GSA Office of Technology Transformation Services

Managing and Curating the AI Grand Challenge for Resilience

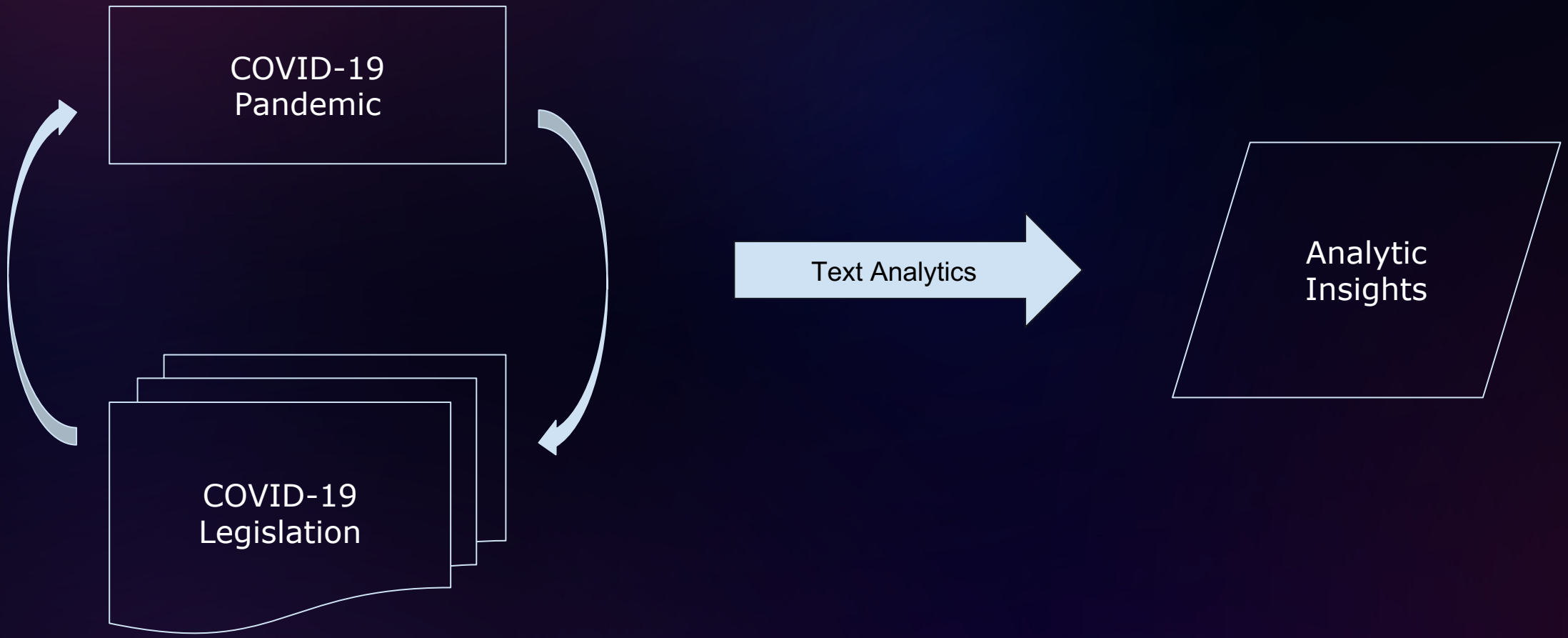
Bryan Lane, Maria Danilova, Steve Babitch, Jarah Meador, Joseph Raeteno, Melisa Marcovitz

