

**UNITED STATES  
PATENT AND TRADEMARK OFFICE**



# Patent Public Advisory Committee Quarterly Meeting

## IT Update

David Chiles

Acting Chief Information Officer

Debbie Stephens

Acting Deputy Chief Information Officer

Thomas A. Beach

Portfolio Manager & Data Strategist

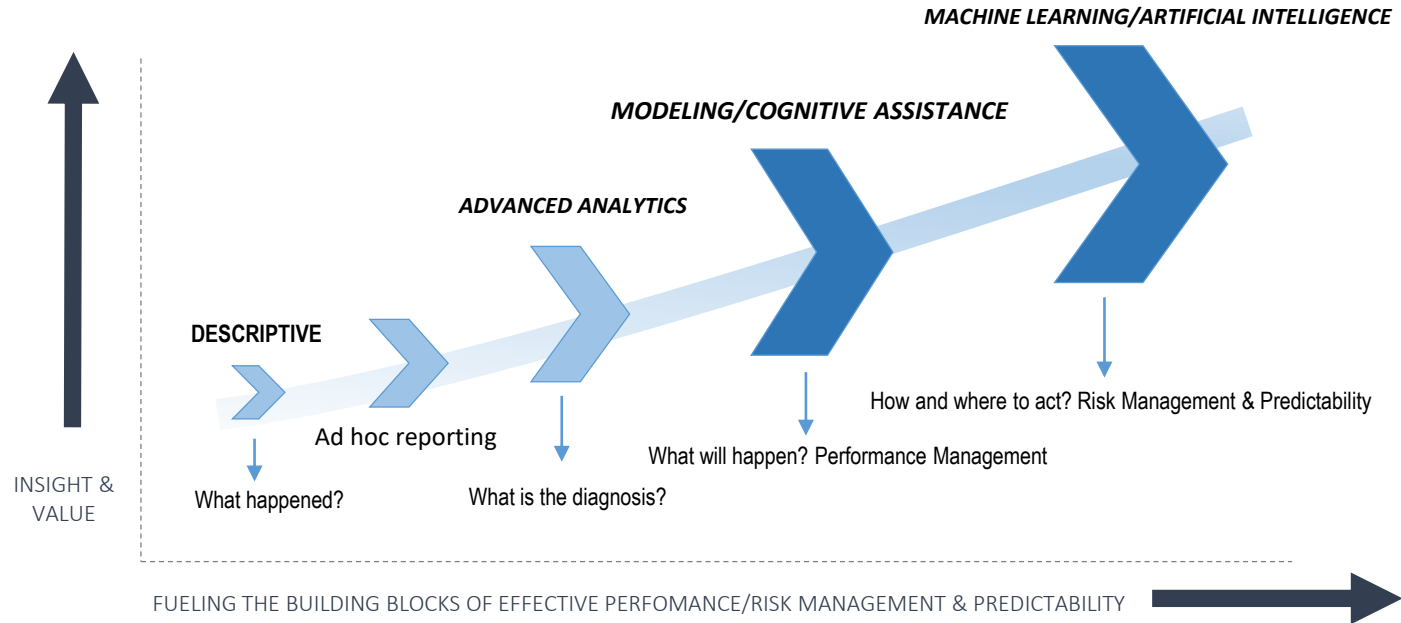
August 2, 2018

UNITED STATES  
PATENT AND TRADEMARK OFFICE



# Where are we headed?

## SCALING THE VALUE PROPOSITION OF DATA SCIENCE and QUALITY & PREDICTABILITY



# Definition of Artificial Intelligence

---

**Artificial Intelligence:** any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. Colloquially, the term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving".[1]

## **Areas of Focus in searching patent documents and technical literature:**

- *Natural Language Processing* - is the ability of a computer program to understand human language as it is spoken as a component of AI.
- *Deep (Machine) Learning* - a subset of machine learning in Artificial Intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled.
- *Word Embedding* –language modeling and feature learning techniques in natural language processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers and create Synonyms and concepts analysis related to Patents.

