

LSA ANNUAL MEETING 2025 - CHICAGO

CRN41 - Economic Crime and Corporate Compliance

ROUNDTABLE SESSION

Tech-Powered Disclosures: The Role of Technology and Artificial Intelligence in Countering Organizational Misconduct

Organizer:

- Stephen Holden (De Montfort University, UK)

Chair:

- Diane Ring (Boston College Law School, USA)

Participants:

- Douglas Arner (University of Hong Kong, China)
- Vivienne Brand (Flinders University, Australia)
- Gaia Fiorinelli (Scuola Superiore Sant'Anna, Pisa, Italy)
- Elisabetta Pietrocarlo (Luiss University, Italy)
- Glenn Sorrentino (Science & Design, Inc., USA)

Date: May 24, 2025

Suggested Citation

BLUEBOOK: [Name Surname], LSA Annual Meeting Chicago, CRN41 Economic Crime and Corporate Compliance, Roundtable Session 'Tech-Powered Disclosures: The Role of Technology and Artificial Intelligence in Countering Organizational Misconduct,' CORPORATE CRIME OBSERVATORY, p. [0] (May 24, 2025), <https://corporatecrime.co.uk/crn41-economiccrime#chicago2025>

APA: [Surname, Initial] (2025, May 24). LSA Annual Meeting Chicago, CRN41 Economic Crime and Corporate Compliance, Roundtable Session 'Tech-Powered Disclosures: The Role of Technology and Artificial Intelligence in Countering Organizational Misconduct'. *Corporate Crime Observatory*, p. [0]. Retrieved from <https://corporatecrime.co.uk/crn41-economiccrime#chicago2025>

HARVARD: [Surname, Initial] (2025) LSA Annual Meeting Chicago, CRN41 Economic Crime and Corporate Compliance, Roundtable Session 'Tech-Powered Disclosures: The Role of Technology and Artificial Intelligence in Countering Organizational Misconduct'. *Corporate Crime Observatory*, p. [0]. Available at: <https://corporatecrime.co.uk/crn41-economiccrime#chicago2025>

OSCOLA: [Name Surname], LSA Annual Meeting Chicago, CRN41 Economic Crime and Corporate Compliance, Roundtable Session 'Tech-Powered Disclosures: The Role of Technology and Artificial Intelligence in Countering Organizational Misconduct', (*Corporate Crime Observatory*, 24 May 2025), p. [0]. <<https://corporatecrime.co.uk/crn41-economiccrime#chicago2025>>

Table of Contents

1. The Role and Tension of Predictive Analytics in Misconduct Prevention.....	3
A Board-Level Perspective: The Double-Edged Sword of Data.....	3
An Industry-Specific View: The Financial Sector Case Study.....	4
Practical and Technical Challenges.....	4
The European Regulatory Framework: "High-Risk" vs. "Prohibited" AI.....	5
2. Repurposing Cybersecurity Monitoring as a Compliance Asset.....	5
Cybersecurity Violations as Proxies for Broader Misconduct.....	6
The Challenge of Distinguishing Malice from Error.....	6
Data Controls, Business Incentives, and Technical Realities.....	6
The Impact on Corporate Culture and Trust.....	7
3. The Globalization of AI Regulation: EU vs. US Approaches.....	8
A Surprising Convergence in Regulatory Expectations.....	8
The Critical Divergence: Enforcement and Penalties.....	9
The Future of Harmonization and a New Enforcement Model.....	9
4. AI-Powered Whistleblowing Systems and Anonymity.....	10
The Technical and Practical Challenges of Anonymity.....	10
Culture as a Greater Barrier Than Technology.....	11
Using Technology to Enhance Trust and Feedback.....	11
5. The Human in the Loop: Roles, Training, and Ethical Judgment.....	12
Corporate Structure and the Human Pipeline Problem.....	13
The Challenge of Training and Explainability.....	13
Ethical Trade-offs and the Impact on Human Judgment.....	14
6. Final Question: Metrics for Success.....	15

