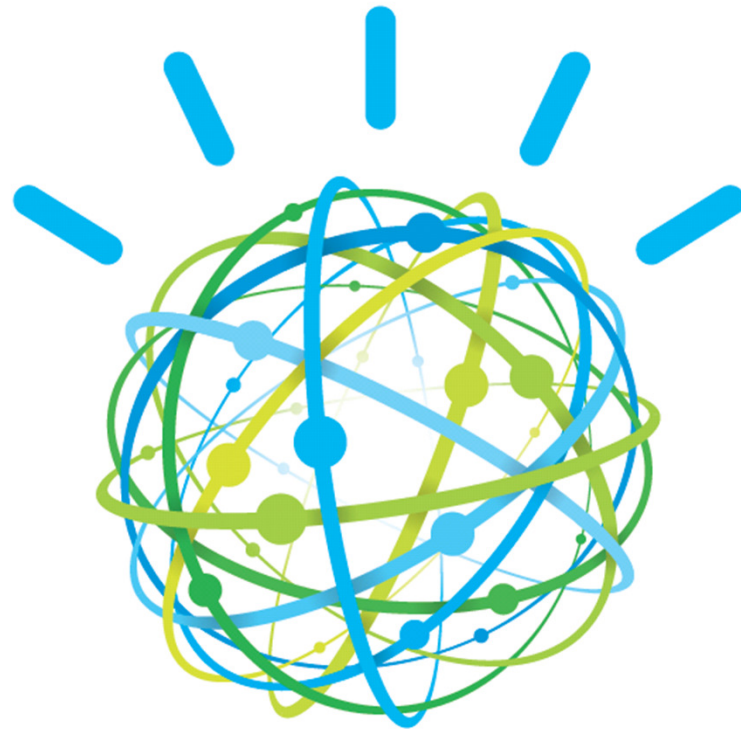


# *IBM “Watson”: Beyond playing “Jeopardy!”*

Katharine Frase, VP Industries Research, IBM

*with thanks to:*

David Ferrucci, Principal Investigator,  
DeepQA Team @ IBM Research



April 24, 2012

© 2011 IBM Corporation

# Want to Play Chess or Just Chat?



- Chess

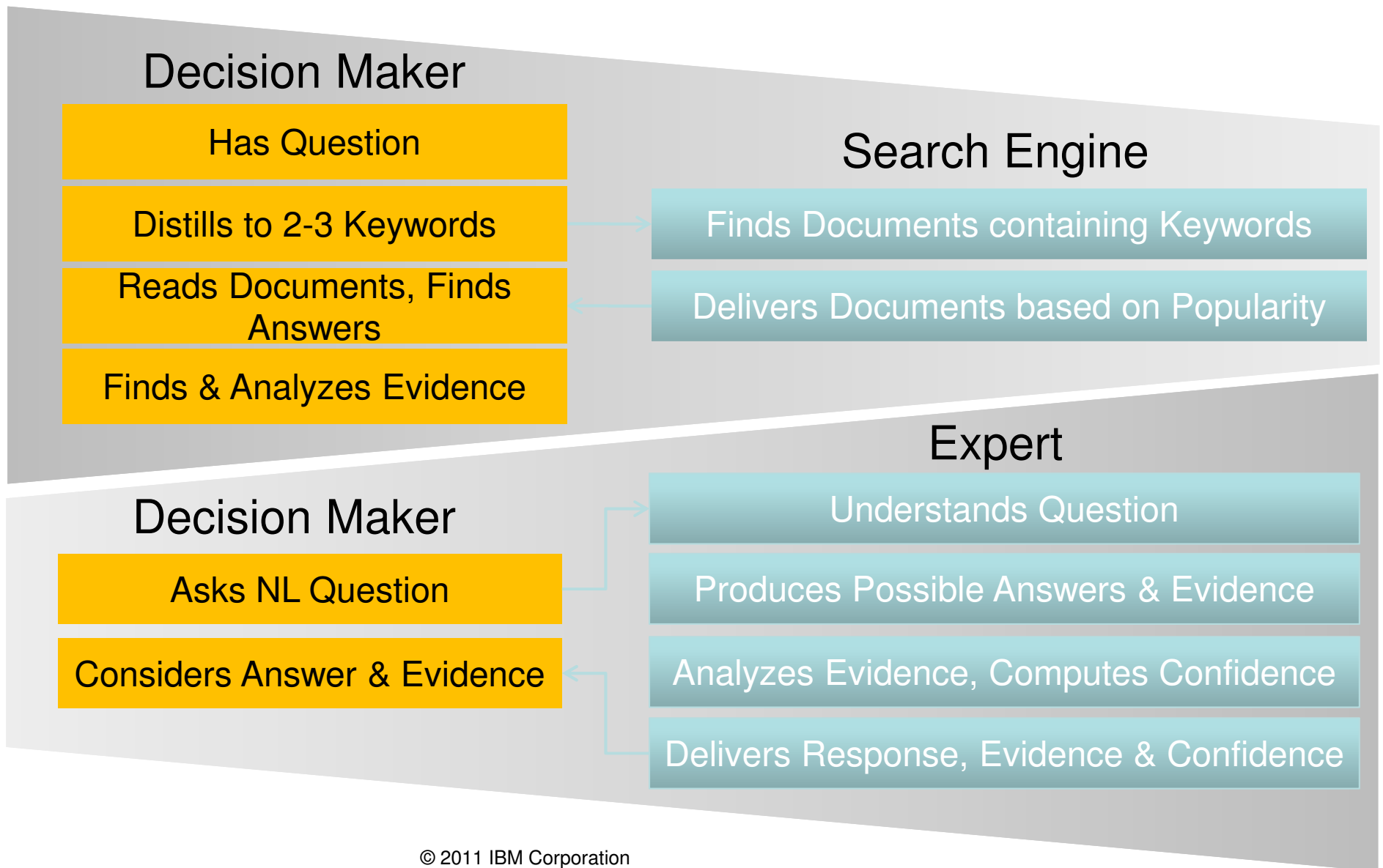
- A finite, mathematically well-defined search space
- Limited number of moves and states
- All the symbols are completely grounded in the mathematical rules of the game

- Human Language

- Words by themselves have no meaning
- Only grounded in **human cognition**
- Words navigate, align and communicate an infinite space of intended meaning
- Computers can **not** ground words to human experiences to derive meaning



# Informed Decision Making: Search vs. Expert Q&A



# Easy Questions?

$$\ln((12,546,798 * \pi)) ^ 2 / 34,567.46 = \mathbf{0.00885}$$

Select *Payment* where *Owner*="David Jones" and *Type(Product)*="Laptop",

Owner	Serial Number
David Jones	45322190-AK

Serial Number	Type	Invoice #
45322190-AK	LapTop	INV10895

Invoice #	Vendor	Payment
INV10895	MyBuy	\$104.56

David Jones  
↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓  
David Jones

=

Dave Jones ≠  
↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓  
David Jones

# Hard Questions?

Computer programs are natively **explicit**, **fast** and **exacting** in their calculation over numbers and symbols....But **Natural Language** is implicit, highly contextual, ambiguous and often imprecise.