# Unique Opportunities for Artificial Intelligence

---

## **People**

- Highly structured workforce
- Larger well-trained workforce where change takes time

## **Process**

- Rapidly changing legal landscape
- Complexity of examination procedures that require expertise in specific arts/technologies, legal writing skills and understanding of rules, law and guidance around patent prosecution

## **Technology**

- Existing Legacy systems/infrastructure and Next Generation to support Artificial Intelligence adoption
- Nature of the data (text and image)
- Industry is still maturing

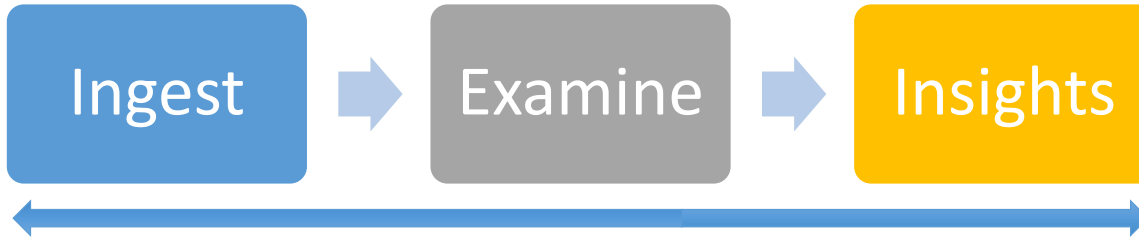
# Related People/Process Efforts

---

- Executive Actions/Roundtables on use of Crowdsourcing to identify prior Art (2014)
- Software partnership – Prior art search roundtable (2015)
- STIC Awareness Campaign (2015)
- Patent Litigation and USPTO Trials: Implications for Patent Examination Quality (January 2015)
- Enhanced Quality Initiatives (2016); including Automated Pre-Search Examination effort and Post Grant Outcomes (<https://www.uspto.gov/patent/initiatives/post-grant-outcomes>)
- Continuing to partner with Patents Business on Internal engagement opportunities and requirements gathering for path forward planning
- Continuing to engage IP5 Offices regarding the use of AI capabilities in patent examination
- Continuing to AI outreach engagements with Industry, Academia and Patent Search Firms

# Technology AI Efforts - Patents

---



**Opportunity  
Areas of**

- Classification Automation/Pre-Classification
- Patent Searching
- Work product consistency/quality
- Knowledge Management
- Academic/International/Industry of current capabilities

# Pre-Classification/Automation

---

## Data Sources

- CPC Scheme Definition
- Patents/PGPubs
- QA validated classification data
- Ensemble models of over 300k CPC symbols



## Machine Learning/AI Algorithms

- Supervised Machine Learning Classification Algorithms to train on data sources
- Unsupervised Machine Learning to intelligently parse information about the incoming applications
- An Ensemble Algorithm to combine various algorithms to achieve best performance



Example CPC Symbol

**E21B**                      **19**                      /                      **18**

**Sub-Class Level**   **Group Level** / Subgroup

(~94%)                      (~87%)                      (~70%)

*\* Accuracy Varies by CPC Sections*



# Patent Searching (Boolean vs AI-Based)

---

## Boolean or text based search (current system):

- Historically, the approach to **automated prior art searching** is to determine a **few keywords** from the patent application and, based on simple text frequency matching of these keywords, retrieve published applications and patents.
- Simple keyword and classification codes searches may have **limited utility** in the patent prosecution context.
- The high prevalence of uncommon language patterns and creation by patent applications of **'abstract vocabulary'** specific to their claimed invention can make text based searching challenging, especially as terminology and meaning of word evolve over time.

## Vision of AI-Based Prior Art Search:

- Augmenting more **data sources** and gathering of relevant documents as a **potential pre-search capability** and **an AI-based search platform** leveraging recommendation engines based on document traits and search string analytics
- **Leveraging Traits with AI-Based Search:**
  - Synonyms/Topical Concepts extraction
  - Automated CPC classification
  - Design/Utility Image similarity search
  - Federated Similarity prior art search
  - Knowledge management features
  - Office Action Search and identifying Prior Art use in similar examination

# Additional AI IT Efforts Underway

---

## I. Project Initiatives at the USPTO

### ❖ Patent Enriched Citation Data

- International data exchange effort

### ❖ Browser Based Cognitive assistants for Patents

- Augmenting existing tools with AI/ML

## II. Exploratory Projects at the USPTO

- AI based Patent Term Library Generator

- AI based Image Search

- Deep Machine Learning Chat Bots

# Questions and Comments

**David Chiles, Acting Chief Information Officer**

(571) 272-9400

[David.Chiles@uspto.gov](mailto:David.Chiles@uspto.gov)

**Debbie Stephens, Acting Deputy Chief Information Officer**

(571) 272-9400

[Deborah.Stephens@USPTO.GOV](mailto:Deborah.Stephens@USPTO.GOV)

**Thomas A. Beach, Portfolio Manager & Data Strategist**

(571) 272-6988

[Thomas.Beach@USPTO.GOV](mailto:Thomas.Beach@USPTO.GOV)

