

AI as Mirror and Engine – governing from within in an algorithmic age

René de Baaij, Groesbeek, 19 January 2025

Reader's Guide

- **For whom:** RvB (Executive Board) / RvC (Supervisory Board), MT (management team) and public sector leaders; focus on strategy, legitimacy and duties of care.
- **How to read:** sections 1–7 build the argument (human & system); sections 9–11 provide the instrumentarium. Read the Prologue (section 0) for context; for implementation feel free to jump to 9–11 and leaf back for depth.
- **What you'll take away:** three decision lines (value, risk, legitimacy), one cadence (monitor–challenge–audit) and five artefacts (model card, logbook, risk file, user guide, exit plan).

Prologue – The room, the screen, the gaze

It's early. The screen lights up before the meeting room fills. The first question of the day is not a greeting but a prompt. An answer rolls out that holds together better than yesterday's notes.

Someone sighs with relief, someone else frowns, a third feels inexplicable irritation. In those three micro-reactions the playing field is exposed: AI as promise, as threat, as mirror. What happens here is not merely technical or procedural; it is also psychological, relational, managerial and legal. It touches how you set direction, how teams make meaning and how the organisation carries responsibility.

This longread connects two lenses that rarely meet in a single story: the psychodynamic, relational and humanistic foundation **and** the managerial, business and legal perspective. We look inward and outward at the same time, at undercurrent and uppercurrent. At what AI does with people, and what people do with AI; at steering, structure and accountability.

The core message is simple and sharp: **AI does not change humans into another species; it amplifies what is already there – and makes the consequences administratively and legally explicit.**

Core statements (with brief explanations)

- **AI amplifies what exists.**
It strengthens patterns, speeds up rhythms and makes implicit norms explicit in code. That yields clarity, and also enlarged blind spots if counter-force is missing.
- **Neutrality is an illusion without governance.**

A model is not an arbiter but a design choice with assumptions, data and thresholds.

Secure revocability and contestability, or 'efficiency' becomes an alibi.

- **Relationship precedes technology.**

What the system does to people and people do to the system shapes culture.

That requires dialogue, exception and repair as structural functions.

- **Legitimacy is evidenced action.**

Care counts only when it is traceable in logbooks, cards and decisions.

Explanation and remediation are not side issues; they are part of the product.

1. Inner world and uppercurrent – projections onto the algorithm

In classic hierarchies omnipotence, certainty and omniscience were projected onto the leader. Now the algorithm slides into that role. It appears as the new "object": smart, fast, apparently neutral. That appearance of neutrality makes it attractive as a carrier of idealisation and denial. "The model says so" has grown into a modern defence mechanism – a way to avoid shame, doubt or moral burden. Psychodynamically this is not aberrant but predictable: transference and counter-transference occur as much **between people and systems** as between people. We speak to the system, but really to our need for certainty. We listen to the score, but really to our fear of falling short.

Administratively this carries through. Value creation shifts from individual expertise to the **configuration** of data, processes, models and governance. Legally the question shifts with it: not "who did this?", but "how is the system designed, who owns it, which duties of care applied and is remediation possible?"

In terms of motivation this touches autonomy, competence and relatedness (Self-Determination Theory; see note 1).

Three simpler, therefore harder questions belong to every AI file from now on:

1. **What does this system do to our roles and relationships?** Who gets a voice, who falls silent?
2. **Where do we use data and models to avoid contact rather than deepen it?**
3. **How do we stay rooted in empathy and proportionality precisely when decisions are being 'optimised'?**

Those who avoid them gain speed without direction, and compliance without legitimacy.

2. AI as system knot – patterns, power and

variety

AI is not a loose tool; it is a **system knot** that creates new feedback loops, dependencies and power structures. From systems thinking and complexity theory the core question shifts: not whether AI “works”, but **which patterns it amplifies or suppresses**.

Requisite variety becomes concrete (see note 8). Organisations that want to handle complex reality need variety in perspectives. Many models do the opposite: they score, sort and normalise. That is efficient, but also a source of **cultural drift** toward conformity and risk aversion. Less deviation, less dissent, less creativity – with a managerial side effect: blind spots that only surface in incidents and complaints.