Person	Birth Place
A. Einstein	ULM

Structured

Unstructured

- Where was X born?

*One day, from among his city views of Ulm, Otto chose a water color to send to Albert Einstein as a remembrance of Einstein's birthplace.*

Person	Organization
J. Welch	GE

- X ran this?

*If leadership is an art then surely Jack Welch has proved himself a master painter during his tenure at GE.*

The Jeopardy! Challenge: *A compelling and notable way to **drive and measure** the technology of automatic Question Answering along 5 Key Dimensions*

**Broad/Open  
Domain**

**Complex  
Language**

**High  
Precision**

**Accurate  
Confidence**

**High  
Speed**

**\$200**

If you're standing, it's the direction you should look to check out the wainscoting.

**\$1000**

The first person mentioned by name in 'The Man in the Iron Mask' is this hero of a previous book by the same author.

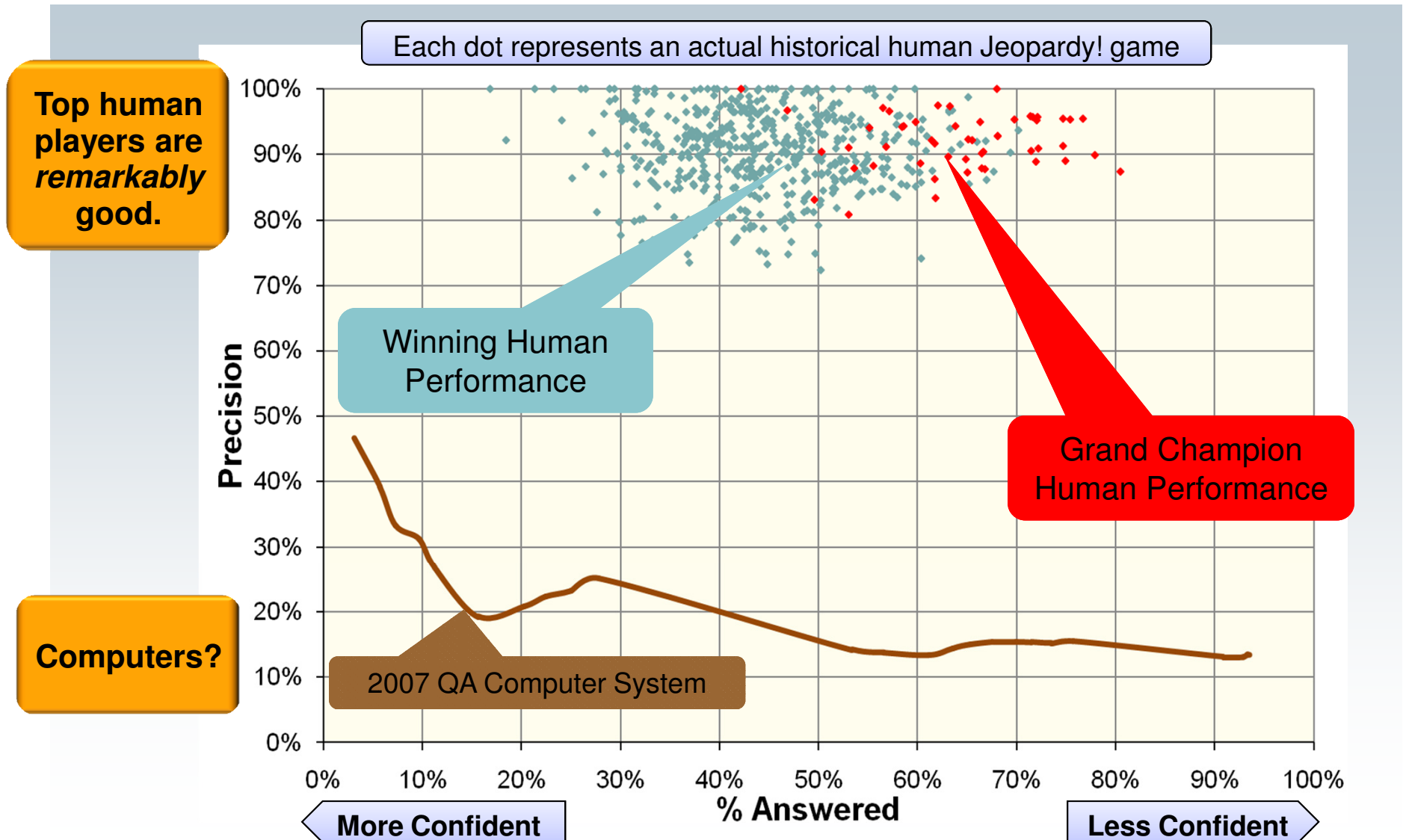
**\$600**

In cell division, mitosis splits the nucleus & cytokinesis splits this liquid *cushioning* the nucleus