SUMMARY OF THE DISCUSSION

1. The Role and Tension of Predictive Analytics in Misconduct Prevention

The roundtable's first discussion, initiated by moderator Diane Ring, centered on the capacity of technology, particularly machine learning and predictive analytics, to prevent corporate misconduct. Ring framed the conversation by highlighting the growing ability of technology to scan vast amounts of corporate data, such as purchase orders and travel claims, to flag potential problems. This capability shifts compliance into a realm of proactive, internal corporate surveillance, distinct from state surveillance. She noted the diverging regulatory landscapes, pointing out that the EU's AI Act has designated certain risk-scoring systems as "high risk," requiring specific assessments. The core question posed to the panel was whether these predictive scores actually prevent misconduct or merely label it after the fact, exploring the tension between the speed and scope offered by technology and the inherent concerns about its accuracy and potential for bias.

Key Observations and Ideas:

A Board-Level Perspective: The Double-Edged Sword of Data

Vivienne Brand, adopting the viewpoint of a corporate board member, articulated a nuanced and cautious stance on the integration of AI into compliance. She described two immediate thoughts upon being presented with new AI capabilities. The first reaction is positive, based on the principle that more information is beneficial. Even historical data provides valuable insights, revealing much about the organization and what is likely to occur in the future. However, this optimism is immediately tempered by a significant risk: the danger of "drowning in the data". Brand emphasized that a persistent and major problem for governance at the board level is the challenge of filtering information into a format that is both "condensable and meaningful." Consequently, while AI offers more data, it also presents "more risk, more danger of just confusing the message." Brand also raised the concern that AI-generated material might not be interrogated with the same critical rigor as in the past. This leads to the well-documented phenomenon of people defaulting to trusting AI, even when it is wrong. She noted that in early AI applications like search, professionals could be led to believe the AI even if it advised them to do something professionally wrong. For a board, this presents a complex challenge:

welcoming the potential of more information while grappling with the problem of how to "manage this intelligently."

An Industry-Specific View: The Financial Sector Case Study

Douglas Arner offered a contrasting perspective from the financial sector, a "heavily regulated industry" where large-scale data analysis is already deeply embedded in operations. He noted that compliance has been the biggest growth area in finance for the last three decades, with systems often driven by specific regulatory requirements. These systems serve a dual purpose: they are designed to deter misconduct internally while also enabling both the institution and its regulator to identify issues as they arise. The sheer volume of data in this context is fundamentally different from other sectors. Arner gave the example of the UK's Financial Conduct Authority, which now receives approximately [7 billion transaction reports per year](#). Financial institutions must intensively analyze this data not only to identify potential issues like market manipulation before regulators do, but also to monitor operational considerations. With this volume of data, it is now possible to identify subtle changes in trading behavior that might indicate a trader is acting improperly. Arner drew a powerful analogy, comparing the monitoring of traders to that of Olympic athletes, where "every keystroke, every entry, every movement" is tracked to maximize performance while also managing downside risks.

Practical and Technical Challenges

Glenn Sorrentino brought the discussion to a practical level, stressing that "AI it's not monolithic" and its implementation is fraught with trade-offs. Drawing from his experience with OCR technology for mortgage agents, he highlighted accuracy as a major hurdle, using the example of voice assistants that often struggle with accents. Privacy is another of the "biggest thing[s]," with questions around the appropriateness of using voice recognition in the workplace. The stakes are high, as incorrect predictions, such as in sales forecasting, can have "huge implications" for a business. Sorrentino asserted that at this stage of development, "most times humans are supplementing the AI," and any implementation requires "a lot of oversight by a human right now". He also commented on the market dynamics in Silicon Valley, where there is a rush to integrate and sell AI, forcing other companies to follow suit to remain competitive, regardless of the technology's maturity.

