

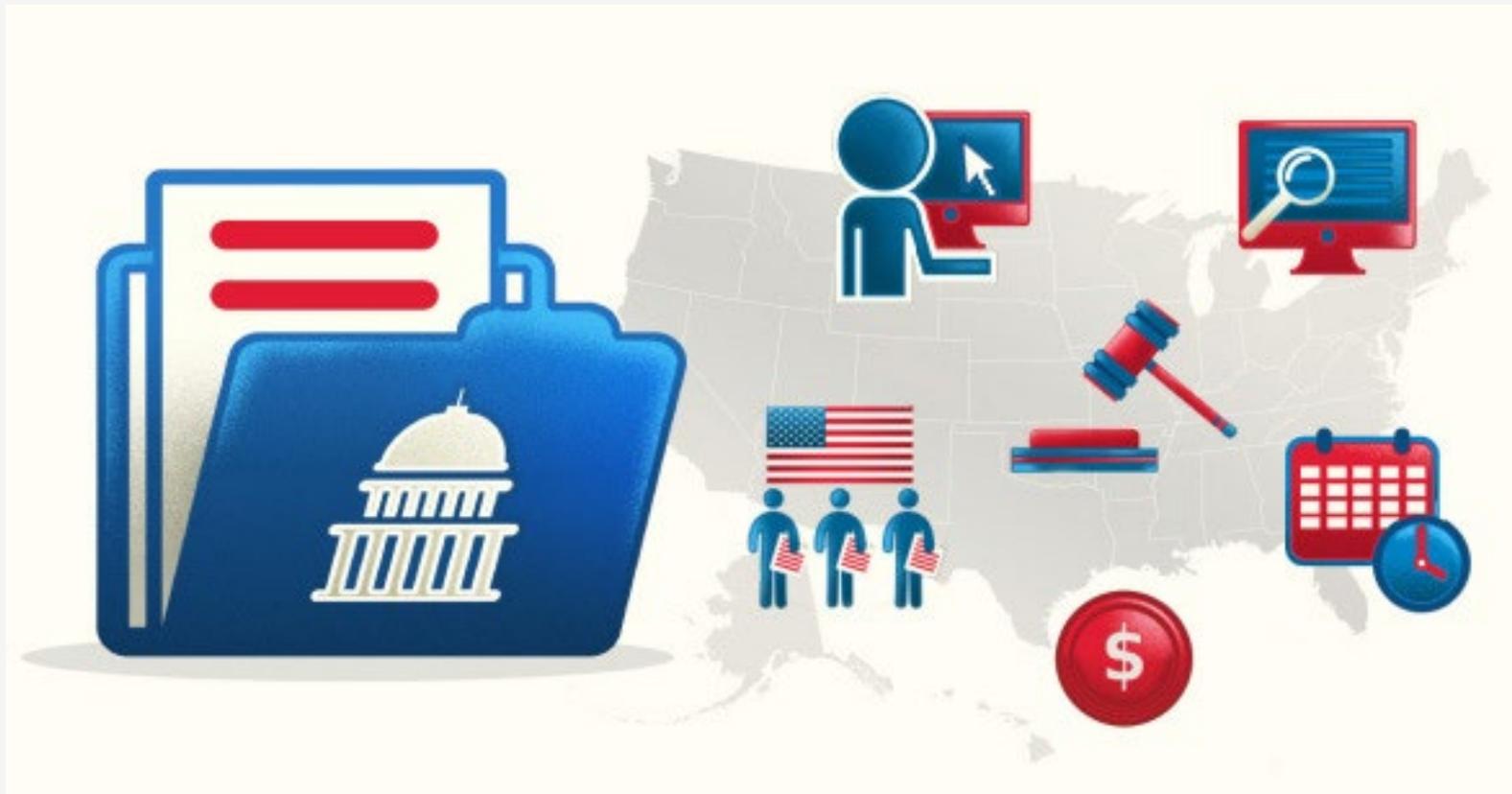
*CHIEF FOIA  
OFFICERS COUNCIL  
- TECHNOLOGY  
COMMITTEE –  
ARTIFICIAL  
INTELLIGENCE  
WORKING GROUP*

*AI IN FOIA OVERVIEW*

*Nick Wittenberg - Chair*

*Michelle McKown*

*Jennifer MacDonald*



*FOIA AI Overview:  
Backlogs, Boredom, Bodies*



L5/5 (A)

Barged Closed Files  
71399110

CONTACTS  
555  
FROM TO  
DESTROY SEE NO.

