

Module 2

Cognitive Computing

Foundation of Cognitive Computing

- Cognitive computing is a technology approach that enables humans to collaborate with machines.
- A cognitive system uses the data to train, test, or score a hypothesis.
- Cognitive computing is the use of computerized models to simulate the human thought process in complex situations where the answers might be ambiguous and uncertain.

Computers are faster than humans at processing and calculating, but they have yet to master some tasks, such as understanding natural language and recognizing objects in an image. Cognitive computing is an attempt to have computers mimic the way a human brain works.

Cognitive computing is an attempt to have computers mimic the way the human brain works. To accomplish this, cognitive computing uses artificial intelligence ([AI](#)) and other underlying technologies

- Expert systems, Neural networks, Machine learning.
- Deep learning, Natural language processing ([NLP](#)).
- Speech recognition, Object recognition, Robotics.
- It uses self learning algorithms, data analysis and pattern recognition.
- Learning technology can be used for speech recognition, sentiment analysis, risk assessments, face detection and more.
- IBM's cognitive computer system, Watson
- Examples SIRI, Google Assistant, Cortana , Alexa
- Applications: healthcare, banking, finance and retail, logistics

A cognitive system has three fundamental principles to work as described below:

- 1> **Learn**:- The system leverages data to make inferences about a domain, a topic, a person, or an issue based on training and observations from all varieties, volumes, and velocity of data.
- 2> **Model** :- To learn, the system needs to create a model or representation of a domain (which includes internal and potentially external data) and assumptions that dictate what learning algorithms are used.
- 3> **Generate Hypothesis**:- A cognitive system assumes that there is not a single correct answer. The most appropriate answer is based on the data itself.

cognitive computing systems must have the following attributes/features:

- 1> adaptive
- 2> interactive
- 3> iterative and stateful
- 4> contextual

- **Adaptive:** CC devices must be adaptive systems that are capable of learning as information changes and as goals and requirements evolve. They should be able to resolve ambiguity and tolerate unpredictability.
- **Interactive:** These devices should be able to interact with other processors, devices, cloud services, and humans as well.
- **Iterative and stateful:** They are able to define a problem by posing questions or finding additional source input if a problem statement is ambiguous or incomplete. They are also able to recall previous interactions in a process and return information that is suitable for the specific application at a given point in time.
- **Contextual:** CC devices should be able to understand, identify, and extract contextual elements such as syntax, meaning, time, location, appropriate domain, regulations, user profiles, process, task and goal. They can pull from multiple sources of information, including both structured and unstructured digital information, as well as sensory inputs (visual, gestural, auditory, or sensor-provided).

- With cognitive computing, we are bringing together two disciplines:

Cognitive Computing = Cognitive Science + Computer Science

- Cognitive science—The science of the mind.
- Computer science—The scientific and practical approach to computation and its applications. It is the systematic technique for translating this theory into practice.

Cognitive Computing Vs AI

Cognitive Computing vs. AI

COGNITIVE COMPUTING		ARTIFICIAL INTELLIGENCE
TECHNOLOGIES	Machine learning, natural language processing, neural networks, deep learning, sentiment analysis	Machine learning, natural language processing, neural networks, deep learning
CAPABILITIES	Simulate human thought processes to assist humans in finding solutions to complex problems	Find patterns in big data to learn and either reveal hidden information or deliver solutions to complex problems
PURPOSE	Augment human capabilities	Automate processes
INDUSTRIES	Customer service, healthcare, industrial sector 	Finance, security, healthcare, retail, manufacturing, government 

- **Benefits or Advantages of Cognitive Computing**
- Following are the benefits or **advantages of cognitive computing**:
 - It helps in improvement of customer engagement and service. This enhancement of customer experience is possible due to use of cognitive applications such as cognitive assistants, social intelligence, personalized recommendations, behavioral predictions.
 - It helps in enhancing employee productivity and quality of service/product outcomes.
 - Every second 1.7 MB of data is generated by each person on the earth. Out of the total, about 99.5% of data is not being analyzed. Once this data is analyzed, it can help in unlocking many business opportunities by identifying right markets , new customer segments, new products to launch and so on.
 - Enterprises can enhance operational efficiency by implementation of cognitive applications such as predictive asset maintenance, automated replenishment systems etc.
 - It helps in providing very accurate data analysis. Hence cognitive systems are employed in healthcare industry.

- **Drawbacks Or Disadvantages Of Cognitive Computing**

- Security is one of the major concern as digital devices manage critical information in cognitive computing.
- The other big hurdle is its voluntary adoption by enterprises, government and individuals.
- Change management is another challenge as this technology has power to learn like humans and behave like natural humans. Hence people are fearful as they feel machines would replace humans in future.
- The cognitive computing based systems/products require lengthy development cycles.

