

‘We’re Not Doing Ethical AI, We’re Doing Legal AI’: How Compliance Shapes Responsible AI Practice

Tallal Ahmad¹, Arshia Dutta², Corey Jackson¹, Abraham Mhaidli³

¹The Information School, University of Wisconsin–Madison

²Information Security Group, Royal Holloway, University of London

³Max Planck Institute for Security and Privacy

tallal.ahmad@wisc.edu, arshia.dutta.2023@live.rhul.ac.uk, corey.jackson@wisc.edu, abraham.mhaidli@mpi-sp.org

Abstract

As AI systems increasingly shape high-stakes decisions, ethics frameworks and toolkits have been introduced to promote fairness, accountability, and transparency. Yet in practice, developers work in environments where regulatory compliance fundamentally reshapes and often displaces deliberation about ethics. Drawing on 20 interviews with AI/ML practitioners across North America, Europe, and the UK, we provide an empirical account of how compliance operates not merely as a floor for ethical practice but frequently as its ceiling, and in some cases, as a substitute that forecloses deliberation about ethics. Across our interviews, ethics work persisted in development workflows only where it could be rendered as auditable evidence embedded in existing process gates; the underlying values that could not be converted into artifacts were routinely dropped. We further show that whether practitioners resist or absorb compliance logic depends not on individual values but on structural position within organizations. Our findings have implications for how AI ethics resources are designed, how accountability is assigned across development organizations, and how AI governance scholarship should account for the gap between jurisdiction-bound regulation and cross-jurisdictional AI deployment.

Introduction

In 2019, the Dutch government’s use of AI to detect fraudulent childcare benefit claims wrongly accused thousands of families, disproportionately targeting immigrants and low-income households, ultimately toppling the sitting government (Hadwick and Lan 2021; Arts and van den Berg 2025). In the United States, Amazon abandoned an AI hiring tool after it was found to systematically discriminate against women (Dastin 2018), and biased diagnostic models have been shown to exacerbate racial disparities in healthcare outcomes (Jindal 2022). These incidents demonstrate not only the social risks of poorly governed AI systems but also the limits of abstract ethical principles when they are not meaningfully operationalized in technical workflows.

To address such concerns, researchers have developed a wide array of responsible AI (RAI) resources: toolkits for auditing bias, documentation templates such as datasheets and model cards, and frameworks for articulating fairness

and accountability (Sandvig et al. 2014; Gebru et al. 2021; Mitchell et al. 2019; Crisan et al. 2022; Madaio et al. 2020; Kroll 2021). Throughout this paper we distinguish between two types of resources. AI ethics *frameworks* are high-level documents that articulate principles such as fairness, accountability, or transparency. AI ethics *toolkits*, by contrast, translate those commitments into practical methods, metrics, and checklists that developers can apply directly to models and datasets. This literature has largely concentrated on two strands: articulating responsible AI measures and frameworks, and evaluating the usability and effectiveness of existing toolkits (Wong, Madaio, and Merrill 2023; Atkins, Badrie, and Otterloo 2021). While valuable, these approaches often treat responsible AI tools in isolation, abstracted away from the organizational and regulatory environments that shape how practitioners actually use them.

By contrast, scholarship in AI governance emphasizes regulation and compliance as central to how AI is shaped in practice. Legal requirements such as the EU AI Act (European Parliament and Council of the European Union 2024), GDPR (European Parliament and Council of the European Union 2016), and U.S. sector-specific mandates from the FDA (U.S. Food and Drug Administration 2025) and HHS (U.S. Department of Health and Human Services 2025) create powerful incentives for organizations to prioritize compliance above all else (Butt 2024; de Almeida, dos Santos, and Farias 2021; Lau et al. 2023). For many organizations, regulatory adherence is not a parallel consideration but the baseline condition for product development, the “priority number one” against which all other design and ethical choices are negotiated. Yet far less is known about how compliance and ethics intersect in the day-to-day workflows of AI practitioners, how toolkits and frameworks are taken up within compliance-shaped development environments, and what the structure of that intersection means for AI governance.

To investigate these dynamics, we conducted a qualitative study with 20 AI/ML developers and practitioners working across Europe, North America, and the UK, guided by the following research question:

- How do AI practitioners translate external regulatory requirements into internal frameworks, toolkits, workflows, and compliance evidence, and what tensions emerge between regulatory compliance and ethics work?

Our findings reveal a more differentiated picture than prior accounts of compliance-as-constraint suggest. Compliance operated in two distinct modes across our data. In the first, *sequencing*, compliance comes first and ethics work gets whatever time and resources remain, since regulatory obligations are met before any voluntary ethical exploration can begin. In the second, *substitution*, compliance actively displaces ethical deliberation altogether. As one participant put it, following GDPR simply “removes my need to deal with” ethics. Both patterns produce marginalised voluntary ethics work, though substitution has more consequential effects on ethical deliberation. Within these compliance-first workflows, ethics persisted only where it could be rendered as an auditable artifact embedded in an existing process gate (bias evaluations, model cards, explainability justifications), while values that could not be converted into evidence were routinely dropped. We further show that whether practitioners resist or absorb compliance logic depends not primarily on the strength of their individual values but on their structural position: founders, freelancers, and consultants with exit options pushed back; employees embedded in organizational hierarchies did not. Finally, we document how compliance cultures differ across jurisdictions (EU, US, Canada) in ways that have direct implications for the design of international AI governance frameworks.

This paper contributes an empirical account of how compliance functions as the operative logic of ethics work in AI development, identifying the conditions under which ethics either persists or gets displaced and the structural and jurisdictional factors that shape that outcome. We show that because compliance functions as both floor and ceiling for ethical practice, the specificity of regulatory requirements determines whether the resulting practice is substantive or symbolic. Regulators are therefore not only setting minimums but shaping the operative scope of responsible AI development itself. Our findings carry implications for how ethics resources are designed, how accountability is distributed across development organizations, and how AI governance scholarship engages with the gap between jurisdiction-bound regulation and cross-jurisdictional AI deployment.

Related Work

We synthesize literature on (1) AI ethics frameworks and toolkits and their adoption barriers, and (2) AI governance mechanisms and compliance labor, to clarify our terminology, position our contribution relative to prior empirical work, and motivate our focus on how practitioners operationalize ethics under compliance-shaped constraints.

AI Ethics Toolkits & Frameworks

Despite the proliferation of both, a persistent gap remains between ethical commitments and day-to-day practice. Most frameworks articulate high-level values but provide limited process-level guidance for integrating them into development workflows (Robles Carrillo 2020; Rakova et al. 2021). Empirical and scoping studies document that existing RAI guidelines are experienced as abstract, difficult to reconcile with business incentives, and underspecified in terms

of success metrics and responsibilities (Morley 2021; Sadek et al. 2025). At the toolkit level, misalignment with real-world workflows is a recurring finding: many toolkits are not designed with active practitioner participation, resulting in limited organizational uptake (Madaio et al. 2020). Practitioners frequently repurpose toolkits for cross-functional communication and internal documentation rather than as normative guidance (Yildirim et al. 2023), and adopt piggy-backing tactics to attach fairness work onto existing institutionalized procedures in the absence of dedicated resources (Deng et al. 2023). Ethics-focused resources often lack the ecologically detailed, action-focused scaffolding practitioners need to move from awareness to concrete action (Gray et al. 2023). Beyond toolkit usability, recent empirical work documents that the burden of making ethics work in organizations is individualized and politically fraught. Studies of ethics entrepreneurs, privacy champions, and AI practitioners show that those who take up ethics responsibilities must navigate indifferent cultures, misaligned incentives, moral stress, and diffuse responsibility across distributed teams (Ali et al. 2023; Tahaei, Frik, and Vaniea 2021; Rattay et al. 2025; Drage, McInerney, and Browne 2024; Orr and Davis 2020; Widder and Nafus 2023). While this work powerfully documents the affective and organizational costs of ethics labor, it has yet to fully examine how ethics frameworks and toolkits become entangled with emerging AI regulatory regimes in practitioners’ day-to-day workflows, leaving a critical gap in understanding how to make these resources actionable under compliance pressures.

