

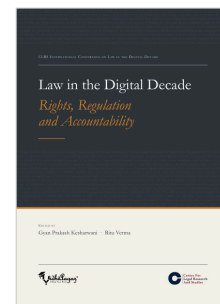
CHAPTER 14

Two Black Boxes: Reassessing Judicial Deference in India's Emerging AI-Driven Security Apparatus

Mahek Motwani

in *Law in the Digital Decade: Rights, Regulation and Accountability*, edited by Gyan Prakash Kesharwani and Ritu Verma (VidhiAagaz 2026), pp. 173–179

DOI <https://doi.org/10.63108/VAB.LDD.1.14>



CITATIONS

BLUEBOOK (21ST ED.)

Mahek Motwani, *Two Black Boxes: Reassessing Judicial Deference in India's Emerging AI-Driven Security Apparatus*, in *LAW IN THE DIGITAL DECADE: RIGHTS, REGULATION AND ACCOUNTABILITY* 173 (Gyan Prakash Kesharwani & Ritu Verma eds., VidhiAagaz 2026).

OSCOLA (4TH ED.)

Mahek Motwani, "Two Black Boxes: Reassessing Judicial Deference in India's Emerging AI-Driven Security Apparatus" in Gyan Prakash Kesharwani and Ritu Verma (eds), *Law in the Digital Decade: Rights, Regulation and Accountability* (VidhiAagaz 2026) 173.

CHICAGO (17TH ED.)

Motwani, Mahek. "Two Black Boxes: Reassessing Judicial Deference in India's Emerging AI-Driven Security Apparatus." In *Law in the Digital Decade: Rights, Regulation and Accountability*, edited by Gyan Prakash Kesharwani and Ritu Verma, 173–179. Maihar: VidhiAagaz, 2026. <https://doi.org/10.63108/VAB.LDD.1.14>.

APA (7TH ED.)

Motwani, M. (2026). Two black boxes: Reassessing judicial deference in India's emerging AI-driven security apparatus. In G. P. Kesharwani & R. Verma (Eds.), *Law in the digital decade: Rights, regulation and accountability* (pp. 173–179). VidhiAagaz. <https://doi.org/10.63108/VAB.LDD.1.14>

SOURCE

<https://doi.org/10.63108/VAB.LDD.1.14>

LICENCE

Open access under the Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0): it may be copied, shared and adapted for non-commercial purposes, with credit to the authors and the source.

Citations are given as a guide; check them against the style manual you follow. This page was added by VidhiAagaz to the downloaded file. The chapter follows, with the page numbers of the printed book.

CHAPTER 14

Two Black Boxes: Reassessing Judicial Deference in India's Emerging AI-Driven Security Apparatus

MAHEK MOTWANI*

The rise of artificial intelligence over the last decade has profoundly transformed the character of national security. Threats once confined to land, sea, air, and space are now identified and acted upon across data networks, satellite systems, and financial infrastructures, increasingly by AI systems rather than human analysts. Constitutional democracies, including India, have long relied on judicial deference doctrines to reconcile national security secrecy with the rule of law. Mechanisms such as state secrets privilege and in-camera review assume that classified decisions, though hidden from public view, rest on human reasoning a court could understand if given access. This paper argues that AI-driven security decisions, from predictive threat scoring to automated watchlist generation, break this assumption by introducing a second, distinct layer of opacity. Classification conceals information that remains knowable to a cleared observer; AI conceals reasoning that may be inscrutable even to its own operators, a problem sharpened where systems are foreign-built or privately developed and thus doubly opaque to the state itself. Courts equipped only for the first opacity are unequipped for the second, leaving algorithmic determinations effectively unreviewable regardless of clearance.

Using India as its primary case study, this study contends that this conflation strains rights to privacy, expression, and equal protection under Articles 21, 19, and 14 of the Indian Constitution. Through a comparative analysis of the Anglo-American doctrine of state secrets privilege and the EU AI Act, it argues that traditional tests of legality, necessity, and proportionality cannot be applied when algorithmic logic resists reconstruction. The paper proposes a bifurcated standard of review separating classified-content review from algorithmic-methodology review, allowing courts to evaluate system interpretability independently of classified inputs. It concludes by addressing executive and technical objections to this reform.

