

## M2 thesis projects

The projects are at the intersection of machine learning, law, and causal inference. Applicants should be familiar with or interested in learning about new tools and comfortable with computing.

Example papers are listed in [1](#) and presented in [2](#).

1. Applying [1](#), [2](#), and [3](#) to the judicial texts or speech (phrases, citations, audio) and analyzing the causal impacts on their votes and society-wide outcomes.
2. [Deep learning](#) of judicial decisions (improving on Lasso [1](#), low-dimensional [2](#), and high-dimensional [3](#)) for useable research and policy [app](#) (that uses exogenous variables in law due to random assignment of judges to measure the causal effect of policy choices on socioeconomic outcomes, like environmental regulation on pollution and labor regulation on wages). Auxiliary project on automated case classification and citation, dissent, or sentence length prediction.
3. Historically extend [1 2](#) back to 1955, [3](#) for women, combine audio with textual data [4 5](#), model social dynamics in audio (not just textual [6](#)) data in oral arguments [7](#). Structural models of network formation [8](#).
4. Examine influence of financial payments, [markets](#), [politics](#), and campaign [ads](#) on judicial text, development of [law](#), and legal [schools](#) of [thought](#).
5. Characterize judicial [learning](#), influence (adoption of legal innovations), ambiguity, or cognitive decline, using repeated random peer assignment. Estimate modes of reasoning - obligation, authority, rights, and duties.
6. Production of justice: time of day [1](#), peer effects [2](#), residual deterrence [3](#), sequential contrast [4](#), and prosecutorial discretion [5](#). Identify characteristics noisy to human prosecutors and what prosecutors maximize [6](#).
7. Theoretically and experimentally develop new ways [1 2](#) of conducting surveys and revisit questions like what shapes policy preferences and how to model [them](#).
8. SAS project involves physician conflicts of interest and malpractice with 24 million physician payments identified by doctor name and NDC-9 drug code and patient outcomes (in U.S. Medicare database). Analysis of scientific reproducibility crisis linking "[p-curve](#)" of scientific articles by NDC-9 drug code to payments, subsequent FDA warnings, and patient outcomes.

Up to 12 terabytes of data are available, mostly not analyzed before.