

Beyond the black box: the new security governance

July 12, 2026

AI is not a passing fad but a structural force transforming work processes and creating new risk profiles. For this reason, AI governance cannot remain confined to IT departments, nor can it be treated as a bureaucratic add-on to the Risk Assessment. What is needed is an integrated governance model that bridges the traditional silos between security, privacy, and fundamental rights.

In the second article in our series, Does AI help us work better?, our conversation with Federico Cabitza, Director of the FBK Research Center for Digital Health and Wellbeing, explores the new competencies required of professionals responsible for workplace safety and prevention.

Designing the integration of AI into the workplace means, first and foremost, ensuring that its use is compatible with workers' dignity, professional autonomy, and responsibility.

This is not merely a technical challenge. The real innovation lies in breaking down bureaucratic "silos" by integrating occupational safety (RAD), privacy (DPIA), and the protection of fundamental rights (FRIA) into a single governance strategy.

AI is a structural force that reshapes roles and competencies. As such, it requires the Prevention and Protection Service (SPP) to evolve into a strategic function capable of interpreting not only data, but the entire socio-technical context.

In this second article, we explore the skills needed to translate the opacity of algorithms into understandable and assessable risk factors.

GS: You argue that the Prevention and Protection Service (SPP) should become a strategic player in AI governance, moving beyond technical "silos" to work closely with IT, HR, and legal departments. What are the priority non-technical skills that workplace safety professionals should acquire today to translate the opacity of an algorithmic "black box" into understandable and assessable risks within the Risk Assessment Document (RAD)?

FC: "I believe the starting point is recognizing that the black box is not merely a technical problem. It is also an organizational, cognitive, and decision-making challenge. Of course, technical skills are

necessary—or at least a basic level of AI literacy. People need to understand what predictive models and generative systems are, and be familiar with concepts such as error, uncertainty, bias, model drift, and validation. But for the Prevention and Protection Service, that alone is not enough.

The first non-technical competency is the ability to conduct a socio-technical analysis of work. An AI system should never be evaluated as an isolated object, but as part of a broader work process: who uses it, when, with what degree of autonomy, under what time pressures, with what opportunities to challenge its recommendations, and with what effects on responsibilities, competencies, and working relationships. In many cases, the risk lies not in the model itself, but in how it is embedded within the organization.

The second competency involves understanding human factors. Safety professionals must be able to recognize phenomena such as automation bias, overreliance, deskilling, loss of situational awareness, cognitive fatigue, stress caused by continuous monitoring, and dependence on automated recommendations. These risks may be less visible than physical hazards, but they are no less significant. If workers become accustomed to no longer making decisions—or feel unable to challenge an algorithmic recommendation—we are already witnessing a transformation of occupational risk.

The third competency is interdisciplinary translation. The SPP does not need to become a data science department, but it must know how to ask the right questions of those who develop, purchase, or deploy AI systems. Questions such as: What data was the system trained on? Under what conditions was it validated? What types of errors are expected? How are updates managed? How can operators challenge the system's recommendations? What happens if the system fails? Are there logs, audit trails, intervention thresholds, and escalation procedures?

The fourth competency is legal and organizational. This does not mean replacing the legal department, but understanding the relationship between technological risk, accountability, documentation, traceability, and governance. In the Risk Assessment Document, it is not enough simply to state that 'an AI system has been introduced.' We must explain how that system changes work activities, what risks it introduces or amplifies, what technical, organizational, and training measures have been adopted, and who is responsible for ongoing oversight.

Finally, a methodological competency is essential: the ability to evaluate evidence. A technology is not safe simply because a vendor presents it as innovative, nor because it performs well in a demonstration. It must be assessed in its real context of use. This is one of the central principles of the DHWB Center's mission: helping companies, institutions, policymakers, and decision-makers understand whether, how, and under what conditions a digital technology or AI system can generate value for health—clinical, organizational, and social—in a safe, sustainable, and responsible way."

Silent warning signs

Modern governance must be able to mitigate emerging risks such as techno-complexity and the stress created by opaque decision-making processes. Governing AI also means recognizing the subtle, often invisible changes it can produce in workers' well-being. In the concluding article of this series, we will explore how to identify the early warning signs of mental distress associated with working alongside "intelligent" machines.

<https://magazine.fbk.eu/en/news/beyond-the-black-box-the-new-security-governance/>

TAGS

- #digitalhealthwellbeing
- #prevention
- #security

AUTHORS

- Giancarlo Sciascia