AI Governance and Compliance Labor

AI governance spans hard law (binding legislation such as the EU AI Act (European Parliament and Council of the European Union 2024), GDPR (European Parliament and Council of the European Union 2016), and U.S. sector-specific mandates from the FDA (U.S. Food and Drug Administration 2025) and HHS (U.S. Department of Health and Human Services 2025)) and soft governance mechanisms including industry standards, certifications, and audits (Floridi et al. 2018; Kroll 2018; Lewis et al. 2020). While self-regulation at the organizational level has been deemed insufficient to mitigate AI deployment risks (Roski et al. 2021), soft governance mechanisms such as ISO/IEC standards and the IEEE P7000 series translate governance principles into auditable benchmarks for transparency and traceability (Kroll 2021; Larsson and Heintz 2020). These hard-and-soft-law regimes make evidence actionable by requiring traceability, documentation, and explanation, pressing teams to organize work around artifacts and process gates.

This landscape has expanded rapidly. ISO/IEC 42001 established a certifiable management system standard for organizational AI governance, complemented by ISO/IEC 42005 on AI system impact assessment and ISO/IEC 42006 on requirements for certification bodies, with ISO/IEC 42003, offering implementation guidance for 42001, currently under development (Seet 2026). Certification against these standards is increasingly requested in enterprise procurement. In the EU, CEN/CENELEC is developing harmonised standards for the AI Act; once published and assessed by notified

bodies, conformity with them will confer a presumption of compliance, effectively converting standards adherence into legal defensibility (CEN-CENELEC 2026). In Canada, the Treasury Board’s Directive on Automated Decision-Making requires federal institutions to complete an Algorithmic Impact Assessment before deploying an automated decision system, though it does not bind private-sector developers; the proposed Artificial Intelligence and Data Act, introduced as part of Bill C-27, died on the order paper when Parliament was prorogued in January 2025 and has not been reintroduced (Innovation, Science and Economic Development Canada 2023; Treasury Board of Canada Secretariat 2025). In the United States, sub-federal instruments such as New York City’s Local Law 144, which requires bias audits of automated employment decision tools, illustrate how obligations arrive through fragmented municipal and state channels rather than a unified federal regime (Wright et al. 2024). Together these instruments extend the evidence-production demands described above into new organizational and jurisdictional settings.

Empirical studies of practitioners implementing these regimes reveal that compliance is not a straightforward translation of rules into practice but a situated, often contested form of labor. Grover describes developers as street-level bureaucrats who creatively interpret vague GDPR and CCPA requirements, frequently treating compliance as a one-off risk mitigation project rather than an ongoing obligation (Grover 2024). Waldman’s analysis of privacy law implementation argues that organizations often produce symbolic structures policies, trainings, and checklists, that satisfy formal expectations while offering limited substantive protection for users (Waldman 2019). Related work shows how responsibility for ethical and legal alignment is fragmented across supply chains and organizational hierarchies, with individual engineers shouldering disproportionate and precarious responsibility within fast-paced development cultures (Tahaei, Frik, and Vaniea 2021; Ali et al. 2023; Drage, McInerney, and Browne 2024; Orr and Davis 2020; Widder and Nafus 2023). Recent scholarship further emphasizes that compliance functions not only as a legal constraint but as a force that shapes workflows, technical design choices, and organizational accountability (Hacker et al. 2022). Beyond binding rules and voluntary standards, governance operates through operational mechanisms that demand evidence of responsible practice even when that evidence risks becoming largely symbolic (Waldman 2019; Grover 2024). Prior work identifies three levers mapping to day-to-day development: data governance (lawful collection, provenance, access control), model transparency and interpretability (traceability, explanation duties), and system-level process controls (documentation, audits, risk registers) (Kroll 2018; Larsson and Heintz 2020; Lewis et al. 2020; Kroll 2021). In practice, these mechanisms materialize as artifacts (datasheets, model cards, bias-audit reports, explainability justifications) and as process gates that condition release decisions.

Synthesis and contribution. Prior work richly catalogs principles, frameworks, toolkits, and standards, and increasingly documents how ethics and compliance work are frag-

mented, individualized, and sometimes reduced to symbolic exercises (Morley 2021; Sadek et al. 2025; Waldman 2019; Grover 2024; Widder and Nafus 2023; Ali et al. 2023). However, this scholarship pays less attention to how ethics frameworks and toolkits are *operationalized as evidence practices* in day-to-day AI development. Less still is known about the conditions under which ethics either persists or gets displaced by compliance logic and how those conditions vary by structural position and regulatory context across jurisdictions. We contribute an empirical account of these operational dynamics across regions, identifying the mechanism by which ethics is converted into compliance evidence, what that conversion drops, and the structural conditions under which practitioners resist rather than absorb compliance logic. In doing so, we contribute to empirical accounts of AI ethics-in-practice, responsible AI resource design, and the critique of compliance-driven governance as a sufficient substitute for substantive ethical judgment.

Methodology

Interviews

We conducted semi-structured interviews designed to elicit how AI/ML practitioners encounter, interpret, and operationalize ethics frameworks and toolkits within compliance-shaped workflows. Following informed consent and role/background questions, core prompts probed how frameworks and toolkits fit into development processes, how legal and regulatory requirements are translated into technical obligations and artifacts, and what organizational resources and incentives shape ethics work. The protocol used transitions and optional probes to adapt ordering and depth while keeping coverage consistent across interviews. A note-taker was present for all sessions.

All interviews were conducted remotely over Zoom lasting between 45 and 75 minutes. Participants received \$70 USD compensation reflecting the specialized nature of AI/ML development work. Our University’s Institutional Review Board approved the study (IRB #2024-1554).

Recruitment and Participants

We recruited participants using a combination of professional networks and open calls in online communities, including LinkedIn groups focused on AI development and governance and Reddit communities including r/Machine-Learning and r/datascience. Consistent with prior observations that recruiting skilled software developers through conventional channels can be challenging (Kaur et al. 2022), these posts yielded limited responses. We therefore expanded recruitment to freelance platforms Upwork and Fiverr, which prior work has shown provide access to participants with substantial software development experience and are suitable for studies requiring technically skilled participants (Kaur et al. 2022). We explicitly targeted independent practitioners and consultants and used project descriptions and profiles to pre-screen for relevant AI/ML experience. We also employed snowball sampling, asking interviewees to connect us with other practitioners.

