# Comments on the Draft Regulations for Use of Artificial Intelligence in Courts, 2026

*Submitted to the Member Secretary, AI Committee, Supreme Court of India*

**Submitted by:** Bharath Reddy, Associate Fellow, The Takshashila Institution, Bengaluru.
**Date:** 20th June 2026
**Contact:** bharath@takshashila.org.in

## 1. Introduction and General Observations

The draft is a thoughtful and substantially well-constructed instrument. Five targeted amendments would make them enforceable: independent audit under confidentiality constraints (Clause 38); testing for overreliance on AI in human oversight assessments (Clause 35); risk-tiered verification for generative AI (Clauses 3 and 8); auditable bias-detection standards (Clause 6); and private-sector incentive design (Clause 46). These will help reduce the gap between the draft's stated principles and its operational instruments in practice.

The forward-looking, opportunity-oriented framing is commendable. The draft does not treat AI primarily as a threat to be contained. The presumption in favour of responsible adoption (Clause 16) and the explicit 'Innovation over Restraint' principle (Clause 17), frame AI as an instrument for relieving the judiciary's most pressing structural constraint: the case backlog and the state-capacity limits (the judiciary's limited administrative and technical capacity) that sustain it. This is a welcome initiative that recognises that the cost of under-adoption is as real as the cost of careless adoption.

The clear enumeration of permissible and prohibited uses offers a clear road-map. By setting out an illustrative list of permitted applications (Clause 19) alongside a list of prohibitions (Clause 20), the draft gives courts, litigants, and vendors a legible map of what AI is meant to do and what it must never do. Clear prohibitions on algorithmic adjudication, risk scoring, and predictive profiling are precisely the high-stakes applications where automation has caused demonstrable harm in other jurisdictions, and excluding them at the outset protects the legitimacy of the wider adoption effort. This clarity of expectation is valuable and should be preserved through subsequent revisions of the guidelines.

With that foundation acknowledged, we offer five recommendations for improvement.

## 2. Observations and Recommendations

## 1. Audits Should Be Independent as Well as Confidential (Clause 38)

Clause 38(2) requires that all audits be conducted in-house and that architectural information never be shared with any external party for audit purposes. It bars external audit absolutely and requires that source code, algorithms, and datasets never leave court premises for audit by any third party. The motivation of protecting the confidentiality and security of sensitive judicial systems is sound. This conflates two distinct objectives — confidentiality and auditor independence. This may undermine the credibility of the audit rather than strengthen it.

Independent third-party audit is a foundational element of trust. When the body that procures and deploys a system is also the only body permitted to audit it, the audit cannot provide the external assurance on which public trust depends. Security-cleared, NDA-bound external auditors can satisfy both confidentiality and independence requirements. The draft's legitimate confidentiality concern can be met without sacrificing independence.

**Recommendation.** We suggest Clause 38(2) be revised to permit audit by independent, security-cleared third-party auditors operating under binding confidentiality obligations and, where necessary, within the Controlled Environment Testing facility the draft already establishes in Clause 36. This preserves the security of judicial systems while supplying the external assurance that in-house audit alone cannot provide. The AI Register and incident-reporting machinery already in the draft can accommodate independent audit findings without further structural change.

## 2. The Impact Assessment Must Test Human Oversight Against Overreliance on AI (Clause 35)

Clause 35 requires a Technical and Ethical Impact Assessment (TEIA) before any AI System is approved, Clause 35(3)(e) expressly lists 'compliance with Human-in-the-Loop requirements' among the matters the assessment must evaluate, and Clause 8 states that accountability for any decisions made with AI rests exclusively with the officer using the system.

A well-documented failure mode of human + AI systems is overreliance on AI which includes mechanisms such as automation bias, confirmation bias, ordering effects and overestimating explanations[1]. Supervisors of automated systems systematically over-trust them, approve outputs without genuine scrutiny, and lose vigilance precisely as the system becomes more reliable[2]. A judicial officer reviewing hundreds of case summaries is not an effective safeguard. The TEIA needs to go beyond confirming a human is nominally present, and test whether that human can realistically exercise independent judgment.