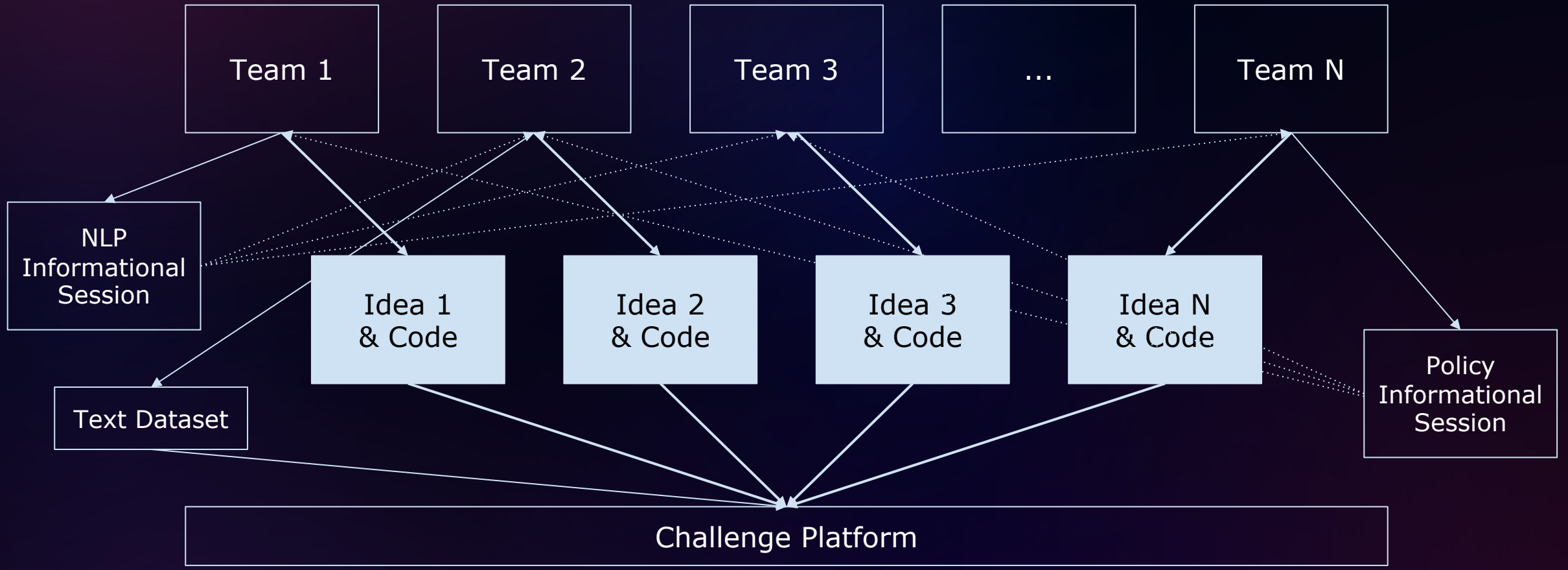
Presented on: 08/13/2021

Presented by:

Akhil Kondepudi -- University of Michigan, 2022

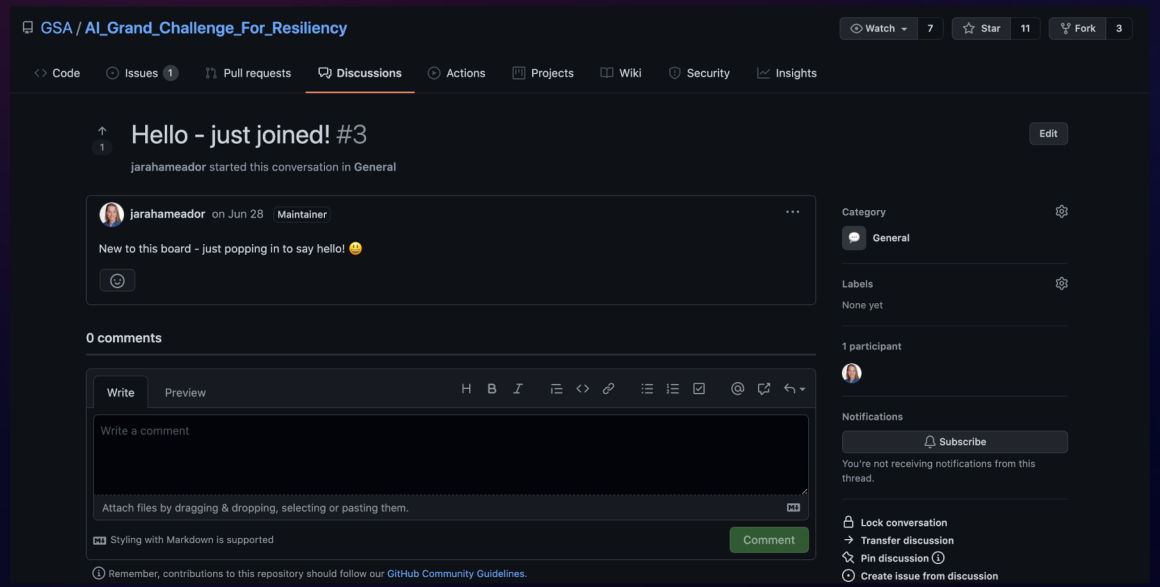
AI GRAND CHALLENGE FOR RESILIENCE

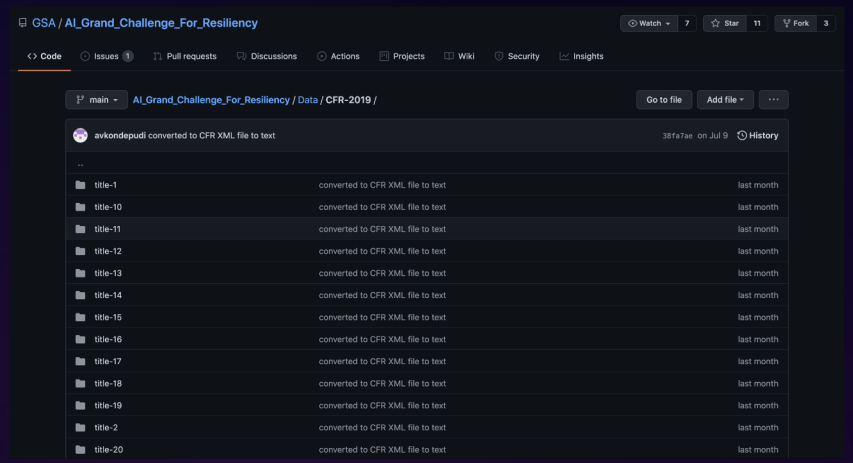

COVID-19
Pandemic
COVID-19
Legislation
Text Analytics
Analytic
Insights

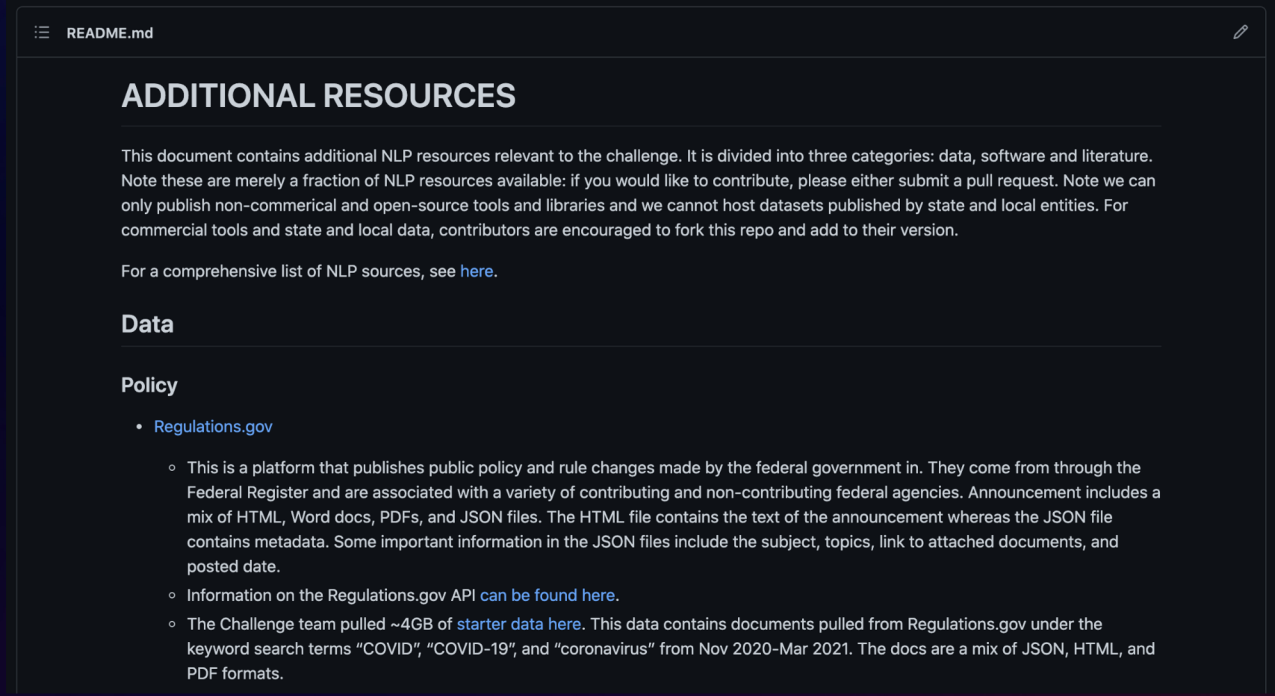

Team 1
Team 2
Team 3
...
Team N
NLP Informational Session
Idea 1 & Code
Idea 2 & Code
Idea 3 & Code
Idea N & Code
Policy Informational Session
Text Dataset
Challenge Platform

Hosting Platform: GitHub

AI GRAND CHALLENGE FOR RESILIENCE







Natural Language Processing

AI GRAND CHALLENGE FOR RESILIENCE

Data	Methods	Software
COVID-19 legislation, executive orders, titles	LDA, Word2Vec, deep learning	spaCy, Stanza, NLTK

01 Data

Problem

AI GRAND CHALLENGE FOR RESILIENCE

Unstructured

Not “machine-readable”