- **Challenges of Cognitive Computing**

- **Data Privacy:** Cognitive computing relies heavily on data analysis, raising concerns about the privacy and security of sensitive information.
- **Complexity:** Implementation of cognition solutions can be complex and may require significant integration efforts with existing systems.
- **Ethical and Bias:** Perpetuate biases present in the training data, leading to unfair or discriminatory outcomes.

- **How does cognitive computing work?**
- Cognitive computing works through a combination of technologies and processes that aim to simulate human-like intelligence and decision-making.
- **Collection:** The initial step in cognitive computing involves the gathering of extensive datasets from various sources, encompassing both structured and unstructured data like text, images, videos, and sensor readings.
- **Ingestion:** Following this, the acquired data undergoes ingestion into the cognitive computing system, wherein it is systematically organized, categorized, and stored in a format conducive to effective analysis.
- **NLP:** A pivotal aspect of this process is Natural Language Processing (NLP), a fundamental component enabling the system to comprehend and interpret human language, encompassing both written and spoken communication. NLP algorithms are employed to process textual data, extracting meaning and identifying relationships between words and concepts.

- Subsequently, cognitive computing heavily relies on **Machine Learning Algorithms** to scrutinize and glean insights from the ingested data. Two primary types of machine learning are employed in this context:
- Supervised Learning, where the system is trained on labeled data associating inputs with known outputs.
- Unsupervised Learning, where the system identifies patterns and relationships within the data without predefined labels.
- **Analysis:** This machine learning capability is instrumental in facilitating pattern recognition within the cognitive system.
- **Predictions:** Through the analysis of patterns, correlations, and trends within the data, the system gains a comprehensive understanding of complex relationships, enabling it to make accurate predictions.
- This iterative process underscores the dynamic and evolving nature of cognitive computing, where continuous learning and adaptation are integral to enhancing system capabilities over time.

- **Applications of Cognitive Computing**

- **Retail Industry**
- It helps the marketing team to collect more data and then analyse to make retailers more efficient and adaptive.
- These help companies to make more sales and provide personalized suggestions to the customers.
- E-commerce sites have integrated cognitive computing very well, they collect some basic information from the customers about the basic details of the product they are looking for and then analyse the large available data and recommend the products to the customer.
- Through demand forecasting, price optimization, and website design, cognitive computing has provided retailers with the tools to build more agile businesses.
- Apart from e-commerce sites, cognition can be very useful for on-floor shopping as well. It will help retailers to provide customers personalized products – what they want, when they want, and how they want to derive meaningful experiences, opportunities to reduce the wastage and losses by providing the fresh products by predicting the demand priorly, and by automating areas it will reduce the cycle time, effort and improve the efficiency.

- **Logistics**
- Cognition is the new frontier in the Transportation, Logistics, and Supply chain.
- It helps at every stage of logistics, like Designing Decisions in Warehouse, Warehouse Management, Warehouse Automation, IoT, and Networking.
- In the warehousing process, cognition helps in compiling storage code, automatic picking with the automated guided vehicle, and use of warehouse robots will help to improve work efficiency.
- Logistics distribution links use cognition to plan the best path improving the recognition rate which will save a lot of labour.
- IoT will help in warehouse infrastructure management, optimizing inventory, enhancing operations in the warehouse and the autonomous guided vehicle can be used for picking and putting operations.
- Apart from IoT, the other important technology is Wearable Devices, which helps to convert all the objects to sensors and augments human decision-making and warehouse operations. These devices have evolved from smartwatches to smart clothes, smart glasses, computing devices, exoskeletons, ring scanners, and voice recognition.

- **Banking and Finance**
- Cognition in the banking industry will help to improve operational efficiency, customer engagement, and experience and grow revenues.
- Cognitive banking will completely reshape the banking and financial institutions on three dimensions: Deeper contextual engagement, New analytics insights, and Enterprise transformation.
- We are already experiencing examples of such transformation for tasks like performing various banking transactions digitally, opening a new retail account, processing claims and loans in minutes. This technology has proved to be very helpful in the areas of product management and customer service support.
- Cognitive banking will provide customized support to the customers, it will help in deciding personalized investment plans based on the customer being risk-averse or risk-taker.
- Also, it will provide personalized engagement between the financial institution and the customer by dealing in the individual fashion with each customer and focusing on their requirements.
- Here, the computer will intelligently understand the personality of the customer based on the other content available online authored by the customer.

