The Challenges of Data Mining and A.I. for State Laws and Legislative Research: Recommendations for the Development of Legislative A.I.

Sarah E. Ryan, JD, PhD, Esq   & Lingzi Hong, PhD
Departments of Information Science & Data Science

Sarah Ryan, JD, PhD, Esq

Mixed methods research design for law and social science; criminal and veterans law expert

Lingzi Hong, PhD

Mixed methods research design for AI application development; RAG-RL, LLM development expert

Topics

1 **Statutes: The Skeleton of our Legal System**
Statutes as the skeleton of law, foundation of cases ♦ Statutory construction: The First Step Act

2 **Building an (AI) Judge: Scalia or Ginsburg?**
Textualists vs. Purposivists ♦ Textual and inter-textual differences and pathways

3 **Pilot Project**
Six state law problems ♦ Classifying law types (budget, public, private)

4 **Furthering the Conversation about DM & AI for Statutes**

Adobe stock image

Statutes, the Skeleton of Law

- Article I of U.S. Constitution: Legislative powers

- Statutes set the elements of crimes and torts, the boundaries of property law, define groups under the law, and more
  - Veterans
  - Seniors
  - Disabled
  - Children
  - Married

  Groups defined by statute

- Statutes gird most **judicial** filings and opinions
  - Civil complaints list the statutes under which the Plaintiff seeks redress
  - Criminal charges are brought under statutes, such as 18 U.S. Code § 1659 - Attack to plunder vessel

- Statutes create and sunset **executive** agencies and prescribe the boundaries of their authority

UNITED STATES DISTRICT COURT
DISTRICT OF CONNECTICUT

GILEAD COMMUNITY SERVICES, INC.,   :   Case No.:   3:17-cv-627
RAINBOW HOUSING CORP. and THE   :
CONNECTICUT FAIR HOUSING CENTER, INC.:
                                    :
                                    :
PLAINTIFFS,   :
                                    :
v.   :
                                    :
TOWN OF CROMWELL, ENZO FAIENZA,   :
individually and in his official capacity as Mayor :
of the Town of Cromwell,   :
ANTHONY SALVATORE, individually and in his :
official capacity as Town Manager of the Town of :
Cromwell, and JILLIAN MASSEY, in her official :
capacity as the Zoning Enforcement Officer of the :
Town of Cromwell   :
                                    :
DEFENDANTS.   :   April 17, 2017

COMPLAINT

I. INTRODUCTION

1.   Plaintiffs Gilead Community Services, Inc. and Rainbow Housing Corp. (hereinafter collectively "Gilead") and the Connecticut Fair Housing Center ("CFHC" and, collectively with Gilead, "Plaintiffs") bring this civil rights action, pursuant to the federal Fair Housing Act, the Americans with Disabilities Act, and the Rehabilitation Act of 1973, to remedy discriminatory actions by the Defendants that resulted in the denial of housing to Gilead's clients, who are people with disabilities, injured Gilead's reputation, finances and ability to assist its clients in moving out of segregated institutional settings and into housing in the community, caused CFHC to divert scarce resources to address this discrimination and frustrated CFHC's mission of ensuring equal access to housing for all. Plaintiffs seeks declaratory, injunctive and monetary relief.

1

"Plaintiffs... bring this civil rights action, pursuant to the federal Fair Housing Act [**42 U.S.C. § 3601 *et seq.***], the Americans with Disabilities Act [**42 U.S.C. § 12131 *et seq.***], and the Rehabilitation Act of 1973 [**29 U.S.C. § 794**]"

-*excerpt of federal civil rights complaint*

"defendants did conspire to distribute and possess... five grams or more of cocaine base (crack cocaine) in violation of **Title 21 United States Code, Section(s) 846, 841(a)(1) and 841(b)(1)(B)(iii)**"

-excerpt of fed. criminal complaint

Judicial Authority Under the First Step Act:
What Congress Conferred Through Section 404