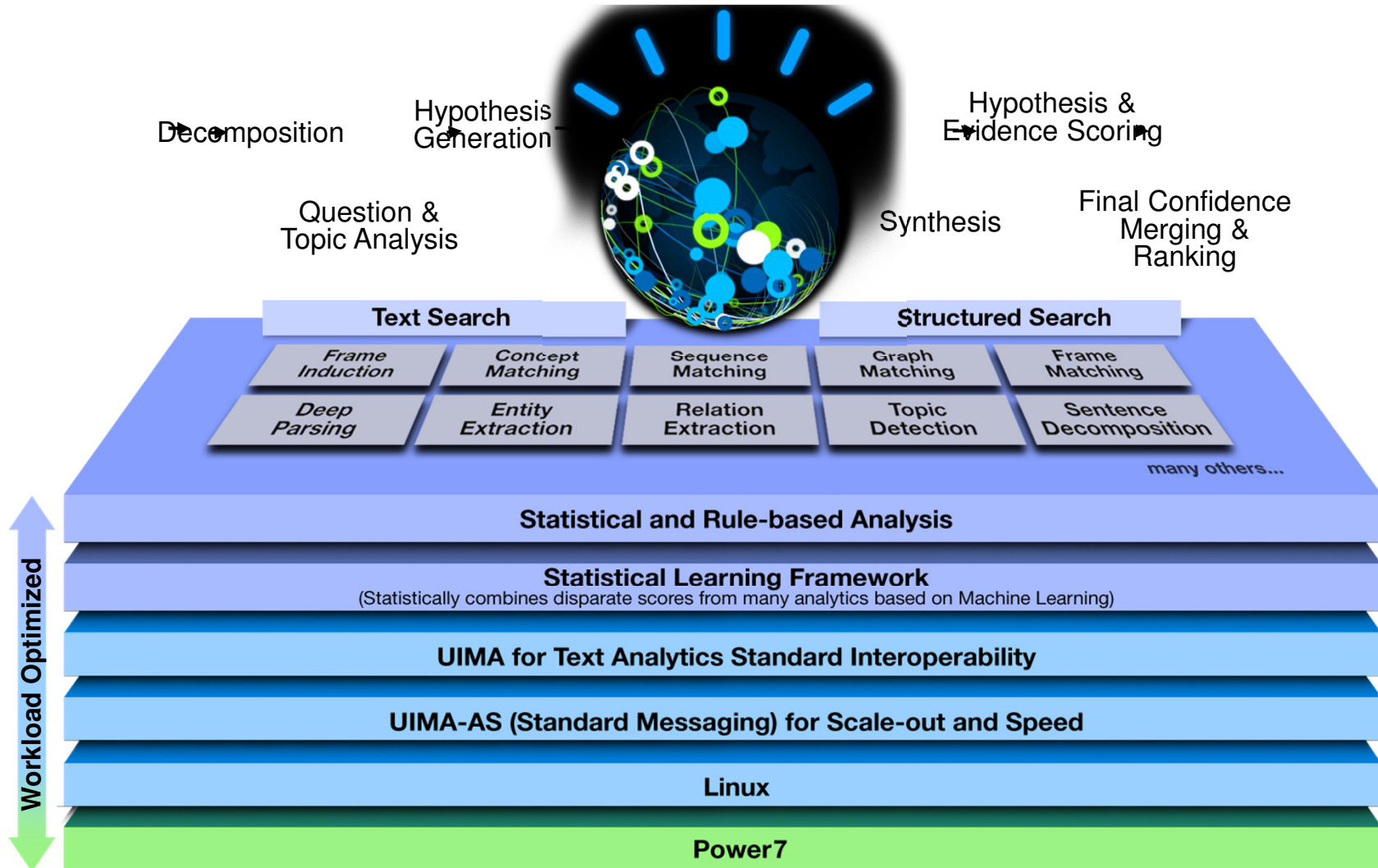
**\$2000**

Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that's farthest north

# The Best Human Performance: *Our Analysis Reveals the Winner's Cloud*



# What is Watson?



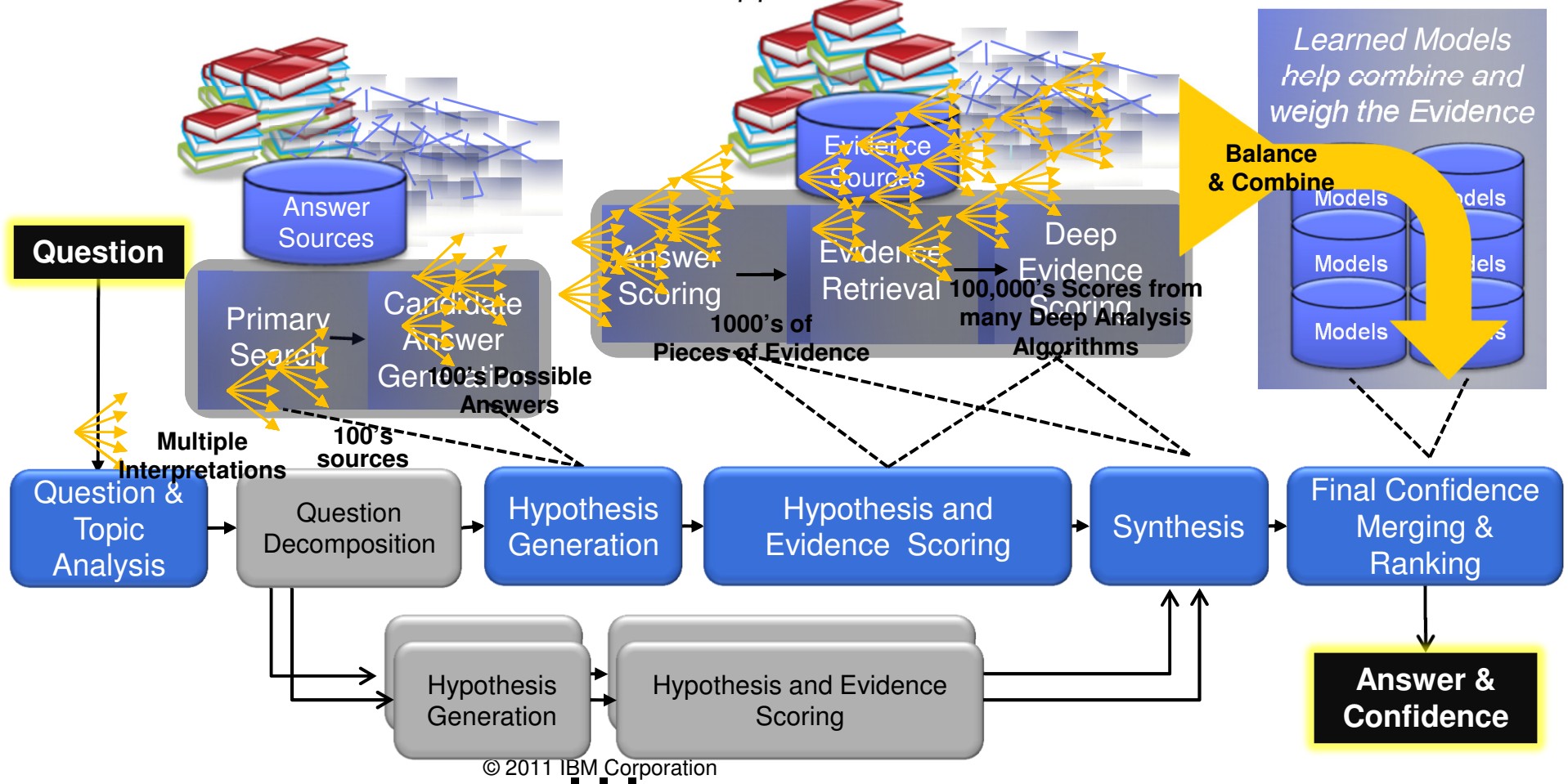


# DeepQA: The Technology Behind Watson

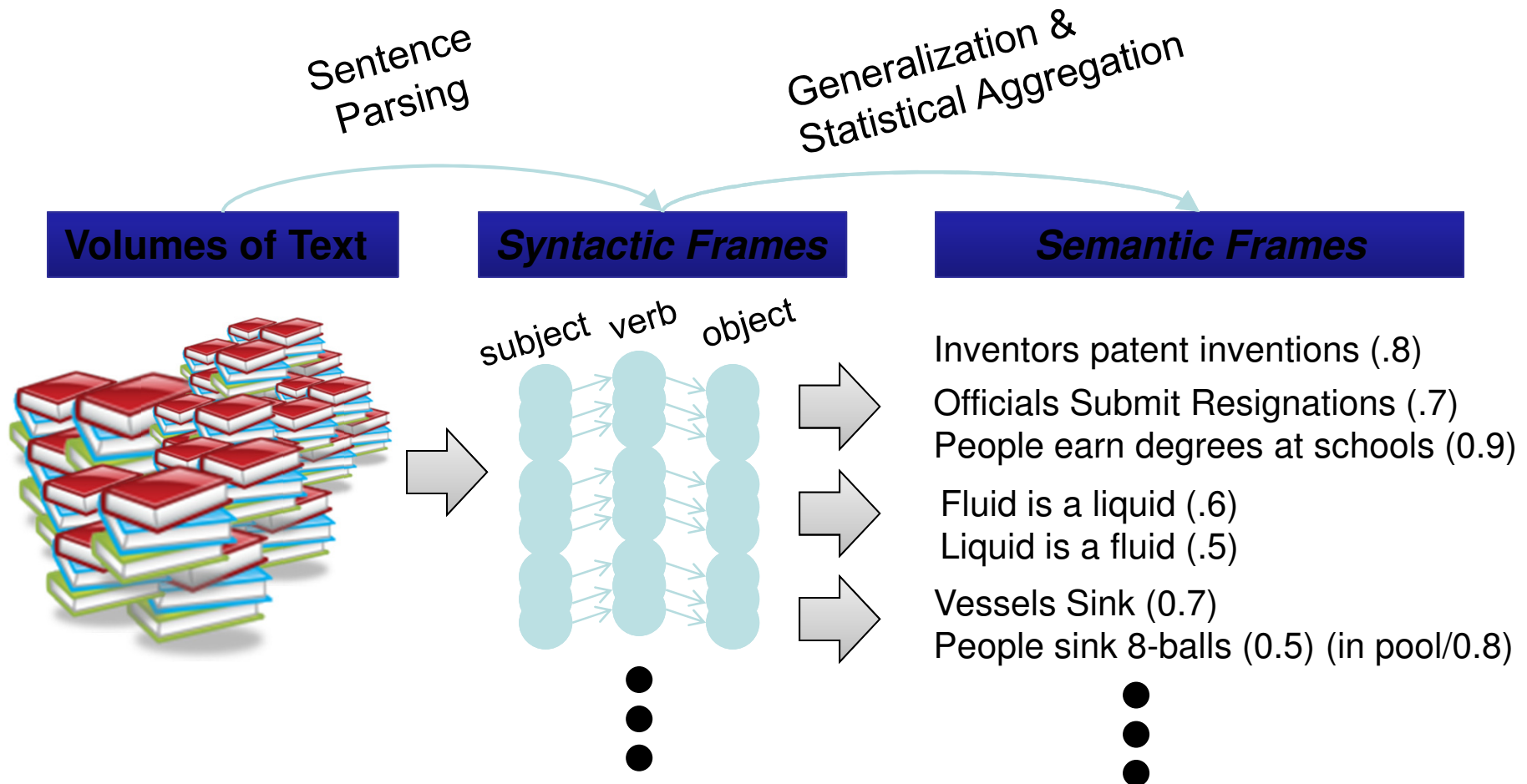