---

[1] Samir Passi & Mihaela Vorvoreanu. 2022. Overreliance on AI: Literature Review. Microsoft Technical Report MSR-TR-2022-12. Microsoft Corporation. https://www.microsoft.com/en-us/research/publication/overreliance-on-ai-literature-review/.

[2] ibid.

**Recommendation.** We suggest Clause 35(3) be expanded so that the assessment of Human-in-the-Loop compliance under sub-clause (e) tests the quality of oversight, not merely its presence. Specifically, the TEIA should require that each system:

- surfaces its own confidence or uncertainty alongside every output, so the reviewing officer can calibrate scrutiny rather than treat all outputs as equally reliable;
- is designed to require active engagement by the reviewer rather than passive approval, with stronger engagement requirements for higher-risk uses;
- logs override and acceptance rates, to be reviewed in the periodic audit under Clause 38. An officer who never overrides an AI recommendation is a signal of rubber-stamping to be investigated; and
- is supported by oversight training that explicitly addresses automation bias.

## 3. The Treatment of Generative AI Should Be Risk-Calibrated, Not Absolute (Clause 3 and 8)

The draft defines Generative AI separately (Clause 3(y)) and attaches heightened disclosure and verification obligations to it, reflecting legitimate concern about hallucinations(Clause 3(z)). While the concern is valid, the draft does not fully reconcile this caution with its own list of permitted uses.

Several of the permissible applications in Clause 19 — translation of judgments and pleadings (19(c)), transcription of proceedings (19(b)), and legal research and document summarisation (19(d)) — are performed by generative AI systems. There is no meaningful non-generative method of producing fluent translation or summarisation at scale. The draft therefore relies on generative AI to deliver many of its efficiency use cases while simultaneously imposing heavy constraints on it. This results in a tension that will create uncertainty for courts trying to apply the regulation.

The problem is that the verification burden and the generative-AI disclosure machinery in Clause 8(3) and Clause 3(y) are applied uniformly, without calibration to the risk of the underlying task. Verifying an AI-generated case citation in a legal-research memo is essential, because a hallucinated precedent can corrupt a judicial outcome. Verifying every line of an AI-generated translation or transcript of a routine notice imposes a prohibitive cost.

**Recommendation.** We suggest that the regulation tier its verification and disclosure requirements according to the risk profile of the task, consistent with the proportionality principle the draft already endorses in Clause 12. Generative outputs that bear on substantive rights, evidence, or legal reasoning should attract the full verification standard; generative outputs confined to low-stakes administrative communication should attract a lighter, class-based verification regime of the kind Clause 8(3) already contemplates for certified administrative tools. We further suggest a clarifying cross-reference be added to Clause 19 itself, acknowledging that several of the listed

permissible uses are generative in nature and are governed by the tiered verification standard, so that the tension is resolved on the face of the permitted-use list rather than left to interpretation.

## 4. The Non-Discrimination Standard Needs a Test for Residual Bias (Clause 35)

Clause 6 requires that AI systems be fair and non-discriminatory. Clause 14 adds a qualifier that AI is trained on data 'to the extent feasible, free from discriminatory bias.'

However, the requirement in Clause 6 does not acknowledge that bias is not a bug that careful design can eliminate; it is an inherent feature of systems trained on historical data, and judicial data in particular encodes the historical patterns of the society that produced it. The draft's 'to the extent feasible' qualifier is welcome, but needs a measurable test to make it operational.

What is needed is not a promise of neutrality but a disciplined, documented process for detecting residual bias and making a conscious, recorded judgment about whether it is tolerable for the specific use to which the system is put.