The European Regulatory Framework: "High-Risk" vs. "Prohibited" AI

The European panelists detailed the specific and layered approach of the EU AI Act in regulating these technologies. Elisabetta Pietrocarlo explained that whether a compliance tool is deemed "high-risk" ultimately "depends" on a specific impact assessment for each system, as a "one size fits all" approach does not work. Formally, high-risk systems are those with potential impacts on safety, security, and fundamental rights, as listed in Annex 3 of the regulation, with examples including AI for credit scoring or employee monitoring. However, she provided a critical distinction: certain systems can cross a line into being prohibited entirely under Article 5. For instance, a system used to monitor employees that also "is meant to evaluate emotions" would pose an "unacceptable" risk and thus be considered a prohibited practice, not merely a high-risk one. Furthermore, any system that engages in profiling will always be classified as high-risk.

Gaia Fiorinelli expanded on this, confirming that corporate compliance programs can fall under the "prohibited practices" of Article 5. She gave the example of social scoring, noting that while companies have a legitimate interest in evaluating fraud risk, the data used must be directly relevant to that output. A key principle is that individual risk assessments should be focused on "activities on behaviors on data rather than on the specific features of the person". Echoing Pietrocarlo, Fiorinelli stated explicitly that "emotional recognition systems are banned in the workplace." The overarching guidance for companies implementing these tools is to ensure the focus remains on monitoring processes and activities, not on profiling people. The discussion also clarified that liability is distributed along the supply chain; the "deployer" of the AI is responsible for its use, while the "provider" is responsible for ensuring third parties can use it responsibly.

2. Repurposing Cybersecurity Monitoring as a Compliance Asset

Following the initial discussion on predictive analytics, Diane Ring shifted the focus to the second major topic: the intricate link between cybersecurity and compliance. She framed this issue as an opportunity to reimagine "insider threat toolkits" as proactive compliance assets. The premise is that organizational misconduct often leaves a "digital trail," such as "logins at odd hours or privilege escalations" and "unauthorized database inquiry." While cybersecurity controls are already designed to be an "early warning system" for issues like insider fraud or corruption, the central question became whether this internal monitoring can be repurposed as a formal compliance asset. This pivot raises critical practical

questions about what data can be captured, how to differentiate between "malicious conduct and just a legitimate error," and whether "overzealous digital compliance" might create a "chilling effect" on employees' willingness to report issues in good faith.

Key Observations and Ideas:

Cybersecurity Violations as Proxies for Broader Misconduct

Gaia Fiorinelli suggested that cybersecurity violations could serve as effective proxies for identifying corruption or other misconduct. She referenced the ReSPA report "[Abuse of Information Technology \(IT\) for Corruption](#)" of 2013, which, while focused on the public sector, found that IT violations like illegal logins or unauthorized data access were often integral parts of a larger "crime of corruption." In that context, the violation of IT rules represented an "abuse of powers of the public official." Fiorinelli proposed that this same perspective could be shifted to the corporate sector, allowing companies to view and investigate cybersecurity breaches by their own officials in a similar light.

The Challenge of Distinguishing Malice from Error

Douglas Arner highlighted that the line between "malicious behavior versus error is becoming more problematic," largely due to the increasing sophistication of "deepfake and other sorts of threats." He provided a stark example where biometric data was used to create an entirely deepfaked video conference call, including the company's CFO and senior management, which successfully directed a fraudulent transfer of a large sum of money. Arner also pointed to a growing problem related to generative AI, where employees copy and paste sensitive "internal data in external Gen AI systems." This act, which is difficult to identify, makes proprietary corporate data publicly exposed. To address the challenge of separating intentional acts from mistakes, Elisabetta Pietrocarlo proposed deploying "behavior techniques, analytic techniques based on machine learning." Such systems could analyze the "normal behavior" of users and systems within the organization and, through self-learning, establish a baseline. When the system detects anomalies, it could assign a risk score to help prioritize red flags and distinguish genuine misconduct from simple errors, thereby reducing the "noise" within the organization.

