IIGF
INDIA INTERNET GOVERNANCE FORUM

IGAP INDIAN GOVERNANCE AND POLICY PROJECT

EVENT REPORT
November 2025

From Trust to Risk:

THE CASE FOR PRIVILEGE IN AI CONVERSATIONS

Index

01 Background

The Indian Governance & Policy Project ***(IGAP)*** organized a panel discussion titled *"From Trust to Risk: The Case for Privilege in AI Conversations"* on Day 2 of the India Internet Governance Forum 2025 (November 28, 2025) at India International Centre, Max Mueller Marg, New Delhi.

This first-of-its-kind session, aimed to map the current legal and policy landscape governing AI-mediated communications, highlighted the gaps created by rapid adoption of conversational AI tools, and explored whether new protections or frameworks may be required as these systems become increasingly embedded in personal, professional, and sensitive contexts. More broadly, the event sought to generate early insights which could guide future research, regulatory thinking, and ethical design choices to supplement the adoption of AI systems.

Moderated by Dhruv Garg, Founding Partner at IGAP, the discussion encouraged robust dialogue among the panellists on critical issues. The panel featured a distinguished group of experts who brought diverse perspectives to the table:

- **Professor (Dr.) Anup Surendranath**
  Professor of Law and Executive Director,
  The Square Circle Clinic, NALSAR University of Law
- **Kapil Chaudhary**
  Partner, Dentons Link Legal
- **Kirti Mahapatra**
  Partner, Shardul Amarchand Mangaldas & Co.

Disclaimer

The views and opinions expressed by panelists during this event were shared in their personal capacity and do not reflect or represent the official positions, policies, or views of the organizations, institutions, or entities with which they are affiliated. All statements and perspectives were offered within the specific context of the event's themes.

02 Framing the Issue

Privilege, as understood in common-law, has traditionally been described as an evidentiary rule. However, the Supreme Court of India in its recent judgment underscored that it encompasses a broader substantive-right dimension.[1] At its core, privilege ensures that when a person approaches a legal advisor, they can freely disclose information without fear of unauthorized disclosure. Under the Bharatiya Sakshya Adhiniyam, now Sections 122–124 (previously 126–129 of the Indian Evidence Act), this protection operates dually:

- a lawyer cannot disclose communication without the client's consent (subject to limited exceptions), and
- a client cannot be compelled to disclose confidential communications with their lawyer.

As millions of people now turn to AI tools for sensitive personal matters, including legal questions, mental health concerns, medical information, and intimate disclosures, these conversations currently carry no confidentiality protections comparable to those shielding human professional communications from compelled disclosure.

The urgency of the issue was highlighted by recent events, such as the decision of a New York federal judge to order OpenAI to produce millions of chat logs from ChatGPT users for the ongoing high-profile copyright litigation with the New York Times.[2] Recently, OpenAI CEO Sam Altman publicly suggested that such interactions should be protected like communications with human professionals, he also acknowledged that no jurisdiction currently recognises any form of "AI privilege."[3]

This brings into focus a foundational question at the intersection of law and technology:

> *Should communications between users and AI systems receive any form of legal privilege akin to attorney–client or doctor–patient privilege?*

Our expert panel examined whether such privilege could or should extend to AI communications. The discussions explored whether an "AI privilege" might serve legitimate aims, such as broadening access to justice, protecting vulnerable users, and enabling candid disclosure, while also identifying the risks that the recognition, or absence, of such a right could introduce.

03 Issues Discussed

Deliberations among the panellists generated critical insights on whether AI systems, particularly chatbots, should be accorded any form of privilege, and if not, what alternative frameworks might protect sensitive user communications.

1. Foundational Principles: Decisional Autonomy and the Justification for Privilege

The panel examined the normative foundations of privilege, emphasizing that its justification lies not merely in promoting candid communication but in safeguarding 'decisional autonomy'. This refers to the ability of an individual to make meaningful choices about their life, and is grounded in a Kantian conception of human dignity. Dr. Surendranath explained that India's existing jurisprudence and regulatory framework was intentionally conservative in granting 'privilege' as this right carves out 'zones of exclusion' where otherwise relevant evidence was inadmissible. He reflected that this represents a careful balance between truth-seeking and other competing public-interest principles essential to ensuring the sanctity of legal processes. This is also reflected in other exclusionary rules of evidence, such as those on confessions and hearsay. Within this framework, the panel questioned whether AI tools can meet the normative conditions traditionally required for establishing privileged relationships. Certain domains of human functioning required protection for autonomy to exist at all, and indicated a conceptual link between privilege, privacy, and autonomy.