In systemic language **AI encodes implicit norms**. Power and habitus shift to datasets, features and thresholds. Without attention you entrench exclusion as “objective logic”. That is why governance is not primarily a document but a **relational design**: who defines the data, who selects the models, who can make exceptions and who can overrule? This perspective is sociomaterial in nature (see note 6) and aligns with complex responsive processes (see note 7).

For you as a leader a simple rule with hard consequences follows: **AI programmes are change programmes**. They intervene in roles, processes, identity and accountability. Those who run them as IT (information technology) projects get technical deliverables and social collateral damage.

3. Strategy and operations – speed, meaning and yardsticks

AI can lift productivity, shorten lead times and make quality more predictable. But without a **target architecture** it dissolves into loose pilots and demo wins. Strategically the order is: **problem → value → risk → design**. First define clearly which business problem and which value formula, *then* tools.

Otherwise technology becomes solution and the human becomes problem.

Operationally AI requires a visible **chain design** from question to effect. Intake, design, validation, go-live, monitoring, retraining and sunseting form one cadence. Thresholds and retrain criteria are set in advance. Prompts, features, hyperparameters and versions become **assets with ownership**.

You see: this is no hobby corner; this is service management for decision-making.

Learning is the weak spot. AI excels at **single-loop**: getting better at what we already do.

Double-loop – whether we measure the right things and use the right success criteria – remains a human and organisational process. If that layer is skipped, blindness grows with high precision. The result: beautiful dashboards, wrong decisions (Kolb; see note 3; Argyris & Schön; see note 4; Mezirow; see note 5).

Therefore learning organisations put three practices in place: (1) **reflection on outcomes** – not only

“is it correct?”, but “what does it mean?”; (2) **dialogical decision-making** – data as conversation starter, not as arbiter; (3) **co-evolution of capabilities** – digital literacy plus judgement, moral sensitivity and legal alertness.

4. Lawfulness and legitimacy – the normative infrastructure

AI operates within a **normative landscape** in Europe that is layered and in motion. The basic principles of data protection have stood for years: lawfulness, purpose limitation, data minimisation, accuracy, storage limitation, integrity/confidentiality and accountability.

Around these sit rules for platforms, data access and portability, safety and duties of care in essential sectors (see notes 9–13). AI-specific frameworks ask for controlled risk analysis, documentation, monitoring and options for correction.

The administrative verb is **evidencable**. Not only being careful, but being able to show that you *were* careful: model cards, logbooks, versioning, impact assessments, override decisions with reasoning, incident notes with remediation. Fix counter-force: who may stop the system, and when?

Legitimacy is more than legality. It rests on **explainability, contestability and repair**. Whoever automates must also make visible **what is not** automated. Draw red lines: no behavioural steering without necessity and proportionality; no black box in contexts with high error costs and low explainability; no data collection without necessity and a clear legal basis. Make this explicit, in language that customers, clients, citizens and employees understand.

5. Identity and embodied work – the ability to carry tension

AI touches professional identity. If systems analyse, advise and create, what does it mean to be an expert or a leader? The task shifts from **knowing more** to **being able to carry more**: tension, uncertainty, slowness, conflict and moral ambiguity.

Spiritual and humanistic language finds a concrete function here: staying oriented in a world that seems mechanically makeable. The core questions become: **what will we not do, even if we can?** Which applications do we refuse because they hollow out relationships, dignity or future?

Embodied work is a precondition. AI accelerates rhythm and raises information pressure. The body pays first. Teams that organise rhythm, silence, recovery and reflection stay on course. Teams that skip that layer burn out – and that is not only a wellbeing issue but also a managerial risk: drop-out, errors, eroding trust.

6. Methodical anchoring – inquiry as a practice of awareness

Most AI conversations strand in tool choice, pilots and risk lists. What is missing is **methodical anchoring**: a way of working where reflection, dialogue and meaning-making are taken as seriously as accuracy and uptime.

Qualitative methods – interview, casework, discourse and narrative analysis – are indispensable for understanding what AI does to experience and interaction. Numbers show patterns; stories show meaning. Combine this with **action research**: formulate hypotheses, follow effects, adjust, decide again. Every AI programme becomes a learning process.

On the formal side this calls for an **AI management system** (see notes 14–19): policy and standards; processes for risk, validation, monitoring, incident response, retirement; competencies (model risk, data ethics, legal interpretation); assurance (internal audit, external review, maturity measurement). And – crucial – **due diligence** toward suppliers: data sources and licences, evaluation sets and bias, red-teaming, content authenticity, portability and exit options. Contract audit rights and sanctions for non-compliance.

