Tech Ethics LAB

RAISE THE BAR

Raise the Bar 2025: Key Takeaways and What's Next

Detailed Analysis

Sara E. Berger and Megan McDermott

Trust Is a Leadership Choice

How are people thinking about it?

Across panels and case studies, one message was clear: trust is not a technical property, it's a human one. While things like transparency, security, or auditability are critical ingredients of responsible AI, they're not enough on their own. Real trust doesn't start with the system itself; it grows from human leadership and organizational culture. So, while technical rigor may sustain confidence, human accountability creates it. To illustrate this point, speakers noted growing evidence that employees often hesitate to adopt AI tools (or at least organizationally-sponsored tools), not because they lack the skills to use them well, but because they lack confidence in how those tools are being explained, governed, and/or used by those at the top of the org chart.

What this clearly illuminates is that trust is less about what a system does than what it signals. A model that works perfectly but is introduced without clarity or accountability can still erode confidence. By contrast, even imperfect systems can earn trust when teams see that leadership is transparent about risks, proactive in managing them, and honest about what's still uncertain. A panelist summed it up plainly by stating: AI failure isn't an 'if', it's a 'when'.

Accountability, then, must be personal as well as procedural. Several speakers warned that when everyone is responsible, no one truly is. True accountability doesn't mean everyone carries equal blame when things go wrong, it means individuals are clearly empowered - and expected - to own ethical outcomes within their sphere of influence. This accountability should be reinforced by aligning success metrics and rewards with the long-term health of the organization more holistically - both its stakeholders and its systems. The organizations that will be best prepared to manage change and uncertainty in this new area are those that make integrity visible and reward it accordingly.

What are people doing about it?

Organizations are beginning to take the language of trust and translate it into operating models. Several speakers described the creation of cross-functional governance groups that bring together leaders with diverse roles and functions - risk, legal, technology, ethics - to assess new systems and ensure that decisions about AI are made in full view of their human and organizational impacts. These structures mark an early but important shift: from compliance as a box-checking exercise to trust as a shared practice.

At the same time, panelists noted that most organizations are only partway through this journey. While governance frameworks are becoming more common, incentives and culture have not yet caught up. Some companies are experimenting with adding responsible AI goals to leadership KPIs or embedding transparency and accountability measures into product reviews and employee training. Others are piloting communication strategies that create psychological safety, inviting employees to raise concerns and ask questions without fear of reprisal. The common thread across these efforts is recognition that trust cannot be delegated—it must be lived, modeled, and reinforced daily through decisions and behaviors.

Key Questions

- How can we measure trust as a dynamic behavior, not just a perception?
- What organizational structures best sustain psychological safety and accountability?
- How can organizations design incentives that reward integrity as much as innovation?
- What balance of transparency and discretion best strengthens trust around AI for stakeholders within and external to the organization?
- How do leaders communicate about change and uncertainty that reinforce rather than undermine confidence?

The main take-away? Trust cannot be engineered; it must be continuously demonstrated and earned. The most forward-looking enterprises understand that technical reliability and human integrity reinforce each other, and that one without the other is not just incomplete, but inherently dangerous. By making accountability visible, aligning incentives with ethical outcomes, and treating transparency as a daily behavior rather than a periodic report, organizations can build the kind of trust - internally and externally - that accelerates both innovation and confidence. As one panelist aptly put it, "The better the brakes, the faster the train."

Reliability by Design

How are people thinking about it?

"AI, like a cat, behaves according to its nature", is the memorable comparison used by one of the moderators to make the point that responsible AI isn't about expecting perfection - or being surprised by system flaws - but about designing for predictability and creating the systems and safeguards that help technology work within clear, human-defined bounds.

The discussions focused less on controlling specific types of data or algorithms and more on designing data and model ecosystems responsibly and intentionally, with an eye toward outcomes and inferences. Across panelists, there was a shared sentiment that emerged over the discussion: while the complexity of AI systems is real, many failures in responsible AI stem from neglecting the basics. Organizations often pour resources into sophisticated defenses against adversarial threats - undeniably important - but overlook simple validation and human-in-the-loop quality checks that could prevent critical issues earlier in the process.

Panelists also emphasized that risk cannot be managed uniformly across sectors. The tolerance for error in a recommendation engine is not the same as in a diagnostic tool or defense system. Responsible AI therefore demands proportional oversight systems designed with a clear understanding of what level of uncertainty is acceptable, and where it is not.