- **Power and Energy**
- ‘Smart Power’ is the new intelligent future. The oil and gas industry faces huge cost pressure to find, produce and distribute crude oil and its byproducts. Also, they face a shortage of skilled engineers and technical professionals. Energy firms take various critical decisions where huge capital is involved, like which site to explore, allocation of resources, and quantity of production. For a long time, this decision was taken based on the data collected and stored and the expertise and intuition of the project team.
- With cognitive computing, technologies process volumetric data to support decisions and learn from those results. This technology will help us to make various important decisions of the future like commercially viable oil wells, ways to make existing power stations more efficient, and will also give a competitive advantage to existing power companies.

- **Cyber Security**
- Cognitive Algorithms provides end-to-end security platforms and detects, assesses, researches, and remediate the threats. It will help to prevent cyber Attacks (or cognitive hacking), this will make customers less vulnerable to manipulation as well as provide a technical solution to detect any misleading data and disinformation.
- With the increase in volumetric data, and rise in [cyber attacks](#), and the shortage of skilled cybersecurity experts we need modern methods like cognitive computing to deal with these cyber threats. Major security players in the industry have already introduced cognitive-based services for cyber threats detection and security analytics. Such cognitive systems not only detect threats but also assess systems and scan for vulnerabilities in the system and propose actions.
- The other side of the coin is that for cognitive computing we need huge volumetric data, now securing the privacy of the data is also of utmost importance. To take full advantage of cognitive computing we need to build a large database of information, and at the same time also maintain its confidentiality and prevent data leakage.

- **Healthcare**
- We have already briefly discussed the application of Cognitive Computing in healthcare. Recent advancement in cognitive computing has helped medical professionals to make better treatment decisions and improves the efficiency of the Medical professionals and also improves the outcomes of the patients. It is a self-learning algorithm that uses Machine learning algorithms, data mining techniques, visual recognition, and natural language processing, dependent on real-time patient information, medical transcripts, and other data. The system processes an enormous amount of data instantly to answer specific questions and make intelligent recommendations. Cognitive computing in healthcare links the functioning of humans and machines where computers and the human brain truly overlap to improve human decision-making. This will empower doctors and other medical professionals to better diagnose and treat their patients and above all helps in planning customized treatment modules. For example, Genome Medicine is one such area that has evolved by cognitive computing.

- **Education**
- Cognitive computing is going to change how the education industry has been working. It has already started bringing a few of the changes. It will change how the schools, colleges, and universities have been functioning, and it will help to provide personalized study material to students. Can you even imagine how fast a cognitive system can search the library or the journals and research papers from a digital library?
- A cognitive assistant can provide personal tutorials to students, guide them through the coursework, and can also help students to understand certain critical concepts at their own pace. It can also guide students in selecting the courses depending upon their interest. It can act as a career counsellor. Cognitive computing will not only help students but will also help teachers, support staff, and administrative staff for delivering better service, preparing student reports and feedback.