Variability within class

Not all information is useful for ML purposes

Data Access

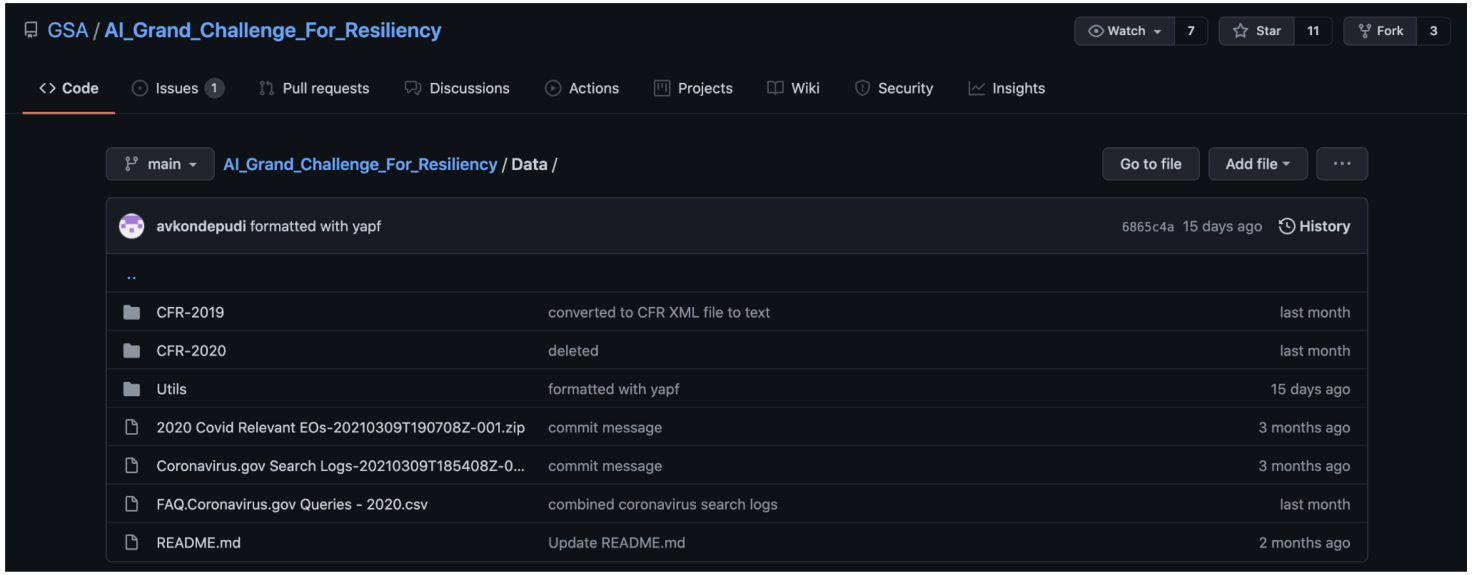
All data not present in one location

Exists across variety of formats (PDF, XML, etc.)

Hurdles to access & use data

Aggregated Data

Aggregated starter data for participants on GitHub



02 Methods

Problem

AI GRAND CHALLENGE FOR RESILIENCE

Natural language processing has had tremendous advances recently

However, applications to legislative data are mostly unexplored

- Methods may not be appropriate for use across multiple domains
- Results may require policy knowledge for interpretation

Literature Conferences

Compiled all resources and posted to GitHub page

A Frustratingly Easy Approach for Entity and Relation Extraction

Zexuan Zhong **Danqi Chen**
Department of Computer Science
Princeton University
{zzhong, danqic}@cs.princeton.edu

LEGAL-BERT: The Muppets straight out of Law School

Ilias Chalkidis [†‡] **Manos Fergadiotis** [†‡]
Prodromos Malakasiotis [†‡] **Nikolaos Aletras** [*] **Ion Androutsopoulos** [†‡]
[†] Department of Informatics, Athens University of Economics and Business
[‡] Institute of Informatics & Telecommunications, NCSR "Demokritos"
[*] Computer Science Department, University of Sheffield, UK
[ihalk,fergadiotis,rulller,ion]@aueb.gr
n.aletras@sheffield.ac.uk

JURIX
The Foundation for Legal Knowledge Based Systems

Policy & NLP Experts

Helped curate list of experts to hold information sessions, give domain-specific advice to participants, etc.

Regulations.gov API

Office of Government-wide Policy Overview

Sample Submission

Provide example of implementation of grading rubric, ideas

Part I: Team Information & Overview

Team Information
Bryan Lane[1], bryan.lane@gsa.gov
Akhil Kondepudi[1], akhil.kondepudi@gsa.gov
Asitang Mishra[2], asitang.mishra@jpl.nasa.gov
Masha Danilova[1], maria.danilova@gsa.gov

[1] Technology Transformation Services, General Services Administration
[2] Jet Propulsion Laboratory, NASA

Contributions
Akhil Kondepudi designed and wrote the submission. Mr. Bryan Lane and Mr. Asitang Mishra reviewed the submission. Ms. Masha Danilova was instrumental in facilitating collaborations for our project.