> *"For me, the core justification for privileged communication is decisional autonomy. Certain zones of human functioning require protection for autonomy to exist at all; this comes from a very Kantian understanding of human dignity."*
>
> **Dr. Anup Surendranath**

2. Nature of Interaction: Fiduciary Relationships and the Commercial Functionality of AI Systems

The panel examined whether a doctrine historically rooted in 'fiduciary relationships' (i.e. where one party must act primarily in the interest of the other) could extend to AI systems that possess no distinct legal personality, ethical duties, or obligations of loyalty. Privilege traditionally arises in relationships where the recipient is bound to act solely in the communicator's interests, with confidentiality duties that persist despite strong public-interest pressures. Under existing legal frameworks, some members of the panel felt that AI systems lacked the capacity to bear fiduciary obligations. It was noted that a number of popular AI systems and tools operated through commercial infrastructure and were shaped by other incentives in addition to the user's immediate interest. Members of the panel also noted a deep information asymmetry, where platforms deploying AI invariably knew far more about end-users than users knew about the systems handling their data. Within this conceptual framework, the establishment of a fiduciary relationship between AI systems/tools and end-users was perceived to be difficult.

This also prompted a broader conceptual inquiry regarding the end goal of AI privilege. Panellists noted that clarity was needed on whether the goal was to protect the sensitivity of the information, or the nature of the relationship through which it is disclosed. Extending privilege to AI would require, at minimum, imposing enforceable duties, professional ethics standards, and 'licensing-style' accountability on the entities operating AI systems. This would reflect the necessary regulatory commitments that accompanied privileged professional relationships.

> *"Privilege rests on the existence of a fiduciary relationship between two individuals. So when we ask whether privilege protects a relationship or a function, my instinct, as a regulatory lawyer, is that functionality cannot have privilege."*
>
> **Kirti Mahapatra**

> *"AI does not have a legal personality. We are stepping into a space where the entity in question is not a legal person and does not owe fiduciary obligations."*
>
> **Kapil Chaudhary**

> *"A fiduciary relationship represents vulnerability. The professional must act with absolute loyalty. How do we ensure that the only thing driving the interaction between the user and the chatbot is the user's interest?"*
>
> **Dr. Anup Surendranath**

3. Reframing the Core Issue: Privacy versus Privilege

Members of the panel also stressed that while privacy and privilege were related, they served distinct purposes and operated at different stages of a legal process. It was suggested that the push for 'AI privilege' reflected broader discomfort with how digital systems handled personal information, rather than a principled extension of 'privilege doctrine'. The panel emphasized that the focus should remain on individuals' expectations and autonomy when engaging with AI systems, rather than framing the issue as a contest between the State and digital platforms. Even when data was lawfully obtained, evidentiary rules determined its admissibility. This underscored the observation that the principles of privacy governed collection, processing and disclosure, while privilege governed admissibility in legal processes. Recognition of this distinction was perceived as crucial to avoiding the collapse of the two concepts.

> *"There is a fundamental distinction between privacy and privilege. Privilege is narrow, and AI platforms do not have fiduciary obligations, yet users assume protections that do not exist."*
>
> **Kapil Chaudhary**

> *"The issue must be grounded in individual autonomy, not framed primarily as a tug-of-war between the State, platforms, and their respective data interests. In the context of confidential medical communication, the privilege question is ultimately one of courtroom admissibility and evidentiary use, which is distinct from the broader information-protection concerns governed by privacy law."*
>
> **Kirti Mahapatra**

4. Technical and Regulatory Concerns: Data Access, Retention and Deletion

The panel discussed the technical and regulatory constraints that complicated the extension of privilege to AI-mediated conversations. Some of these constraints emanated from the Digital Personal Data Protection Act, 2023, which required retention of essential data for at least one year. This created a mandatory period during which even sensitive user-AI exchanges would need to be stored and accessible through lawful processes. During this period, platforms would not be able to delete data upon a user request, and any privilege, if recognized, would only be applicable once State or court access to it was sought. Further, it was noted that AI systems routinely stored, processed, and repurposed user inputs for safety, quality improvement, integrations, and backend operations. This raised fundamental questions about what user data 'deletion' meant in practice.