Potential participants completed a screening questionnaire covering current role, years of experience with AI technologies, domains and lifecycle stages worked in, familiarity with specific AI ethics frameworks and toolkits, and organizational location and size. To be eligible, respondents had to be at least 18 years old, consent to audio recording, hold a role in AI or software development (software developer, AI practitioner, data scientist, researcher, or manager/team lead), report prior experience with AI technologies, have participated in at least one stage of the ML/AI development lifecycle, and work for an organization located in the EU, UK, US, or Canada. Of 32 respondents, 12 did not meet these criteria and were not interviewed. We conducted 20 interviews with AI/ML developers and practitioners based in Europe, North America, and the United Kingdom. While the sample included many consultants and startup founders, three participants were employed at large technology companies. We sought diversity across roles (engineers, consultants, founders), sectors (healthcare, NLP, generative AI, finance), and organizational scales (startups, SMEs, large firms), rather than demographic diversity per se. Table 1 provides a full overview of participant roles, locations, and experience.

Inductive Thematic Analysis

We analyzed the interview data using an inductive thematic analysis approach (Braun and Clarke 2006). The research team read through all transcripts and iteratively generated codes based on what participants discussed, allowing salient concepts, practices, and challenges to surface without being constrained by the interview structure. Data collection and analysis proceeded iteratively; by approximately 18 interviews, new data no longer introduced substantively new categories, and we concluded recruitment at 20 participants having reached thematic saturation.

All interviews were transcribed with Otter.ai, manually reviewed for accuracy, and imported into NVivo for analysis. Following Braun and Clarke's six-phase approach (Braun et al. 2022), two authors independently coded portions of the data, met regularly to compare interpretations, and refined the coding scheme through discussion. We consolidated overlapping codes into broader themes spanning seven major categories: personal ethical views, the framework-toolkit distinction, limitations of current resources, regulatory compliance, operationalization of ethics, country-specific adaptation, and organizational friction. These themes did not map one-to-one with the interview guide but reflected patterns cutting across structured questions and emergent issues, grounded in participants' lived experiences.

Researcher Reflexivity

Our analytic lens is shaped by training in information science, computer science, and HCI, and by engagement with the literature on responsible AI, governance, and compliance. This predisposed us to notice how regulation, documentation, and audit requirements structure development work. To avoid overweighting this perspective, we used open, non-leading prompts, explicitly invited counterexamples, and treated tensions and contradictions as core analytic

material rather than noise. We emphasized confidentiality with participants, welcomed omission of sensitive details, and reiterated that there were no right answers.

Recruitment through English-language online communities and freelance platforms likely privileges practitioners who are public-facing, comfortable discussing their work online, and situated in contexts where compliance and documentation are salient. This shapes transferability: our findings travel most readily to similar settings, with experiences in other geographies, languages, or organizational forms particularly large, and highly structured in-house teams potentially underrepresented. Our focus on North America, Europe, and the UK was deliberate rather than incidental: these regions are the primary sites of AI legislation and responsible AI framework development, and understanding how compliance shapes ethics work in these dominant contexts is a necessary precondition for comparative work across regulatory environments with different levels of AI governance maturity. Throughout analysis we engaged in reflexive memoing, actively sought disconfirming evidence, and revised the codebook when categories reflected our priors more than participants' accounts.

Findings

Ethics as Practice

Participants framed ethics as a mix of organizational requirements and personal commitments, with individual convictions sometimes pushing practices beyond what policies demanded. P1 (UK) advocated explicitly informing users about new data collection in a mobile app even when company policy did not require disclosure: *"for this mobile application... it needs to say what kind of new additional data we are taking... and at least ask them for the permission."* Others described ethics in terms of personal red lines. P13 (Canada) rejected contracts to build AI for weapons or mass layoffs, calling such work "not only... not ethical, it's immoral," while P7 (Turkey) refused a depression-related dataset, questioning whether consent from vulnerable individuals could be meaningful even when legally obtained. Yet participants consistently emphasized the tension between personal conviction and institutional incentive. P6 (US) observed that in most firms "there's not really anyone whose job it is to think about the ethical side of things," and P18 (UK) offered a blunter diagnosis: "what you need to do (at work) is save your job, not get to the truth." Several described ethics as an ongoing reflexive practice. P12 (US) spoke of "trying to wake up tomorrow less wrong", while others linked their ethical commitments to prior experiences with data privacy (P17, US) or concerns about labor displacement (P11, US).

Some participants sat explicitly at the seam of ethics and law. P16 (US), co-founder of an AI compliance consultancy, framed ethics as architectural rather than remedial: *"if you don't approach your product with (AI) ethics at the outset, you can't get the cat back in the bag."* These accounts present ethics as both an organizationally structured requirement and a site of individual negotiation. Importantly, practitioners who successfully pushed ethical commitments be-

ID	Jurisdiction	Role(s)	Field / Specialization	Experience
P1	UK	Full Stack Engineer	Energy, Web Development	< 1 year
P2	US	PhD Candidate	Human-Centered AI, AI Ethics	~3 years
P3	EU	Engineer	Home Appliances, Machine Learning	2 years
P4	UK	ML Engineer, AI Developer	Generative AI	> 1 year
P5	US	Engineer / Developer	AI Hardware, Text-to-Image Gen.	3 years
P6	US	ML Engineer	NLP	~5 years
P7	Turkey	Consultant, Freelance R&D	NLP	> 2 years
P8	EU	AI Developer	Machine Learning, Deep Learning	> 1 year
P9	UK	Data Engineer	Data Engineering, AI Training	3 years
P10	US	Predictive Modeler	Pharmaceuticals	> 5 years
P11	US	AI / ML Engineer	Large Language Models (LLMs)	~2 years
P12	US	Founder, Consultant	XR, Financial Services, AI Ethics	Extensive
P13	Canada	Startup Founder	Ethical AI, Operations Research	~7 years
P14	UK	Manager & AI Consultant	Logistics, E-commerce, AI Strategy	Extensive
P15	Canada	Founder, Consultant	Healthcare, Medical Devices, VR	15 years
P16	US	Co-founder, Consultant	Data Privacy, AI Compliance, Law	< 5 years
P17	US	AI Project Manager, Consultant	AI Ethics	~3 years
P18	UK	Founder, Speaker	AI for Social Good, AI Ethics	~10 years
P19	US	Researcher	AI Security & Safety	1 year
P20	EU	Entrepreneur, Consultant	AI Automation & Development	2 years

Table 1: The participants in our research study.

yond compliance minimums shared a common structural feature: they occupied positions of relative independence as founders, freelancers, or consultants rather than salaried employees subject to organizational performance pressures. P7’s dataset refusal, P13’s contract rejection, P12’s project exit, and P20’s insistence on AI disclosure to users were all made from positions where the practitioner retained discretion over their own work. By contrast, participants embedded in organizational hierarchies described ethics work as precarious, informal, and subordinated to release timelines. This pattern, structural position as the primary condition enabling ethical pushback, not the strength of individual values recurs throughout our findings and has direct implications for how AI governance frameworks should assign accountability.

How Practitioners Understand Frameworks and Toolkits

Participants generally distinguished frameworks and toolkits along a single axis: abstraction versus practice. Frameworks were described as high-level, principle-based documents produced by governments, standards bodies, or professional organizations; toolkits as developer-facing resources offering tangible methods, metrics, and libraries that can be integrated into technical workflows. As P4 (UK) put it: “*AI ethics frameworks are kind of like high-level principles and guidelines... but responsible AI toolkits are technically more practical... hands-on implementation that we make sure.*” P15 (Canada) drew perhaps the sharpest version of this line, describing frameworks as qualitative and aspirational and toolkits as quantitative instruments for evaluating, benchmarking, and demonstrating compliance, pointing to the NIST AI Risk Management Framework (AI 2023) and FDA guidance on AI-enabled device software (U.S. Food and Drug Administration 2025) as examples.