Keywords: Judicial Deference, Algorithmic Opacity, National Security Law, Indian Constitutional Jurisprudence, Proportionality Test

1 INTRODUCTION

Artificial intelligence has increasingly been incorporated in decisions once reserved for human judgment within the machinery of national security. It has entered the space of threat scoring, watch-listing, and surveillance in ways that touch upon and in some cases directly concern constitutional law. This shift has occurred with little corresponding change in the legal architecture meant to hold such decisions accountable. Courts across constitutional democracies, including India, continue to rely on judicial deference doctrines built for a different era: one in which national security decisions, however secretive, were understood to rest on human reasoning that a court could, in principle, examine if granted appropriate access.

This paper contends that this situation represents a diminution of rights under the Indian Constitution. Mechanisms such as state secrets privilege and in-camera review were designed to manage the

* Student at SNDT Women's University, Juhu, Mumbai, Maharashtra, India.

concealment of information, a problem of access, not comprehension. Artificial intelligence introduces a second, structurally distinct problem: the *concealment of reasoning*. Where classified information remains knowable to a sufficiently cleared observer, the internal logic of many AI systems, particularly those built on deep learning architectures, may be unreconstructable even to the engineers and agencies that deploy them. Courts equipped only to negotiate access to secrets are institutionally unprepared to confront a system that offers no comprehensible secret to access in the first place. This problem deepens further where such systems are foreign-built or privately developed, denying even the state itself full insight into their design and function.

The consequence is a quiet but significant erosion of constitutional accountability. Decisions that materially affect due process under Article 21, that raise concerns of arbitrary state action under Article 14, and that touch upon the boundaries of executive power, may now be functionally unreviewable, not because courts have chosen to defer, but because deference has been rendered meaningless by the absence of anything left to defer to.

This paper proceeds in six parts. Part 2 describes the traditional architecture of judicial deference in India in matters of national security. Part 3 develops the paper's central theoretical contribution: a distinction between two forms of opacity, classification and algorithmic inscrutability, that current doctrine conflates. Part 4 grounds this distinction in concrete case studies drawn from India's evolving security apparatus. Part 5 examines the constitutional stakes this conflation raises. Part 6 proposes a reform: a bifurcated standard of judicial review that separates classified-content review from algorithmic-methodology review. Part 7 addresses potential objections.

2 THE TRADITIONAL ARCHITECTURE OF NATIONAL SECURITY DEFERENCE

Judicial deference in national security matters has never meant the absence of a legal framework; it has meant the presence of a particular kind of framework, one built to manage secrecy rather than to dissolve it. In India, this framework developed principally around the interception and surveillance powers vested in the executive under Section 5(2) of the Indian Telegraph Act, 1885,¹ which permits interception of communications in the interest of public safety or during a public emergency. In *People's Union for Civil Liberties v. Union of India* (1997),² the Supreme Court, while declining to strike down the provision, read into it a set of procedural safeguards: interception orders were to be reviewed by a committee of senior bureaucrats rather than by courts, and communicated only in summary form. The judgment did not open the executive's reasoning to judicial scrutiny; it substituted an internal, executive-led review for an external, judicial one, on the understanding that the underlying decision, however secret, remained a reasoned act attributable to an identifiable official who could, in principle, account for it.

The same structural logic appears, in a different form, in the Anglo-American doctrine of state secrets privilege. In *United States v. Reynolds* (1953),³ the U.S. Supreme Court permitted the executive to withhold evidence from a civil proceeding on the ground that disclosure would endanger national security, requiring only that a responsible official formally claim the privilege and that a court satisfy itself, without necessarily examining the evidence in detail, that the claim was not implausible on its face. Subsequent

¹ Indian Telegraph Act, No. 13 of 1885, § 5(2) (India)

² *People's Union for Civil Liberties v. Union of India*, (1997) 1 SCC 301

³ *United States v. Reynolds*, 345 U.S. 1, 10–11 (1953)

practice, including in-camera review and the use of security-cleared special advocates in the United Kingdom's closed material proceedings, elaborated this basic model without displacing it: a court forgoes independent assessment of classified material, trusting that a human decision-maker examined it and could defend the conclusion reached, if pressed.