Data Controls, Business Incentives, and Technical Realities

Glenn Sorrentino approached the topic from a pragmatic standpoint, stating that a great deal of compliance is about ensuring proper "visibility into data" and

permissions, such as controlling which employees can access historical sales data. An audience member questioned whether privacy was a simple trade-off, suggesting other variables like access controls and data retention periods could be adjusted. Sorrentino responded that data retention is a "tricky situation" because businesses are incentivized to retain as much data as possible, unless compelled by regulations like GDPR or lawsuits. He described the extensive level of monitoring that is technically possible, including tracking the first thing a user clicks on a site, how long they dwell there, and even their mouse movements. He argued that a combination of "willful ignorance, business justification, and just not having a legitimate interest in all forms of privacy and security" often prevents the implementation of perfect systems. Furthermore, because new business priorities always take precedence, companies rarely go back to perfectly secure their systems later on. To underscore the point that barriers to ideal compliance are often a matter of will rather than technology, Glenn Sorrentino noted that while an economic crime expert like Prof. Costantino Grasso could design a "perfectly, like, compliant system," the reality is that "most businesses don't want perfect systems." Professor Grasso, who was present in the audience, agreed and expanded on this, stating that governments also do not want these systems, and that such a reluctance stems from the fear that imposing such "perfect" systems in a global context characterized by asymmetric efforts to tackle corporate crime would put their companies at a competitive disadvantage against others that are not bound by the same stringent rules.

The Impact on Corporate Culture and Trust

Vivienne Brand focused her comments on the crucial "culture piece," directly addressing the concern about a "chilling effect." She posed the fundamental question of whether increased automation and monitoring make employees "more or less likely to trust the system." She noted that strong empirical evidence shows cultures of trust are "essential to effective whistleblowing." Brand outlined two potential, and opposing, cultural outcomes. On one hand, enhanced monitoring could increase the "expressive function" of the internal systems, signaling to employees that the company is attentive and supportive. On the other hand, it could be perceived as a "destructive cultural element." She drew a powerful historical parallel to post-reunification East Germany, where a deep cultural resistance to informing persisted for generations due to the societal memory of the Stasi. Glenn Sorrentino provided a modern anecdote illustrating the reality of this monitoring, recounting how he and a colleague installed the Tor Browser on a work phone, only to receive a call from security "within 30 seconds" offering to come and uninstall it for them. The incident

served as a stark confirmation of the type of intensive monitoring that was actually in place.

3. The Globalization of AI Regulation: EU vs. US Approaches

The roundtable's third major topic, introduced by the moderator, Prof. Diane Ring, addressed the "globalization of efforts to regulate in this space". Ring set the stage by contrasting the European Union's draft AI Act, which imposes comprehensive "life cycle duties" such as risk classification and human oversight, with the distinct regulatory path being taken by the United States. This divergence presents a significant challenge for multinational corporations, which must design their compliance toolsets to satisfy multiple, and sometimes conflicting, legal regimes that have extraterritorial reach. The initial question posed was where the obligations of the EU's high-risk framework exceed US expectations, and vice versa.

Key Observations and Ideas:

A Surprising Convergence in Regulatory Expectations

Despite the different formal approaches, the panel identified a significant trend toward convergence between EU and US expectations. Elisabetta Pietrocarlo pointed to the US Department of Justice's (DOJ) recently updated "[Evaluation of Corporate Compliance Programs](#)" (ECCP), which now explicitly details what is expected of companies using new technologies like AI. The DOJ now expects companies to assess AI-related risks, implement measures to prevent negative consequences, and ensure humans are involved in evaluating AI outputs and taking responsibility for their use. Furthermore, prosecutors will consider the active monitoring and testing of AI systems when evaluating if a compliance program is effective in practice. Pietrocarlo concluded that these DOJ expectations align so closely with the EU AI Act's requirements that while differences remain, "the distance is not too long between the expectation and the obligation."