Statement of Interest
The team has a wide range of experience in both natural language processing and policy. This challenge provided us with an opportunity to test recent NLP advances to the policy domain and observe the results. Additionally, the team is extremely interested in the impact of regulations passed during the COVID-19 pandemic on the pandemic.

Part II: Submission Information

Title
Understanding regulations passed during the COVID-19 pandemic using text embeddings

Link to Forked Repository
<link>

Submission Overview
TTS AI Team's "Text Embedding" submission uses the Code for Federal Regulations, deep metric learning techniques like contrastive learning, and transformer models to characterize government-wide policy changes as a result of the COVID-19 pandemic.

Part III: Narrative Whitepaper

Project Overview
The COVID-19 pandemic spurred multiple regulations designed to protect industries' interests from the ensuing impact. These regulations can be found in the Code for Federal Regulations, a codification of the general and permanent rules published in the Federal Register by the departments and agencies of the Federal Government produced by the Office of the Federal Register (OFR) and the Government Publishing Office. Representation learning in the policy domain has been sparse, so we address this bottleneck by training various models with specific pretexts on this dataset to create embeddings at both the document and sentence level. We then tested the efficacy of these modes by training a one-layer classifier on top of the frozen model. Finally, we plotted these embeddings under various classifications to better understand the effect of the pandemic on regulations.

Related Work

Metric Learning
The type of representation learning used in this submission is metric learning. This method aims to learn embeddings such that similar data are closer to each other in the embedding space, and vice versa. Current metric learning applications in the deep learning space first train deep learning models to extract rich feature representations using specific tasks known as "pretasks." These tasks are not relevant to the main purpose of the model. Next, the model is finetuned on a "downstream" task to test the quality of features learned, often by training a linear classifier on top of the frozen pretrained model.

Contrastive Learning
In recent years, contrastive learning has transformed the field of representation learning, especially in the field of computer vision. The idea of contrastive learning is that using an data point as an "anchor," makes similar, or "positive," data points closer to the anchor in an embedding space and different, or "negative," data points further away from the anchor. When doing unsupervised learning, an example of a positive sample may be a transformation of the anchor, whereas a negative sample would be a transformation of another sample in the anchor's mini-batch. This seemingly facile approach can be compounded upon by using the labels: now, positive samples are transformations of samples in the same class as the anchor and vice versa. This technique has led to significant improvements in the field of computer vision. Its application to NLP has been slow due to the discrete nature of the textual data, but preliminary results have already shown improvements in this space as well. For example, Giorgi et al. utilized a contrastive approach to unsupervised learning on text, which resulted in similar results as supervised and semi-supervised methods on downstream tasks.

The BERT Model
Specifics can be found in Delvin et al. (https://arxiv.org/pdf/1810.04805.pdf)

Applied Metric Learning

<MLM + NSP; Transformers>

03 Software

Current Landscape

AI GRAND CHALLENGE FOR RESILIENCE

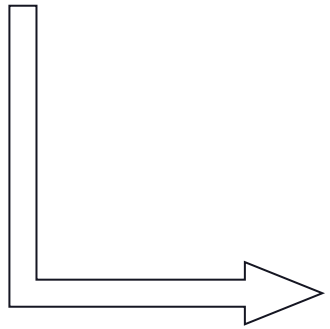
Libraries	Tools	Implementations
General, multipurpose NLP libraries	Perform specific task extremely well	Open-source code for state-of-the-art deep learning models