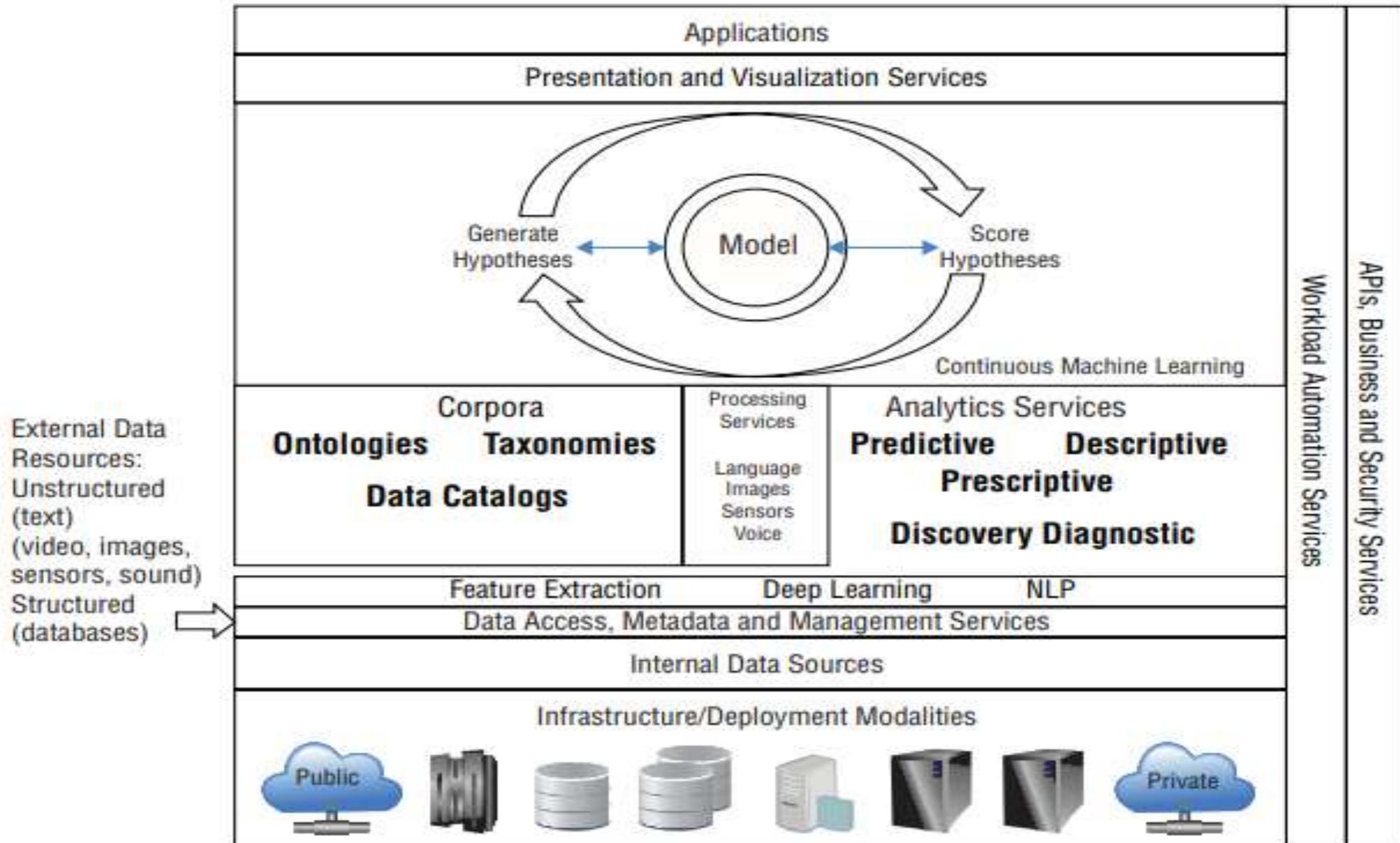


Figure : Elements of a cognitive system

- there are 3 essential components of a cognitive computing system:
- 1. **A way of interpreting input:** A cognitive computing system needs to answer a question or provide a result based on an input. That input might be a search term, text phrase, a query asked in natural language, or it may be a response to an action of some sort (ex: procurement of a product).
- The first thing a system needs to do is understand the *context* of the signal. Examples: location, speed of motion (in case of smartphone personal assistants, for illustration purposes). Such context info will enable the system to narrow down the potential responses to those that are more appropriate.
- Cognitive computing systems need to start somewhere — they need to “know” or expect something about the user to interpret the input. The more contextual clues that can be derived, defined or implied, the easier it will be to narrow the appropriate types of information to be returned.

- **2. A body of content/information that supports the decision:** The purpose of cognitive computing is to help humans make choices and solve problems. But the system does not make up the answer. Even synthesis of new knowledge is based on foundational knowledge.
- The “***corpus***” or domain of information is a key component. The more effectively that information is curated, the better the end results. Knowledge structures are important, taxonomies and metadata are required, and some form of information hygiene is required.
- In order to create a cognitive system, there needs to be organizational structures for the content which provide meaning to rather unstructured content.
- IBM Watson for example ingests many structured (as well as semi-structured) repositories of information: dictionaries, news articles and databases, taxonomies, and ontologies (ex , Wordnet, DBpedia..) These sources provide the information needed to respond to questions, forming the *corpus* of information that Watson draws upon.
- This initial modeling requires investment of time + expertise to build the foundational elements from which the system can then synthesize responses. Each of these actions requires content modeling and metadata structures, use cases (supported by customer engagement approach).

- **3. A way of processing the signal against the content/info corpus:** This is where machine learning for example comes into play. ML has for long been applied to categorization and classification approaches, and advanced text analytics. The processing might be in the form of a query/matching algorithm or may involve other mechanisms to interpret the query, transform it, reduce ambiguity, derive syntax, define word sense, deduce logical relationships or otherwise parse/process the signal against the *corpus*.
- **Machine learning** has many “flavours” — from various types of supervised learning approaches (where a known sample or result is used to instruct the system what to look for), to classes of unsupervised approaches (where the system is simply asked to identify/form patterns and outliers) - and even combinations of these approaches at different stages of the process. (an unsupervised learning approach could, for instance, identify hidden structures so the output could be applied as a “training set” to other data sources).
- The key here is to iteratively improve the system’s performance over time by approximating an output and using that as an input for the next round of processing. In some cases, incorrect answers (as judged by a human or another data source) might be input for the next time the system encounters the problem or question.