What unites these mechanisms, across jurisdictions, is an assumption so basic that it is rarely stated: secrecy conceals information, not reasoning. A classified file might describe a source, a surveillance intercept, or a threat assessment, but the inferential steps connecting that information to a decision, to intercept a call, to detain a suspect, to deny an entry, were until recently drawn by a human analyst using methods a lawyer or judge could reconstruct in principle, even if barred from doing so in practice. Judicial deference, on this view, is not abdication. It is a calculated wager that behind the veil of classification lies an accountable, comprehensible act of judgment. That wager is the doctrinal inheritance now being tested by artificial intelligence.

3 TWO BLACK BOXES: DISTINGUISHING CLASSIFICATION FROM ALGORITHMIC OPACITY

AI in national-security systems introduces a problem that is distinct from the problem of secrecy that has classically been associated with classification. Much of classification, fundamentally, creates a problem of access: the State has information but decides that it cannot be examined by certain parties. With appropriate legal procedure, the information can be accessed and is in principle intelligible and worthy of assessment. In-camera proceedings, confidential submissions and other methods of restricted disclosure are contingent upon this idea.

However, certain AI and machine-learning systems generate outputs by the complex interplay among their models, training data and large numbers of variables. In these systems, understanding the information provided to the system is not sufficient to understand how particular outputs were generated. The issue is not limited to the court's ability to access the information but whether the court can meaningfully assess the process by which the information was fashioned into a recommendation, risk score, threat analysis, or other security assessment. This distinction is important because access and explanation are not necessarily the same thing. A court may, for example, be given access to the relevant data, the model or even technical documentation without being able to determine whether a particular output was sufficiently reliable to justify state action. Whether the output of a system that is used to identify persons as potential security risks is meaningful is a question of both method and accuracy, not whether the system is available to the court.

The quality of the underlying data is therefore an important part of the problem. An algorithm can process information efficiently without that information necessarily being accurate, complete, current or free from systematic bias. If unreliable or inappropriate data contributes to a security assessment, reviewing only the final output may not be sufficient. The court may need to examine the relationship between the data used, the methodology applied and the resulting assessment.

This forms what this paper calls the second black box. In the first black box, information is hidden from access because it is classified. In the second, the form and content of that which is available for analysis by an AI system and its subsequent output which can affect state actions is hidden. These problems can overlap, but are not the same. While a court may be able to regulate access to classified material, it may still struggle to assess the reasoning and reliability of an AI-assisted analysis and subsequent action.

This problem is compounded where the AI system is designed or supplied by a private or foreign entity and the government has limited access to a system's design, training or running. In this scenario, where neither the affected party nor reviewing court can meaningfully assess the confidence or reasoning of a security assessment assisted by AI, what can be used for the purposes of constitutional scrutiny?

4 NATGRID AND THE EMERGING SECURITY ARCHITECTURE

NATGRID provides a useful illustration of the technological changes taking place within India's national-security architecture. It is designed to connect multiple government databases, including immigration, banking, telecommunications and travel records, with authorised security and law-enforcement agencies. According to the Ministry of Home Affairs, NATGRID enables access to information from these different sources to support the identification of suspicious individuals, investigation of terror networks and intelligence-led investigations. As of March 2026, it connects 11 Central User Agencies, police forces across all States and Union Territories, and 11 data-providing organisations.⁴

The system is also developing more advanced analytical capabilities. Government material identifies GANDIVA as an advanced NATGRID analytics tool that enables multi-source data collection and intelligence analysis for counter-terrorism and criminal investigations.⁵ The wider Multi-Agency Centre and Subsidiary Multi-Agency Centre network has also been upgraded with AI- and machine-learning-enabled software for real-time analysis of intelligence inputs. These developments show that the use of computational and AI-assisted analysis in India's security architecture is no longer purely hypothetical.

At the same time, the public information available about these systems does not establish that an AI system independently makes final national-security decisions. That distinction is important. The argument of this paper does not depend on treating AI as an autonomous decision-maker. It is enough that AI-assisted systems increasingly participate in the collection, integration and analysis of information on which security assessments and subsequent state action may depend.