## *Massively Parallel Probabilistic Evidence-Based Architecture*

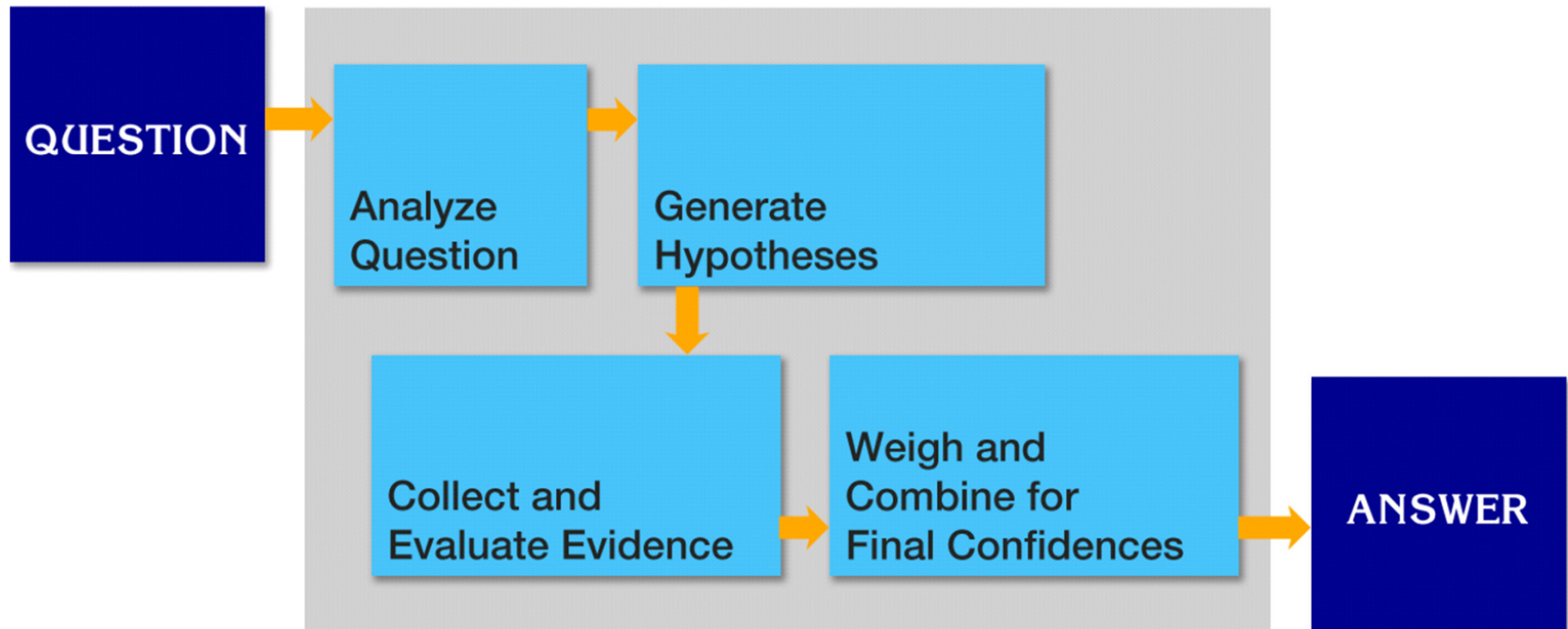
*Generates and scores many hypotheses using a combination of 1000's **Natural Language Processing, Information Retrieval, Machine Learning** and **Reasoning Algorithms**.*

