Stanford University
Human-Centered
Artificial Intelligence

RegLab

Policy Brief
HAI Policy & Society
Stanford RegLab
June 2025

Cleaning Up Policy Sludge: An AI Statutory Research System

Faiz Surani, Lindsey A. Gailmard, Allison Casasola, Varun Magesh, Emily J. Robitschek, Christine Tsang, Derek Ouyang, and Daniel E. Ho

The Problem

THERE IS A GROWING RECOGNITION THAT "POLICY SLUDGE"—outdated, obsolete, or cumbersome legal requirements and regulations—can impede adaptable governance.[1] But reforming a daunting volume of statutes, regulations, and codes can be challenging.

Consider five examples:

- As a law professor in the 1970s, Ruth Bader Ginsburg hired an army of Columbia Law School students to cull through the United States Code for provisions that discriminated on the basis of sex. The final report, using 59 key words, was an "extensive, but not exhaustive" list which provided the blueprint for equal rights litigation.

Key Takeaways

Legal reform can get bogged down by policy sludge that is strewn about millions of words of statutes and regulations. Such sludge can make programs hard to administer for civil servants and even more difficult to navigate for the public.

Stanford RegLab developed the Statutory Research Assistant (STARA), a domain-informed AI system capable of performing accurate and comprehensive statutory surveys that help to identify and eliminate policy sludge.

As an illustration, RegLab partnered with the San Francisco City Attorney's Office to identify all legislatively mandated reporting requirements, many of which are burdensome and can serve little purpose after decades. Based on the collaboration, the city attorney spearheaded a consultative process with city departments, culminating in a proposed ordinance to delete or consolidate over a third of these requirements.

AI systems like STARA enable researchers, advocates, attorneys, and government officials to gain a more comprehensive understanding of often opaque legal mandates, identify policy sludge, and accelerate meaningful reform efforts.

- In the 1980s, the U.S. Department of Justice under the Reagan administration tried to count federal crimes. After two years, they gave up. The responsible official said that one could "die and [be] resurrected three times," and still not know the true number.

- In 2021, California enacted a law that required all county recorder offices to identify and redact racist deed records. Such racial covenants, which prohibit people of particular races from residing on the property, are unenforceable and yet they persist. In Santa Clara County alone, that meant sifting through 80 million pages of deed documents dating back to the 1800s.[2] Los Angeles recently spent $8 million for a contractor to use key words to find such covenants—a process expected to last over seven years.

- Congressionally mandated reports are, according to political scientist Frank Fukuyama, a prime example of how "government is made inefficient by the layers of rules bureaucrats themselves are forced to labour under." Congress has lost track of thousands of these reports, creating a congressional "black hole" that weighs down civil servants, with many reports producing little benefit. As Supreme Court Justice Neil Gorsuch noted, one report—on the Social Security Administration's printing operations—took 95 employees over four months to complete. As the Congressional Research Service has conceded, there is no "search method that can obtain an exact accounting of all reports required," given that the U.S. Code contains some 32 million words.

The law runs across millions of statutes, regulations, deeds, and other documents. And sometimes, the problem facing policymakers, judges, and reformers is simply knowing what the law is.

- In 2024, San Francisco voters approved a ballot measure requiring the city to simplify its sprawling system of advisory bodies and commissions. The San Francisco Municipal Code, along with resolutions by the Board of Supervisors, totals 16 million words, and a civil grand jury found "there [was] no centralized list of commissions."

The law runs across millions of statutes, regulations, deeds, and other documents. And sometimes, the problem facing policymakers, judges, and reformers is simply *knowing what the law is*.[3]

In our paper, "What Is the Law? A System for Statutory Research (STARA) with Large Language Models," we introduce an automated system that aims to address the unique challenges of statutory research by rapidly parsing and compiling legal provisions. It enables researchers, advocates, attorneys, and government

officials to understand the full breadth of legislative mandates. Our work highlights the promise of using large language models (LLMs) to build domain-specific systems that perform well in complex fields such as statutory research, where they can help governments reduce policy sludge and pave the way for meaningful statutory reform.