Sarah E. Ryan

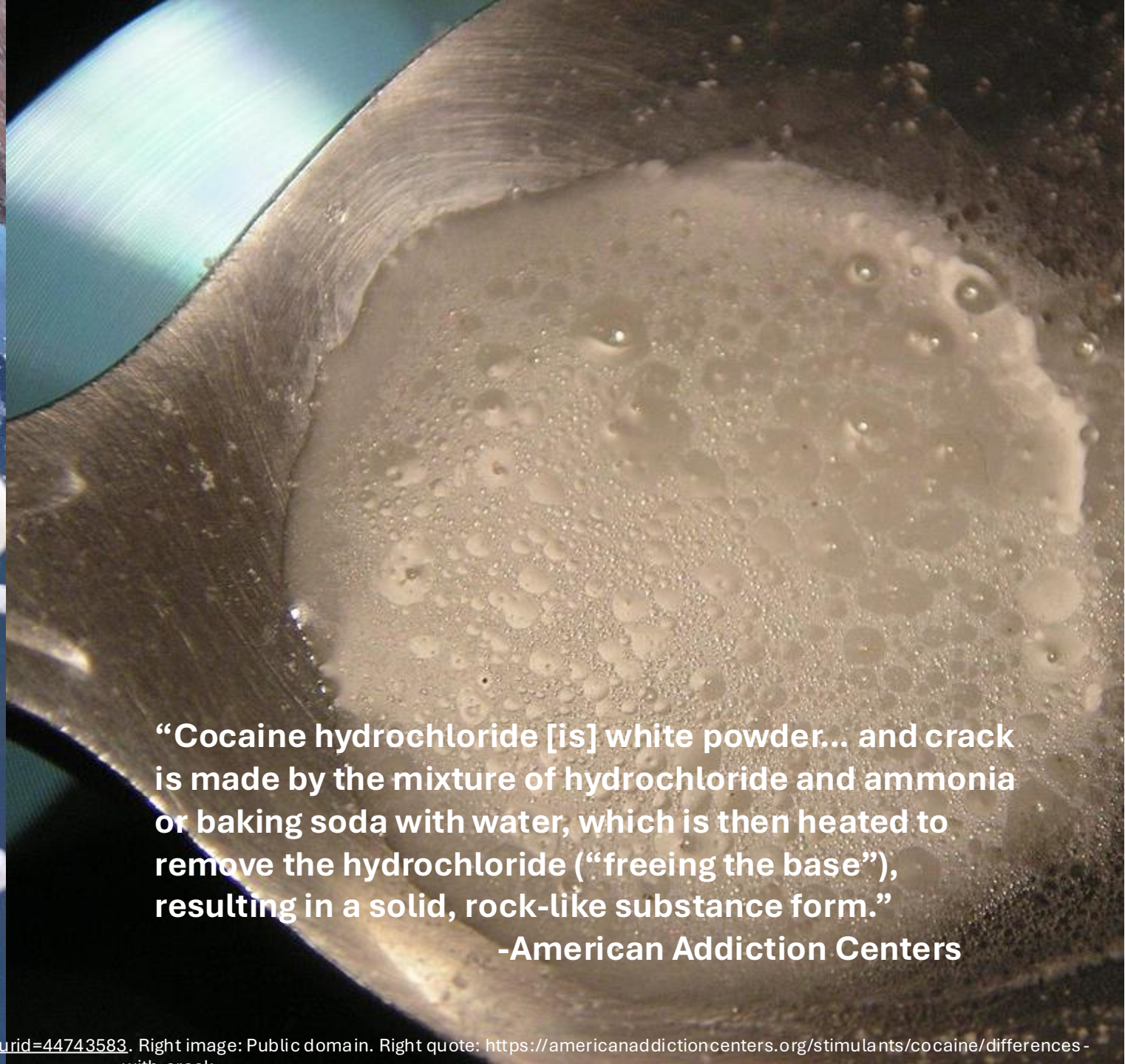
Judicial Authority Under the First Step Act: What Congress Conferred Through Section 404

*Sarah E. Ryan**

The First Step Act of 2018 promised relief to inmates serving disproportionately long sentences for cocaine base distribution. Section 404, the focus of this Article, seemed straightforward. But in the spring and summer of 2019, district judges began reviewing section 404 cases and reaching dissonant results. Appeals followed, focused on four questions of judicial authority: (1) Who may judges resentence?, (2) May judges engage in plenary resentencing or merely sentence reduction?, (3) May judges resentence all concurrent criminal convictions or only crack cocaine convictions?, and (4) Must judges adopt the operative drug quantity from the original sentencing? Today, the law of section 404 remains incomplete in every circuit. This Article reviews the legislative history, text, and legal context of section 404. It finds that Congress intended broad judicial authority in section 404 resentencings.

"Cocaine hydrochloride [is] white powder... and crack is made by the mixture of hydrochloride and ammonia or baking soda with water, which is then heated to remove the hydrochloride ("freeing the base"), resulting in a solid, rock-like substance form."
-American Addiction Centers

The Four Sentences that Set Thousands Free §404 of P.L. 111-220, "First Step Act" (2018)

- (a) Definition of Covered Offense.—In this section, the term "covered offense" means a violation of a Federal criminal statute, the statutory penalties for which were modified by section 2 or 3 of the Fair Sentencing Act of 2010, that was committed before August 3, 2010.