What are people doing about it?

Companies leading in data and model integrity are taking a holistic approach, thinking in systems rather than modules. They recognize that issues like bias and privacy preservation don't end once data has been cleaned; they can easily reappear through human fine-tuning, labeling, review, and feedback. Responsible AI therefore requires vigilance at every point of contact between human and machine, whether end-user, developer, auditor, or annotator.

Leaders also emphasized the importance of confronting trade-offs directly. For example, enacting privacy protections through anonymization may inadvertently erase data signals needed to detect or mitigate inequities. Responsible AI, then, is less about optimizing for single metrics and more about managing interdependencies - the balance between fairness, privacy, accuracy, and efficiency within specific contexts.

Panelists described emerging strategies: some organizations are starting from intended outcomes and working backward, reverse-engineering governance workflows, model architectures, and data structures to achieve those results responsibly. Others are turning to federated architectures to promote interoperability and accountability without sacrificing innovation or core values. Across the board, participants agreed that embedding guardrails, auditability, and cross-functional oversight into AI ecosystems is just as critical as data curation and model development itself.

Key Questions

- What new mechanisms can help ensure responsibility or accountability when models infer rather than collect sensitive information?
- How can enterprises balance simplicity with sophistication of data and model governance - how can they make it easier to do the right thing and help ensure that basic validation remains integral to increasingly complicated AI governance frameworks?
- How can cross-functional review structures and tools scale without slowing innovation, and in instances where it can't, how can slowing down in the short term be convincingly reframed as speeding up and futureproofing in the long term?

- What needs to happen so that data and model governance is seen less as a procedure but also as a product in and of itself, as something that requires iterative planning, key performance indicators, and measurable outcomes beyond compliance?

The main take-away? There are plenty of examples - across history and sectors - that the responsible governance of data and models is not in opposition to technological innovation or business opportunity. When reliability is built through deliberate design and human oversight, organizations create systems that evolve safely within clear, human-defined boundaries. Strong guardrails and transparent governance enable progress that is both confident and controlled, allowing teams to move quickly without losing sight of accountability.

Ethics That Earn Their Keep

How are people thinking about it?

A compelling tension surfaced the course of one particular discussion: 53% of executives say that their AI ethics governing bodies are not effective, yet recent evidence shows that companies investing more in AI ethics see up to 30% higher operating profits than competitors (IBM IBV Study, October 2025). Why? A central through-line across the panelists' discussion was that AI governance is not a technical problem but rather a deeply human one.

Participants emphasized that the success or failure of governance boards or AI ethics programs rarely hinge on the specifics of data or models (nor on the existence of ethical principles) but instead on culture, incentives, and the willingness to make ethical reflection a part of daily practice. It also hinges on developing the skills necessary to anticipate and spot risks early on across the entire data, model, and ecosystem stack and communicate these issues effectively. The challenge for responsible AI generally and AI governance specifically, then, lies not in drafting more policies or compliance checkmarks, but rather in creating the conditions for transparency, inclusion, and shared moral reasoning to thrive.

Accountability was a particularly salient topic of discussion. When the audience was polled about who was accountable for responsible AI outcomes in their organization, the top answer was everyone. While it may sound noble, as one person said, "If everyone is accountable, is anyone accountable?" To govern well, you need more than good intent but also ownership of risks and outcomes, which requires power and often a funded mandate (e.g., resources to enact processes, steer projects, or education, proof of alignment to business needs and regulatory requirements, mechanisms for ensuring return-on-investment, etc.). There needs to be a recognition of the table stakes and critical discussions on not only what risks are acceptable but whose risks the organization is willing to take on. This also necessitates rethinking what it means to actually offer a "seat at the table" and to whom. Where is the table, what does it look like, and when is more than a seat required to have an impact on what happens in the technology landscape?.

What are people doing about it?

Organizations that treat AI governance as a dynamic and proactive capability rather than a static function or reactive mechanism are ahead of the curve. Some are establishing cross-functional ethics boards that include legal, technical, business, and societal perspectives to ensure accountability and coherence across silos and in absence of regulatory guidelines. Others are embedding "ethical checkpoints" into product lifecycles—moments to

pause and assess who benefits, who bears the risk, and how decisions align with broader organizational values and client or consumer needs. Several participants highlighted the growing role of external validation—through third-party audits, public transparency reports, or shared benchmarks—to reduce internal fragmentation, improve trust and unity, and mediate conversations with end-users and the public.