Methodology thus also becomes **ethics**: how do we reach judgement, who is heard, which data count and which do not? Ownership of meaning must remain with people, not with the vendor. This aligns with international principles and guidance (see notes 20–21).

7. Paradoxes and conditions – the necessary discomfort

Managerial reality is a field of tensions. AI makes them visible and unavoidable:

- **Efficiency ↔ resilience.** Faster and cheaper can be more fragile. *Condition:* choose resilience where disruption is costly; choose efficiency where recovery is easy.
- **Centralisation ↔ autonomy.** One platform gives scale, but diminishes local wisdom. *Condition:* central standards, local variation in application.
- **Automation ↔ human dignity.** Delegation relieves, but can de-humanise. *Condition:* clear bounds for delegation, with contestability and repair.
- **Data minimisation ↔ personalisation.** More data is not always better. *Condition:* make purpose limitation explicit, limit retention, demonstrate effect.
- **Explainability ↔ performance.** Complex models sometimes perform better yet are less traceable. *Condition:* match model choice to error costs and context.

- **Openness ↔ security.** Transparency builds trust, and increases attack surface. *Condition:* differentiate public, internal and confidential.

Tensions do not disappear by policy; they require cadence and conversation. Leading is holding course amid contradictory truths.

9. Minimal viable governance – doing a few things very well

Big ambitions succeed by **doing few things consistently**. A working minimum you can start with tomorrow:

Roles

- **System owner (accountable)** – mandate, resources, reporting.
Holds the mandate and bears integral responsibility.
Secures decision lines, resources and reporting toward the board.
- **Model owner** – performance, bias, lifecycle.
Steers performance, bias and ageing of the model end-to-end.
Guards validation, monitoring and timely retraining or retirement.
- **Data owner** – definitions, quality, access.
Defines sources, quality and access rights for data.
Guards purpose limitation, minimisation and data characteristics for models.
- **Risk & compliance** – frameworks, tests, escalations.
Translates laws and standards into workable requirements in the process.
Orchestrates tests, exceptions and escalations with timeliness.
- **Security** – threats, resilience, incidents.
Assesses threats and vulnerabilities across the full chain.
Secures detection, response and recovery including supplier dependencies.
- **Business owner** – purpose, value, effects in practice.
Anchors problem definition, value and effect in operations.
Organises frontline feedback and decides stop/go.

Cadence

- **Monthly:** model monitoring and incident review.
Assess performance, drift and incidents with concrete actions.
Normalise small corrections so large failures do not arise.
- **Quarterly:** challenge session with independent counter-force.

Provides organised dissent and alternative assumptions.

Prevents groupthink and refines thresholds and scope.

- **Semi-annual:** audit of documentation, drift, fairness and rights.

Gives assurance on documentation, fairness and rights protection.

Makes improvement decisions traceable for audit and stakeholders.

- **On every change:** impact assessment, test report, go/no-go by the board with override criteria.

Conducts impact analysis and tests upfront with clear criteria.

Records the reasoning for go/no-go and any overrides traceably.

Artefacts

- **Model card** (purpose, data, bounds, performance, fairness, explainability).

Describes purpose, bounds, data, performance and assumptions.

Forms the basis for explanation, audit and responsible use.

- **Logbook** (versions, parameters, overrides, incidents, remediation).

Records all versions, parameters, overrides and incidents.

Enables review, learning and liability.

- **Risk file** (error costs, context, mitigations, residual risk).

Translates error costs and context into mitigations and residual risk.

Supports decisions about model choice and degree of delegation.

- **User guide** in plain language (rights, redress, contact).

Explains rights, contestation and contact clearly to users.

Raises acceptance and makes contestability practical and low-threshold.

- **Exit plan** (vendor lock-in, portability, escrow).

Defines portability, escrow and migration steps upon termination.

Reduces vendor lock-in and continuity risks.

All this can be kept compact. It is not a paper factory; it is traceable craftsmanship.

10. From principle to practice – three design questions

A. Where do we place the human counter-force?

In the loop (real time), on the loop (periodic oversight) or above the loop (framing). Choose deliberately per use case and explain why.

At-the-line lowers error costs when impact is high and tolerance is low.

Above- and on-the-loop secure proportionality and learnability over time.