- (b) Defendants Previously Sentenced.—A court that imposed a sentence for a covered offense may, on motion of the defendant, the Director of the Bureau of Prisons, the attorney for the Government, or the court, impose a reduced sentence as if sections 2 and 3 of the Fair Sentencing Act of 2010 were in effect at the time the covered offense was committed.

- (c) Limitations.—No court shall entertain a motion made under this section to reduce a sentence if the sentence was previously imposed or previously reduced in accordance with the amendments made by sections 2 and 3 of the Fair Sentencing Act of 2010 or if a previous motion made under this section to reduce the sentence was, after the date of enactment of this Act, denied after a complete review of the motion on the merits. Nothing in this section shall be construed to require a court to reduce any sentence pursuant to this section.
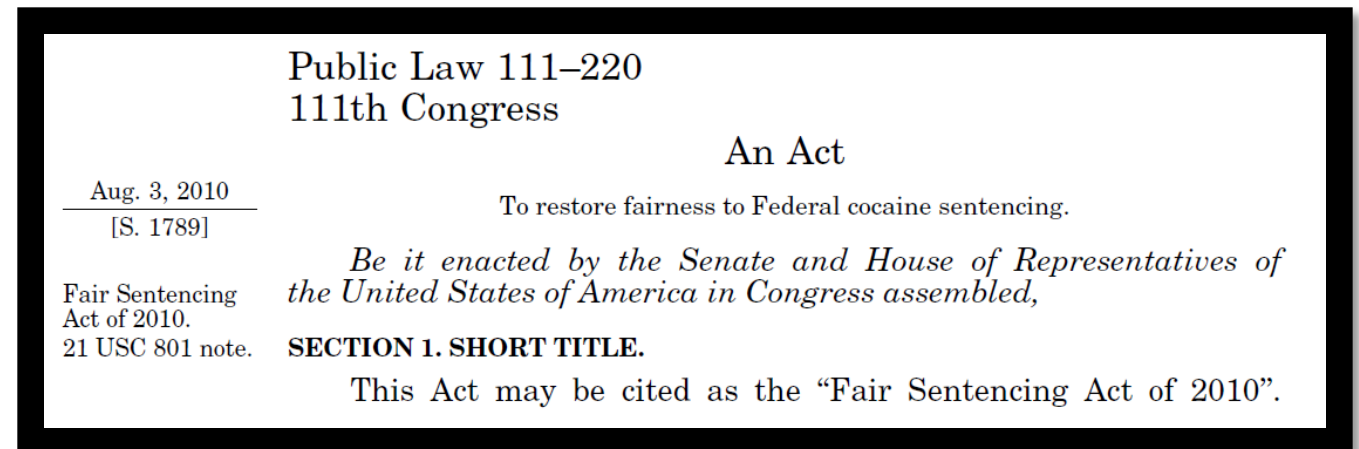
Sentence One

(a) Definition of Covered Offense.—In this section, the term "covered offense" means a violation of a Federal criminal statute, the statutory penalties for which were modified by section 2 or 3 of the Fair Sentencing Act of 2010, that was committed before August 3, 2010.

(a) Definition of Covered Offense.—In this section, the term "covered offense" means a violation of a Federal criminal statute, the statutory penalties for which were modified by section 2 or 3 of the Fair Sentencing Act of 2010, that was committed before August 3, 2010.

ACTION: Read the 2010 FSA law to determine which criminal offenses are included in this 2018 FSA law

Public Law 111–220
111th Congress

Aug. 3, 2010
[S. 1789]

Fair Sentencing
Act of 2010.
21 USC 801 note.

An Act

To restore fairness to Federal cocaine sentencing.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the "Fair Sentencing Act of 2010".

The Fair Sentencing Act of 2010

21 USC 801 note.

SECTION 1. SHORT TITLE.

This Act may be cited as the "Fair Sentencing Act of 2010".

SEC. 2. COCAINE SENTENCING DISPARITY REDUCTION.

(a) CSA.—Section 401(b)(1) of the Controlled Substances Act (21 U.S.C. 841(b)(1)) is amended—

(1) in subparagraph (A)(iii), by striking "50 grams" and inserting "280 grams"; and

(2) in subparagraph (B)(iii), by striking "5 grams" and inserting "28 grams".