- **Infrastructure and Deployment Modalities**
- In a cognitive system it is critical to have a flexible and agile infrastructure to support applications that continue to grow over time. As the market for cognitive solutions matures, a variety of public and private data need to be managed and processed. In addition, organizations can leverage Software as a Service (SaaS) applications and services to meet industry-specific requirements. A highly parallelized and distributed environment, including compute and storage cloud services, must be supported.

- **Data Access , Metadata & Management Services** → Because cognitive computing centers around data, it is not surprising that the sourcing, accessing, and management of data play a central role. Therefore, before adding and using that data, there has to be a range of underlying services. To prepare to use the ingested data requires an understanding of the origins and lineage of that data. Therefore, there needs to be a way to classify the characteristics of that data such as when that text or data source was created and by whom. In a cognitive system these data sources are not static. There will be a variety of internal and external data sources that will be included in the corpus. To make sense of these data sources, there needs to be a set of management services that prepares data to be used within the corpus. Therefore, as in a traditional system, data has to be vetted, cleansed, and monitored for accuracy.
- **Corpus, Taxonomies & Data Catalogs** → *Corpus* is the knowledge base of ingested data and is used to manage codified knowledge. Such data is primarily be text-based (documents, textbooks, patient notes, customer reports, and such) but contemporarily there's support also for unstructured and semi-structured data (ex. videos, images, sounds). Ontologies are often developed by industry groups to classify industry-specific elements such as standard chemical compounds, machine parts, or medical diseases and treatments. In addition, the corpus may include ***ontologies*** that specify entities and their relationships (it is quite often required to use a subset of an industry-based ontology to include only the data that pertains to the focus of the cognitive system). A ***taxonomy*** provides context within the ontology.

- **Data Analytics Services** → these are the techniques used to develop understanding of the data ingested and managed within the corpus. A set of advanced algorithms are applied to develop the model for the cognitive system. Typically, users can take advantage of structured, unstructured, and semi-structured data that has been ingested and begin to use sophisticated algorithms to predict outcomes, discover patterns, or determine next best actions. These services do not live in isolation. They continuously access new data from the data access layer and pull data from the corpus. A number of advanced algorithms are applied to develop the model for the cognitive system.

Continuous machine learning

- Machine learning is the technique that provides the capability for the data to learn without being explicitly programmed. Cognitive systems are not static. Rather, models are continuously updated based on new data, analysis, and interactions. It is the procedure embraced by Cognitive Systems typically involves two sets of dynamics: (a) Hypotheses Generation and (b) Hypotheses Evaluation.

Hypothesis Generation and Evaluation

- *It is a testable assertion based on evidence that explains some observed phenomenon; goal here is to look for evidence to support (or refute) hypotheses*, this is typically accomplished through an iterative process of training the data. Training may occur automatically based on the systems analysis of data, or training may incorporate human end users. After training, it begins to become clear if the hypothesis is supported by the data. If the hypothesis is not supported by the data, the user has several options. For example, the user may refine the data by adding to the corpus, or change the hypothesis. To evaluate the hypothesis requires a collaborative process of constituents that use the cognitive system. Just as with the creation of the hypothesis, the evaluation of results refines those results and trains again.

- **The Learning Process**
- To learn from data you need **tools** to process both structured and unstructured data. For unstructured textual data, NLP services can interpret and detect patterns to support a cognitive system. Unstructured data such as images, videos, and sound requires deep learning tools. Data from sensors are important in emerging cognitive systems. Industries ranging from transportation to healthcare use sensor data to monitor speed, performance, failure rates, and other metrics and then capture and analyze this data in real time to predict behavior and change outcomes.

- **Presentation and Visualization Services**
- To interpret complex and often massive amounts of data requires new visualization interfaces.
- Data visualization is the visual representation of data as well as the study of data in a visual way. For example, a bar chart or pie chart is a visual representation of underlying data.
- Patterns and relationships in data are easier to identify and understand when visualized with structure, color, and such.
- The two basic types of data visualizations are static and dynamic. In either or both cases, there may also be a requirement for interactivity. Sometimes looking at the visualized representation of the data is not enough. You need to drill down, re-position, expand and contract, and so on. This interactivity enables you to “personalize” the views of the data so that you can pursue non-obvious manifestations of data, relationships, and alternatives. Visualization may depend on color, location, and proximity. Other critical issues that impact visualization include shape, size, and motion. Presentation services prepare results for output. Visualization services help to communicate results by providing a way to demonstrate the relationships between data. A cognitive system brings text or unstructured data together with visual data to gain insights. In addition, images, motion, and sound are also elements that need to be analyzed and understood. Making this data interactive through a visualization interface can help a cognitive system be more accessible and usable.