Some participants were uncertain how this distinction applied specifically to AI ethics (P7, Turkey; P6, US), while others saw the two as complementary rather than competing: frameworks to articulate values and open questions early in a project, toolkits to operationalize those values in a specific domain later on (P3, EU; P12, US; P15, Canada). The more consequential question in practice, however, was not whether participants could define the distinction, but which resources they actually reached for, and why compliance demands, rather than ethical principles, typically determined that choice. We examine this in the sections that follow.

Limits of Current Ethics Resources

Participants expressed broad skepticism about the practical applicability of existing ethics toolkits and frameworks. Many were familiar with high-profile initiatives (AI Fairness 360, datasheets, model cards, the What-If Tool) but described them as too vague, fragmented, or research-oriented for everyday development work. P3 (EU) captured a recurring tension: “*for research, it’s useful. I have doubts for industry,*” explaining that tools like AIF360 could not be easily integrated into tightly coupled production pipelines. P4 (UK) identified a related problem with standardization: “*different toolkits measure fairness differently... if you use the same framework under two different ethics frameworks, you tend to get different results*”, a problem that intensified when monitoring models under shifting data distributions.

Beyond usability, participants identified a deeper dysfunction: ethics resources were frequently repurposed as instruments of organizational self-presentation rather than constraints on practice. P18 (UK) described their function as protecting the practitioner’s position rather than the people the tools nominally serve, a tension we return to below. In regulated sectors, this symbolic use was compounded by domain mismatch: generic toolkits often failed to account for

sector-specific compliance obligations, forcing practitioners to reconcile overlapping and inconsistent requirements. These limitations were not experienced as isolated technical shortcomings. Rather, they reflected a structural condition: when ethics resources do not map onto the compliance obligations that actually gate deployment, they get set aside. For most participants, the framework that reliably structured their work was not an ethics toolkit but a regulatory one. We turn to this compliance baseline next.

Compliance-First Workflows as the Baseline Condition

Across domains, participants described regulatory compliance as the primary driver of how AI systems were designed, documented, and deployed. Compliance referred to a mix of legal, sectoral, and industry-mandated requirements (consumer protection tests, data protection law, industry certifications, and emerging AI-specific regulations) and meeting these obligations was “priority number one,” a prerequisite that shaped design decisions, workflows, and timelines before other considerations could be addressed.

In safety-critical and consumer-facing domains, compliance frameworks functioned as explicit gatekeepers. P3 (EU), working on a tumble dryer humidity sensor, described how consumer protection tests structured the entire development process: *“we should do everything we could to please the certification bodies. . . . Once that is done, then you can take care of everything else.”* Similarly, participants in financial and healthcare settings described extensive documentation and review processes tied to GDPR, the EU AI Act, FDA and HHS guidelines, and sectoral regulators. P15 (Canada) was unambiguous about the hierarchy: “between a guideline and a statutory requirement, we always default to statutory requirement.”

Crucially, compliance did not merely sequence ethics work; in several accounts it displaced it entirely. P14 (UK), a logistics and e-commerce AI consultant, described his approach to data ethics plainly: *“I’ve just gone, okay, I’m going to follow GDPR primarily. . . . that removes my need to deal with it.”* This is a stronger dynamic than compliance-as-ceiling: GDPR compliance was deployed instrumentally as an escape mechanism from ethical deliberation. Where P3 describes compliance as a gate that ethics work must pass through first, P14 describes compliance as a substitute that renders ethics work unnecessary. Both patterns, sequencing and substitution, were present in our data, and both produce the same outcome: voluntary ethical exploration is marginalised.

Compliance was also embedded in contractual and organizational structures that left practitioners little design discretion. P9 (UK), a data engineer, described projects beginning not with an ethics workflow but with a client-supplied compliance pack, a checklist specifying which fields must be masked, which machines could store data, and which validation steps must be logged before data could move downstream. His work consisted of wiring these constraints into the pipeline and producing evidence that each step had passed. This made shipping faster and audit-ready but defined both the boundaries and the sequence of work before

any ethical judgment could be exercised.

Taken together, these accounts reveal compliance operating not merely as a floor for ethics work but, in practice, frequently as its ceiling and sometimes as its replacement. The following section examines how, within these compliance-shaped workflows, ethics is nonetheless enacted: not as an independent set of commitments but as a practice of producing auditable evidence.

Operationalizing Ethics as Compliance Evidence

Within compliance-first workflows, ethics was rarely implemented as an independent set of commitments. Instead, participants described operationalizing ethics by producing artifacts and making model and workflow choices that could serve as evidence of responsible practice for internal review, clients, and potential audits. Understanding how this conversion happens, and what it drops, is the central analytical contribution of this paper.

The conversion process followed a recognizable sequence across contexts. At the input stage, practitioners received compliance obligations either as regulatory requirements, client-supplied documents, or organizational modules. P9 (UK) described this entry point precisely: *“the framework was provided. . . and my job is just to integrate those frameworks. . . they had a [Jira] task list, we have to make sure that we are fulfilling those.”* His role was implementation, not ethical design. The checklist arrived as a deliverable, each item was wired into the pipeline, and completion of each item generated a logged artifact. No discretion over what counted as sufficient, who decided, or what the requirements excluded was exercised at the developer level. P4 (UK) described a parallel dynamic on the client-facing end: documentation of all modeling choices was assembled into a report delivered as a procurement artifact *“they also asked for a detailed report so that they can re-evaluate”*, with the client explicitly requesting evidence of responsible practice rather than engaging with the substance of design decisions.

At the model architecture stage, compliance demands shaped technical choices in ways that are rarely visible in toolkit-focused accounts. P6 (US) described favoring boosted tree models over deep learning specifically because *“explainability was a high priority. . . if they get audited, or if the government says, ‘why you made this decision,’ you need to be able to provide that.”* Accuracy was not abandoned, but interpretability as an audit-readiness requirement constrained the solution space before any ethics toolkit was consulted. Similarly, P11 (US) described running bias checks before LLM training not primarily as an ethical commitment but as an economic one: *“once you train the LM, it’s hard to unchange it. . . and it costs a lot of money to change the LM after you have trained it.”* The timing of ethics work was determined by cost asymmetry, not by principled upstream integration.

In more heavily regulated contexts, the evidence-production process involved multiple institutional layers. P10 (US), working in pharmaceutical AI at a major cancer center, described a sequential approval chain: proposed use cases were reviewed first by a direct supervisor, then

by a dedicated AI legal counsel, then by a cybersecurity team, with business associate agreements negotiated with vendors like AWS before any deployment could proceed. *“They would say, yes you can, no you may not. And if they say, no you may not, we have to go back.”* Each layer generated a review artifact; the chain as a whole constituted the compliance evidence that authorized deployment. Ethics in this context was not a set of values practitioners brought to their work but a procedural property of the approval chain itself. P15 (Canada) made the cost structure of this process explicit: *“I have a number sign [dollar cost] next to every requirement. . . how much does it cost for me to build it and to maintain it.”* Compliance decisions were resource allocation decisions, with ethical requirements triaged by fiscal impact.

