

Center for Text Analytic Methods in Legal Studies

- Kevin D. Ashley, School of Law, LRDC, ISP
- James M. Anderson, RAND Institute for Civil Justice
- Daqing He, Informatics and Networked Systems, SCI, ISP
- Rebecca Hwa, Computer Science, SCI, ISP
- Diane J. Litman, Computer Science, SCI, LRDC, ISP

Motivation

- Legal texts now seen as data:
 - Court decisions, statutes, and regulations.
- New sources of legal data are available:
 - Harvard Caselaw Access Project.
 - 6.7 million US cases.
- New higher resolution text analytic techniques
 - Transformer language models (e.g., BERT)
 - Extract case topics, legal areas, timeline events, legal tests, and outcomes.
 - Identify bias in legal language.
- Center connects:
 - Pitt Law & RAND legal domain experts and
 - Pitt NLP/ML/text analytics researchers.
 - Researchers at Duquesne Law and Worcester Polytechnic Inst.

Project Description

- Identify socially relevant empirical legal hypotheses testable via text analytics.
- Adapt and apply text analytics to extended legal data sets to evaluate hypotheses.
- Based on results, publish papers and apply for external funding to sustain Center's research.

Context

- Empirical legal studies were constrained to manual text analyses.
- Pitt team experienced in developing / applying new text analytics:
 - Natural language processing
 - Machine learning
 - Applied to legal cases and statutes.
- Rand team and Pitt Law faculty experienced in identifying / evaluating socially relevant empirical legal hypotheses.

Apply new analytic techniques to new sources of legal text data to evaluate legal hypotheses in ways not previously possible.

Project Deliverables

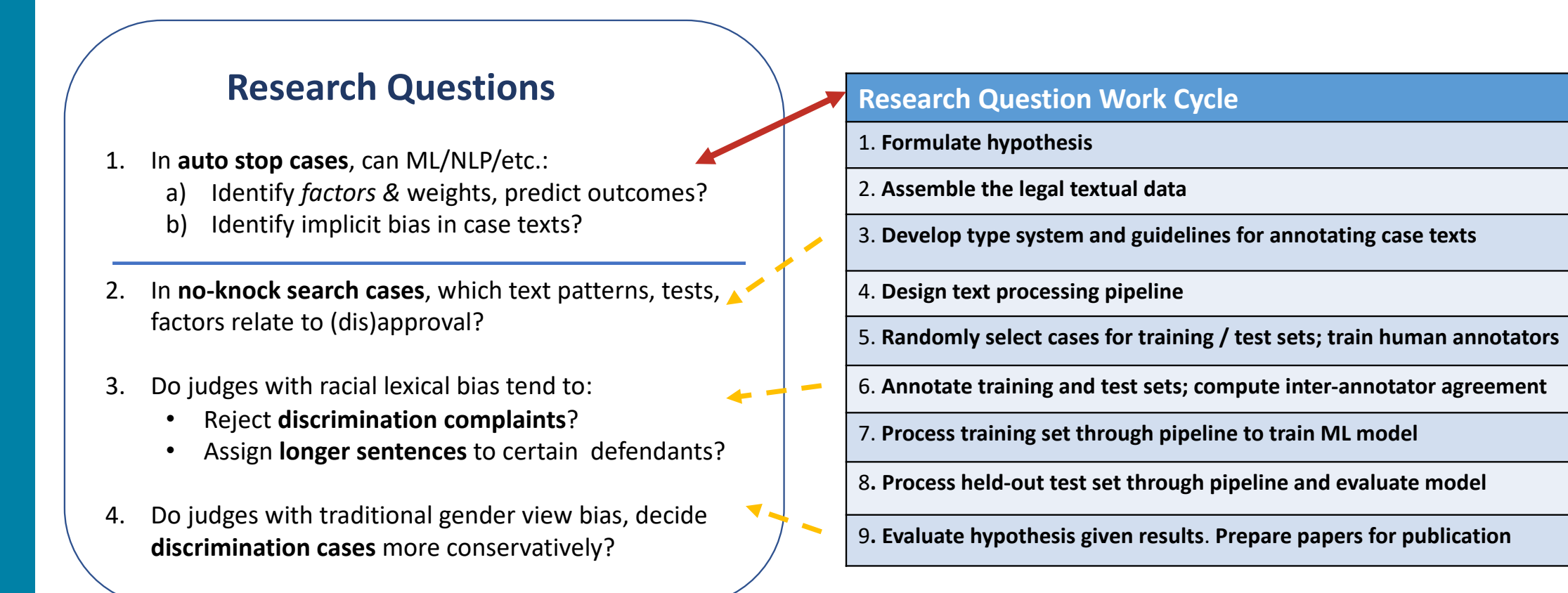


Initial Domain: Constitutionality of police automobile stops to search for drugs.

- Courts assess if police had reasonable suspicion that motorist is transporting drugs.
- “Driving while black” Problem: police may stop motorist due to racial bias.

Hypotheses: Machine learning models can:

- learn to identify factors in opinion texts and
- compute weights that courts assign to factors.
- assess likelihood that court will find facts sufficient for search.
- investigate racial bias in court decision texts.



Potential Impact

- Focuses on important hypotheses.
 - re social issues: racism, gender equality, immigration, public health, crime, or education.
- Policy implications for courts, police, civil society.
- Engages current literature.
 - Empirical legal research field awakening to methodological possibilities of text analytics.
- Pedagogical opportunities.
 - Motivate / engage law and prelaw students.
 - Introduce students to technology and to socially-relevant legal issues.
 - Help students read legal decisions more effectively.
- Focal center for research on text analytics and legal studies in Pittsburgh.

References

Michael Livermore and Daniel Rockmore (eds.) *Law as Data: Computation, Text, & the Future of Legal Analysis*. (2018). Santa Fe Institute of Science.

Acknowledgements

Thanks to the University of Pittsburgh School of Law, School of Computing and Information, RAND Corp., and Pitt Momentum Funds.