(b) IMPORT AND EXPORT ACT.—Section 1010(b) of the Controlled Substances Import and Export Act (21 U.S.C. 960(b)) is amended—

(1) in paragraph (1)(C), by striking "50 grams" and inserting "280 grams"; and

(2) in paragraph (2)(C), by striking "5 grams" and inserting "28 grams".

Sentence One

i.e., Threshold sentence

(a) Definition of Covered Offense.—In this section, the term "covered offense" means a violation of a Federal criminal statute, the statutory penalties for which were modified by section 2 or 3 of the Fair Sentencing Act of 2010, that was committed before August 3, 2010.

ACTION:
Review the judgment to determine **what** crimes were committed and **when** defendant was sentenced

AO 245 B (Rev. 06/05)- Judgment in a Criminal Case

United States District Court
Western District of Michigan

UNITED STATES OF AMERICA	JUDGMENT IN A CRIMINAL CASE
-vs-	Case Number: 1:08-CR-65-01
WALTER GREENKING BOULDING	USM Number: 13637-040
	Lawrence J. Phelan Defendant's Attorney

THE DEFENDANT:

☐ pleaded guilty to_____.

☐ pleaded nolo contendere to Count(s) ___, which was accepted by the court.

☒ was found guilty on Counts 1 and 2 of the Superseding Indictment after a plea of not guilty.

Accordingly, the court has adjudicated that the defendant is guilty of the following offense(s):

Title & Section	Offense Ended	Count No.
21 U.S.C. § 846, 841(a)(1), (b)(1)(A)(iii) and 851	2/28/08	1
21 U.S.C. § 841(a)(1), (b)(1)(B)(iii) and 851	2/28/08	2

Nature of Offense

Ct. 1 - Conspiracy to Distribute and/or Possess With Intent to Distribute 50 Grams or More of Cocaine Base
Ct. 2 - Possession With Intent to Distribute 5 Grams or More of Cocaine Base

The defendant is sentenced as provided in the following pages of this judgment. The sentence is imposed pursuant to the Sentencing Reform Act of 1984.

☒ Original Indictment is dismissed on motion of the United States.

IT IS ORDERED that the defendant shall notify the United States attorney for this district within 30 days of any change of name, residence, or mailing address until all fines, restitution, costs and special assessments imposed by this judgment are fully paid. If ordered to pay restitution, the defendant shall notify the court and United States attorney of any material change in the defendant's economic circumstances.

Date of Imposition of Judgment: April 24, 2009

Dated: _____April 27, 2009_____

/s/ Robert J. Jonker _____
ROBERT J. JONKER
UNITED STATES DISTRICT JUDGE

Sentence One

(a) Definition of Covered Offense.—In this section, the term "covered offense" means a violation of a Federal criminal statute, the statutory penalties for which were modified by section 2 or 3 of the Fair Sentencing Act of 2010, that was committed before August 3, 2010.

BIG ISSUE 1: Many/most of these defendants were charged with more than one crime, e.g., crack distribution + an unlicensed firearm

ACTION 3: Analyze Congressional committee reports, etc.

Approach 1 (minority): Resentence when all charges fall under FSA 2010 (no guns!; *textualism)

Approach 2 (majority): Resentence when any covered charges fall under FSA 2010, including for non-covered offenses (guns) (because of Congress-intent; *purposivism)

Sentence Two
(i.e., Procedure sentence)

(b) Defendants Previously Sentenced.—A court that imposed a sentence for a covered offense may, on motion of the defendant, the Director of the Bureau of Prisons, the attorney for the Government, or the court, impose a reduced sentence as if sections 2 and 3 of the Fair Sentencing Act of 2010 were in effect at the time the covered offense was committed.

ISSUE 1: Getting a rehearing

ACTION:
Request a resentencing via motion or, through some novel procedure developed by the judge, have your case re-opened

Sentence Two

(b) Defendants Previously Sentenced.—A court that imposed a sentence for a covered offense may, on motion of the defendant, the Director of the Bureau of Prisons, the attorney for the Government, or the court, **impose a reduced sentence** as if sections 2 and 3 of the Fair Sentencing Act of 2010 were in effect at the time the covered offense was committed.