These sequences illuminate what the conversion from ethics to compliance evidence systematically drops. First, ethical considerations that cannot be rendered as checkable artifacts disappear from the workflow entirely. P6 (US) noted that annual compliance trainings GDPR, anti-money laundering were treated as boxes to check rather than occasions for substantive engagement: people “just kind of rush through it,” producing a completed-course artifact with no behavioral change. Second, the conversion can degrade the very technical quality it claims to protect. P14 (UK) acknowledged that removing personally identifiable data to satisfy GDPR compliance may have reduced model performance: *“has that meant that my models are not as good? Quite possibly.”* The compliance artifact anonymized data, documented process, satisfies auditors while the underlying system becomes less accurate. Third, and most directly, artifacts that satisfy gatekeepers do not necessarily protect affected communities. P18 (UK) named this plainly: ethics tools are *“good to make sure you keep your job. . . but useless to the communities they supposedly are out to help.”* The dual-legibility problem artifacts readable by auditors but not meaningful to end users is not incidental to compliance-driven ethics work; it is structural.

Across these accounts, ethics persisted only where it could be rendered as an auditable artifact embedded in an existing process gate. Values that could not be converted into evidence, and practitioners without the structural position to insist on their inclusion, were routinely displaced.

Localizing Compliance Across Jurisdictions

The evidence-production dynamics described above were not uniform but varied substantially across jurisdictions, shaped by the character of local regulatory regimes, enforcement cultures, and the maturity of AI-specific governance. Compliance, in other words, is not a single condition but a locally constructed one.

In Europe, compliance was framed as a regulation-driven mandate with clear enforcement mechanisms and a known roadmap of progressive restrictions. P20 (EU) described planning work around both current GDPR obligations and the staged rollout of the EU AI Act: *“They started with phase one, which is some limitations, then phase two is more limitations. . . it’s not that bad now, but it’s going to be in two years.”* This temporal clarity shaped tooling decisions

upstream of model design P20’s team excluded common platforms that did not meet GDPR requirements, including widely-used US-built tools, from their development stack entirely. P3 (EU) noted that the AI Act had made ethical assessment a formal market-access condition for high-risk AI systems, meaning that for European practitioners, ethics evidencing is not voluntary but legally required for deployment. P8 (EU), working in the automotive sector, described structured in-house GDPR compliance training provided to all engineers a qualitatively different model from the annual checkbox training described by US participants.

The United Kingdom presented a distinct post-Brexit position. Practitioners retained GDPR in domestic law but operated under a softened enforcement standard P14 (UK) noted that UK GDPR incorporates a “reasonable man” defense that gives practitioners more interpretive latitude than the EU version. ISO 27001 was used as a GDPR-equivalent signal to clients: *“ISO 27001 says you must anonymize data after a certain period, which is the same as GDPR. . . there are ways of kind of showing [compliance]”* (P14, UK). P18 (UK) described the ICO’s advice hotline as a practical resource, and noted that UK practitioners apply a “jigsaw method” for PII if identity can be assembled from data fragments, it counts as identifiable regardless of whether any single field is sensitive. This represents a more stringent PII standard than US practice, but softer institutional enforcement than the EU.

In the United States, participants described a patchwork, industry-led landscape where compliance was driven less by regulatory force than by fear of lawsuits, reputational harm, or sector-specific pressures. P12 (US) summarized the motivational structure: *“fear when a lawsuit breaks, fear of being that next company is a big thing. Fear of not being able to say that you tried is a big thing.”* Fines were treated as a cost of doing business rather than a structural deterrent, and P16 (US) acknowledged that full compliance was “actually impossible” given the rapidly changing and fragmented state-level legislative landscape. By early 2025, this environment had deteriorated further: P12 observed that “the regulatory picture in the United States has become severely compromised,” accelerating the tendency among US practitioners to index on EU frameworks as de facto global baselines even absent legal obligation to do so.

Canada occupied a liminal position: regulatory intent was clear but operationalization was not, pending the development of jurisprudence. P13 (Canada) described the situation plainly: *“we know the intent, we understand the intent, but it’s vague. . . we’re waiting for the first cases so that we better understand what Canada is wanting from the AI companies.”* In the absence of settled law, ethics committees filled the governance vacuum but P13 reported that this produced disproportionate burdens on small companies, with a one-year ethics committee review process for a one-year project, while large companies faced minimal accountability. Practitioners in Canada monitored EU regulatory developments as proxies for where domestic law was likely to land. Given the small Canadian subsample (n=2), we treat these as illustrative observations rather than characterizations of Canadian practice.

Turkey represented the study's most peripheral regulatory context. P7 (Turkey) described awareness of KVKK, Turkey's data protection law, as general rather than specific: "*we only know that we need to be careful about privacy.*" In practice, compliance was managed at the contract stage by project managers rather than integrated into development workflows, with data handling obligations specified in client agreements rather than through any ethics framework or toolkit. This contractual model compliance as negotiated terms rather than regulatory infrastructure produced a different kind of evidence artifact: the signed contract rather than the audit report.

Together, these accounts demonstrate that compliance-driven ethics evidencing is not a universal practice but a jurisdictionally differentiated one. Across the EU and US, the two contexts with sufficient participant coverage in our sample we observe substantively different framings: ethics-as-legal-obligation with auditable market-access gates in the EU, and ethics-as-liability-shield with reactive, documentation-oriented compliance in the US. Our smaller Canadian subsample suggest further dimensions of variation that warrant dedicated study.

Discussion

Ethics-as-Compliance in Everyday Development

A central contribution of this work is to surface the gap between how ethics toolkits and frameworks are depicted in the responsible AI literature and how they are actually taken up in practice. Prior work often presents toolkits as stand-alone technical remedies capable of embedding values like fairness or transparency into design, and frames ethics as an aspirational, principle-led practice (Holstein et al. 2019; Madaio et al. 2020; Morley et al. 2020). In our study, ethics was far more often enacted as *compliance work*: producing auditable artifacts and satisfying regulatory or contractual checklists (Hacker et al. 2022). This resonates with recent accounts of privacy and AI practitioners whose moral labor is channeled through organizational compliance regimes, boundary constraints, and epistemic hierarchies (Grover 2024; Waldman 2019; Ali et al. 2023; Rattay et al. 2025; Widder and Nafus 2023; Orr and Davis 2020).

Participants described datasheets, model cards, and fairness audits less as free-standing ethical commitments and more as paperwork to show that a system could withstand scrutiny. Artifacts that generate auditable evidence risk assessments, bias-audit reports, certification test results traveled far within organizations, while aspirational principles not tied to a required checkpoint generally did not.

The sequencing and substitution patterns described above are best understood as analytic distinctions rather than mutually exclusive organizational types. The same organization may sequence ethics work on one project and substitute compliance for it on another, and individual practitioners moved between the two depending on client demands and time pressure. What separates them is not the presence or absence of ethical concern but whether compliance is treated as a threshold that ethics work must clear or as a warrant that ethics work has already been done. Substitution is the

more consequential framing because it forecloses deliberation rather than deferring it, though our data do not support claims about which pattern is more prevalent.