This creates the problem identified in the preceding section. When multiple sources of information are integrated and analysed through advanced computational systems, a court reviewing a resulting state action may have to consider more than whether the underlying information is classified. It may also have to ask whether the data used was sufficiently reliable, whether the analytical process was appropriate for the purpose for which it was used, and whether the resulting assessment can be meaningfully examined. The more such systems become embedded in security decision-making, the more important these questions become for constitutional review.

NATGRID therefore illustrates an emerging legal problem rather than an established example of unlawful or autonomous AI decision-making. The question is not whether the system should be presumed defective, but whether existing doctrines of national-security deference are capable of responding if an AI-assisted assessment is later challenged on constitutional grounds.

⁴ Press Information Bureau, Government of India, *National Intelligence Grid (NATGRID)* (2026)

⁵ Press Information Bureau, Government of India, *National Intelligence Grid (NATGRID)* (2026)

5 CONSTITUTIONAL STAKES

The conflation of these two forms of opacity is not merely a procedural inconvenience; it implicates the core commitments of Indian constitutional law. Three are especially salient.

First, the right to privacy and personal liberty under Article 21.⁶ Puttaswamy held privacy to be an intrinsic facet of Article 21 and required that any state incursion upon it satisfy a proportionality standard, encompassing legality, a legitimate state aim, necessity, and a balancing of the measure's impact against its purpose. A proportionality inquiry of this kind presupposes that a court can identify what the measure actually does and why, in order to weigh it against the individual's interest. Where the operative reasoning is algorithmically opaque, courts are asked to conduct a balancing exercise without one side of the scale being visible, testing proportionality against a process rather than a decision.

Second, the guarantee against arbitrary state action under Article 14. Since *E.P. Royappa v. State of Tamil Nadu* (1974),⁷ arbitrariness has been treated as antithetical to equality in the Constitution, and *Maneka Gandhi v. Union of India* (1978)⁸ extended this reasoning to require that procedures affecting personal liberty be fair, just, and reasonable, not merely prescribed by law. An arbitrariness inquiry asks whether a decision follows from an intelligible, non-capricious basis. A determination generated by a system whose own designers cannot fully reconstruct its reasoning sits uneasily within this framework: it may be statistically well-calibrated and constitutionally arbitrary at once, since calibration answers whether a system is accurate on average, not whether any given individual determination is intelligible on its own terms.

Third, the separation of powers. Judicial deference to the executive in security matters has always rested on the premise that courts are ceding review of a human judgment, made by an accountable official who remains, in principle, answerable through other channels, political, administrative, or eventually judicial. Where the operative determination is substantially shaped by a system no official can fully explain, deference risks becoming something else: not a court declining to review executive judgment, but the effective absence of any institution capable of reviewing the judgment at all, since it is not clear the executive itself possesses a reviewable rationale to defer to.

6 PROPOSED REFORM: A BIFURCATED STANDARD OF JUDICIAL REVIEW

If the difficulty is a conflation of two distinct forms of opacity, the remedy this paper proposes is their deliberate separation into two independently reviewable questions.

The first, classified-content review, would proceed largely as it does today: courts, through in-camera examination, cleared special advocates, or comparable mechanisms, assess whether the state's justification for withholding underlying information, sources, raw intelligence, sensitive methods, is genuinely necessitated by security concerns. Nothing in this proposal disturbs that inquiry.

The second, algorithmic-methodology review, would ask a different and presently unasked question: independent of whether the underlying data is classified, can the system's design, training methodology, and general decisional logic be assessed for interpretability, proportionality, and freedom from arbitrary or discriminatory pattern by a qualified, appropriately cleared technical auditor, functioning

⁶ Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1

⁷ E. P. Royappa v. State of Tamil Nadu, (1974) 4 SCC 3

⁸ Maneka Gandhi v. Union of India, (1978) 1 SCC 248, ¶ 48

independently of the deploying agency? Such review would not require disclosing the specific classified inputs that produced a given output; it would require disclosing, in a controlled and if necessary security-cleared setting, how the system in general terms weighs categories of factors, what its known error rates and failure modes are across demographic groups, and whether its confidence thresholds meet a defensible standard before being relied upon for consequential decisions. This is analogous to auditing a decision-making process rather than adjudicating a particular case, a review of the recipe rather than the specific meal.