B. How does exception become normal?

Every rule has edge cases. Make exception-making legitimate: who may do it, how is it reasoned, how is it learned from?

Set procedure, criteria and a duty to state reasons – no ad-hoc leniency.

Document, share patterns and take periodic improvement decisions.

C. What degree of variety is needed?

Standardise where safety and scalability require it. Allow variation where contextual knowledge adds quality. Use governance to **make difference possible and traceable**, not to erase it.

Too much standardisation costs quality; too much variety costs scale.

Define decision space and deviation routes explicitly per context.

11. What now, what later, what not – a practical stance

Now (0–3 months)

- **Inventory** all AI applications and models; appoint owners.
Without visibility of the actual footprint, governance stays theoretical and risk unfocused.
A lean inventory prevents shadow IT (information technology), speeds audits and enables prioritisation.
- **Define risk appetite** and critical quality attributes (safety, fairness, privacy, explainability).
Without explicit boundaries, trade-offs become implicit and arbitrary.
Clear thresholds steer design choices and speed decisions in incidents.
- **Set stop criteria and override authority**; arrange a practical kill-switch.
A system that cannot be stopped safely is not safe to start.
Clear authorities limit damage, speed remediation and raise trust.
- **Start impact assessments** for the top three use cases; document data flows, assumptions and error costs.
Impact concerns people, processes and rights—not only metrics.
Early insight prevents expensive redesigns and enables explainability to the outside world.
- **Set the cadence**: monthly monitoring and a quarterly challenge with independent counter-force.
Without cadence attention drifts to incidents; with cadence learning capacity grows.
Independent dissent prevents groupthink and normalises course correction.

Later (6–12 months)

- **Build the AI management system**: processes, roles, audit trails, maturity measurement.
Quality arises not from intention but from repeatable work practices.

A light, traceable system makes performance scalable without bureaucracy.

- **Vendor due diligence:** licences, datasets, evaluations, red-teaming, content authenticity, exit options.

What you outsource remains your responsibility toward customer and regulator.

Contract audit rights and exit options to limit lock-in and legal exposure.

- **Train leaders and key roles** in judgement under uncertainty and legal interpretation.

Technical knowledge without judgement increases the precision of wrong decisions.

Developing shared language and criteria raises consistency and explainability.

- **Integrate reflection into operations:** case reviews, frontline narratives, improvement decisions with reasoning.

Stories give meaning to data and make effects felt and correctable.

Recording decision rationales creates a reproducible quality culture.

Not (unless exceptionally justified)

- **Irreversible delegation** of high-impact decisions without human counter-force.

Delegation without a brake harms dignity and legal protection when errors occur.

Therefore limit scope and secure contestability and remediation up front.

- **Black box** in contexts with high error costs and low explainability.

Where the price of error is high, traceability is part of safety.

Choose simplicity or hybrid models when explanation is essential for trust.

- **Data collection** without necessity, legal basis and clear retention period.

More data is not automatically better; it increases risks without proportional value.

Formulate purpose limitation, minimise, and ensure timely deletion.

- **KPIs (key performance indicators)** that perfectly measure the unimportant and say nothing about meaning and legitimacy.

Precision without relevance steers toward pseudo-performance and undermines trust.

Tie metrics to value, rights and effects in lived reality.

12. Closing – Leadership as the art of carrying

AI is not an outsider dropping by; it is an **extension of our capacities and our blind spots**. It enlarges our ability and our lack. Leading in this era is the art of **carrying**: carrying tension, carrying plurality, carrying the slowness that quality needs. It asks for imagination **and** discipline, for inner orientation **and** traceable processes.

The invitation is simple and demanding:

Make technology an ally of humanity and the rule of law.

Design counter-force that works when it matters.

Keep the conversation going – especially where it rubs.

If you hold these three lines, AI becomes no black box but a clear mirror. Not machinery that replaces people, but an infrastructure that **amplifies** professional judgement, relational quality and public trust. That is transformation from within: step by step, visibly accountable, with a cadence that can be sustained.

Reflective question to close:

Which decision in your organisation could be better tomorrow by less automation and more explanation – and whom will you invite today to look at it together?