The Solution

We developed the Statutory Research Assistant (STARA),[4] a domain-informed AI system designed to automate statutory and regulatory research. STARA performs comprehensive statutory surveys, i.e., systematic compilations of legal provisions relevant to a given question or policy area. Unlike general-purpose tools, STARA exploits the capabilities of frontier AI models and a domain-specific architecture tailored to the structure of legal codes. It incorporates key elements such as hierarchical organization, cross-references, and definitions—unique features of legal codes that have made manual and automated approaches to statutory research challenging. Intuitively, STARA represents legal codes in the way the basics of statutory interpretation are taught to law students.

STARA's research pipeline operates in three stages. First, it preprocesses and segments statutory codes into enumerable units, annotating each with relevant legal context drawn from headings, cross-references, definitions, and editorial notes. Then, it uses LLMs to reason about statutory language, classify provisions, and extract structured information. Finally, STARA can agentically organize, analyze, and report on the results of its statutory surveys.

This approach not only enables higher precision and recall than general-purpose tools, but it can also be adapted to parse diverse bodies of law—from the U.S. Code to state codes to city municipal codes. We benchmark STARA's performance and show that it makes frontier LLMs as much as three times more accurate on complex statutory research tasks relative to a general-purpose AI system, making previously infeasible research efforts not just possible, but trivial.

Illustration

To explore the power of this approach for cleaning up policy sludge, the Stanford Regulation, Evaluation, and Governance Lab (RegLab) partnered with the San Francisco City Attorney's Office, under the direction of David Chiu. We identified a number of fruitful directions but focused, as a first target, on the problem of legislatively mandated reports. As the Washington Post put it, such reports are a "parable about why some big government systems break down," costing taxpayers $163 million each year (according to the last estimate in 1993) and preventing civil servants from focusing on what matters most.

First, Stanford RegLab ingested the city's municipal codes and resolutions—approximately 16 million words of enacted law. Second, we collaborated on a search for all legislatively mandated reports, resulting in a list of 528 required reports, strewn about the 16 million words of the code. These reports increased sharply over time, as seen in Figure 1. Between 2000 and 2025, the number of reports more than doubled, while public sector employment increased by only 10%.[5] This is just one sign that civil servants have been asked to do more, without a corresponding increase in resources.

"This collaboration is a great example of how governments can use AI to make impactful change." City Attorney David Chiu