AI GRAND CHALLENGE FOR RESILIENCE

Regulations Parser

```
<?xml version="1.0" encoding="UTF-8"?>
<CFRDOC xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="CFRMergedXML.xsd">
  <AMDDATE>May 2, 2018</AMDDATE>
  <FMTR>
    <TITLEPG>
      <CODE/>
      <PRTPAGE P="1"/>
      <TITLENUM>Title 1</TITLENUM>
      <SUBJECT>General Provisions</SUBJECT>
      <REVISED>Revised as of January 1, 2019</REVISED>
      <CONTAINS>Containing a codification of documents of general applicability and future effect</CONTAINS>
      <DATE>As of January 1, 2019</DATE>
      <PUB>
        <P>Published by the Office of the Federal Register National Archives and Records Administration as a Special Edition of
        the Federal Register</P>
      </PUB>
    </TITLEPG>
    <BTITLE>
      <OENOTICE>
        <PRTPAGE P="?ii"/>
        <HD SOURCE="HED">U.S. GOVERNMENT OFFICIAL EDITION NOTICE</HD>
        <HD SOURCE="HED">Legal Status and Use of Seals and Logos</HD>
        <GPH DEEP="54" HTYPE="LEFT" SPAN="1">
          <GID>archives.ai</GID>
        </GPH>
        <P>The seal of the National Archives and Records Administration (NARA) authenticates the Code of Federal Regulations (
        CFR) as the official codification of Federal regulations established under the Federal Register Act. Under the
        provisions of 44 U.S.C. 1507, the contents of the CFR, a special edition of the Federal Register, shall be judicially
        noticed. The CFR is prima facie evidence of the original documents published in the Federal Register (44 U.S.C. 1510).</P
        >
        <P>It is prohibited to use NARA's official seal and the stylized Code of Federal Regulations logo on any republication
        of this material without the express, written permission of the Archivist of the United States or the Archivist's
        designee. Any person using NARA's official seals and logos in a manner inconsistent with the provisions of 36 CFR part
        1200 is subject to the penalties specified in 18 U.S.C. 506, 701, and 1017.</P>
        <HD SOURCE="HED">Use of ISBN Prefix</HD>
        <P>This is the Official U.S. Government edition of this publication and is herein identified to certify its
        authenticity. Use of the 0-16 ISBN prefix is for U.S. Government Publishing Office Official Editions only. The
        Superintendent of Documents of the U.S. Government Publishing Office requests that any reprinted edition clearly be
        labeled as a copy of the authentic work with a new ISBN.</P>
      </OENOTICE>
```

Parses title data from Code of Federal Regulations

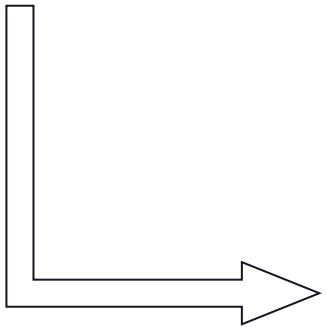


TITLE	DATE	TEXT	SECTNO	PART	PART DESC	PART NUMBER	SECT SUBJECT
Title 1—General Provisions	As of January 1, 2019	Any person may reproduce or republish, v	2.6	Pt. 2	PART 2—GENERAL INF(	2	Unrestricted use.
Title 1—General Provisions	As of January 1, 2019	(a) Documents filed with the Office of the	3.2	Pt. 3	PART 3—SERVICES TO	3	Public inspection of docu
Title 1—General Provisions	As of January 1, 2019	(a) Pursuant to chapter 15 of title 44, Unit	5.1	Pt. 5	PART 5—GENERAL	5	Publication policy.
Title 1—General Provisions	As of January 1, 2019	The following documents are required to b	5.2	Pt. 5	PART 5—GENERAL	5	Documents required to be
Title 1—General Provisions	As of January 1, 2019	Whenever the Director of the Federal Reg	5.3	Pt. 5	PART 5—GENERAL	5	Publication of other docu
Title 1—General Provisions	As of January 1, 2019	The Federal Register serves as a daily su	5.5	Pt. 5	PART 5—GENERAL	5	Supplement to the Code
Title 1—General Provisions	As of January 1, 2019	There shall be an edition of the Federal R	5.6	Pt. 5	PART 5—GENERAL	5	Daily publication.
Title 1—General Provisions	As of January 1, 2019	The Government Printing Office shall distr	5.7	Pt. 5	PART 5—GENERAL	5	Delivery and mailing.
Title 1—General Provisions	As of January 1, 2019	Without prejudice to any other form of cita	5.8	Pt. 5	PART 5—GENERAL	5	Form of citation.
Title 1—General Provisions	As of January 1, 2019	Each document published in the Federal I	5.9	Pt. 5	PART 5—GENERAL	5	Categories of documents
Title 1—General Provisions	As of January 1, 2019	Pursuant to section 1506 of title 44, Unite	5.1	Pt. 5	PART 5—GENERAL	5	Forms of publication.
Title 1—General Provisions	As of January 1, 2019	Each daily issue of the Federal Register s	6.1	Pt. 6	PART 6—INDEXES AND	6	Index to daily issues.
Title 1—General Provisions	As of January 1, 2019	Analytical subject indexes covering the cc	6.2	Pt. 6	PART 6—INDEXES AND	6	Analytical subject indexes
Title 1—General Provisions	As of January 1, 2019	(a) Each daily issue of the Federal Regist	6.3	Pt. 6	PART 6—INDEXES AND	6	Daily lists of parts affecte
Title 1—General Provisions	As of January 1, 2019	A monthly list of sections of the Code of F	6.4	Pt. 6	PART 6—INDEXES AND	6	Monthly list of sections af
Title 1—General Provisions	As of January 1, 2019	(a) The Director of the Federal Register m	6.5	Pt. 6	PART 6—INDEXES AND	6	Indexes, digests, and gui
Title 1—General Provisions	As of January 1, 2019	(a) Pursuant to chapter 15 of title 44, Unit	8.1	Pt. 8	PART 8—CODE OF FED	8	Policy.
Title 1—General Provisions	As of January 1, 2019	(a) Criteria. Each book of the Code shall b	8.3	Pt. 8	PART 8—CODE OF FED	8	Periodic updating.
Title 1—General Provisions	As of January 1, 2019	(a) The Director publishes a special editio	9.1	Pt. 9	PART 9—THE UNITED S	9	Publication required.
Title 1—General Provisions	As of January 1, 2019	(a) The Manual will contain appropriate in	9.2	Pt. 9	PART 9—THE UNITED S	9	Scope.
Title 1—General Provisions	As of January 1, 2019	The Director publishes a special edition o	10.1	Pt. 10	PART 10—PRESIDENTIA	10	Publication required.
Title 1—General Provisions	As of January 1, 2019	The Director of the Federal Register shall	10.1	Pt. 10	PART 10—PRESIDENTIA	10	Publication required.
Title 1—General Provisions	As of January 1, 2019	The Administrative Committee may autho	10.13	Pt. 10	PART 10—PRESIDENTIA	10	Coverage of prior years.
Title 1—General Provisions	As of January 1, 2019	(a) The subscription price for the paper ec	11.2	Pt. 11	PART 11—SUBSCRIPTIC	11	Federal Register.
Title 1—General Provisions	As of January 1, 2019	(a) The online edition of the Manual, issue	11.4	Pt. 11	PART 11—SUBSCRIPTIC	11	The United States Govern
Title 1—General Provisions	As of January 1, 2019	The annual subscription price for the mon	11.7	Pt. 11	PART 11—SUBSCRIPTIC	11	Federal Register Index.