- **Cognitive Applications**

- A cognitive system must leverage underlying services to create applications that address problems in a specific domain. These applications that are focused on solving specific problems must engage users so that they gain insights and knowledge from the system.
- In addition, these applications may need to infuse processes to gain insight about a complex area such as preventive maintenance or treatment for a complex disease.
- An application may be designed to simulate the smartest customer service agent. The end goal is to turn an average employee into the smartest employee with many years of experience.
- A well-designed cognitive system provides the user with contextual insights based on role, the process, and the customer issue they are solving.
- The solution should provide the users insights so they make better decisions based on data that exists but is not easily accessible

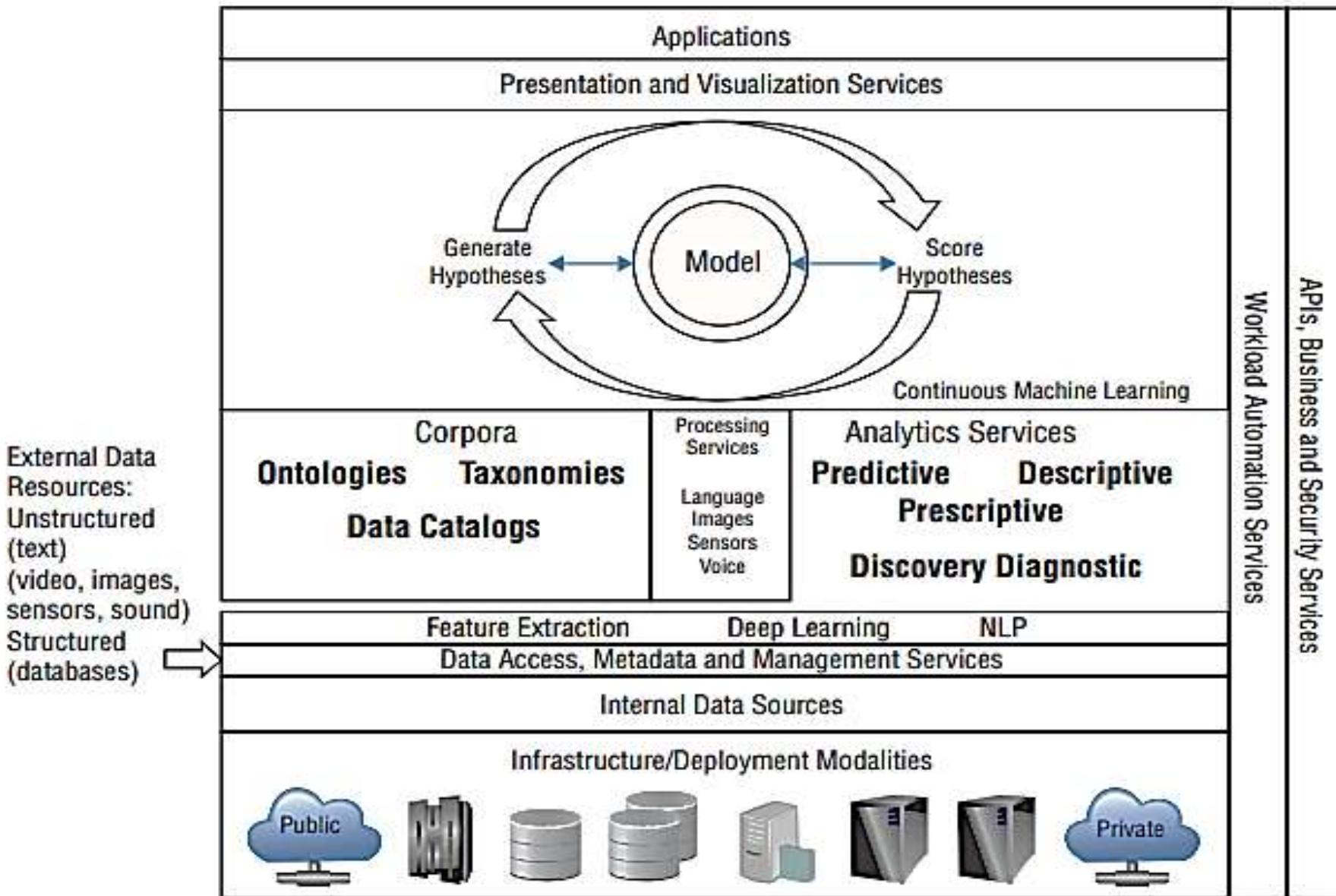


Figure : Architecture of a cognitive system

Building the Corpus

A corpus is a machine-readable representation of the complete record of a particular domain or topic. Experts in a variety of fields use a corpus or corpora for tasks such as linguistic analysis to study writing styles or even to determine authenticity of a particular work.

a number of questions have to be addressed early in the design phase for a cognitive computing system:

- Which internal and external data sources are needed for the specific domain areas and problems to be solved? Will external data sources be ingested in whole or in part?
- How can you optimize the organization of data for efficient search and analysis?
- How can you integrate data across multiple corpora?
- How can you ensure that the corpus is expanded to fill in knowledge gaps in your base corpus?