Vivienne Brand suggested this convergence is driven by several factors beyond formal law. These include the "big four audit firms" developing their own standards that filter down to companies, broader macro-level pressures, and corporate self-interest, as companies seek efficient, standardized approaches to avoid "grief." Gaia Fiorinelli added that the EU AI Act itself is designed to have a global effect through the supply chain. The Act makes AI providers responsible for potential misuse by deployers, requiring them to implement "contractual and

technical safeguards." This effectively projects the EU's standards onto global technology companies that want to operate within the European market.

The Critical Divergence: Enforcement and Penalties

A significant point of divergence emerged during a discussion on enforcement. An audience member expressed distress at comparing the DOJ's guidelines to the EU's formal regulation, noting the historical weakness of US corporate crime enforcement, where less than 5% of such crime is discovered and fines are often a tiny fraction of profits.

In response, Gaia Fiorinelli clarified that the sanctions in the EU AI Act are formally administrative but can be exceptionally high, calculated as a percentage of a company's global income. Elisabetta Pietrocarlo added that because the fines are so severe, they have a strong "punitive effect" and can be considered "criminal in nature". The choice of administrative over criminal penalties was explained as a function of the EU's limited legal competencies, as criminal law is largely left to individual member states. However, member states are free to build on this foundation; for example, Italy has already drafted a law to introduce specific criminal sanctions for the misuse of AI, in addition to the EU's administrative ones.

The Future of Harmonization and a New Enforcement Model

When asked about the role of international standard-setting bodies, Vivienne Brand expressed a pessimistic view, stating that such efforts are "going a little bit backwards at the moment". Douglas Arner, however, offered a different perspective from the financial sector. He observed that any current drive for further harmonization is coming from the industry itself, not from international organizations. He described the approach of global financial institutions, which is to build a centrally harmonized internal system that is flexible enough to meet differing reporting requirements in individual jurisdictions. Arner also identified an evolving and influential enforcement trend in financial services that is beginning to spread to other sectors. When a major compliance failure occurs, regulators are increasingly using a three-pronged approach:

1. Imposing a massive fine, sometimes in the billions.
2. Mandating that the institution spend billions more to build the necessary internal technological systems to prevent a recurrence.
3. Appointing an external monitor, supervised by the enforcement agency, to oversee the implementation of these new systems.

This model represents a shift towards forcing deep, structural technological and procedural change within an organization, going far beyond simple monetary penalties.

4. AI-Powered Whistleblowing Systems and Anonymity

The discussion then transitioned to its fourth topic: the challenges and potential of AI-powered disclosure systems. Moderator Diane Ring framed the issue by asserting that even the most sophisticated algorithm cannot guarantee fairness if the broader system surrounding it fails to inspire trust. She posed a series of critical questions: How can automated whistleblowing portals preserve anonymity without enabling misuse or creating unacceptable "re-identification risk"? And how can organizations maintain transparency when using "black box" systems governed by proprietary vendor agreements? This set up a core tension between accuracy, privacy, and the need for systems to be explainable versus the desire for intellectual property security.

Key Observations and Ideas:

The Technical and Practical Challenges of Anonymity

Glenn Sorrentino addressed the technical aspects, distinguishing between security measures like encryption, which protect the *content* of a message, and privacy measures, which protect the *identity* of the sender. He noted that technologies to enhance anonymity, such as native Tor Onion services, exist but are often not implemented due to restrictive corporate security policies and a general reluctance to fully protect whistleblowers.

Vivienne Brand described re-identification as a "really big problem," expressing skepticism about how a purely automated system could prevent it. She contrasted this with a human-led process she had experienced on a board, where a single, trusted individual acted as a triage point, making constant, nuanced, and judgment-heavy assessments to prevent the whistleblower's identity from being pieced together.