Regulations Parser

```
<?xml version="1.0" encoding="UTF-8"?>
<CFRDOC xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="CFRMergedXML.xsd">
  <AMDDATE>May 2, 2018</AMDDATE>
  <FMTR>
    <TITLEPG>
      <CODE/>
      <PRTPAGE P="1"/>
      <TITLENUM>Title 1</TITLENUM>
      <SUBJECT>General Provisions</SUBJECT>
      <REVISED>Revised as of January 1, 2019</REVISED>
      <CONTAINS>Containing a codification of documents of general applicability and future effect</CONTAINS>
      <DATE>As of January 1, 2019</DATE>
      <PUB>
        <P>Published by the Office of the Federal Register National Archives and Records Administration as a Special Edition of
        the Federal Register</P>
      </PUB>
    </TITLEPG>
    <BTITLE>
      <OENOTICE>
        <PRTPAGE P="?ii"/>
        <HD SOURCE="HED">U.S. GOVERNMENT OFFICIAL EDITION NOTICE</HD>
        <HD SOURCE="HED">Legal Status and Use of Seals and Logos</HD>
        <GPH DEEP="54" HTYPE="LEFT" SPAN="1">
          <GID>archives.ai</GID>
        </GPH>
        <P>The seal of the National Archives and Records Administration (NARA) authenticates the Code of Federal Regulations (
        CFR) as the official codification of Federal regulations established under the Federal Register Act. Under the
        provisions of 44 U.S.C. 1507, the contents of the CFR, a special edition of the Federal Register, shall be judicially
        noticed. The CFR is prima facie evidence of the original documents published in the Federal Register (44 U.S.C. 1510).</P
        >
        <P>It is prohibited to use NARA's official seal and the stylized Code of Federal Regulations logo on any republication
        of this material without the express, written permission of the Archivist of the United States or the Archivist's
        designee. Any person using NARA's official seals and logos in a manner inconsistent with the provisions of 36 CFR part
        1200 is subject to the penalties specified in 18 U.S.C. 506, 701, and 1017.</P>
        <HD SOURCE="HED">Use of ISBN Prefix</HD>
        <P>This is the Official U.S. Government edition of this publication and is herein identified to certify its
        authenticity. Use of the 0-16 ISBN prefix is for U.S. Government Publishing Office Official Editions only. The
        Superintendent of Documents of the U.S. Government Publishing Office requests that any reprinted edition clearly be
        labeled as a copy of the authentic work with a new ISBN.</P>
      </OENOTICE>
```