This separation resolves the doctrinal confusion identified in Part 3 without requiring courts to develop new expertise in machine learning from first principles; it requires them, instead, to route the second-order question to institutions equipped to answer it, whether specialised technical tribunals, court-appointed independent auditors, or a dedicated oversight body with standing technical capacity, much as courts already route valuation disputes to accountants or medical questions to physicians rather than deciding them unaided.

The most immediate objection is that methodology, too, can be sensitive: disclosing how a security system weighs risk factors could itself furnish adversaries with a map for evasion. This concern is real but not, on inspection, a reason to collapse the two forms of review back into one. A methodology audit conducted by a cleared technical body, operating under the same confidentiality obligations that already govern classified-content review, need not become public to be judicially useful; what matters is that someone institutionally positioned to assess algorithmic soundness does so, and that a court can rely on that assessment, much as it already relies on in-camera findings it does not itself read into the public record.

There is emerging, if still developing, comparative support for a review of this kind. The European Union's Artificial Intelligence Act requires that high-risk AI systems be designed so that their operation is sufficiently transparent to enable those overseeing them to interpret the system's output and use it appropriately, and separately requires human oversight mechanisms capable of allowing overseers to correctly interpret a system's output and its limitations.⁹ These obligations remain, in the assessment of much of the surrounding technical literature, more readily achievable for global explanations of a system's general behaviour than for local, case-specific reasoning, precisely the distinction this paper's bifurcated proposal is built to accommodate: global, methodology-level review as a judicially administrable substitute for case-specific algorithmic explanation that many systems cannot presently provide.

7 ANTICIPATED OBJECTIONS AND LIMITATIONS

The institutional competence objection holds that courts, having long deferred to the executive's assessment of security necessity, are still less equipped to assess the soundness of a machine learning system's design. This paper's proposal does not ask courts to acquire this competence directly; it asks them to institutionalise access to those who already possess it, through appointed technical auditors, in the same way courts have long relied on court-appointed experts in patent, medical negligence, or valuation disputes without becoming engineers or physicians themselves.

⁹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), arts. 13(1), 14(4), OJ L, 2024/1689, 12 July 2024

The technical limitation objection is more serious: some architectures may resist meaningful interpretability regardless of audit rigour; no external methodology review can explain a genuinely inscrutable model beyond describing its statistical behaviour. Where this is true, the honest constitutional conclusion may be that such systems are unsuitable for consequential, rights-affecting security determinations until interpretable alternatives exist, a conclusion this paper does not shy from, since a system that cannot be reviewed cannot be reconciled with a constitutional order premised on accountable state action.

Finally, feasibility varies across jurisdictions with different institutional capacities. India's comparatively nascent technical-audit infrastructure means implementation would likely begin narrower than the full model sketched here, perhaps confined initially to the highest-stakes categories of algorithmic security determination, before wider application. This is a reason for phased implementation, not for abandoning the underlying distinction.

8 CONCLUSION

Judicial deference in national security law has always rested on a wager that secrecy conceals reasoning, not the absence of reasoning altogether. Artificial intelligence, deployed at increasing scale within India's security apparatus with only nascent statutory or judicial oversight, threatens to make that wager untenable, not because states have become more secretive, but because some of what they now rely upon may not be explicable to anyone, cleared or not. Treating algorithmic opacity as simply a deeper form of classification risks either abandoning meaningful review altogether or demanding an explanation no system can give. Separating classified-content review from algorithmic-methodology review offers courts a doctrinally coherent path between these failures, preserving legitimate secrecy while insisting that the reasoning beneath a security determination, however produced, remain subject to some independent, competent form of scrutiny. As India's security institutions continue to adopt these tools faster than doctrine has adapted to them, this distinction may prove necessary not only to constitutional principle, but to the basic proposition that the exercise of state power, however sophisticated its instruments, remains an exercise for which someone can be asked to account.

* * *