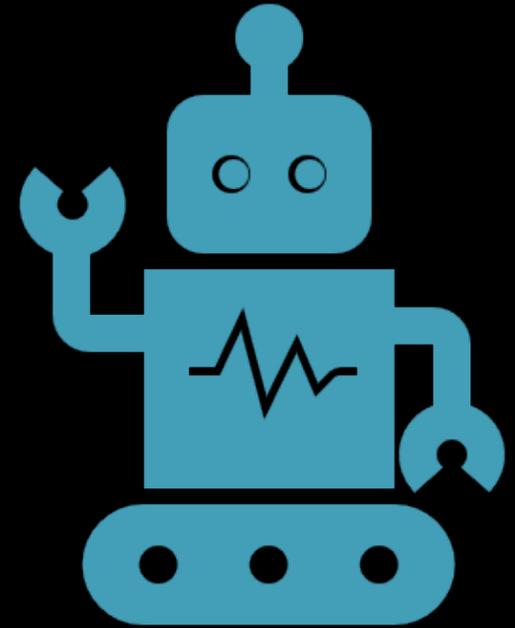
CONTACTS  
601  
FROM TO  
DESTROY SEE NO.

3

3

1368

*FOIA  
PROGRAMS  
LEVERAGE  
AI TO MEET  
GOALS*





ARTIFICIAL INTELLIGENCE:  
AUTOCORRECT



ROBOTIC PROCESS AUTOMATION:  
FIND AND REDACT

*What is AI?*

# *Basic Technological Solutions are Advancing*

---

Boolean

Filters

Custodian

Subject

Keyword

Search Within  
a Search

# *Tools are available to assist-*

---

- “Nearly all information created today is created in digital form, with dizzying arrays of technologies creating dizzying arrays of data types, at rates and in volumes that continue to grow geometrically, with an estimated **89 billion business** emails sent each day and large organizations now measuring data stores in **petabytes**. At the same time, advances in digital forensics, information retrieval, and other disciplines have yielded a plethora of tools that make it possible to conduct discovery in even the largest cases in a manner that is defensible, timely, and cost effective.”



*WHY AI?*

*DATA SIZE HAS GROWN*

*3 MEGAPIXEL CAMERA'S  
VS. XXL MEGAPIX*

*OUTSTRIPS AN INDIVIDUAL  
FOIA OFFICER'S ABILITY TO  
MANAGE DATA WITHOUT  
ROBUST TECH TOOLS*

# *Artificial Intelligence and FOIA*

---

- Artificial Intelligence (AI) has made litigation and FOIA more efficient and accurate
- Typically in litigation 10 - 15% Responsive\* Depends on population size, privileges, exemptions, and other sensitivities.
- Training the Machine:
  - Example: Train machine on 10,000 documents for a population of 100,000 documents. Training set used to code majority of population. However, maybe 20,000 documents that don't get coded because they are foreign language, corrupt files, excel, or other structured data. However, you still saved an incredible amount of time and were way more accurate with the 80,000 coded from the 10,000 document training set.

# *Artificial Intelligence and FOIA*

- Solution = Technology Assisted Review (TAR)
    - Predictive Coding – also known as TAR 1.0
    - Continuous Active Learning – also known as TAR 2.0
    - Cluster Visualization
-