This ordering reframes the ethics-compliance relationship in a way normative discussions rarely acknowledge. Normative accounts often assume that ethical practice should exceed legal minimums (Gardner 2010). Our participants treated regulatory and contractual demands as the *floor* and, in practice, frequently the *ceiling* of what counted as responsible AI. If compliance is both floor and ceiling, then the design of compliance requirements is the effective design of ethical practice regulators are not merely setting minimums, they are setting the entire operative scope of what responsible AI development means in day-to-day work. Toolkit-supported practices such as bias checks, documentation, and explainability reporting persisted because they yielded durable, reviewable traces (Horstmann et al. 2024). Voluntary beyond-compliance efforts were rare unless mappable onto a checkable release step a pattern consistent with checkbox cultures and symbolic compliance documented in privacy and AI governance (Balayn et al. 2023; Waldman 2019; Sadek et al. 2025).

Power, Position, and the Limits of Ethical Practice

Our findings challenge assumptions of universality that often accompany ethics toolkits in the literature (Morley et al. 2020; Qiang, Rhim, and Moon 2024; Madaio et al. 2020; Wong, Madaio, and Merrill 2023). Participants described a reality in which these tools frequently fail when confronted with local regulatory, cultural, or domain-specific requirements. P4, working in healthcare, noted that IBM's AIF360 package did not align with sector-specific compliance expectations, making its outputs difficult to defend in audits. P18 critiqued U.S.-centric demographic categories particularly race as meaningless or misleading outside American contexts. For many developers, responsible AI was never a universal blueprint; it had to be reworked for the jurisdiction, risk profile, and enforcement environment at hand. While prior work documents that practitioners routinely localize and customize ethics resources under organizational constraints (Madaio et al. 2024; Varanasi and Goyal 2023; Rattay et al. 2025), our contribution is to identify why: toolkit universality fails not because of poor design alone, but because compliance obligations are themselves jurisdictionally fragmented, and no universal toolkit can map onto regulatory requirements that differ fundamentally across legal contexts.

Two forms of power shaped what persisted as ethics in practice: organizational leverage and epistemic authority. Ethical stances from P1's push for disclosure to outright refusals by P13, P7, and P12 became actionable only when practitioners held structural independence. Founders, freelancers, and consultants could refuse contracts, exit projects, or insist on disclosure norms precisely because their structural position gave them discretion over their own work. Practitioners embedded in organizational hierarchies, by contrast, described ethics work as precarious, informal, and subordinated to release timelines, echoing research on ethics entrepreneurs and boundary-spanning practitioners

(Ali et al. 2023; Widder and Nafus 2023). This pattern is analytically important: it is not the strength of individual values but structural position that determines whether ethics resists or gets absorbed by compliance logic. P7's dataset refusal, P13's contract rejection, and P12's project exit were all enabled by independence, not by superior ethical commitment.

At the same time, what counted as real ethics was filtered through epistemic regimes that privileged quantified objectivity. Artifacts like model cards, audits, and structured risk reports carried more weight than practitioners' embodied or contextual insights unless those insights were translated into measurable outputs (Widder 2024; Orr and Davis 2020). Organizational power set the bounds of possible action, while epistemic authority specified the forms of evidence that could pass those bounds. Our contribution is to make this mechanism concrete in compliance-shaped workflows: values must be rendered as obligations, and obligations as auditable evidence, to persist in day-to-day development.

The dual-legibility problem identified in our findings has a structural corollary. Ethics artifacts function simultaneously as accountability mechanisms and reputational shields (Waldman 2019; Grover 2024; Balayn et al. 2023), which means the communities most affected by AI systems are systematically excluded from the gatekeeping function those artifacts serve.

These findings also reframe how ethics resources should be evaluated. The responsible AI literature typically assesses toolkits against criteria of usability, principle coverage, and workflow integration (Wong, Madaio, and Merrill 2023; Madaio et al. 2020; Morley 2021). Our data suggest a different evaluative lens is needed: if toolkits function primarily as compliance infrastructure in practice producing artifacts that travel to auditors rather than constraints that change design decisions then evaluating them only for usability or normative completeness misses how they actually operate. A toolkit that scores well on usability but produces artifacts legible only to gatekeepers and not to affected communities may perform worse on the dimensions that matter most. Two cases from our data make the point. P18's toolkit satisfied the auditor while leaving the affected community unprotected, and P6's annual compliance trainings produced completion certificates with no behavioral change. Usability evaluations that measure practitioner experience with a toolkit would rate both positively; an evaluation focused on substantive ethical outcomes would not. Evaluation frameworks for ethics resources should therefore assess whether artifacts produced are meaningful to affected stakeholders, whether toolkit use changes design decisions or merely documents them, and whether compliance evidence produced reflects substantive ethical judgment or procedural completion.

Governance Implications

Our findings carry implications that extend beyond toolkit design to the structure of AI governance itself. Three emerge most directly from the data.

Specification quality determines ethics quality. Because compliance functions as both floor and ceiling, the precision of regulatory requirements determines whether ethical practice becomes substantive or symbolic. Practitioners interpret vague mandates like “ensure no bias” or “document your model” minimally, producing artifacts that satisfy auditors without changing design decisions. Waldman (Waldman 2019) and Grover (Grover 2024) document this pattern in adjacent domains, and our data confirm it directly. Concrete obligations work differently. When regulation specifies what traceability must demonstrate, what bias evaluation must cover, and what post-deployment monitoring must trigger, practitioners have something implementable to work toward. Principle-level mandates like “ensure fairness” do not. The EU AI Act's risk-classification structure illustrates this, and our data show it is already reshaping upstream tooling decisions in ways aspirational frameworks do not.

Structural accountability gaps require structural fixes. Governance frameworks that rely on individual practitioner judgment to fill the gap between regulatory minimums and ethical practice will systematically fail for the majority of the AI workforce. Our data show this clearly: the practitioners who successfully pushed ethical commitments beyond compliance floors; P7 refusing a depression dataset on consent grounds, P13 rejecting weapons contracts, P12 exiting projects on ethical grounds all did so from positions of structural independence as founders, consultants, or freelancers with exit options. P6, by contrast, embedded in a financial services organization with clear performance metrics and no designated ethics role, described ethical work sliding toward symbolism when leverage was thin. This is not a training or awareness problem, P6 was not less ethically committed than P13. It is a structural accountability problem: without designated roles carrying formal authority, ethics responsibilities diffuse across teams and get subordinated to release timelines. Regulation that assigns formal ethics responsibilities to designated roles with protected authority rather than assuming individual engineers will fill the gap, would directly address this failure mode.

Jurisdictional variation requires differentiated governance. Our data most clearly distinguish two compliance framings ethics-as-legal-obligation in the EU (with auditable market-access gates) and ethics-as-liability-shield in the US (with reactive, documentation-oriented compliance). Our smaller Canadian subsample points to further variations Canada as ethics-as-intent pending jurisprudence, though this would require larger samples to characterize with confidence. The broader governance point holds across these cases: regulatory frameworks that assume a uniform compliance baseline or that treat GDPR-style mandates as universally transferable will systematically misread how ethics work is organized across these contexts. The US practitioners in our sample who described the regulatory landscape as “severely compromised” by early 2025 and who increasingly indexed on EU frameworks as de facto global baselines illustrate both the fragility of nationally-scoped governance and the gravitational pull of the most stringent available regime. International governance coordination that ac-

knowledges and works with these jurisdictional differences, rather than assuming harmonization, is more likely to produce durable ethics practice than universal frameworks applied uniformly.