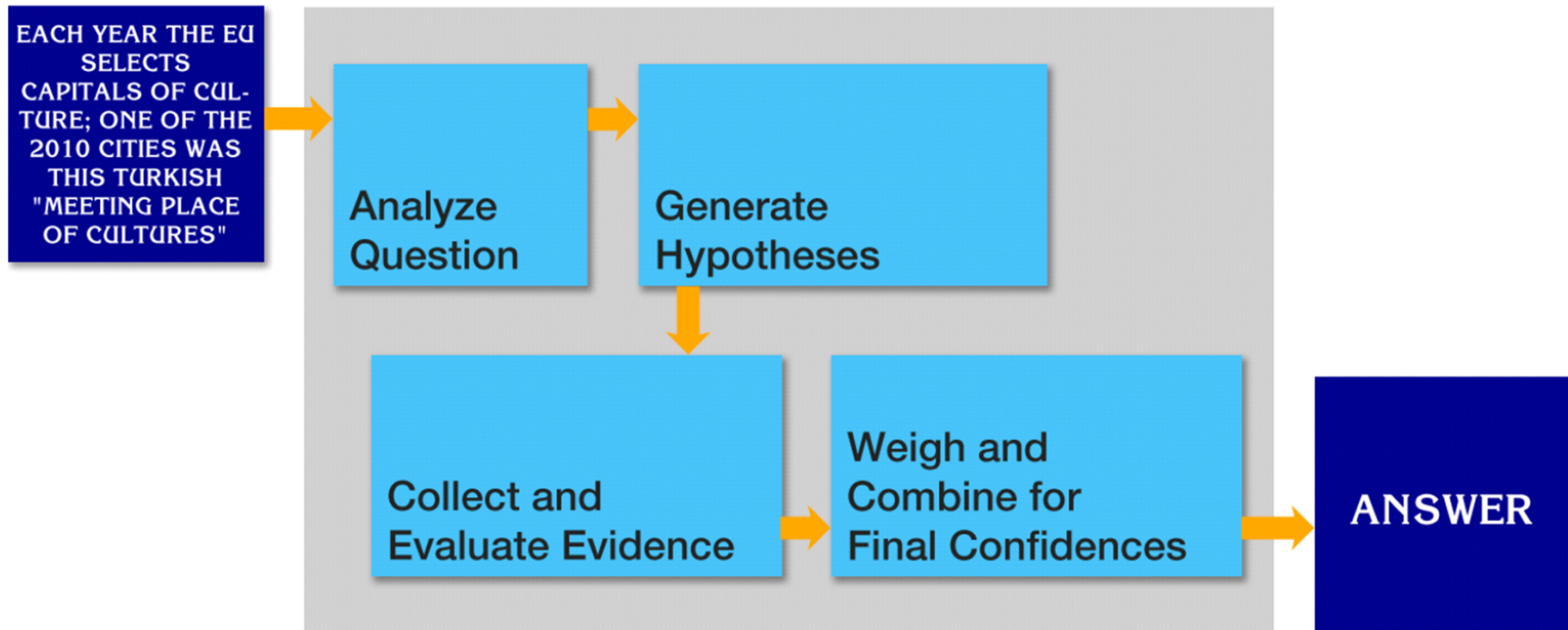
*These gather, evaluate, weigh and balance different types of **evidence** to deliver the answer with the best support it can find.*



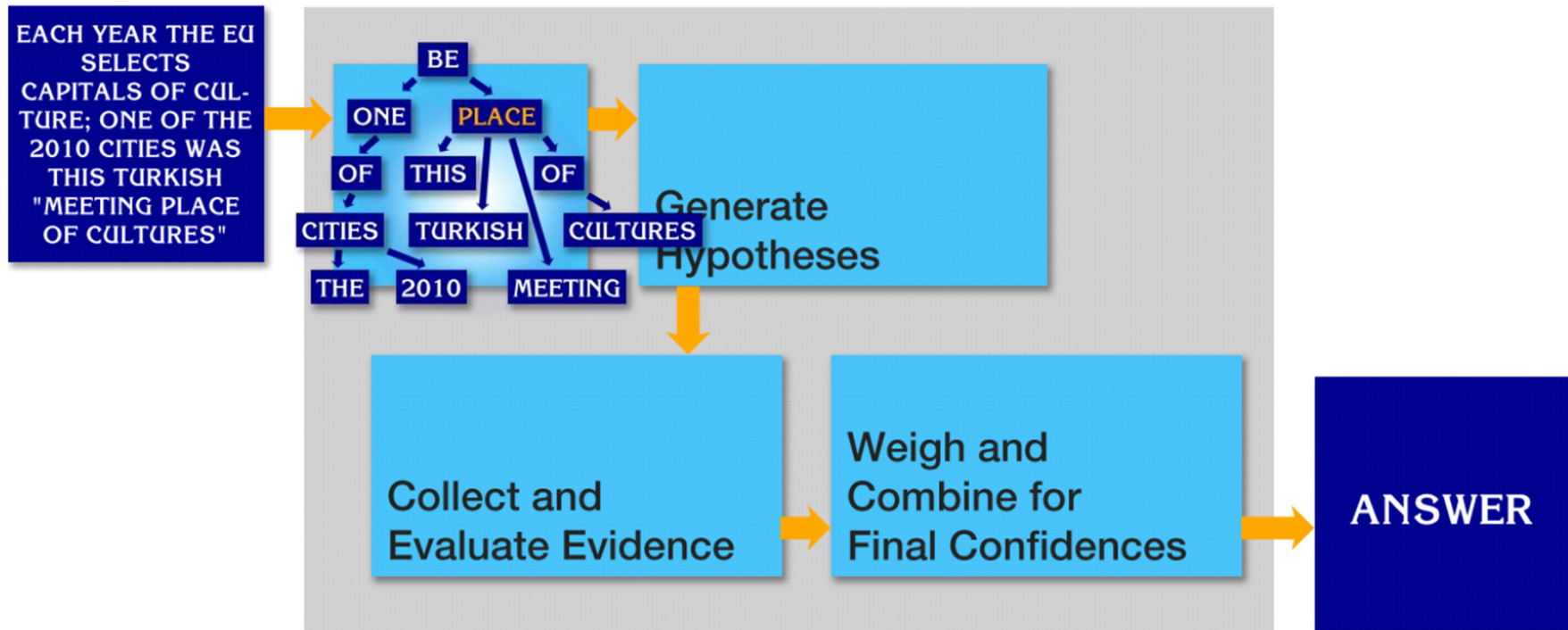
# Automatic Learning From “Reading”