**Recommendation.** We suggest the regulation require, as part of the Technical and Ethical Impact Assessment under Clause 35:

- an explicit statement, recorded in the assessment, that the deploying authority recognises residual bias to be inherent and has actively tested for it;
- mandatory disaggregated performance testing of each system across the categories listed in Clause 6(2) — caste, religion, sex, gender, disability, language, and economic status — with results recorded in the AI Register;
- a use-case-tiered standard for acceptable residual disparity, with the most stringent thresholds applied to any application bearing on personal liberty or substantive rights; and
- identification of the officer or body responsible for certifying that the residual bias is acceptable for the approved purpose.

This converts bias into an auditable obligation, and does so using the institutional machinery the draft has already created.

## 5. Private-Sector Participation Should Be Enabled Under Constraint, Not Deterred by Cumulative Restrictions (Clause 46)

Chapter VI (Clause 46) expressly contemplates the engagement of private entities under prior approval, with a detailed schedule of mandatory contractual safeguards. The safeguards themselves — data-use limits, audit rights, incident reporting, indemnity, sovereign deployment for sensitive data — are appropriate for infrastructure as sensitive as judicial systems, and we support their inclusion.

Our concern is regarding incentive design. The cumulative effect of three provisions may be to deter capable vendors from participating at all, which would leave the judiciary

dependent on in-house capacity it does not currently possess, reintroducing the very state-capacity constraint that AI adoption is meant to relieve. The three provisions are:

- **Clause 46(9)**, which requires that the court retain ownership of, or a perpetual royalty-free licence to, any tool developed using court data, and bars vendors from claiming exclusive intellectual property. For a vendor whose commercial value lies in its model, a blanket no-exclusive-IP rule may make participation commercially unviable.
- **Clause 46(4)(k)**, which prohibits any retraining or fine-tuning (adapting a general AI model to a specific domain using specialised data) on court data without express written approval. This might discourage the iterative improvement on which model quality depends.

Taken together, they may select for vendors who are willing to accept terms that capable AI vendors will decline, leaving the judiciary with a thinner and less capable field of bidders precisely where capability matters most. The objection that judicial data is public infrastructure, and therefore that any IP derived from it should vest in the public, is legitimate. The question is not ownership in principle but incentive design in practice. A blanket no-exclusive-IP rule may achieve the public-ownership objective while simultaneously excluding the vendors whose models would best serve the public interest. A calibrated allocation (public ownership of tools built on court data, licensed access to vendor-owned foundation models) serves both goals.

**Recommendation.** We suggest the draft be revised to enable robust private participation under proportionate constraint rather than under cumulative restriction. Specifically:

- consider replacing the blanket no-exclusive-IP rule with a more calibrated allocation that distinguishes tools built primarily on public judicial data (where public ownership is justified) from vendor-owned large language models (large general-purpose AI systems) merely adapted for court use (where a licence to use, rather than ownership of the model, might offer better incentives to builders); and
- provide a defined, time-bound approval pathway for supervised fine-tuning on anonymised court data, so that quality improvement is possible within guardrails.

The objective is to make participation attractive to capable vendors while preserving the court's control over its data and its outcomes.

## 3. Conclusion

The draft Regulations are a serious and commendable effort to bring AI into the judicial system in a manner that is both ambitious about opportunity and clear about risk. Its optimism about relieving state-capacity constraints, and its clear enumeration of permitted and prohibited uses, are genuine strengths that should be retained.

The recommendations discussed here help build an operational mechanism around the stated principles to ensure they are achievable or enforceable: independent audit under confidentiality constraints (Clause 38); testing for overreliance on AI in human oversight assessments (Clause 35); risk-tiered verification for generative AI (Clauses 3 and 8); auditable bias-detection standards (Clause 6); and private-sector incentive design(Clause 46). They aim to close the gap between the draft's aspirations and its instruments. We, at the Takshashila Institution are available to discuss any of these recommendations with the Committee in further detail.

*Disclosure: AI was used in the analysis and preparation of this response.*