Converts to human-readable format



May 2, 2018

Title 1
General Provisions
Revised as of January 1, 2019
Containing a codification of documents of general applicability and future effect
As of January 1, 2019

Published by the Office of the Federal Register National Archives and Records Administration as a Special Edition of the Federal Register

U.S. GOVERNMENT OFFICIAL EDITION NOTICE
Legal Status and Use of Seals and Logos

archives.ai

The seal of the National Archives and Records Administration (NARA) authenticates the Code of Federal Regulations (CFR) as the official codification of Federal regulations established under the Federal Register Act. Under the provisions of 44 U.S.C. 1507, the contents of the CFR, a special edition of the Federal Register, shall be judicially noticed. The CFR is prima facie evidence of the original documents published in the Federal Register (44 U.S.C. 1510).

It is prohibited to use NARA's official seal and the stylized Code of Federal Regulations logo on any republication of this material without the express, written permission of the Archivist of the United States or the Archivist's designee. Any person using NARA's official seals and logos in a manner inconsistent with the provisions of 36 CFR part 1200 is subject to the penalties specified in 18 U.S.C. 506, 701, and 1017.

Use of ISBN Prefix

This is the Official U.S. Government edition of this publication and is herein identified to certify its authenticity. Use of the 0-16 ISBN prefix is for U.S. Government Publishing Office Official Editions only. The Superintendent of Documents of the U.S. Government Publishing Office requests that any reprinted edition clearly be labeled as a copy of the authentic work with a new ISBN.

gpologo2.eps

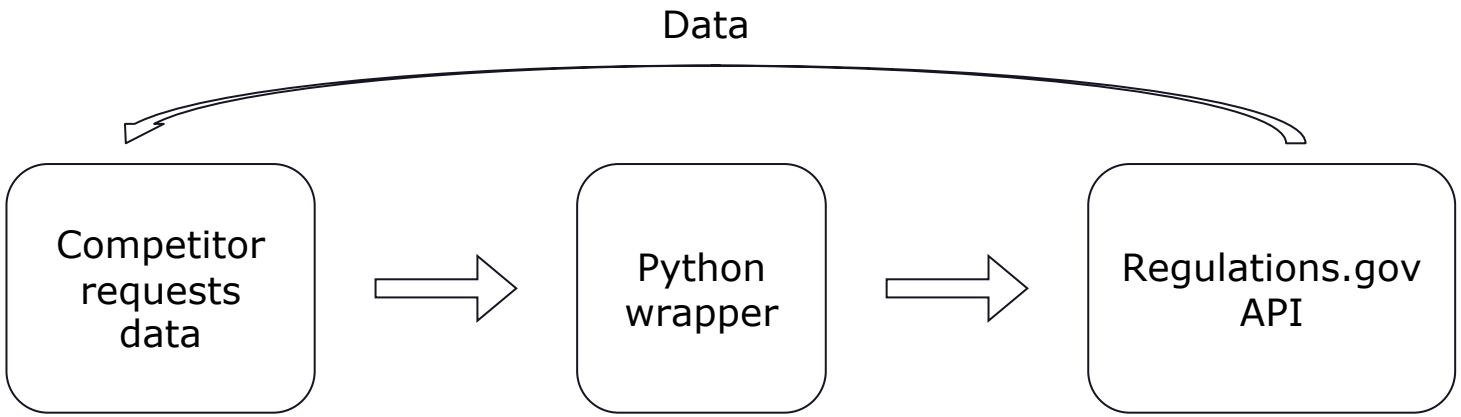
U.S.GOVERNMENTPUBLISHINGOFFICE

U.S. Superintendent of Documents • Washington, DC 20402-0001
http://bookstore.gpo.gov
Phone: toll-free (866) 512-1800; DC area (202) 512-1800

Regulations.gov API

Built Python wrapper around Regulations.gov API

Easier for participants to pull data, integrate into existing workflows



04 The Experience

Lessons I’ve Learned

AI GRAND CHALLENGE FOR RESILIENCE

Government

Slow but necessary process

- Contractors
- Administration
- Checkpoints

Impact is undeniable, fulfilling

Work/life balance emphasized

Data Science

Drawing meaningful conclusions from results

Working alongside others (e.g., policy experts) to implement changes as a result of analyses

- Results from challenge will change how legislation occurs, lead to increased funding for specific initiatives, etc.

My work would not have been possible without the following people:

- The AIGC Team (Bryan Lane, Masha Danilova, and everyone else!)
- The CIF Team (Rachel Dodell, Ariana Soto, DJ Jain)
- TTS (Annabel Lombard, Molly McIntyre)

A big thank you to all of you!

AI GRAND CHALLENGE FOR

RESILIENCE

Questions?