# *Predictive Coding*

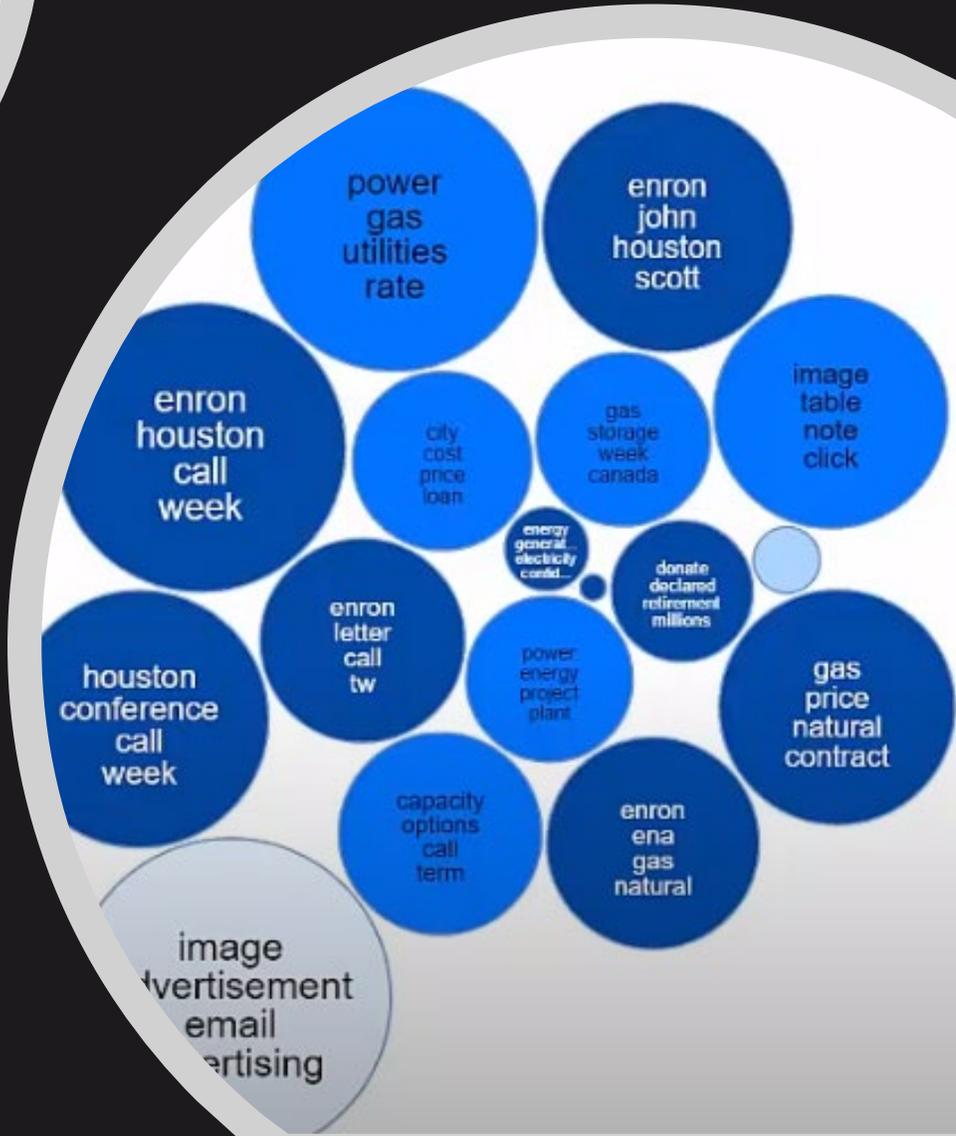
---

- Also know as Tar 1.0
- Decade old tech
- Accepted in Courts in 2012.
- Sample of review
- Need very senior person to help make decisions