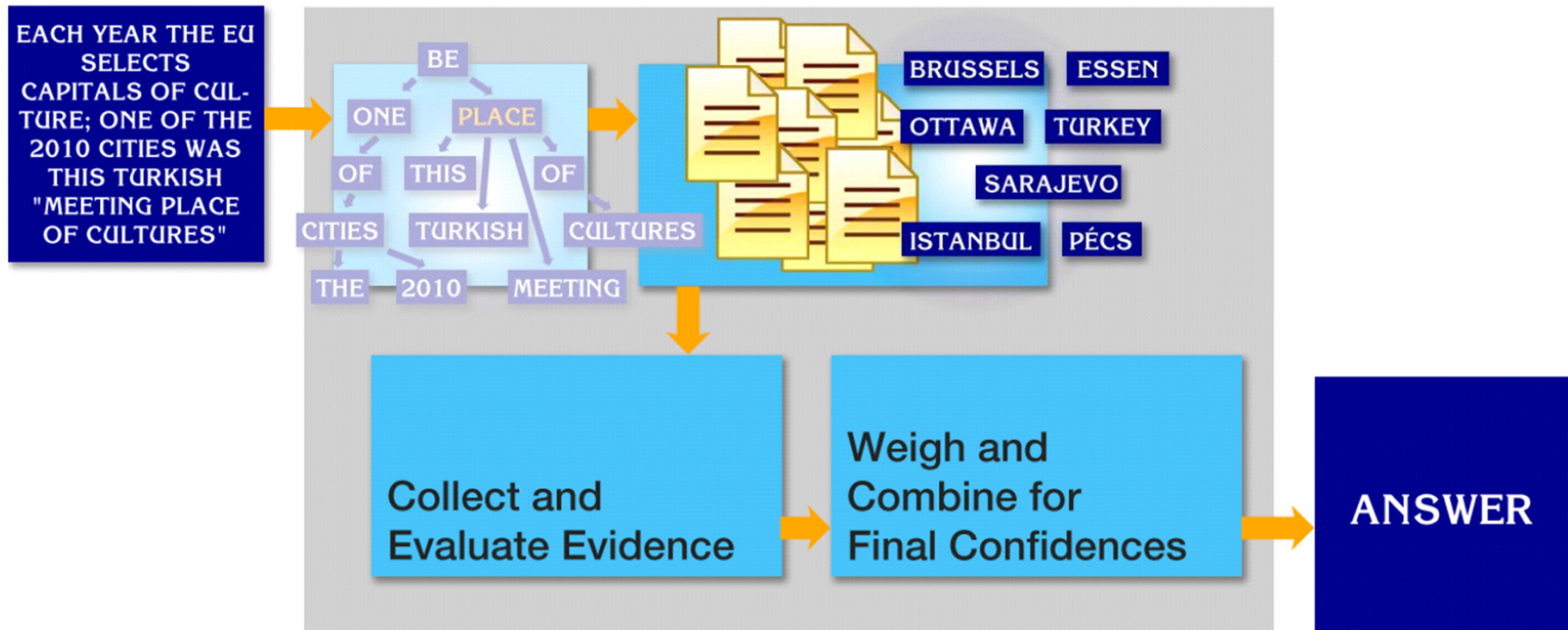












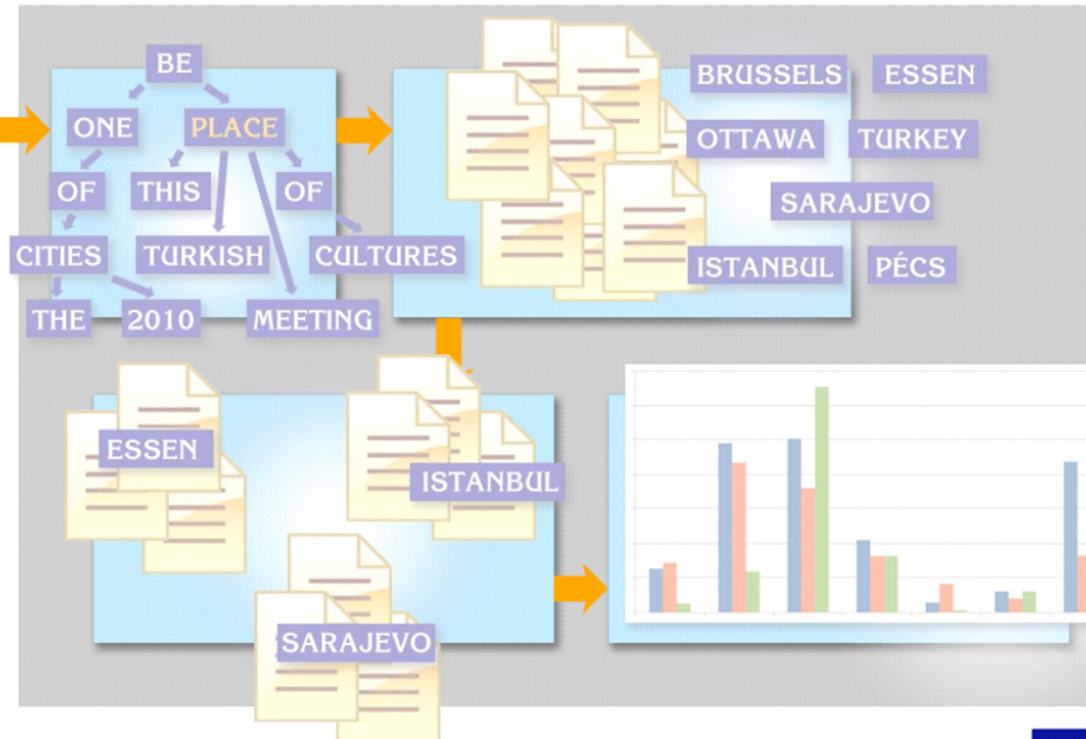


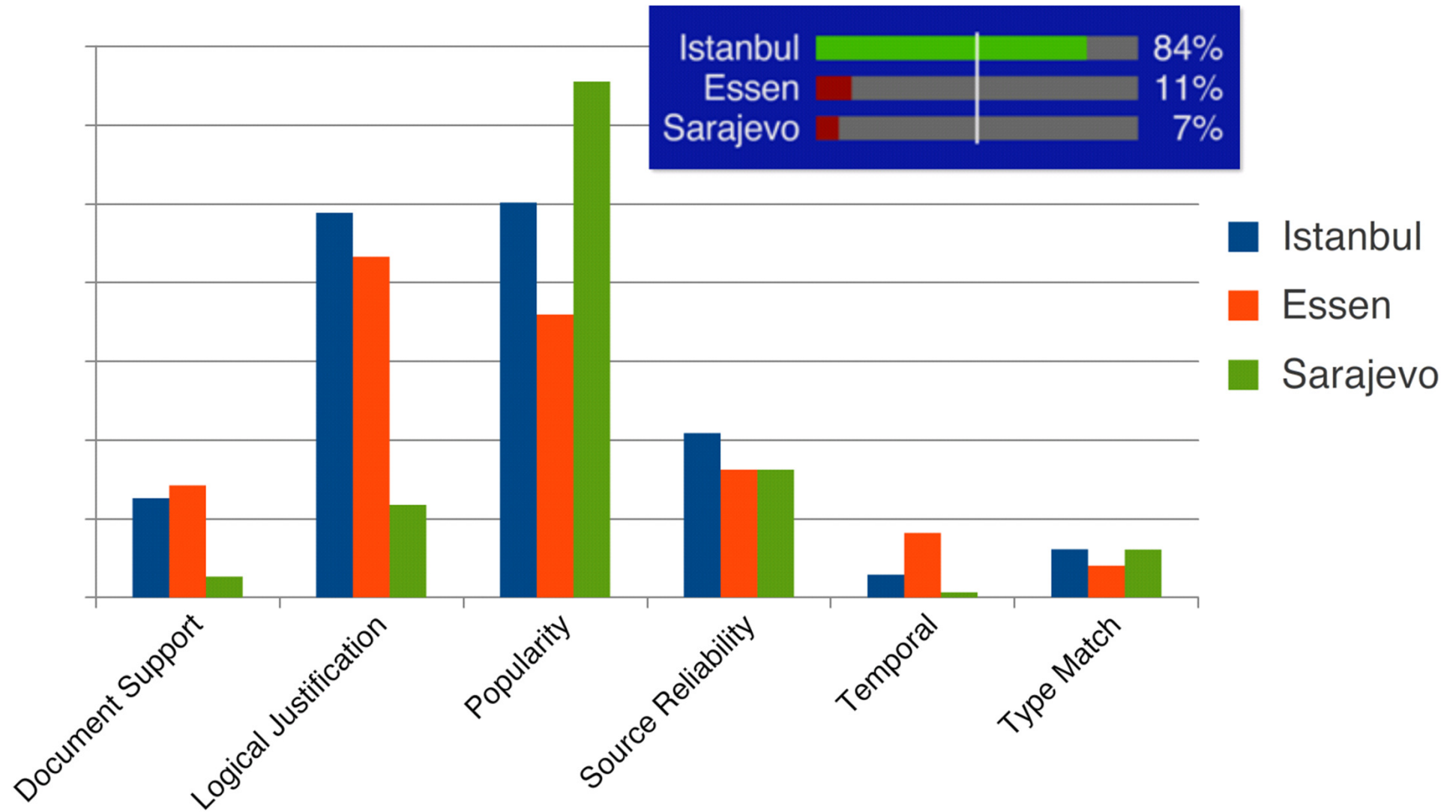






EACH YEAR THE EU  
SELECTS  
CAPITALS OF CUL-  
TURE; ONE OF THE  
2010 CITIES WAS  
THIS TURKISH  
"MEETING PLACE  
OF CULTURES"





# Some Basic Jeopardy! Clues

The *type* of thing being asked for is often indicated but can go from specific to very vague

- This **fish** was thought to be extinct millions of years ago until one was found off South Africa in 1938
- Category: ENDS IN "TH"
- Answer: **coelacanth**
  
- When hit by electrons, a phosphor gives off electromagnetic energy in this **form**
- Category: General Science
- Answer: **light (or photons)**
  
- Secy. Chase just submitted **this** to me for the third time--guess what, pal. This time I'm accepting **it**
- Category: Lincoln Blogs
- Answer: **his resignation**

Categories are not as simple as they seem

Watson uses statistical machine learning to discover that **Jeopardy! categories are only weak indicators of the answer type.**

U.S. CITIES	Country Clubs	Authors
St. Petersburg is home to Florida's annual tournament in this game popular on ship decks ( <b>Shuffleboard</b> )	From India, the shashpar was a multi-bladed version of this spiked club ( <b>a mace</b> )	Archibald MacLeish based his verse play "J.B." on this book of the Bible ( <b>Job</b> )
