

South Korea's AI Basic Act took effect on January 22, 2026. For product and governance teams, the practical question is not whether the law exists. It is which of your systems fall under the high-impact AI category, what you must document, and who inside your company answers for the output when something goes wrong. This post walks through the changes, the oversight obligations, and a working approach to the threat assessment problem that regulated teams keep running into. Treat this as an operational guide, not legal advice. Confirm specifics against the enacted text and the enforcement decree with qualified counsel in Korea.



What the Act Is and When It Applied

The law is formally the Framework Act on the Development of Artificial Intelligence and Establishment of Trust. It sets a national baseline for AI development and use, with particular attention to systems that affect people's rights, safety, and access to services. The effective date of January 22, 2026 is the point at which operator obligations became enforceable, so any system you were still classifying in late 2025 needed a documented position by then.

Three things changed for most organizations:

- A defined category of high-impact AI now carries specific operator duties.
- Generative AI services must give users notice that they are interacting with or viewing AI-generated output, and synthetic content must be labeled.
- Companies meeting certain thresholds, including many foreign providers, must appoint a domestic representative in Korea.

Enforcement is expected to be phased in with a grace period on fines, as announced by the Ministry of Science and ICT. Do not read a grace period as a grace period on obligations. The duties apply now. Only the penalty timing is softened.

What Counts as High-Impact AI

High-impact AI is defined by the domain where the system is used and the effect it has on people. Based on the statutory framework and the enforcement guidance published around it, the areas of concern include:

- Energy, drinking water, and other critical infrastructure operations
- Healthcare, including medical device and diagnostic support
- Transportation and safety-critical physical systems
- Hiring, promotion, and employee evaluation
- Credit, loan, and insurance decisions
- Education assessment that affects admission or advancement
- Public-sector decisions that affect rights or benefits
- Biometric identification used in criminal investigation contexts

The classification question is functional. Ask what the system's output does in the real workflow, not what the product page says it does. A text summarization tool that quietly ranks job applicants has crossed into high-impact territory, regardless of how the vendor labels it.

High-Impact AI Requirements

For systems classified as high-impact, operators face obligations that cluster into a small number of workstreams. The exact wording matters, so map each item to the enacted text, but the working categories are consistent across the framework:

1. **Risk management.** A documented plan that identifies foreseeable harms, how they are measured, and how they are mitigated.
2. **Explanation of outputs.** The ability to explain, at a meaningful level, the main basis for a decision or recommendation that affects a person.
3. **User protection.** Mechanisms that let affected people understand that AI is involved and, where applicable, seek a review.
4. **Human oversight and management.** Defined human involvement in the process, with authority to override or stop the system.
5. **Documentation and records.** Retained evidence of how the system was built, tested, deployed, and monitored.
6. **Safety measures.** Controls that address accuracy, robustness, and misuse.

Item four [suprmind.ai](#) is where most teams are weakest, and it is the subject of the next section.

Human Oversight Rules: What They Actually Require

"Human in the loop" has become a phrase that covers almost anything. Under a regime like this one, it has to mean something specific and checkable. A workable definition has three parts:

- A named person, by role and by name, owns each high-impact decision path.
- That person has the information, time, and authority to disagree with the system.
- The organization can show, after the fact, that the review happened and what the reviewer decided.

A reviewer who clicks "approve" on 400 cases a day is not providing oversight. They are providing a signature. Regulators and courts will ask whether the review was meaningful, and the answer will come from your logs, your staffing model, and your override data.

Threat Assessment: The False Positive and False Negative Tradeoff

Many high-impact deployments involve some form of risk or threat assessment. Examples include workplace violence screening, fraud flagging, insider risk, and safety triage in public-facing services. These systems make two kinds of error, and they pull in opposite directions.

Error type	What happens	Who pays the cost	Typical control
False positive	A person is flagged who posed no real risk	The flagged individual, through investigation, restriction, or reputational harm	Secondary human review, appeal path, limits on automated action
False negative	A real risk is missed	Potential victims, and the organization in incident and liability terms	Escalation thresholds, second-pass review, monitoring for missed signals

Tightening the threshold to reduce false negatives increases false positives. Loosening it does the reverse. There is no setting that eliminates both. The governance decision is which error you are willing to live with, for which population, and who signed off on that choice. In our reviews, the most common failure is that nobody made the choice explicitly. The threshold was inherited from a vendor default, and the risk owner never saw the tradeoff written down.