Notes & literature (selection)

Theory & concepts (human, learning, system)

1. Ryan, R. M., & Deci, E. L. (2000). *Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being*. **American Psychologist**, **55**(1), 68–78. https://selfdeterminationtheory.org/SDT/documents/2000_RyanDeci_SDT.pdf
2. Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). *Job Burnout*. **Annual Review of Psychology**, **52**, 397–422. https://dspace.library.uu.nl/bitstream/handle/1874/13606/maslach_01_jobburnout.pdf
3. Kolb, D. A. (1984). **Experiential Learning: Experience as the Source of Learning and Development**. Prentice-Hall. (Overview: https://books.google.com/books/about/Experiential_Learning.html?id=zXruAAAAMAAJ)
4. Argyris, C., & Schön, D. A. (1978). *Organizational Learning: A Theory of Action Perspective*. Addison-Wesley. (Public version: <https://archive.org/details/organizationalle00chri>)
5. Mezirow, J. (1991). *Transformative Dimensions of Adult Learning*. Jossey-Bass. (Summary: <https://eric.ed.gov/?id=ED353469>)
6. Orlikowski, W. J. (2007). *Sociomaterial Practices: Exploring Technology at Work*. ****Organization Studies**, **28**(9), ****1435–1448**. (Open access variant: <https://www.dhi.ac.uk/san/waysofbeing/data/data-crone-orlikowski-2007.pdf>)
7. Stacey, R. D. (2001). **Complex Responsive Processes in Organizations: Learning and Knowledge Creation**. Routledge. (Overview: <https://www.routledge.com/Complex-Responsive-Processes-in-Organizations-Learning-and-Knowledge-Creation/Stacey/p/book/9780415249195>)
8. Ashby, W. R. (1956). *An Introduction to Cybernetics*. Chapman & Hall. (Open access: <https://archive.org/download/AnIntroductionToCybernetics/AnIntroductionToCybernetics.pdf>)

EU framework (reference date 19 January 2026, Europe/Amsterdam)

1. **GDPR** — Regulation (EU) 2016/679. Official publication: EUR-Lex. <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
2. **DSA** — Regulation (EU) 2022/2065 (Digital Services Act). Official publication: EUR-Lex. <https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng>
3. **Data Act** — Regulation (EU) 2023/2854. Official publication: EUR-Lex. <https://eur-lex.europa.eu/eli/reg/2023/2854/oj/eng>
4. **NIS2** — Directive (EU) 2022/2555. Official publication: EUR-Lex. <https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng>
5. **EU AI Act** — Implementation timeline and application dates. European Parliament Research Service (At-a-Glance, 10 June 2025; overview updated July 2024). https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA\2025\772906
– In brief: entered into force 1 August 2024; prohibitions applicable from 2 February 2025; broad application date 2 August 2026 (source: EPRS and multiple legal updates). See also <https://connectontech.bakermckenzie.com/eu-ai-act-published-dates-for-action/> and <https://www.dlapiper.com/en-us/insights/publications/2025/08/latest-wave-of-obligations-under-the-eu-ai-act-take-effect>

Standards & frameworks (management, risk, assurance)

1. **ISO/IEC 42001:2023** — *Information technology — Artificial intelligence — Management system*. ISO. <https://www.iso.org/standard/42001>
2. **ISO/IEC 23894:2023** — *Artificial intelligence — Guidance on risk management*. ISO. <https://www.iso.org/standard/77304.html>
3. **ISO 31000:2018** — *Risk management — Guidelines*. ISO. <https://www.iso.org/standard/65694.html>
4. **NIST AI RMF 1.0** — *Artificial Intelligence Risk Management Framework*. NIST (26 January 2023). <https://www.nist.gov/itl/ai-risk-management-framework> (PDF: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>)
5. **COSO ERM (2017)** — *Enterprise Risk Management—Integrating with Strategy & Performance*. COSO. <https://www.coso.org/enterprise-risk-management>
6. **IIA (2024)** — *Global Internal Audit Standards*. The Institute of Internal Auditors. <https://www.theiia.org/en/standards/2024-standards/global-internal-audit-standards/>

Principles & guidance

1. **OECD (2019, updated 2024)** — *Recommendation of the Council on Artificial Intelligence (AI Principles)*. OECD. <https://oecd.ai/assets/files/OECD-LEGAL-0449-en.pdf> (Overview: <https://oecd.ai/en/ai-principles>)

2. **EU High-Level Expert Group on AI (2019)** – *Ethics Guidelines for Trustworthy AI*. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>