ISSUES/ACTIONS 2-4

Re-sentence the defendant using:

2a. The old **files/evidence** or

2b. Old files/evidence **+** e.g., prison conduct (rehabilitation)

3a. **Standards** outlined in the 2010 FSA only, including:
 1. reduced disparities in cocaine vs. crack sentencing
 2. no mandatory minimum for simple possession

3b. New standards, trends, ways of sentencing drug crimes

4a-b. Sentence **reduction** only or plenary (fresh slate) approach

Sentences Three and Four
(i.e., Limitations sentences)

(c) Limitations.—No court shall entertain a motion made under this section to reduce a sentence if the sentence was previously imposed or previously reduced in accordance with the amendments made by sections 2 and 3 of the Fair Sentencing Act of 2010 or if a previous motion made under this section to reduce the sentence was, after the date of enactment of this Act, denied after a complete review of the motion on the merits. Nothing in this section shall be construed to require a court to reduce any sentence pursuant to this section.

ACTION 1: Review the defendant's court records to ensure: haven't already received an FSA 2010 reduction, or been considered for an FSA 2010 reduction and denied on the merits. i.e., No second bite at the apple.

ACTION 2: May deny any/all reductions (e.g., under 18 U.S.C. § 3553)

3553: "Sentencing Factors"

18 U.S. Code § 3553 - Imposition of a sentence

U.S. Code   Notes   Authorities (CFR)

(a) Factors To Be Considered In Imposing a Sentence.—The court shall impose a sentence sufficient, but not greater than necessary, to comply with the purposes set forth in paragraph (2) of this subsection. The court, in determining the particular sentence to be imposed, shall consider—

> (1) the nature and circumstances of the offense and the history and characteristics of the defendant;

> (2) the need for the sentence imposed—

Minimum Law Needed to Interpret § 404 of FSA

First Step Act of 2018

(FSA 2018)

Fair Sentencing Act of 2010

(FSA 2010)

Sentencing Factors

(18 U.S.C. § 3553)

[Controlled Substances Act?]

(21 U.S.C. § 841)

[Congress committee reports, floor debates, etc. ?]

Building a Judge via Statutory AI

- Article III of U.S. Constitution: Judicial powers

- Authorizes a federal Supreme Court and inferior federal courts established by Congress

- Supreme Court has some original jurisdiction and appellate j- over **"Law and Fact"** (Art III.2.2). i.e., **Appellate judges are the interpreters of law**

- The rest is history… Article III structure expands, states largely copy it

Law (fact)

Law (fact)

Fact (law)

"No Vehicles in the Park"

Image created in Gemini AI:.... put all the vehicles in a park. that is: make me an image of at least ten types of wheeled vehicles including electric and regular bicycles, Vespas and scooters, a mid-sized car, a pickup truck, a semi-truck, and a unicycle AND make it all in a park

Textualists

Purposivists

What kind of (AI) judges are you building?

Words vs. People

Textualists

What the law **says**

What is linguistically more likely?

Purposivists

What the lawmakers were trying to **do or solve**

What is legislatively more likely?

Caveat

- Disappearing distinction, see, e.g., Harvard Law Today, 2025; Molot, 2006

- Still one of the best ways to frame LLM statutory interpretation pathways, choices, and political implications

Reference Document

Congressional Research Service. (2023, March 10). *Statutory interpretation: Theories, tools, and trends*. CRS: R45153.

*Historical and contemporary legislatures

Sarah's take

TEXTUALISTS

PURPOSIVISTS

| 1. Ordinary meaning (*common dictionaries) |
| 2. Statutory Context |

| 1. Ordinary meaning (*recent legis. docs.) |
| 2. Statutory Context |

2a. w/in this statute; 2b. in stat. scheme (18 U.S.C...)

| 3. Canons of construction |

3a. Semantic (grammar)
3b. Substantive (constitutionality)

| 3. Canons of construction |

| 4. Legislative history |

| 4. Legislative history |

*Legislature that passed this bill (coherence)

| 5. Statutory implement- |

| 5. Statutory implement- |

*Recent implementation by agency

Shared Documents (Steps 1 & 2)

Congressional Research Service. (2023, March 10). Statutory interpretation: Theories, tools, and trends. CRS: R45153.

TEXTUALISTS & PURPOSIVISTS:

U.S. Constitution and…

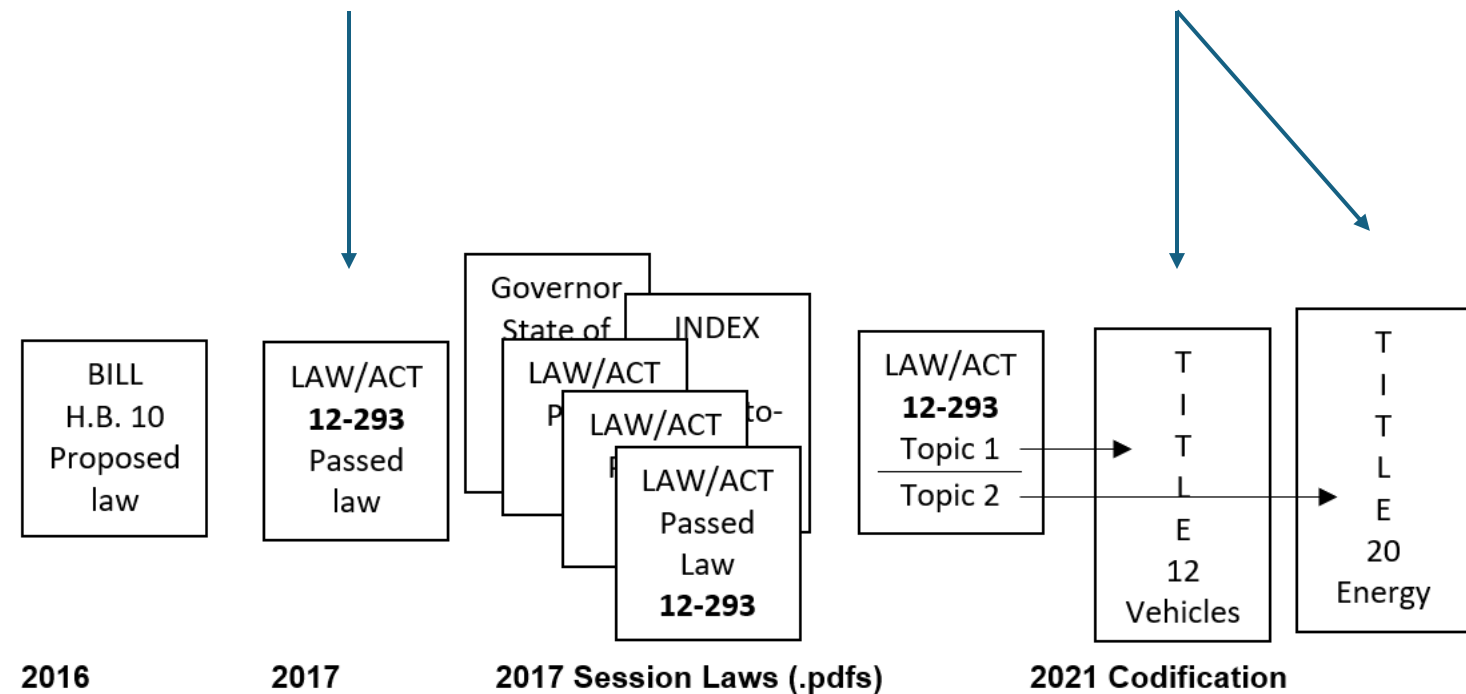Other sections of same law

U.S. Code section(s) where law will live

BILL H.B. 10 Proposed law

LAW/ACT **12-293** Passed law

Governor State of... LAW/ACT ... INDEX LAW/ACT ...co- LAW/ACT Passed Law **12-293**

LAW/ACT **12-293** Topic 1 — Topic 2 —

TITLE 12 Vehicles

TITLE 20 Energy

2016   2017   2017 Session Laws (.pdfs)   2021 Codification

Fig. 1: Process for a Single Law from Bill to Act to Collection in Session Laws to Topical Codification

50+ Canons of Construction (Step 3)

"**Semantic Canons**

- 1. "**Artificial-Person Canon**": "The word person includes corporations and other entities, but not the sovereign."

- 2. *Casus Omissus*: A matter not covered by a statute should be treated as intentionally omitted (*casus omissus pro omisso habendus est*).

- 3. "**Conjunctive/Disjunctive Canon**": "And" usually "joins a conjunctive list," combining items, while "or" usually joins "a disjunctive list," denoting alternatives." (copied w/out footnotes)

"**Substantive Canons**

- 1. **Canon of Constitutional Avoidance**: "[W]here an otherwise acceptable construction of a statute would raise serious constitutional problems, the Court will construe the statute to avoid such problems unless such construction is plainly contrary to the intent of Congress."

- 2. "**Dog that Didn't Bark**" **Presumption**: A "prior legal rule should be retained if no one in legislative deliberations even mentioned the rule or discussed any changes in the rule."

