India IGF (IIGF-2025) Workshop: 11

28 November 2025

12:00 -1:00PM

Session Title	**From Trust to Risk: The Case for Privilege in AI Conversations**
Primary Theme	AI for People, Planet and Progress (Balance between AI innovation and regulation)
Session Format	Panel Discussion
Details of Organisers	
Name of Primary Organiser	Dhruv Garg
Affiliation	Indian Governance and Policy Project **(IGAP)**
Primary Stakeholder Group	**Civil Society**
Organizing Partners (if any)	NA

Session Details :

Specific Issues of Discussion (In 200 words)

The proliferation of AI-based chatbots has reshaped social and behavioral patterns, with users increasingly confiding intimate and confidential details that often carry implications for emotional well-being and mental health. Other conversations may relate to legal or medical issues. Experts have noted that, unlike attorney-client or doctor-patient interactions, the information contained within chatbot conversations lacks 'legal privilege' or confidentiality. Concerns have been raised about this substantial lack in privacy protection. Specifically, deployers of chatbots have been asked to produce the contents of user interactions in various legal contexts. A controversial precedent in this regard was set

by a New York federal court in where the plaintiffs requested the indefinite retention of all user content on ChatGPT, and a preservation order was granted.

The disclosure of AI chat messages creates risks for users. These conversations contain intimate personal details, and may involve admissions to unlawful conduct. Users may assume conversations are private or protected, leading them to overshare sensitive details without understanding the risks. The proposed panel will explore the privacy and regulatory gaps resulting in the lack of privileges for chatbot conversations. The panel will also address questions of accountability, cross-border implications, safeguards and ethical design principles to protect vulnerable users.

Describe the Relevance of Your Session to IIGF2025 Theme* (In 200 words)

This panel sits squarely within the sub-theme of the 'balance between AI innovation and regulation' under 'AI for People, Planet and Progress'. It highlights how the rapid proliferation of AI chatbots has unlocked new forms of digital interaction and support. However, this has also exposed users to unprecedented risks. While innovation has made conversational AI accessible within emotional support, medical, and even legal contexts, the absence of legal privilege, limited privacy protections, and unclear accountability frameworks raise pressing regulatory questions. An expert discussion on these issues would provide a critical first step within the Indian jurisdiction towards addressing these concerns.

Methodology / Agenda* (In 200 words)

This session will adopt an interactive format focused on emerging concepts of 'AI privileges' and the absence of confidentiality protections in AI-enabled interactions.

Introduction (10 Minutes): The discussion will begin with an introduction by the Moderator, who will explain the event flow and introduce the panellists. The Moderator will set the stage by outlining the issues at stake: whether communications with AI systems should attract protections similar to doctor-patient or attorney-client privilege, and what risks arise in their absence.

Speaker Inputs (30 Minutes): The following panel segment will feature each panellist speaking for up to 8 minutes. Topics covered in this segment range from legal and regulatory frameworks to technology deployment, ethical design, and user rights. Panellists will explore cross-border challenges and potential models for recognising privilege in interactions with AI-tools.

Moderated Exchange (15 Minutes): Subsequently, a moderated question and answer session will allow the audience to pose questions and engage directly with the panellists, creating space for practical insights and debate.

Concluding Remarks (5 Minutes): Finally, the Moderator will offer 5 minutes of concluding remarks, highlighting key takeaways and underscoring how this issue reflects the broader need to balance AI innovation with regulatory safeguards for user trust and protection.

- Moderators & Speaker Details

	Name	Affiliation	Designation	Gender	Stakeholder Group
Moderator	Dhruv Garg	Indian Governance and Policy Project (IGAP)	Founding Partner	Male	Civil Society
Moderator 2 (optional)	Shachi Solanki	Indian Governance and Policy Project (IGAP)	Associate Director	Female	Civil Society
Speaker 1	Anup Surendranath	NALSAR University of Law	Professor of Law and Executive Director, The Square Circle Clinic,	Male	Academia
Speaker 2	Kapil Chaudhary	Dentons Legal Legal	Partner	Male	Private
Speaker 3	Kirti Mahapatra	Shardul Amarchand Mangaldas	Partner	Female	Private
Speaker 4					

Session Report (TO BE COMPLETED AFTER SESSION)

This part is given here for the information of session organisers, who will need to provide the requested report within 48 hours after the end of the conference.

- **Brief Summary of Presentations**

  No presentations were made during the panel discussion.

- **Substantive Summary of the key Issues Raised and Discussed**

  The discussion started from the gap that presently AI privilege, as a legal framework, dis not exist, even as people increasingly use chatbots for legal, medical, financial and deeply personal support. Traditional privilege rests on a fiduciary relationship, professional ethics, and evidentiary exclusion. AI systems lack legal personality, licensing and duties of loyalty, and are embedded in commercial, data-driven business models. This raised the question of whether protection should be based on function (what AI is doing) or form (who is providing the service), and whether we risk a two-tier regime where only those who can access human professionals receive the benefit of privilege. Speakers distinguish privacy, surveillance and platform power from the narrower doctrine of privilege, and foreground decisional autonomy (not just privacy or "chilling effect") as the deeper justification for privilege. They also highlight information asymmetry, mandatory data retention under DPDP, law-enforcement access, and the special risks for mental health users and children as key concerns.

- **Conclusions and Suggestions for the Way Forward**

  The panel was broadly skeptical that full attorney-style privilege could be granted to AI systems. Conceptually, it was hard to justify the existence of legal privilege in the absence of a genuine fiduciary relationship. Practically, it risked obstruction of justice, evidence laundering, platform capture, and shielding powerful entities rather than users. At the same time, the panel recognised that AI was mediating highly intimate, autonomy-relevant conversations, especially in mental health and quasi-legal guidance. The panel felt the way forward on addressing the foundational concerns was:
  (i) strengthening privacy and data-governance rules (clear retention limits, deletion rights, transparent training practices);

(ii) sectoral regulation and liability for sensitive domains like digital health;

(iii) UX-level controls (opt-out of training, meaningful consent, defaults protective for vulnerable users and children);

(iv) building judicial doctrine incrementally through cases on admissibility of AI chat logs; and

(v) investing in user literacy about what AI can and cannot protect.

- **Number of Participants in the session:** 30 participants

- **Gender Balance in Moderators/Speakers (Please fill in numbers):** 3 men and 2 women

- **How were gender perspectives, equality, inclusion or empowerment discussed? Please provide details and context.**

  While the panel treated 'AI privilege' primarily as a doctrinal and design question, the issue of its gendered and inclusion-related consequences was also posed to the panel to consider. This included the possibility that marginalised groups may rely more heavily on low-cost AI tools for basic legal or health guidance. Without confidentiality safeguards, it was considered whether these users would face disproportionate risks, potentially deepening existing inequalities. However, fundamental issues with the fiduciary role of AI systems, and awareness on data-use, were seen as critical to determining even the social impact of AI-user interactions with, or without, privilege.

- **Was the Indian perspective of the topic discussed? If yes, how does it compare or contrast with local or global perspectives?:**

  Yes, during the session, the Indian perspective on the law of privilege, as part of the laws of evidence, was discussed. It was also noted that a dedicated legal framework for handling privilege for user-AI interactions had not yet been established in any jurisdiction, including India. The panellists assessed the first principles of the introduction of such a legal framework in India.