

From Principles to Practice: Governing High-Stakes AI Across Institutions and the State

Editors' Introduction:

Muhammad Khattak, Maryam Khalil, Sayo Stefflbauer, and Alex Nguyen

Artificial intelligence is now embedded in the institutions that allocate opportunity, rights, and care: schools and universities, labor markets, healthcare systems, law enforcement, transportation infrastructure, and national security. In each of these domains, AI systems increasingly shape consequential outcomes, often through tools that are difficult for affected people to see, understand, or contest. Governance has failed to keep pace with deployment.

This issue of *GRACE* is organized around a practical claim. The most urgent AI question today is how democratic institutions can establish enforceable guardrails for high-stakes systems that are already in use. The ten essays collected here treat AI governance as an institutional design problem with concrete levers: funding conditions, audit requirements, transparency rules, documentation standards, safety benchmarks, and clear lines of human accountability.

Across domains, a shared architecture emerges. First, disclosure. People must know when AI materially influences decisions about education, employment, liberty, healthcare, or security. Second, human accountability. In high-stakes settings, a named professional, supervisor, or decision-maker must remain responsible for outcomes, rather than outsourcing judgment to a tool. Third, independent auditing and documentation. Without auditable records, performance testing, and credible external review, neither civil rights enforcement nor public trust can survive at scale.

Clear guardrails can accelerate responsible deployment by reducing legal uncertainty, preventing headline failures that trigger public backlash, and creating predictable standards that innovators can design toward. In this sense, disclosure, accountable human oversight, and audit-ready documentation function less like brakes and more like rules of the road that allow high-stakes AI to progress. These essays argue for a uniform floor, not a ceiling. The goal is to set baseline protections while leaving room for institutions and developers to exceed them. Effective regulation can also level the playing field by preventing a race to the bottom in which the least transparent systems gain advantage through opacity.

The issue opens with education, where the governance problem is also a capacity problem. Reichfeld's "Implementing AI Literacy into Education" argues for a national baseline of AI literacy in K–12 education through a standardized annual module tied to federal funding eligibility. It treats AI literacy as a civic competency and frames public competence as a precondition for long-term democratic governance.

From there, the issue turns to the university as a governance laboratory. Escobar et al.'s "Governing Generative AI at Stanford" examines institutional approaches to guiding responsible use across a research university. It emphasizes the need for policies that are clear, implementable, and aligned with academic practice rather than framed as narrow compliance rules. Alongside it, Langevine's "Protecting Student Data in the Age of AI" focuses on higher education data governance and argues that the rapid adoption of AI-enabled educational platforms has outpaced clear federal requirements for data lifecycle governance, transparency, and auditability. By foregrounding how de-identified or derived data can be repurposed beyond students' reasonable expectations, it presses toward enforceable rules that reduce compliance gray zones and support accountability.

From education systems, the issue moves into the labor market, where AI systems influence hiring, promotion, evaluation, and termination at high volume. Mgbahurike and Ocran's "Automated Employment Bureau (AEB)" proposes creating a federal regulator within the Department of Labor to establish baseline standards for AI use in employment decisions. The memo's core diagnosis is structural. Existing civil rights protections depend on transparency and investigability, yet algorithmic systems often operate as opaque intermediaries that are hard for workers to challenge and hard for regulators to assess without dedicated technical capacity.

The issue also addresses two domains where AI governance intersects directly with physical infrastructure, safety, and sustainability. Bempong's "Standardized Autonomous Vehicle Evaluation and Deployment (SAVED) Framework" proposes a national approach to safety evaluation and public accountability for self-driving systems, emphasizing standardized testing expectations, reporting, and deployment thresholds. Agarwal and Beharry's "A Transparency-Based Approach to Regulating the Resource Footprint of U.S. Data Centers" argues that energy consumption, water use, and environmental impacts should become first-class objects of AI governance. The memo presses for transparency-based regulation that enables benchmarking, public accountability, and enforceable efficiency incentives.

The next cluster centers on civil rights and the administration of justice. Sinchi's "Facial Recognition Used in Suspect Identification" develops a federal framework that conditions DOJ grant eligibility on three pillars: disclosure, accountability, and independent auditing, paired with standardized public reporting. The memo treats wrongful arrest as a predictable failure mode when probabilistic matches are treated as determinate evidence without corroboration and documentation. Here the governance stakes are immediate. Transparency is a prerequisite for access to justice.

Healthcare governance occupies the center of the issue because it is where algorithmic power meets vulnerability at scale. Gundlapalli's "Health Professional Shortage Areas (HPSAs)" addresses the ethical deployment of AI and large language models in medically underserved and rural communities, where provider shortages and infrastructure constraints increase both the

potential benefits and the risks. It argues that deployment in shortage areas requires equity evidence before scale, patient-facing transparency that supports trust, and clear governance and liability rules that distinguish attended from unattended systems. Patrick and Adamala's "A National Framework for Regulating AI in the Health Insurance Space" focuses on AI-influenced coverage decisions and proposes enforceable requirements for disclosure whenever AI influences determinations, clinician accountability for adverse decisions, and independent auditing of tools and workflows. Taken together, the healthcare pieces sharpen a recurring theme. High-stakes automation does not merely risk error. It risks eroding the conditions of trust, explanation, and individualized judgment that legitimate institutions depend on.

The issue closes with national security. Solomon's "Regulating LLMs in Warfare: A U.S. Strategy for Military AI Accountability" addresses the rapid integration of large language models into military operations and argues that current governance remains too technology neutral to manage LLM-specific risks. It proposes safeguards centered on human oversight for high-stakes uses, human approval for AI-generated information operations, strengthened data protection with adversarial testing and continuous evaluation, and formal accountability mechanisms that document, audit, and assign responsibility for AI-influenced decisions. Placed at the end of the volume, this contribution underscores the full arc of the issue. From classroom governance to command accountability, AI challenges democratic institutions to build procedures that preserve human responsibility under conditions of automated power.

These ten essays argue that ethical AI becomes meaningful only when translated into rules people can follow, records institutions must keep, and checks that independent reviewers can verify. This first policy issue of *GRACE* offers a blueprint for that translation: disclosure that enables contestation, accountability that remains human, and auditability that makes enforcement possible. In a policy environment crowded with principles, these pieces begin with practice. All the essays here reflect student work in courses at Stanford. *GRACE*'s second and third policy issues invite a larger conversation across institutions and global contexts. The next submission deadlines are April 30, 2026, and August 31, 2026.