How can you determine which data sources need to be updated and at what frequency?

The choice of which sources to include in the initial corpus is critical. Sources ranging from medical journals to Wikipedia may now be efficiently imported in preparation for the launch of a cognitive system. In addition, it may be equally important to ingest information from videos, images, voice, and sensors. These sources are ingested at the data access layer Other data sources may also include subject-specific structured databases, ontologies, taxonomies, and catalogs

- **Corpus Management Regulatory and Security Considerations**
- Data sources and the movement of that data are increasingly becoming heavily regulated, particularly for personally identifiable information. Some general issues of data policies for protection, security, and compliance are common to all applications, but cognitive computing applications learn and derive new data or knowledge that may also be subject to a growing body of state, federal, and international legislation. When the initial corpus is developed, it is likely that a lot of data will be imported using extract-transform-load (ETL) tools. These tools may have risk management, security, and regulatory features to help the user guard against data misuse or provide guidance when sources are known to contain sensitive data. The availability of these tools doesn't absolve the developers from responsibility to ensure that the data and metadata is in compliance with applicable rules and regulations. Protected data may be ingested (for example, personal identifiers) or generated (for example, medical diagnoses) when the corpus is updated by the cognitive computing system.

- **Bringing Data into the Cognitive System**
- Unlike many traditional systems, the data that is ingested into the corpus is not static. You need to build a base of knowledge that adequately defines your domain space. You begin populating this knowledge base with data you expect to be important. As you develop the model in the cognitive system, you refine the corpus. Therefore, you will continuously add to the data sources, transform those data sources, and refine and cleanse those sources based on the model development and continuous learning.
- d be considered and potentially ingested. However, this does not mean that all sources will be of equal value.

Leveraging Internal and External Data Sources

- Most organizations already manage huge volumes of structured data from their transactional systems and business applications, and unstructured data such as text contained in forms or notes and possibly images from documents or corporate video sources. Although some firms are writing applications to monitor external sources such as news and social media feeds, many IT organizations are not yet well equipped to leverage these sources and integrate them with internal data sources.
- Most cognitive computing systems will be developed for domains that require ongoing access to integrated data from outside the organization. Just as an individual learns to identify the right external sources to support decision making—from newspapers to network news to social media on the Internet—a cognitive computing system generally needs to access a variety of frequently updated sources to keep current about the domain in which it operates.
- Also, like professionals who must balance the news or data from these external sources against their own experience, a cognitive system must learn to weigh the external evidence and develop confidence in the source as well as the content over time.
- For example, a popular magazine with articles on psychology may be a valuable resource, but if it contains data that is in conflict with a refereed journal article on the same topic, the system must know how to weigh the opposing positions.

- **Data Access and Feature Extraction Services**
- The data access level depicts the main **interfaces** between the cognitive computing system and the outside world.
- Any data to be imported from external sources must come through processes within this layer. Cognitive computing applications may leverage external data sources in formats as varied as **natural language text, video images, audio files, sensor data, and highly structured data formatted for machine processing**.
- The analogy to human learning is that this level represents the senses. The feature extraction layer has to complete two tasks.
- **First, it has to identify relevant data that needs to be analyzed. The second task is to abstract data as required to support machine learning.**
- The data access level is shown as separate but closely bound to the feature extraction level to reinforce the idea that **some data must be captured and then analyzed or refined before it is ready to be integrated into a corpus suitable for a particular domain**. Any data that is considered unstructured—from video and images to natural language text—must be processed in this layer to find the underlying structure.
- **Feature extraction and deep learning refer to a collection of techniques—primarily statistical algorithms**—used to transform data into representations that capture the essential properties in a more abstract form that can be processed by a machine learning algorithm.

- **Analytics Services**

- Analytics refers to a collection of techniques used to find and report on essential characteristics or relationships within a data set.
- In general, the use of an analytic technique provides insights about the data to guide some action or decision.
- A number of packaged algorithms such **regression analysis** are widely used within solutions.
- Within a cognitive system, a wide range of standard analytics components are available for **descriptive, predictive, and prescriptive** tasks within statistical software packages or in commercial component libraries.
- A variety of tools that support various tasks within cognitive computing systems are available.
- A cognitive computing system generally has additional analytical components embedded in the machine learning cycle algorithms.

