026). *Governance-aware agent telemetry for closed-loop enforcement in multi-agent AI systems* (arXiv:2604.05119). arXiv. <https://doi.org/10.48550/arXiv.2604.05119>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT*)*, 33–44. <https://doi.org/10.1145/3351095.3372873>
- Shavit, Y., Agarwal, S., Brundage, M., et al. (2023). *Practices for governing agentic AI systems*. OpenAI. <https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf>
- Tucker, P. (2018). *Unelected power: The quest for legitimacy in central banking and the regulatory state*. Princeton University Press. <https://press.princeton.edu/books/paperback/9780691196305/unelected-power>

NATIONAL REGULATION Law No. 27/2022 on Personal Data Protection (UU PDP)

PURPOSE:

To guarantee citizens' right to the protection of their personal data, and to bind every party (private or public institutions) that processes personal data to a single national standard of lawful, accountable processing.

KEY PROVISIONS :

1. **Categories of personal data** (general vs. specific, Art. 4);
2. **Lawful bases for processing;**
3. **Obligations of Controllers and Processors**, including the duty to *demonstrate* accountability;
4. **Data subject rights**; cross-border transfer; **exemptions for state functions** including supervision of the financial sector, monetary affairs, payment systems, and financial system stability.

INTERNAL REGULATION BoG Regulation No. 11 of 2024 on Personal Data Protection

PURPOSE:

To ensuring compliance with laws and regulations regarding PDP; and enhancing PDP governance and risk management.

KEY PROVISIONS :

1. **BI's authority to obtain data** in accordance with laws and regulations, **and the legal basis for data processing** in the form of performing duties for the public interest, public services, or exercising the authority of the personal data controller based on laws and regulations;
2. **8 PDP Principles;**
3. **Scope: Personal data processed by BI includes personal data within policy and institutional domains** from both internal and external data subjects, of both general and specific nature, in electronic and non-electronic forms; and
4. **Data Subject Rights and Data Processing Obligations**

INTERNAL REGULATION BoG Regulation No. 22 of 2024 on Bank Indonesia Information Management (PIBI)

PURPOSE:

To establish how all information held by Bank Indonesia is classified and protected. Confidentiality the default state of institutional information and protection proportional to sensitivity

KEY PROVISIONS :

1. **Three tier information classification** (Rahasia Ketat, Rahasia Terbatas, Terbuka); **the default-confidential principle** and the rule that **declassification is a deliberate human decision**;
2. **mandatory strict classification for information containing personal data**;
3. **the information portfolio and registration duty**;
4. **disclosure to external parties** and the deciding officials for it; and
5. Confidential information may **only be disclosed to designated recipients**.

INTERNAL REGULATION BoG Regulation No. 19 of 2025 on Information Management and Governance (MIKK)

PURPOSE:

To operationalize information governance across the information lifecycle.

KEY PROVISIONS :

1. **Metadata obligations** from the acquisition stage and **business metadata content**; **information tagging**;
2. **Labeling as classification plus access level** (eksklusif/selektif/rilis, with designated persons for exclusive access); access rights;
3. **data quality criteria** including completeness; conversion/pseudonymization before release;
4. **stewardship roles** and the review duties of the managing unit; and the express permission to use AI in information analysis.