(CRS, 2023; footnotes removed)

Preferred, Different Documents (All Steps, esp. 4 & 5)

(Sarah's take)

TEXTUALISTS

-Common dictionaries (Merriam-Webster)

-Legal dictionaries (Black's, Ballentine's)

-Grammar, syntax authorities

PURPOSIVISTS

-Legislative history (committee hearings, legislative sponsor statements, legislator floor testimony)

-Enacting/enforcing agency language, rules, and procedures

-Statistics/empirical research

-News

"No Vehicles in the Park"

One possible outcome:

TEXTUALIST: No wheels
Merriam-Webster
vehicle **1:** a means of carrying or transporting something

PURPOSIVIST: No motors
Legislative history
-Committee report on motorized vehicle accidents, pollution

"No Vehicles in the Park"

Another possible outcome:

TEXTUALIST: <u>No motors</u>
American Heritage
vehicle 1b**:** b. A self-propelled conveyance that runs on tires; a motor vehicle.

PURPOSIVIST: <u>No motors</u>
Legislative history
-Committee report on motorized vehicle accidents, pollution

Textualists   Purposivists

Are you designing a Scalia or a Ginsburg?

Our Project: Solving Six State Law Problems

1. No uniform drafting rules (e.g., what is the first word of a law?)

2. Legislatures publish laws capriciously (e.g., PDFs, via vendors, etc.)

3. Open/public sites under and over-report laws in "session laws"
   a. Under: Only "public" law not private law, resolutions
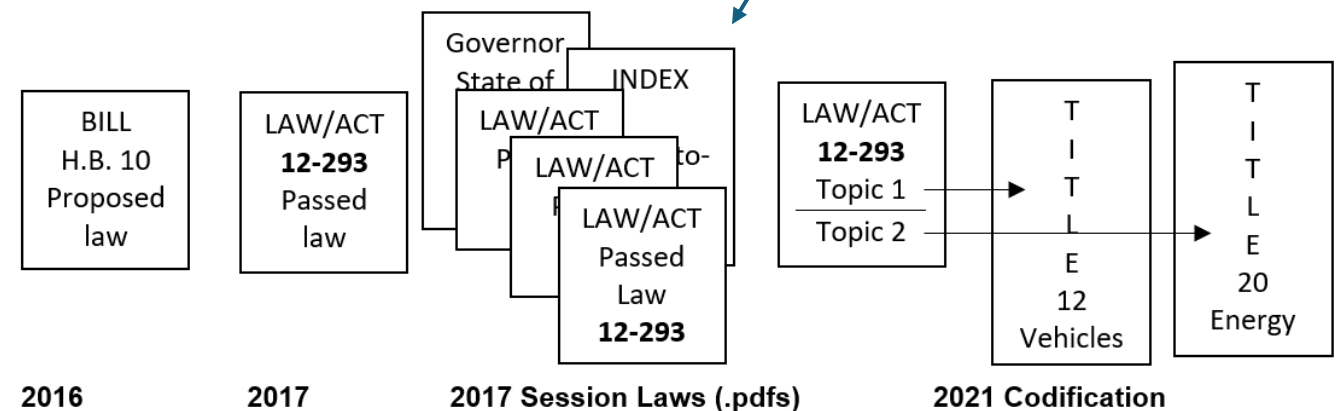   b. Over: Everything, even Governor's address to legislature

Fig. 1: Process for a Single Law from Bill to Act to Collection in Session Laws to Topical Codification

Our Project: Solving Six State Law Problems

4. Individual laws are complex:
    a. Groups/beneficiaries,
    b. Benefits,
    c. Conditions

5. Individual laws exist in concert with other laws (e.g., in the Code)
    a. Definitions section
    b. Rules, procedures
       e.g., agency X enforces)
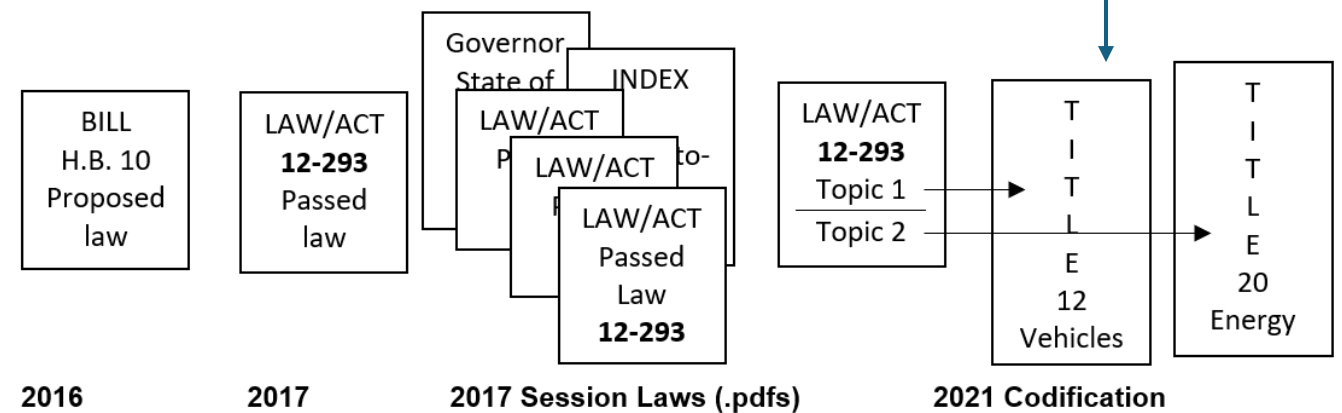   c. Close "cousins"