Beyond formal statutes, our data point to para-regulatory forces that effectively govern practice in the absence of binding law procurement checklists, client security questionnaires, assurance schemes (ISO/IEC families, SOC 2), platform review policies, and insurer requirements tied to cyber risk coverage (Horstmann et al. 2024). These soft mechanisms import obligations into development even where formal AI law is absent or nascent, and they shape what counts as sufficient evidence as concretely as any statutory requirement. Governance that engages with these para-regulatory structures standardizing what procurement checklists must require, what assurance schemes must certify would reach the majority of development contexts that formal AI regulation currently does not.

Limitations and Future Work

Our account is based on qualitative interviews with 20 AI/ML practitioners across multiple sectors and regions. Three limitations follow. First, the sample centers developer-adjacent roles; we did not include regulators, auditors, procurement officers, or in-house legal and compliance teams who define review gates and determine what counts as sufficient evidence. Our account therefore represents one side of the compliance boundary how developers experience and translate requirements, not how those requirements are produced. Second, evidence is primarily self-report rather than ethnographic observation of practice, relying on descriptions of artifact use and workflow decisions rather than direct observation of those decisions being made. Third, the study is cross-sectional amid rapidly evolving regulatory regimes. Some of what we document is likely specific to this transitional moment: the anticipatory planning around the EU AI Act's staged rollout, the shifting U.S. federal posture, and the Canadian regulatory vacuum following the lapse of the proposed Artificial Intelligence and Data Act. Other patterns appear more structural and less contingent on any particular regime, including the conversion of ethical commitments into auditable artifacts, the dependence of ethical pushback on structural position, and the dual-legibility gap between what satisfies auditors and what protects affected communities. These follow from how organizations produce evidence under any compliance requirement, and we expect them to persist as specific regulations change.

Our sample's skew toward consultants, freelancers, and startup founders with only three participants from large technology companies likely over-represents client-facing compliance work and under-represents internal governance dynamics in large, highly structured organizations. This skew is analytically consequential: it may amplify the visibility of resistance cases (founders and freelancers have more structural independence to push back) while under-representing the absorption dynamics that characterize employee-level compliance work at scale. A further limitation concerns how we treat jurisdiction. We characterize compliance cultures by where practitioners and their organizations are located,

but many participants worked across borders, and our data show US practitioners indexing on EU frameworks for market access. Multinational operation means firms often comply upward to the most stringent applicable regime regardless of domicile, which complicates any clean mapping from location to compliance culture. Our jurisdictional observations should therefore be read as describing regulatory environments practitioners orient toward, not bounded national practices. Future work should deliberately sample compliance officers, legal counsel, procurement reviewers, and employees at large technology companies to examine the other side of the compliance boundary. Longitudinal and comparative research examining how jurisdictional differences in compliance culture evolve as the EU AI Act matures and U.S. regulation remains fragmented would further develop the governance implications identified here.

Conclusion

This paper examined how AI practitioners translate regulatory compliance requirements into internal workflows, toolkit use, and compliance evidence, and what tensions emerge between compliance and ethics work. Our findings show that compliance operates not merely as a floor for ethical practice but frequently as its ceiling, and in some cases as a substitute that forecloses ethical deliberation. In our data, ethics persisted in development workflows only where it could be rendered as an auditable artifact embedded in an existing process gate; values that could not be converted into evidence, and practitioners without the structural independence to insist on their inclusion, were routinely displaced.

These dynamics have implications that extend beyond responsible AI toolkit design to the structure of governance itself. Because compliance requirements define the operative scope of ethics work in practice, the specificity of those requirements determines whether ethical practice becomes implementable or remains symbolic. Vague mandates produce symbolic completion; specific evidence obligations produce traceable practice. Regulators are therefore not only setting minimums but determining whether the resulting practice is substantive in any meaningful sense. Governance frameworks that rely on individual practitioner judgment to fill the gap between regulatory minimums and ethical ideals will systematically fail for practitioners embedded in organizational hierarchies, who lack the structural position to push back. And governance that assumes a uniform compliance baseline will misread the different compliance cultures our jurisdictional analysis reveals. Bridging the ethics-regulation gap requires not better toolkits alone, but governance structures that specify meaningful evidence obligations, assign accountable roles with protected authority, and engage with the jurisdictional variation that shapes how compliance work is actually organized across different legal and institutional contexts.

Acknowledgments

This research was funded by the Max Planck Institute for Security and Privacy. We thank our participants for generously sharing their time and their candid accounts of ethics

and compliance work in their organizations.

References

- AI, N. 2023. Artificial intelligence risk management framework (AI RMF 1.0). URL: [https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.1\(0\):0-1](https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.1(0):0-1).
- Ali, S. J.; Christin, A.; Smart, A.; and Katila, R. 2023. Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 217–226. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-0192-4.
- Arts, J.; and van den Berg, M. 2025. What the Dutch benefits scandal and policy's focus on 'fraud' can teach us about the endurance of empire. *Critical Social Policy*, 45(1): 177–187. Publisher: SAGE Publications Ltd.
- Atkins, S.; Badrie, I.; and Otterloo, S. 2021. Applying Ethical AI Frameworks in practice: Evaluating conversational AI chatbot solutions. *Computers and Society Research Journal*.
- Balayn, A.; Yurrita, M.; Yang, J.; and Gadiraju, U. 2023. "Fairness Toolkits, A Checkbox Culture?" On the Factors that Fragment Developer Practices in Handling Algorithmic Harms. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, 482–495. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-0231-0.
- Braun, V.; and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2): 77–101.
- Braun, V.; Clarke, V.; Hayfield, N.; Davey, L.; and Jenkinson, E. 2022. Doing Reflexive Thematic Analysis. In *Supporting Research in Counselling and Psychotherapy*, 19–38. Palgrave Macmillan, Cham. ISBN 978-3-031-13942-0.
- Butt, J. S. 2024. Analytical Study of the World's First EU Artificial Intelligence (AI) Act, 2024. *International Journal of Research Publication and Reviews*, 5(3): 7343–7364.
- CEN-CENELEC. 2026. Artificial Intelligence. European Committee for Standardization and European Committee for Electrotechnical Standardization. Accessed: 2026-08-04.
- Crisan, A.; Drouhard, M.; Vig, J.; and Rajani, N. 2022. Interactive Model Cards: A Human-Centered Approach to Model Documentation. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 427–439. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9352-2.
- Dastin, J. 2018. Insight - Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*.
- de Almeida, P. G. R.; dos Santos, C. D.; and Farias, J. S. 2021. Artificial Intelligence Regulation: a framework for governance. *Ethics and Information Technology*, 23(3): 505–525.
- Deng, W. H.; Yildirim, N.; Chang, M.; Eslami, M.; Holstein, K.; and Madaio, M. 2023. Investigating Practices and Opportunities for Cross-functional Collaboration around AI Fairness in Industry Practice. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 705–716. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-0192-4.
- Drage, E.; McInerney, K.; and Browne, J. 2024. Engineers on responsibility: feminist approaches to who's responsible for ethical AI. *Ethics and Information Technology*, 26(1): 4.
- European Parliament and Council of the European Union. 2016. Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union, OJ L 119, 1–88.
- European Parliament and Council of the European Union. 2024. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, OJ L, 2024/1689.
- Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; Schafer, B.; Valcke, P.; and Vayena, E. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4): 689–707.
- Gardner, J. 2010. Ethics and Law. In *The Routledge Companion to Ethics*. Routledge. ISBN 978-0-203-85070-1. Num Pages: 10.
- Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.
- Gray, C. M.; Chivukula, S. S.; Carlock, T. V.; Li, Z.; and Duane, J.-N. 2023. Scaffolding Ethics-Focused Methods for Practice Resonance. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, 2375–2391. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9893-0.
- Grover, R. 2024. Encoding Privacy: Sociotechnical Dynamics of Data Protection Compliance Work. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-0330-0.
- Hacker, P.; Naumann, F.; Friedrich, T.; Grundmann, S.; Lehmann, A.; and Zech, H. 2022. AI Compliance – Challenges of Bridging Data Science and Law. *J. Data and Information Quality*, 14(3): 21:1–21:4.
- Hadwick, D.; and Lan, S. 2021. Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany. *World Tax Journal*, 2021(4): 609–645.
- Holstein, K.; Wortman Vaughan, J.; Daumé, H.; Dudik, M.; and Wallach, H. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-5970-2.
- Horstmann, S. A.; Domiks, S.; Gutfleisch, M.; Tran, M.; Acar, Y.; Moonsamy, V.; and Naiakshina, A. 2024. "Those