Sorrentino added that in a corporate environment, re-identification of a whistleblower can be "trivial". He explained that a company's IT department does not need to use sophisticated internal IP tracking; they can often simply see that the work computer assigned to a specific person, for instance "Sorrentino," is accessing a known whistleblowing website. He further illustrated the sophistication of de-anonymization by mentioning the tendency of embedding subtle, individualized changes within email content to track

recipients (e.g., adding unique patterns, such as extra spaces in sentences, to emails), which is known as content fingerprinting or content-based watermarking. This creates a digital watermark allowing them to know exactly who received a particular version of a message if it were ever to be leaked. To combat this vulnerability, he suggested several solutions. The first line of defense is education and changes to the corporate environment to normalize access to reporting channels. This includes low-tech solutions like placing QR codes for the reporting site "in every bathroom and every entryway," making it less conspicuous for an individual to visit the page. An even more proactive approach would be to make the whistleblowing page the one that "automatically launches first thing when you open your browser so literally everyone is visiting it," completely removing any stigma or suspicion associated with the web traffic. He also emphasized the importance of point-of-contact education, noting that his own platform presents users with a pop-up containing crucial advice, such as "don't use a work computer, don't use a work network". Beyond these practical measures, he floated a potential high-tech solution involving AI. He described how an LLM (Large Language Model) could be used to process a whistleblower's report to "remove lingo" and "change certain fingerprints" that might inadvertently reveal their identity. The AI could, for example, identify unique words or phrasing patterns and "genericize" the text, adding a sophisticated layer of protection to obscure the author's identity.

Culture as a Greater Barrier Than Technology

A key counterpoint emerged from the audience, suggesting that cultural resistance is a more significant obstacle than any technological challenge. An audience member pointed out that during the drafting of the EU Whistleblower Directive, anonymous reporting was ultimately left to member states to decide upon because there was "so much pushback" and "culture backlash" against it. This stigma persists, with some jurisdictions treating anonymous reporters as "second tier whistleblowers" with fewer protections, highlighting a deep-seated cultural mistrust that technology alone cannot solve.

Using Technology to Enhance Trust and Feedback

Gaia Fiorinelli pivoted the conversation by asking if technology, rather than just posing risks, could enhance trust by giving whistleblowers a way to follow up on what happens after they make a report. While Glenn Sorrentino was initially skeptical, believing fear of retaliation is the primary concern and that chatbots might lower trust, Vivienne Brand saw potential. She noted that recent advances in generative AI are changing people's emotional responses to technology. Referencing a report which found that a major source of distress for

whistleblowers was not being informed of the outcome of their case, she argued that an automated, anonymized feedback loop via a chatbot could be a powerful tool to address this frustration.

An important asymmetry in the European Union's approach to whistleblowing was highlighted by an audience member, revealing a pragmatic and context-dependent stance on anonymity. They explained that the general EU Whistleblower Directive contains "two important deadlines": an acknowledgment of the report must be made within seven days, and feedback on the case must be provided within three months. These procedural requirements complicate anonymous reporting and provide a practical justification for resistant member states to argue against it, asking "how are we going to do it if it's anonymous so we're just not going to do it". However, the audience member pointed out a notable exception that reveals a different set of priorities: the European Commission's own policy for the enforcement of competition law. In this specific domain, the Commission "very much... are fully on support [of] anonymous reporting". This strong encouragement stems from a direct enforcement need, as the Commission is "really struggling from an enforcement perspective to be the guardian" of the market and requires "bottom-up reporting" that it cannot otherwise obtain. Vivienne Brand immediately intervened to confirm that this is not a unique situation, noting it was the "Same in Australia," where in competition law it "was just a standard thing to rely upon tips from competitors". This asymmetry demonstrates that while the general EU directive accommodates member states' cultural and procedural reservations about anonymity, the Commission itself will champion and rely on anonymous channels when the strategic need for information is great enough to overcome institutional resistance.

5. The Human in the Loop: Roles, Training, and Ethical Judgment

The roundtable's final topic, introduced by Diane Ring, focused on the critical role of the "human in the loop" within increasingly automated compliance systems. Ring asserted that even with the most advanced technology and laws, people remain indispensable. She questioned where these individuals fit within the corporate structure—in IT, compliance, or audit—and what this integration looks like. The discussion explored whether boards should create a dedicated AI compliance officer, what training is necessary for roles like auditors to interpret AI outputs, and ultimately, what metrics could be used to define the success of these new tech-powered systems.

