

Senior Data Engineer, Legal Insights

Location: Hybrid – NYC metro. 4 days remote; one in-person collaboration day each week. Remote 4 days; one in-person collaboration day each week.

Role Type: Full-time

Compensation: \$115 k–\$145k

Visa sponsorship: Not available

Experience: 3–6+ years of technical data work

About Scrutinize

Scrutinize uses data and AI to increase transparency and accountability in state courts. We collect court records, interpret them, and publish accessible data. Our work helps decision-makers, voters, journalists, and advocates to evaluate judges, their decisions, and impacts.

Launched in 2023, our work has already influenced [judicial reappointments](#) and shaped debates on court reform across New York. We got our seed funding from [Emergent Ventures](#), are part of Fast Forward's 2025 [cohort](#), and received the inaugural [Legal AI Innovators Award](#) by Ironclad.

People who thrive at Scrutinize embrace hard work, rigorous thinking, and clear-eyed honesty. They communicate directly and transparently, welcome feedback, and are comfortable operating in an environment where resources are lean but goals are set high. They are drawn to mission-driven work, thrive in interdisciplinary teams, and care deeply about real-world impact.

To learn more, visit www.scrutinize.org.

About The Role

You will manage our entire data pipeline—from data collection through analysis to public release. This is an integral role on our founding team that will help us expand our proof of concept and take it national.

Core Responsibilities

- **Collect & parse.** Scrape legal documents and extract structured data using regex and AI prompt-based approaches.
- **Analyze.** Collaborate with attorneys to define metrics and surface trends in judicial behavior and court outcomes from the data.
- **Error-proof.** Design validation tests and manually review edge cases to ensure accuracy and reproducibility.
- **Improve infrastructure.** Evaluate tools, guide migrations, and make our workflows faster, cleaner, and less fragile.
- **Outreach.** Present results clearly to lawyers, reporters, and advocates—most of whom don't speak code.
- **Document.** Write documentation that others can actually use—whether it's for onboarding or troubleshooting.

Wild Success (First Year)

- **Court Data Ownership:** You cleaned, structured, and documented one state's messy court data to prep it for public release.
- **AI Prompting:** You developed reliable GPT prompts that replicate how expert lawyers analyze appellate law.
- **Metric Innovation:** You designed and launched 3–4 original metrics that now power public reporting and advocacy.
- **Public Leadership:** You represented Scrutinize in a major public setting—speaking at a conference, briefing a legislator, or contributing to a media piece.
- **Legal Fluency:** You built trusted relationships with attorneys and deepened your knowledge of legal systems and data.

Minimum Qualifications

- 3–6 years of experience working with large, messy datasets, especially text-heavy ones.
- Advanced fluency in regex, Python, SQL, and R.
- Proven experience scraping complex or protected sites (rate limits, JavaScript, captchas).

- Up to date knowledge and strong experience in AI prompting and knowledge of the various platforms and models' strengths and weaknesses.
- Strong communicator—clear, direct, and collaborative. You speak up when you have a view, listen when others do, and follow direction when the team makes a call.
- Meticulous attention to detail; you take data accuracy seriously.
- Comfortable with ambiguity and fast-changing priorities.
- Comfortable owning high-impact projects in a fast-paced environment, and willing to put in extra effort when a deadline or mission-critical moment calls for it.
- Strong alignment with our mission of increasing judicial accountability through data.
- Bachelor's degree or related experience in computer science, engineering, economics, statistics, or a related field.

Preferred Qualifications

- Experience in a startup or small-team setting, especially with resource-constrained workflows.
- Prior management experience or desire to grow into a technical management role.
- Familiarity with legal texts, court data, or justice-related policy issues.

Compensation

We use a transparent band tied to role scope. The range for this position is \$115k–\$145k.

Benefits

Scrutinize is part of the Urban Justice Center (UJC), a nonprofit organization based in New York City. You will have full access to UJC's benefits, including: 401(k) match, exceptional medical and dental insurance, paid vacation days, paid personal days, paid sick days, paid holidays, and an additional paid week off between Christmas and New Year's Day.

Work Style

- We operate remotely by default.
- One weekday each week we meet in person—usually 3-4 hours in our Wall Street office or a Manhattan café.
- We believe in thoughtful intensity: high responsibility, clear priorities, and space to recharge. You'll have autonomy in managing your time, and we prioritize sustainable work over fire drills.

More About The Role

Your primary tasks will involve scraping large volumes of court documents from state judicial websites; converting unstructured text into clean, structured datasets (judge names, appeal status, etc.); using regex and custom prompts with GPT to classify data based on legal issues and outcomes (constitutional violations; excessive sentences; reversals); and performing thorough quantitative analyses on trends in bail decisions, sentencing patterns, and appellate outcomes.

You'll also design validation tests and perform manual reviews to catch and fix errors, ensuring accuracy and reliability. On infrastructure, you'll evaluate, implement, and document better solutions for storage, versioning, and workflow management as we scale our operations.

This role sits at the intersection of engineering, data science, and practical legal analysis. You'll regularly consult attorneys to ensure accuracy of legal categorizations, communicate your findings clearly to journalists and advocates, and iteratively refine outputs based on stakeholder feedback. Comfort with ambiguity, a rapid workflow, and problem-solving are crucial to succeeding in this position.

Application Process

Send one PDF named Lastname_Firstname_LDE.pdf to info@scrutinize.org with the subject line "Senior Data Engineer, Legal Insights – [Your Name]".

The PDF, in this order, should contain:

- Cover letter (≤ 1 page). Tell us why Scrutinize, why this role, and what aspect of our work most appeals to you.
- Resume (≤ 2 pages).
- Portfolio: link to GitHub, Kaggle, or a public repo or attach ≤ 200 lines of representative code showing relevant work.

- References: names, titles, and emails for three people. We will contact them only after the final interview.

Scrutinize recruits, employs, trains, compensates and promotes regardless of race, religion, color, national origin, sex, sexual orientation, disability, age, veteran status, ancestry, citizenship, marital status, gender identity and all protected status as required by applicable law. We consider qualified applicants regardless of criminal histories, consistent with legal requirements. If you have a special need that requires accommodation, please let us know in your application. Even if you don't think you meet all the criteria above, drop your resume, and we'll take a look – you might be great for another role or another time!