Rochester, New York grew because of its location on this ( <b>the Erie Canal</b> )	A French riot policeman may wield this, simply the French word for "stick" ( <b>a baton</b> )	In 1928 Elie Wiesel was born in Sighet, a Transylvanian village in this country ( <b>Romania</b> )

# Evaluating Possibilities and Their Evidence

In cell division, mitosis splits the nucleus & cytokinesis splits this **liquid** cushioning the nucleus.

- Organelle
- Vacuole
- Cytoplasm
- Plasma
- Mitochondria
- Blood ...

- Many candidate answers (CAs) are generated from many different searches
- Each possibility is evaluated according to **different dimensions of evidence**.
- **Just One** piece of evidence is if the CA is of the **right type**. In this case a “liquid”.

Is(“Cytoplasm”, “liquid”) = 0.2↑

Is(“organelle”, “liquid”) = 0.1

Is(“vacuole”, “liquid”) = 0.2

Is(“plasma”, “liquid”) = 0.7

“Cytoplasm is a **fluid** surrounding the nucleus...”

Wordnet → Is\_a(Fluid, Liquid) → ?

Learned → Is\_a(Fluid, Liquid) → yes.

# Divide and Conquer (*Typical in Final Jeopardy!*)

**Must identify and solve  
sub-questions from different  
sources to answer  
the top level question**

**Lyndon B Johnson**

**In 1968**

this man was U.S. president.

When "60 Minutes" premiered this man was U.S. president.

?