Write the tradeoff into the risk management plan. State the chosen threshold, the expected error rates on your own data, the cost of each error type as your organization understands it, and the named owner who accepted those costs.

Multi-LLM Disagreement Is a Signal, Not a Vote

Some teams now run several language models against the same case and compare results. Multi-LLM orchestration tools make this easy, and side-by-side model outputs can be useful for review. The temptation is to treat agreement as proof, or to average scores as though the models were an independent panel of experts. Both habits are mistakes. Models trained on overlapping data, prompted with the same instructions, and sharing similar blind spots can agree confidently and still be wrong. Averaging numerical scores from different models produces a number with no clear meaning, because each model's scale and calibration differ. A three-model "vote" is not a poll of three independent judges. Disagreement is more useful than agreement. When models diverge sharply on a case, that often means the input is missing something: an ambiguous fact, a conflicting record, a context the prompt did not supply. Treat the divergence as a prompt to gather information or escalate to a human, and record it as such. In practice, we flag any case where model outputs fall into different decision categories and route it to manual review with the side-by-side outputs attached. Do not average outputs. Do not present the majority position as the answer. Present the disagreement, the reason it likely exists, and the information needed to resolve it.



Defensible Documentation and Named Accountability

If a regulator or a court asks why a high-impact decision was made, "the AI said so" is not an answer. It is an admission that no one owned the decision. Defensible documentation answers four questions for every consequential case:

- What information did the system have?
- What did the system output, and at what confidence or with what disagreement?
- Who reviewed it, and what did that person decide and why?
- What was the decision and how could the affected person challenge it?

Here is a concrete example. Suppose a hiring team uses an assessment tool to screen applicants. The tool recommends rejecting a candidate. Under a defensible process, the recommendation goes to Dae-ho Kim, Head of Talent Operations, who is named in the decision policy as the accountable owner for that screening step. Kim reviews the summary, the input data, and any model disagreement. If he overrides the recommendation, that is logged with a reason. If he upholds it, that is logged too. The candidate has a documented path to request review. The name, role, and reasoning are all retrievable. (The names here are illustrative.) Accountability has to be assigned to a person, not a team inbox. Teams fail audits when the answer to "who decided?" is "the review committee" with no individual attached.

What We Still Do Not Know

Every case note we write keeps a running list of open questions. For this law, the list currently includes the following:

- How enforcement inspectors will weigh the difference between documented process and actual reviewer independence.
- How broadly the high-impact categories will be interpreted at the edges, such as internal productivity tools that influence personnel decisions.
- Exact timing of fine enforcement after the grace period, and whether it will vary by sector.
- How the domestic representative requirement applies to specific corporate structures.

- How explanation standards for complex models will be assessed in practice.

Revisit this list on a fixed schedule, not only when something breaks. A 30-day review cadence is enough for most teams in the first year.

A Short Working Checklist

1. Inventory every AI system and flag any that touch the high-impact domains above.
2. For each flagged system, name the accountable human owner by role and person.
3. Document the false positive and false negative tradeoff and the threshold chosen.
4. Define what a meaningful override looks like and log every override with a reason.
5. Route multi-model disagreements to human review. Do not average or vote.
6. Confirm user notice and review paths for affected individuals.
7. Confirm whether generative outputs need labeling and implement it.
8. Assess whether a domestic representative obligation applies to your entity.
9. Set a 30-day follow-up on the open questions list.

Where Vendors and Consultants Fit

Teams evaluating tools for this work, including platforms such as Suprmind.ai, should ask vendors to show their logging, override tracking, and side-by-side output features against the requirements above. Suprmind.ai offers a Start Free Trial with no credit card needed. The source material we reviewed does not list pricing, so confirm plan costs directly with the vendor before budgeting. Consulting firms such as Four Dots can help with classification and documentation work, but the accountability decision cannot be outsourced. Ask any vendor or adviser, including the team led by Radomir Basta, CEO & Founder of Suprmind.ai, which named person in your organization will sign for each high-impact decision path.

The Act's core demand is simple to state and hard to meet: a person with authority, information, and a record of their reasoning stands behind every consequential AI decision. Build the process around that person, and the rest of the compliance work becomes far more manageable.