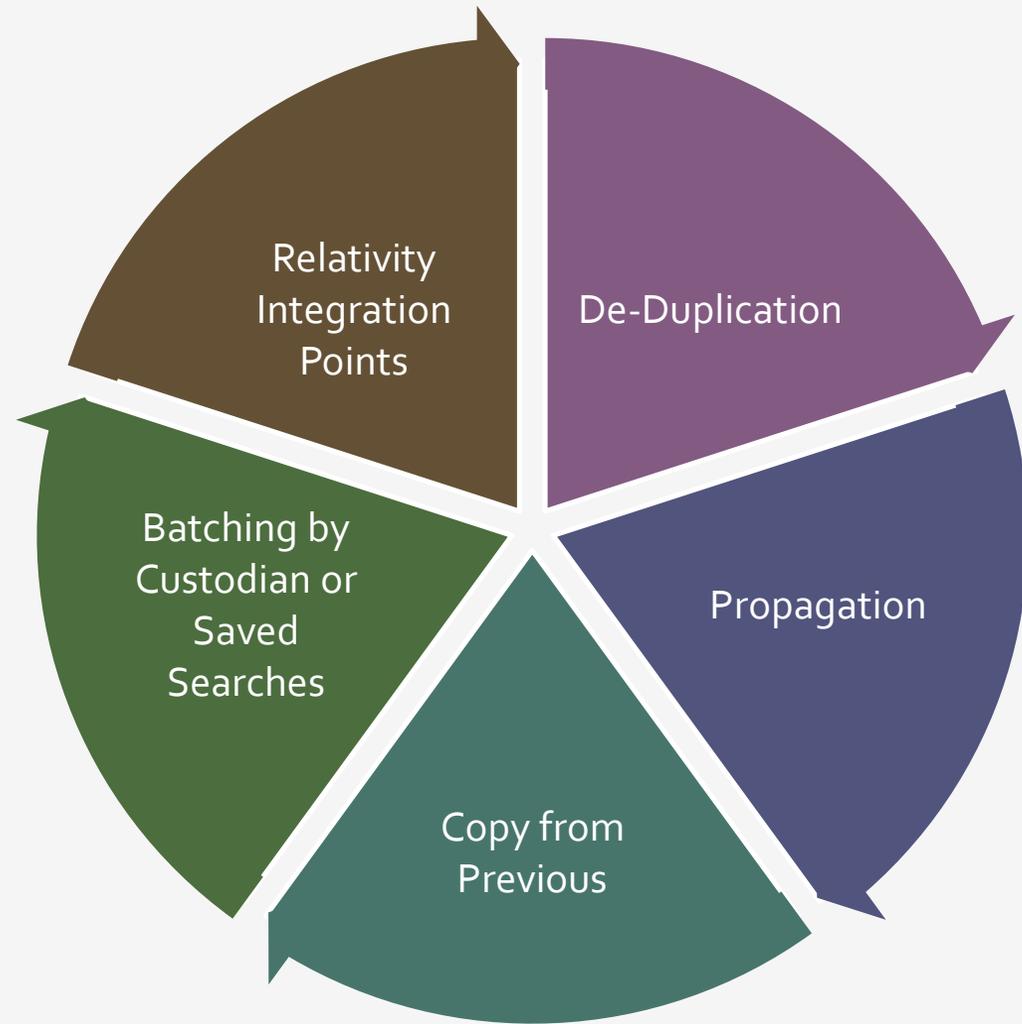
# *Continuous Active Learning*

---

- Also known as TAR 2.0
- Judge Peck- courts would allow continuous active learning if a party requested it
  - Start at doc 1
  - Doesn't take random samples throughout population like Tar 1.0
  - Based on decisions
  - Content is separated into piles



# Cluster Visualization



## *Features to Power FOIA Reviews*

---



*WHO ELSE IS  
USING IT, WHY?*



# *eDiscovery & FOIA*

- eDiscovery has been used in the legal realms for over a decade-Judges accept it
- Large Data cases caused a need for electronic discovery review where manually going through bankers boxes, files, and documents could not be done in a timely fashion
  - Enron, WorldCom, Arthur Andersen\*
- Software is dramatically updated over the years to make process more efficient and accurate
- Using the advancements appreciated in the legal world so too can FOIA appreciate the results

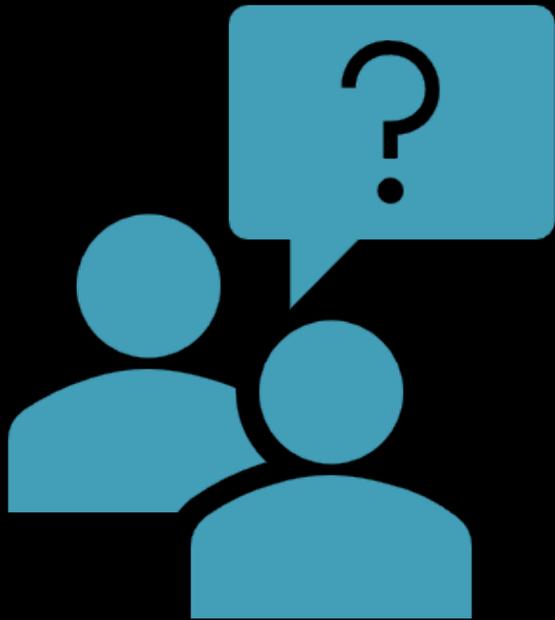
- Case – culling down to focus review.
  - 1,000 in batch, 200 likely responsive, 800 likely not
- Report to show
  - Responsiveness
  - Privilege
  - Exemptions
- Auto Redaction Software
- Exemption 1 and CBI productions can be updated at a later time when matters are determined to be downgraded.
- Reviewed and Produced can maintain same coding decisions and redactions
  - Future request – receive a pre-case assessment that states how many records are in this collection and how many have been reviewed and produced

# *Further Technological Search and Previously Produced Solutions*

*“I REALLY THINK THE FUTURE OF FOIA IS TAKING IT TO THE NEXT LEVEL AND USING ARTIFICIAL INTELLIGENCE, AND USING SOFTWARE THAT CAN DO THINGS LIKE GROUP RECORDS TOGETHER, EITHER BY CONCEPT OR RELATIONSHIP, THOSE KINDS OF SOFTWARE.”*

*MELANIE PUSTAY, DIRECTOR,  
DEPARTMENT OF JUSTICE’S  
OFFICE OF INFORMATION POLICY  
(OIP), MAY 03, 2019*





# *QUESTIONS*

*Nick Wittenberg*

*Nicholas.D.Wittenberg@ostp.eop.gov*