The panel also noted that AI companies necessarily maintain internal access to user conversations due to the architecture of modern systems. Multi-team access for safety, quality review, and model improvement raised a foundational question: *Can confidentiality, a prerequisite for privilege, exist when multiple parties within a corporation have ongoing access to the communications?*

> *"We need to address the black-box problem. AI tools retain access internally for multiple purposes: safety, integration, storage. So if a platform says your data is deleted, is it really deleted?"*
>
> **Kapil Chaudhary**

> *"Entities offering these tools must provide much greater transparency about what happens to the data, where it goes, how they themselves use it, not just what they share with courts and regulators."*
>
> **Kirti Mahapatra**

5. Risks of to Users in the Absence of AI Privilege:

The panel noted that conceptualizations of AI privilege would also vary across its specific use-cases. For instance, factors such as the low ratio of psychologists as compared to the population, and high-rates of self-harm among Indian youth, were argued to contribute towards the uptake in mental-health AI tools. In such situations, users would have an expectation of privacy, and the critical need to protect interactions of vulnerable populations reliant on these platforms was noted. Further, the panel stressed on the need to increase awareness among end-users and enterprise-users about the risks to their data, and the need to identify safeguards in relation to AI-mediated communications, in the absence of an 'AI privilege'.

> *"Until we see judicial pronouncements that clarify how far these protections can go, we will need guardrails to protect the interest of the users: strong confidentiality norms, governance frameworks, and protections for minors and vulnerable populations.Users must start asking platforms explicitly: What are your data retention practices? How can I delete my data?"*
>
> **Kapil Chaudhary**

> *"For users, the answer is not 'Stop using AI tools.' Users must become more aware and demanding about where their data is going and under what conditions it might be disclosed."*
>
> **Kirti Mahapatra**

04 Way Forward

In summary, the panel discussion articulated the structural underpinnings of the right of legal privilege and degree of their application in the context of AI. This underscored that India's approach to AI-mediated communications would need to move beyond the binary question of whether 'AI privilege' akin to legal privilege should exist. The panel indicated a more realistic pathway existed through the development of layered safeguards that protected user autonomy, dignity, and decisional freedom without distorting the narrow legal doctrine of privilege.

It was also articulated that policymakers would need to identify how sensitive interactions between users and AI systems (especially in legal and mental-health contexts) would be treated when sought by State authorities. Finally, the panel acknowledged that end-users required additional practical guidance to understand the risks inherent in AI conversations. Awareness was also needed on the options available to users under India's emerging governance framework for AI systems and data handling. With these concluding remarks the panel made the case for improved privacy, transparency, and accountability mechanisms to protect individuals even where AI systems did not have the attributes of fiduciary actors.

Endnotes

1 In re: Summoning Advocates who give legal opinion or represent parties during investigation of cases and related issues, 2025 INSC 1275.

2 OpenAI loses fight to keep ChatGPT logs secret in copyright case, The Reuters,

https://www.reuters.com/legal/government/openai-loses-fight-keep-chatgpt-logs-secret-copyright-case-2025-12-03/

3 What is 'AI privilege'? OpenAI CEO says talking to ChatGPT should be as private as a doctor's visit, The Economic Times,

https://economictimes.indiatimes.com/magazines/panache/what-is-ai-privilege-openai-ceo-says-talking-to-chatgpt-should-be-as-private-as-a-doctors-visit/articleshow/121782379.cms?from=mdr

EVENT REPORT | NOVEMBER 2025

From Trust to Risk: The Case for Privilege in AI Conversations

IIGF
INDIA INTERNET GOVERNANCE FORUM

IGAP INDIAN GOVERNANCE AND POLICY PROJECT

For more details visit: www.igap.in

Contact us: relations@igap.in