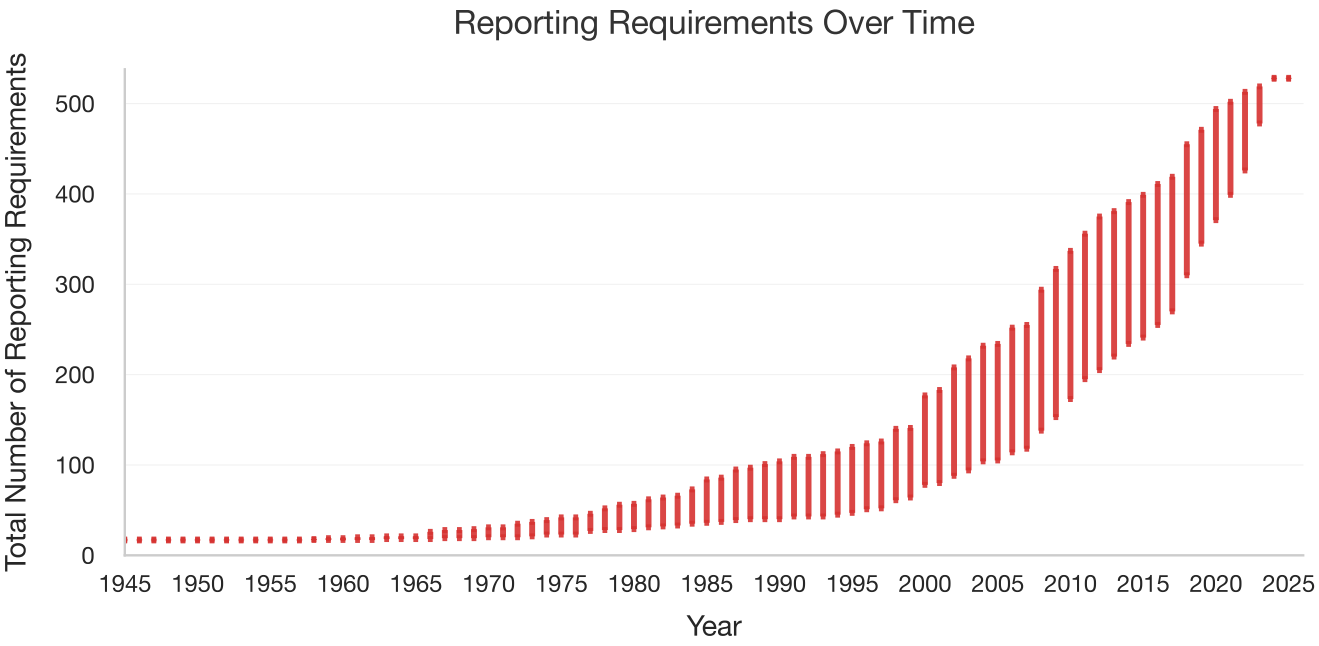

Figure 1: Reporting requirements over time in San Francisco's Municipal Code. The figure illustrates the substantial growth in reporting requirements over time. The vertical lines account for uncertainty in enactment dates. Note: Because historical time-stamped versions of the San Francisco Code are not available, this tabulation does not include reporting requirements that have been previously deleted, but such known instances are rare prior to this effort.

Third, the City Attorney's Office kicked off a consultative process with city departments to understand which requirements were obsolete, capable of simplification, or failed to serve a continuing policy function. The result of this partnership and collaborative process was the introduction of a 351-page omnibus resolution to delete or consolidate 36% of amendable reporting requirements. These included reporting requirements for defunct programs and reports of little present-day relevance, such as a biennial report by the director of public works on newspaper racks (stands that once were a common way to sell copies of newspapers but "are going the way of the dinosaurs"). Another example: an 80-year old provision for quarterly reports from the Redevelopment Agency, an entity that has undergone significant restructuring and already has duplicative local and state reporting obligations. In other instances, the proposal is to consolidate reporting, such as having a single report on housing inventory, rather than a scattered set of overlapping reports submitted by different teams at different times.

"For too long, City Hall has been weighed down by outdated and unnecessary code, slowing its ability to serve residents effectively. This legislation will cut through that clutter, freeing up staff time so departments can focus on what matters most—serving San Franciscans." San Francisco Supervisor Bilal Mahmood

This process illustrates important lessons for code reform and simplification, powered by AI. First, as David Chiu remarked: "Because of the length of our code … it's likely a project we would never have undertaken" without the STARA system. Until the collaboration, the SF Municipal Code and resolutions were searchable only by crude key words—precisely the bottleneck for Ruth Bader Ginsburg and the Justice Department in the 1980s.

Second, the agency consultation process exemplifies the critical role of human judgment and review, a core tenet of human-centered AI. STARA drove down the search costs to make this effort feasible, but ultimately the City Attorney's Office launched a process to leverage the deep expertise with the office and city agencies to prioritize, identify, and propose simplification. That also meant leaving important reports in place.