6. Legislative history too...

Fig. 1: Process for a Single Law from Bill to Act to Collection in Session Laws to Topical Codification

Advancement: Type Identification

To identify likely **budget laws**, we applied a **tokenization function** to count the number of tokens in each file. Noting that states have either annual or biennial budgets, and based upon **manual test coding, we estimated that about 5% of our collection would be budget acts**. Using the tokenization results, we calculated word count quantiles and set a threshold of greater than **40,000 words, which identified 331 files, or 4.69% of files, as likely to be budget acts.**

Advancement: Type Identification

.... After removing **331 files that were likely budget act**, we had **6,723 remaining files**.

To identify likely **private laws**, we .... deployed deep learning models to recognize named entities, specifically person names from texts. The **PERSON entities** were used to construct features such as number of PERSON entities, proportion of PERSON entities, and proportion of the most frequent PERSON entity. We also processed **pronoun features** such as "she, her, hers" and "he, him, his" (i.e., we did not use "they" pronouns as those would appear in public laws)....

Advancement: Type Identification

Of the remaining files, **3,210 did not contain any pronouns, so we concluded that they were highly likely to be public laws**.

That left us with **3,513 files** that were less likely to be budget acts **and could be... public or private laws**. We then applied a **K-means algorithm to separate the 3,513 law files according to their feature patterns. After the K-means clustering, we identified two clusters (C1: 533 files, C2: 2980 files**).

We then **manually-coded all Alabama files** (n=109) as public, private, or other, and compared the results to the machine output. The result was **>90% agreement** between the human and machine coding on which files were likely private laws.

Next Steps for Data, AI, and (State) Law Work

- Solve technical problems like "public law" identification
- Teach developers that legislative sentences can affect great change
- Teach developers that legislative text requires intertextual reading
- Carefully consider the corpus underlying the model (e.g., common dictionaries, continual refreshing with legislative committee reports)
- Acknowledge that a black box/probabilistic AI judge will serve up different outcomes for the same law and facts, and it is likely to skew Scalia
- Focus on what AI does well: brainstorming, identifying potential legal issues (i.e., versus yielding consistent outcomes or bending to rules)

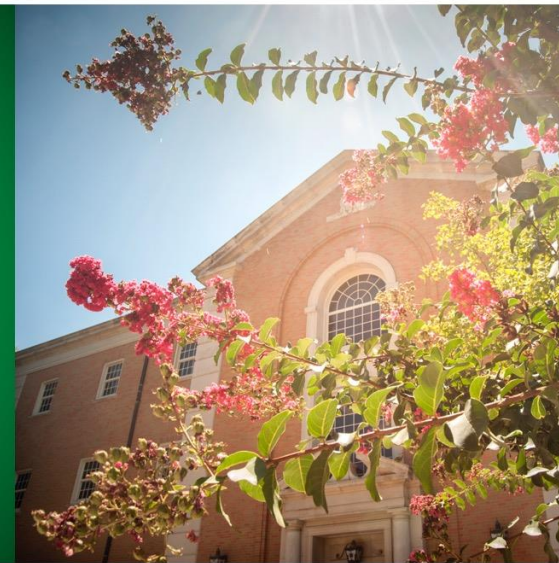Just like every other domain: legal AI is not neutral/apolitical

Statutes are the most important "next" frontier because they are the skeleton of law, including case law

Contact us

sarah.ryan@unt.edu

lingzi.hong@unt.edu

Our Departments

Sarah:
https://informationscience.unt.edu/index.html

Lingzi:
https://datascience.unt.edu/index.html

The Challenges of Data Mining and A.I. for State Laws and Legislative Research: Recommendations for the Development of Legislative A.I.

Sarah E. Ryan, JD, PhD, Esq & Lingzi Hong, PhD
Departments of Information Science & Data Science