In parallel, training programs are evolving from AI literacy and compliance training to ethical fluency: helping leaders and teams navigate ambiguity, manage trade-offs, and understand the socio-technical nature of AI risk. These initiatives signal a broader shift from governance as prevention to governance as possibility—creating systems that not only minimize harm but actively support human agency and situated knowledge to make highly contextual, nuanced decisions.

Key Questions

- What mechanisms and cultural changes can be created or strengthened to ensure that accountability is simultaneously shared (from an inclusion standpoint) but still owned and enforceable?
- What should we be measuring and verifying to know if organizations are being accountable in their approaches, requirements, and outcomes?
- How can organizations better align definitions and goals of 'responsibility" across different ethical, social, legal, technical, and business domains?
- How can we make people across different roles and with different levels of expertise and power more proficient at identifying and reacting to ethical issues - what must be changed about internal and external AI ethics educational programs to better prepare people for doing this work?

The main take-away? Applied technology ethics is more than compliance alone. If compliance defines the foundation or the floor, ethical deliberation and ethically aligned actions define the ceiling. The organizations that will lead in the coming decade will be those that move beyond 'ethics and trust by design' to also include ethics and trust through practice, with an emphasis on the follow-through, maintenance, and updating of actions to match intentions and principles. Ethics in practice - be that through education, governance processes, tools and frameworks, beneficial impacts, or collaborations - must become a continuous dialogue about people, purpose, power, and real-world outcomes.

Humans at the Center of Change

How are people thinking about it?

Perhaps not a novel take, but one that surfaced consistently across panels and from every type of participant, is the notion that responsible AI is fundamentally about people, not tools. The organizations that stand to gain the most from this current technological revolution are those that treat human judgment, creativity, and adaptability as critical, strategic assets - not inefficiencies to automate away. As cited by one of the panelists, a former Harvard Business

Professor, Shoshana Zuboff, observed in her work that one of the most important byproducts of the introduction of the internet wasn't technical capability, but the ability of individuals in every aspect of their lives - but especially at work - to ask better questions. The underlying point here being that technology's true value lies not in automating work, but in deepening human understanding of it.

In the same vein, speakers emphasized that the future of work will depend less on mastering discrete technical skills and more on cultivating deep domain expertise, contextual awareness, and systemic thinking - capabilities that enable humans to piece together unique connections in ways that AI cannot. For leaders, this shift requires a reframing of workforce resilience as a form of strategic resilience. It's not about cutting costs through automation; it's about building long-term value through human-AI collaboration. True leadership, participants argued, increases revenue by unlocking inspiration and innovation, not by shrinking the workforce.

What are people doing about it?

Forward-looking organizations are already redesigning roles, incentives, and learning systems to support continuous adaptation. Some are experimenting with new approaches to compensation and career development that reward curiosity and skill growth, not just output. Others are rethinking job architecture altogether, moving from static hierarchies to more fluid, project-based teams that blend technical expertise with ethical and creative problem-solving.

Education and workforce development programs are also starting to move from strictly focusing on "AI literacy" toward more holistic "resilience literacy", preparing workers to better evolve with technology rather than chase it. Overall, across these approaches runs a common thread: a shift from optimizing human workers for machines to designing systems and structures that optimize for human potential

Key Questions

- How do we value learning as a measurable and rewarded outcome within organizations?
- What lessons can be drawn from models (e.g. rotational programs in East Asia) that institutionalize lifelong learning and cross-disciplinary growth?
- How do we measure the value of creativity and innovation in ways that recognize exploration, not just efficiency?
- What organizational structures and designs best promote ethical decision-making and sound human judgment alongside automation?

The main take-away? Responsible AI begins and ends with people, and keeping humans at the center of innovation means redefining progress itself. The organizations that will lead in this new era are those that see human judgment, creativity, and learning as core business capabilities, not peripheral values. Designing systems and structures that strengthen rather than sideline human capability, is the surest way to build resilient, ethical, and future-ready enterprises.

Ultimately, the question shouldn't be how AI will change work, but how we will shape work - and ourselves - to ensure that technology advances our individual and collective capacities to flourish.