**The DeepQA architecture attempts different *decompositions* and  
recursively applies the QA algorithms**

## Some Growing Pains (early answers)

the People

### THE AMERICAN DREAM

Decades before Lincoln, Daniel Webster spoke of government "made for", "made by" & "answerable to" them

No One

WW I

### NEW YORK TIMES HEADLINES

An exclamation point was warranted for the "end of" this! In 1918

a sentence

Apollo 11 moon landing

### MILESTONES

In 1994, 25 years after this event, 1 participant said, "For one crowning moment, we were creatures of the cosmic ocean"

the Big Bang

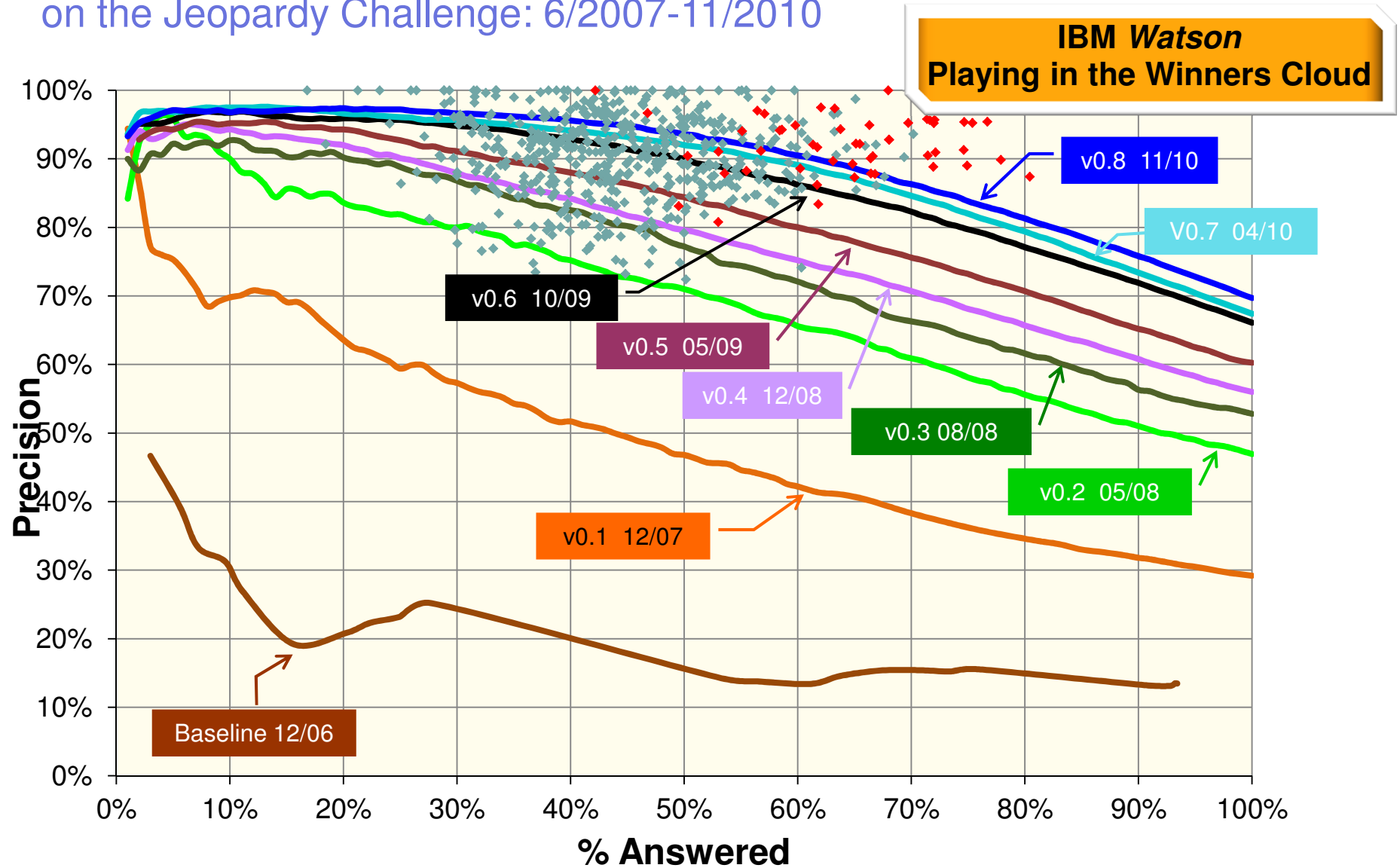
Call on the phone

### THE QUEEN'S ENGLISH

Give a Brit a tinkle when you get into town & you've done this

urinate

## DeepQA: Incremental Progress in Answering Precision on the Jeopardy Challenge: 6/2007-11/2010





One Jeopardy! question can take **2 hours on a single 2.6Ghz Core**  
Optimized & Scaled out on 2,880-Core Power750 using UIMA-AS,  
*Watson* is answering in ~3 seconds



- 90 x IBM Power 750<sup>1</sup> servers
- 2880 POWER7 cores
- POWER7 3.55 GHz chip
- 500 GB per sec on-chip bandwidth
- 10 Gb Ethernet network
- 15 Terabytes of memory
- 20 Terabytes of disk, clustered
- Can operate at 80 Teraflops
- Runs IBM DeepQA software
- Scales out with and searches vast amounts of unstructured information with UIMA & Hadoop open source components
- SUSE Linux provides a cost-effective open platform which is performance-optimized to exploit POWER 7 systems
- 10 racks include servers, networking, shared disk system, cluster controllers

# Watson brings many innovations to apply to business issues

Watson Capabilities	Best Fit for Watson	Watson Solution Area and Industry Application
<ol style="list-style-type: none"> <li><b>Natural language understanding</b></li> <li><b>Broad domain of unstructured data</b></li> <li><b>Parallel hypothesis generation and confidence scoring</b></li> <li><b>Iterative Question/Answering to refine results</b></li> <li><b>Machine learning</b></li> </ol>	<ul style="list-style-type: none"> <li>Analysis of unstructured data</li> <li>Critical questions that require decision support with prioritized recommendations and evidence</li> <li>High value in decision support</li> <li>Leverage scale to maximize machine learning and improve outcomes over time</li> </ul>	<p><b>Assisting Research</b> (<i>e.g., Financial Services</i>)</p> <p><b>Assisting research</b> of macroeconomic trends, current events and historic data to support investment decisions</p> <p><b>Results:</b> increased alpha and risk management due to improved investment decision making and customer insights</p> <p><b>Other Applications:</b> Customer Insights, Knowledge Mgmt</p> <p><b>Assisting Discovery</b> (<i>e.g., Legal Services</i>)</p> <p><b>Assisting research</b> of historical and current documents to discover relevant references and identify relationships.</p> <p><b>Results:</b> increased confidence in due diligence, prior art or precedent evaluations, due to context understanding of full passages and learned value of sources, vs “keyword search” based methods.</p> <p><b>Other Applications:</b> R&amp;D Mgmt, Compliance Mgmt</p> <p><b>Assisting Diagnosis</b> (<i>e.g., Healthcare</i>)</p> <p><b>Assisting diagnosis and suggestions</b> for patient condition based on patient data, family history, medical journals etc.</p> <p><b>Results:</b> improved patient care and payer/provider cost management based on more efficient patient outcomes</p> <p><b>Other Applications:</b> Contact Center, Technical Support</p>

# THANK YOU