- things are written by lawyers, and programmers are reading that.” Mapping the Communication Gap Between Software Developers and Privacy Experts. *Proceedings on Privacy Enhancing Technologies*.
- Innovation, Science and Economic Development Canada. 2023. The Artificial Intelligence and Data Act (AIDA) — Companion document. Government of Canada. Published February 17, 2023; archived. Accessed August 4, 2026.
- Jindal, A. 2022. Misguided Artificial Intelligence: How Racial Bias is Built Into Clinical Models. *Journal of Brown Hospital Medicine*, 2(1). Publisher: Department of Medicine, Brown University.
- Kaur, H.; Klivan, S.; Votipka, D.; Acar, Y.; and Fahl, S. 2022. Where to Recruit for Security Development Studies: Comparing Six Software Developer Samples. 4041–4058. ISBN 978-1-939133-31-1.
- Kroll, J. A. 2018. The fallacy of inscrutability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133): 20180084. Publisher: Royal Society.
- Kroll, J. A. 2021. Outlining Traceability: A Principle for Operationalizing Accountability in Computing Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’21, 758–771. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8309-7.
- Larsson, S.; and Heintz, F. 2020. Transparency in artificial intelligence. Publisher: Alexander von Humboldt Institute for Internet and Society gGmbH.
- Lau, D.; Samy, G. N.; Rahim, F. A.; Maarop, N.; and Hassan, N. H. 2023. Review of The Governance, Risk and Compliance Approaches For Artificial Intelligence. *Open International Journal of Informatics*, 11(2): 25–35.
- Lewis, D.; Hogan, L.; Filip, D.; and Wall, P. J. 2020. Global Challenges in the Standardization of Ethics for Trustworthy AI. *Journal of ICT Standardization*, 123–150.
- Madaio, M. A.; Chen, J.; Wallach, H.; and Wortman Vaughan, J. 2024. Tinker, Tailor, Configure, Customize: The Articulation Work of Contextualizing an AI Fairness Checklist. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW1): 214:1–214:20.
- Madaio, M. A.; Stark, L.; Wortman Vaughan, J.; and Wallach, H. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6708-0.
- Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I. D.; and Gebru, T. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* ’19, 220–229. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6125-5.
- Morley, J. 2021. Operationalising AI ethics: barriers, enablers and next steps | AI & SOCIETY.
- Morley, J.; Machado, C. C. V.; Burr, C.; Cowls, J.; Joshi, I.; Taddeo, M.; and Floridi, L. 2020. The ethics of AI in health care: A mapping review. *Social Science & Medicine*, 260: 113172.
- Orr, W.; and Davis, J. L. 2020. Attributions of ethical responsibility by Artificial Intelligence practitioners. *Information, Communication & Society*, 23(5): 719–735. Publisher: Routledge eprint: <https://doi.org/10.1080/1369118X.2020.1713842>.
- Qiang, V.; Rhim, J.; and Moon, A. 2024. No such thing as one-size-fits-all in AI ethics frameworks: a comparative case study. *AI & SOCIETY*, 39(4): 1975–1994.
- Rakova, B.; Yang, J.; Cramer, H.; and Chowdhury, R. 2021. Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1): 7:1–7:23.
- Rattay, S.; Vakkuri, V.; Rozendaal, M. C.; and Shklovski, I. 2025. “Why do we do this?”: Moral Stress and the Affective Experience of Ethics in Practice. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI ’25, 1–15. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-1394-1.
- Robles Carrillo, M. 2020. Artificial intelligence: From ethics to law. *Telecommunications Policy*, 44(6): 101937.
- Roski, J.; Maier, E. J.; Vigilante, K.; Kane, E. A.; and Matheny, M. E. 2021. Enhancing trust in AI through industry self-governance. *Journal of the American Medical Informatics Association*, 28(7): 1582–1590.
- Sadek, M.; Kallina, E.; Bohné, T.; Mougenot, C.; Calvo, R. A.; and Cave, S. 2025. Challenges of responsible AI in practice: scoping review and recommended actions. *AI & SOCIETY*, 40(1): 199–215.
- Sandvig, C.; Hamilton, K.; Karahalios, K.; and Langbort, C. 2014. Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms.
- Seet, M. 2026. *ISO 42001 and Legal Compliance: A Principled Implementation of the AI Management System*. Springer Nature.
- Tahaei, M.; Frik, A.; and Vaniea, K. 2021. Privacy Champions in Software Teams: Understanding Their Motivations, Strategies, and Challenges. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21, 1–15. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8096-6.
- Treasury Board of Canada Secretariat. 2025. Directive on Automated Decision-Making. Government of Canada. Effective April 1, 2019; current version June 24, 2025.
- U.S. Department of Health and Human Services. 2025. Strategic Plan for the Use of Artificial Intelligence in Health, Human Services, and Public Health.
- U.S. Food and Drug Administration. 2025. Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations. Publisher: FDA.
- Varanasi, R. A.; and Goyal, N. 2023. “It is currently hodgepodge”: Examining AI/ML Practitioners’ Challenges during

Co-production of Responsible AI Values. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, 1–17. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9421-5.

Waldman, A. E. 2019. Privacy Law's False Promise. *Washington University Law Review*, 97(3): 773–834.

Widder, D. G. 2024. Epistemic Power in AI Ethics Labor: Legitimizing Located Complaints. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 1295–1304. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-0450-5.

Widder, D. G.; and Nafus, D. 2023. Dislocated accountabilities in the “AI supply chain”: Modularity and developers’ notions of responsibility. *Big Data & Society*, 10(1): 20539517231177620. Publisher: SAGE Publications Ltd.

Wong, R. Y.; Madaio, M. A.; and Merrill, N. 2023. Seeing Like a Toolkit: How Toolkits Envision the Work of AI Ethics. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW1): 145:1–145:27.

Wright, L.; Muenster, R. M.; Vecchione, B.; Qu, T.; Cai, P.; Smith, A.; Investigators, C. . S.; Metcalf, J.; Matias, J. N.; et al. 2024. Null Compliance: NYC Local Law 144 and the challenges of algorithm accountability. In *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*, 1701–1713.

Yildirim, N.; Pushkarna, M.; Goyal, N.; Wattenberg, M.; and Viégas, F. 2023. Investigating How Practitioners Use Human-AI Guidelines: A Case Study on the People + AI Guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9421-5.