Third, STARA illustrates the need for tailoring AI systems to the domain. Crudely using off-the-shelf systems for reform and simplification is an approach bound to result in serious shortcomings.[6] Honing AI research tools to fit the particular domain, however, requires openness and benchmarking to produce tools that deliver accuracy, consistency, and reliability.[7]

The Promise of Reform

Policy sludge isn't just a San Francisco problem. STARA and the partnership with San Francisco provide a template for how AI systems could aid legal reform in many domains and jurisdictions. Consider other examples:

- **Congressionally mandated reports:** Federal agencies are required to submit thousands of reports to Congress each year. STARA identified more than 11,000 reporting mandates strewn across 6,814 sections of the U.S. Code, including thousands that do not appear on the House clerk's official listing. These findings provide a starting point for easing the worsening reporting burden on federal agencies—3,359 reporting requirements were enacted in the 115th Congress alone.

- **Process and commissions:** A ballot initiative required San Francisco to enumerate and potentially streamline a bevy of advisory bodies and commissions to improve administration of city government. Attorneys could spend days on such a research task, given the volume and

dated search methods for the municipal code. STARA, however, helped to identify all 111 known commissions, and 33 more. It was "outrageously helpful," in the words of one official faced with this task, and the results informed the work of the Commission Streamlining Task Force.

- **Fees and permitting:** Like other jurisdictions, San Francisco has aimed to support small businesses by cutting down license, permit, and registration fees, especially as they continue to recover from the pandemic. STARA can surface hundreds of fee requirements, coinciding with an ordinance to reduce or eliminate fees. STARA could similarly help to surface permit requirements, identifying overlapping provisions, highlighting inconsistencies, and detecting repealed or outdated provisions. For instance, inconsistencies between state and local housing laws have stymied attempts to respond to California's housing crisis.

- **Outdated laws:** A system like STARA can also help surface more general categories of outdated or illegal provisions. For instance, the San Francisco code requires any person working for a public works contract to be a citizen, a provision that likely violates a federal law prohibiting citizenship discrimination. Another provision calls for individuals under the age of 21 to show written consent of a parent or guardian to get a library card. Prior efforts to identify such provisions, like those of Ruth Bader Ginsburg, have entailed extensive human effort; STARA lowers those search costs dramatically.

- **Legislation and interpretation:** Tools like STARA may enable legislative staff and policymakers to identify how proposed legislation affects existing statutory and regulatory provisions. Similarly, lawyers and judges may often need to understand how a decision affects existing law.[8] By clarifying how rules intersect and interact, STARA may ultimately support more informed, deliberate, and durable legislation and interpretation.

STARA can accelerate reform efforts, support policymaking, and reduce the volume of policy sludge that can hamstring government.

By surfacing these insights with a high degree of accuracy and at scale, AI systems like STARA can accelerate reform efforts, support policymaking, and reduce the volume of policy sludge that can hamstring government—and build state capacity in the process.[9]

Endnotes

1. See Ezra Klein and Derek Thompson, *Abundance* (2025); Nicholas Bagley, "The Procedure Fetish," 118 Mich. L. Rev. 345 (2019); Jennifer Pahlka, *Recoding America: Why Government Is Failing in the Digital Age and How We Can Do Better* (2023); Jennifer Pahlka and Andrew Greenway, "The *How* We Need Now: A Capacity Agenda for 2025 and Beyond," Niskanen Center (Dec. 2024), https://www.niskanencenter.org/wp-content/uploads/2024/12/Niskanen-State-Capacity-Paper_-Jen-Pahlka-and-Andrew-Greenway-2.pdf.

2. The Regulation, Evaluation, and Governance Lab and a collaborator at Princeton University partnered with the Santa Clara County Clerk-Recorder's Office to prototype, develop, and integrate a machine-learning-based pipeline to identify racial covenants. See Faiz Surani, Mirac Suzgun, Vyoma Raman, Christopher D. Manning, Peter Henderson, and Daniel E. Ho, "AI for Scaling Legal Reform: Mapping and Redacting Racial Covenants in Santa Clara County," J. Legal Analysis (forthcoming, 2025).

3. We do not mean to trigger long-standing philosophical debates about what the law is. Instead, here we mean identifying legal requirements, when those have been expressly enumerated in a large body of statutes or regulations.