Corporate Structure and the Human Pipeline Problem

When asked if companies should create a dedicated AI Compliance Officer, Vivienne Brand responded that the approach is highly "company dependent." While a massive, tech-heavy entity might create such a role, smaller organizations lack that luxury. The consensus, she noted, is that the responsibility must be "pervasive" throughout the organization and should not be siloed within the IT department, as the challenges are "bigger than IT". Brand also highlighted a significant "pipeline problem" : while boards and senior leaders are "upskilling like crazy" on AI, they are often starting from a low base, and the internal talent may not yet exist to fill a senior AI-focused compliance role.

The Challenge of Training and Explainability

The conversation then turned to the practicalities of training and the need for AI systems to be explainable. Vivienne Brand pointed out that organizations like the Australian Institute of Company Directors are creating targeted training courses, and the "big four" accounting firms are seizing the market opportunity to offer their own training programs.

From a technical perspective, Glenn Sorrentino argued for "responsible software development" that provides feedback at the right time. He stressed the importance of explainability, asking, "why did the AI make that decision?" and suggesting the need for "well designed robust human interfaces" that allow a user to see the system's "chain of logic". However, Douglas Arner cautioned that explainability is "not necessarily as simple from a technical standpoint as is often portrayed". Often, it requires two distinct systems: one that performs the task and a second that attempts to explain the first one's actions. Sorrentino added that even when "explainable data" is provided, it often comes in the form of "tables and spreadsheets of raw data," which creates a new "interpretability problem" requiring a technologist and a lawyer to decipher it together.

Gaia Fiorinelli grounded the discussion in regulation, detailing the specific requirements for "human oversight" under the EU AI Act. For oversight to be meaningful, a human user must be able to understand the system's capabilities and limitations, monitor its functioning, be aware of the risk of "automation bias," interpret the output correctly, decide in which cases *not* to use the AI, and ultimately be able to "intervene or stop the operation of the system". Fulfilling these requirements, she concluded, means bridging a "huge gap" in the digital literacy of all users.

Ethical Trade-offs and the Impact on Human Judgment

An audience member raised a profound ethical question regarding the cultural impact of over-relying on AI for compliance, drawing a parallel to the classic business ethics debate between a rules-based and a principles-based approach. Understanding this parallel is key to seeing the risk they highlighted.

A rules-based approach functions like a detailed instruction manual, providing employees with a comprehensive list of explicit do's and don'ts. Its goal is strict compliance with the letter of the law or policy. However, its major flaw is that it can encourage a "check-the-box" mentality, stifling personal moral judgment as employees focus on what is technically forbidden rather than what is ethically right. This fosters a culture where any action not explicitly covered by a rule is considered permissible. In contrast, a principles-based approach acts as a compass, guiding behavior through broad values like integrity and fairness. It requires employees to use their own judgment to navigate complex situations, fostering a more resilient culture where individuals take personal responsibility.

The audience member's insight was to equate an AI compliance system with the ultimate rules-based system. No matter how sophisticated, the AI is a highly advanced rule-follower, operating on the patterns it has been trained to detect. By implementing such a tool, an organization risks two dangerous cultural shifts:

The Atrophy of Human Judgment: The audience member questioned whether relying on AI risks dulling our own "observational capacities" and "moral judgment." When employees know an AI is constantly scanning for misconduct, they may subconsciously outsource their own ethical vigilance. Ethical reasoning is like a muscle; if unused, it weakens. Over time, the ability of employees to spot subtle red flags or question grey-area situations diminishes because they assume the system is handling it.