4. Faiz Surani, Lindsey A. Gailmard, Allison Casasola, Varun Magesh, Emily J. Robitschek, and Daniel E. Ho, "What Is the Law? A System for Statutory Research (STARA) with Large Language Models," in *20th International Conference on Artificial Intelligence and Law* (ICAIL 2025) (forthcoming).

5. See City and County of San Francisco, Comprehensive Annual Financial Report (CAFR) (2010) at 218, https://www.sf.gov/sites/default/files/2024-08/CCSF%20CAFR%20FY2010%20final.pdf (reporting 29,610 total annually funded positions in 2001); Grow SF, "San Francisco's Budget Is Bigger than 17 U.S. States," https://growsf.org/research/2025-04-28-sf-city-budget/ ("The city employs approximately 33,000 people in total…").

6. Emaan Hariri and Daniel E. Ho, "AI for Statutory Simplification: A Comprehensive State Legal Corpus and Labor Benchmark," in *20th International Conference on Artificial Intelligence and Law* (ICAIL 2025) (forthcoming).

7. See, e.g., Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning, and Daniel E. Ho, "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools," 22 J. Empirical Legal Stud. 216 (2025); Matthew Dahl, Varun Magesh, Mirac Suzgun, and Daniel E. Ho, "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models," 16 J. Legal Analysis 64 (2024).

8. For instance, in *Free Enterprise Fund v. Public Company Accounting Oversight Board*, 561 U.S. 477 (2010), Justice Breyer, in dissent, compiled 573 agency arrangements that might be affected by the majority's ruling.

9. See Nicholas Bagley, "Using Artificial Intelligence to Build State Capacity," Divided Argument (Jun. 9, 2025), https://blog.dividedargument.com/p/using-artificial-intelligence-to; Jennifer Pahlka and Andrew Greenway, "The *How* We Need Now: A Capacity Agenda for 2025 and Beyond."

Reference: The original article is accessible at Faiz Surani, Lindsey A. Gailmard, et al., "**What Is the Law? A System for Statutory Research (STARA) with Large Language Models**," *20th International Conference on Artificial Intelligence and Law* (ICAIL 2025) (Forthcoming). A prior, extended version of the original article is accessible at https://dho.stanford.edu/wp-content/uploads/STARA.pdf.

For more information about STARA, see the project website at https://reglab.github.io/stara/.

Stanford University's Institute for Human-Centered Artificial Intelligence (HAI) applies rigorous analysis and research to pressing policy questions on artificial intelligence. A pillar of HAI is to inform policymakers, industry leaders, and civil society by disseminating scholarship to a wide audience. HAI is a nonpartisan research institute, representing a range of voices. The views expressed in this policy brief reflect the views of the authors. For further information, please contact **HAI-Policy@stanford.edu**.

Faiz Surani is a data scientist at the Regulation, Evaluation, and Governance Lab (RegLab) at Stanford University.

Lindsey A. Gailmard is a postdoctoral scholar at the RegLab at Stanford University.

Allison Casasola is an incoming research fellow at the RegLab at Stanford University.

Varun Magesh is a research fellow at the RegLab at Stanford University.

Emily J. Robitschek is a research fellow at the RegLab at Stanford University.

Christine Tsang is the executive director of the RegLab at Stanford University.

Derek Ouyang is the senior research manager at the RegLab at Stanford University.

Daniel E. Ho is the William Benjamin Scott and Luna M. Scott Professor of Law, Professor of Political Science and Computer Science (by courtesy), Senior Fellow at the Stanford Institute for Human-Centered Artificial Intelligence (HAI), Senior Fellow at the Institute for Economic Policy Research (SIEPR), and director at the RegLab.

Note: The authors thank Jen Pahlka for sparking the collaboration between RegLab and the City of San Francisco, and Andrea Bruss, Nancy King, and Caroline Meinhardt for providing feedback on this policy brief.

Stanford University
Human-Centered
Artificial Intelligence