The "AI's Silence as Endorsement" Problem: The most significant risk is that the AI's silence can be misinterpreted as approval, creating a culture where anything not flagged by the AI is considered "okay" or "permitted." A novel form of misconduct that the AI has not been trained to detect will go unnoticed. Worse, even if a human employee feels something is ethically questionable, the fact that the powerful AI raised no alarm can lead them to dismiss their own intuition. The machine's perceived "all-clear" can override human moral concern, creating a massive blind spot where new types of misconduct can flourish.

The panelists agreed this was a significant concern, offering different perspectives on the problem:

Vivienne Brand saw this as a "real issue," noting that the "lack of nuance" in first-generation AI is particularly problematic because its convenience and low cost make it easy to default to. She expressed further concern that as future generative AI develops a "very convincing veneer of a nuanced and sophisticated analysis," the temptation for people to "default to it when we shouldn't" will only increase. She concluded that this reliance on AI "does tend to take us away from" the kind of "principles thinking" that is essential for robust ethics.

Gaia Fiorinelli contextualized the issue using organizational theory, explaining that tech-driven compliance models often rely on a "mechanical view of a corporation". This view treats an organization like a machine where actions are simply parts of a process to be optimized. She contrasted this with viewing the organization as a "cultural... environment," a perspective that would necessitate a different, less tech-centric approach to developing compliance programs.

Glenn Sorrentino provided a pragmatic perspective on human behavior, arguing that people are conditioned to take the path of least resistance. He argued that people will almost always "optimize for efficiency," using this analogy: "if you've ever seen a sidewalk and a path in the grass, it's because it's faster to go that way." He explained that this natural tendency to "take shortcuts when they can" is intensified in a corporate environment by external pressures like "unrealistic quotas". Whether its service agents needing to close 500 cases a week or sales teams chasing multimillion-dollar targets, the pressure to perform encourages finding the quickest route to a goal, which can mean uncritically accepting an AI's suggestions. He did note that this can depend on the person's investment in their role, suggesting that long-term service agents who see their work as a career and are dedicated to helping people will likely apply "much more oversight". To provide a concrete example of this behavior, he described his experience implementing two-factor authentication. He explained that if there is a "skip button, you're going to hit the skip button" because the user's immediate goal is not to set up security; it is to "log in to get that phone number that you need". From the user's perspective, securing the data is the company's responsibility, not theirs. This illustrates the human tendency to prioritize immediate tasks over procedural requirements, which is central to the ethical concern about over-relying on AI.

6. Final Question: Metrics for Success

Prof. Diane Ring concluded by asking each panelist what metrics they would use to determine if tech-powered disclosure systems are working successfully.

Glenn Sorrentino: He focused on a combination of operational and cultural metrics. Operationally, he suggested tracking the number of "cases closed that have been resolved" and framing success in a "financial way" by calculating the value of potential fines avoided. However, he stressed that the ultimate and most important measure is cultural: "whistleblower retention". Success, in his view, would be a system that ensures employees who report wrongdoing in good faith can not only stay in their jobs but also have a fair chance at career advancement afterward.

Vivienne Brand: Her approach centered on comparative data and benchmarking. She emphasized the critical need to establish "Baseline data before implementation". This would allow an organization to concretely measure the changes and impacts brought about by the new system. Furthermore, she advocated for collecting "industry-imparative data," which would enable the company to benchmark its performance against that of its peers.

Gaia Fiorinelli: She defined success in terms of process efficiency and regulatory integration. For her, a key metric is not just the data itself, but the system's "ability to integrate in a single procedure... different regulatory requirements". The success of a tech-powered system would be demonstrated by its capacity to achieve a "Streamlining of processes," making compliance more coherent and manageable.

Douglas Arner & Elisabetta Pietrocarlo: They both converged on system accuracy as the most crucial metric. While Elisabetta Pietrocarlo first raised the issue of managing "false positives," Douglas Arner elaborated on the critical balance required. He explained that a system is useless if it misses all the actual instances of misconduct, but it is equally useless if it generates so many false positives that "it gets completely ignored" by the organization. Therefore, the key measure of success is achieving that essential combination of high accuracy and a low false positive rate.