- **Machine Learning** :Continuous learning without reprogramming is at the heart of all cognitive computing solutions. Although the techniques used to acquire, manage, and learn from data vary greatly, at their core most systems apply algorithms developed by researchers in the field of machine learning. Machine learning is a discipline that draws from computer science, statistics, and psychology.
- **Finding Patterns in Data** :A typical machine-learning algorithm looks for patterns in data and then takes or recommends some action based on what it finds. A pattern may represent a similar structure (for example, elements of a picture that indicate a face), similar values (a cluster of values similar to those found in another data set) or proximity (how “close” the abstract representation of one item is to another). Proximity is an important concept in pattern identification or matching. Two data strings representing things or concepts in the real world are “close” when their abstract binary representations have similar characteristics.
- **Supervised Learning** :Supervised learning refers to an approach that teaches the system to detect or match patterns in data based on examples it encounters during training with. The training data should include examples of the types of patterns or question-answer pairs the system will have to process. Learning by example, or modeling, is a powerful teaching technique that can be used for training systems to solve complex problems. After the system is operational, a supervised learning system also uses its own experience to improve its performance on pattern matching tasks. In supervised learning, the job of the algorithm is to create a mapping between input and output. The supervised learning model has to process enough data to get to the wanted level of validation, usually expressed as accuracy on the test data set. Both the training data and independent test data should be representative of the type of data that will be encountered when the system is operational.

- **Reinforcement Learning**
- Reinforcement learning is a special case of supervised learning in which the cognitive computing system receives feedback on its performance to guide it to a goal or good outcome.
- Unlike other supervised learning approaches, however, with reinforcement learning, the system is not explicitly trained with sample data. In reinforcement learning, the system learns to take next actions based on trial and error.
- Some typical applications of reinforcement learning include robotics and game playing. The machine learning algorithms assess the goodness or effectiveness of policies or actions, and reward the most effective actions.
- **A sequence of successful decisions results in reinforcement, which helps the system generate a policy that best matches the problem being addressed.**
- Reinforcement learning for cognitive computing is most appropriate where a system **must perform a sequence of tasks and the number of variables would make it too difficult to develop a representative training data set.**
- For example, reinforcement would be used in **robotics or a self-driving car**. The learning algorithm must discover an association between the reward and a sequence of events leading up to the reward. Then the algorithm can try to optimize future actions to remain in a reward state.

- **Unsupervised Learning**
- Unsupervised learning refers to a machine learning approach **that uses inferential statistical modeling algorithms to discover rather than detect patterns or similarities in data**.
- An unsupervised learning system can **identify new patterns, instead of trying to match a set of patterns it encountered during training**. Unlike supervised learning, unsupervised learning is based **solely on experience with the data rather than on training with sample data**. Unsupervised learning requires the system to discover which relationships among data elements or structures are important by analyzing attributes like **frequency of occurrence, context (for example, what has been seen or what has occurred previously), and proximity**. Unsupervised learning is the best approach for a cognitive computing system when an expert or user cannot give examples of typical relationships or question-answer pairs as guides to train the system.
- This may be due to the complexity of the data, when there are too many variables to consider, or when the structure of the data is unknown (for example, evaluating images from a surveillance camera to detect which person or persons are behaving differently from the crowd) **Unsupervised learning is also appropriate when new patterns emerge faster than humans can recognize them so that regular training is impossible**. For example, a cognitive computing system to evaluate network threats must recognize anomalies that may indicate an attack or vulnerability that has never been seen before

- **Hypotheses Generation and Scoring**
- A hypothesis in science is a testable assertion based on evidence that explains some observed phenomenon or relationship between elements within a domain.
- The key concept here is that a hypothesis has some supporting evidence or knowledge that makes it a plausible explanation for a causal relationship. It isn't a guess. When a scientist formulates a hypothesis as an answer to a question, it is done in a way that allows it to be tested.
- **The hypothesis actually has to predict an experimental outcome. An experiment or series of experiments that supports the hypothesis increases confidence in the ability of the hypothesis to explain the phenomenon.**
- This is conceptually similar to a hypothesis in logic, generally stated as “if P then Q”, where “P” is the hypothesis and “Q” is the conclusion. In the natural sciences we conduct experiments to test hypotheses.
- Using formal logic we can develop proofs to show that a conclusion follows from a hypothesis (or that it does not follow). **In a cognitive computing system we look for evidence—experiences and data or relationships between data elements—to support or refute hypotheses. That is the basis for scoring or assigning a confidence level for a hypothesis. If a cognitive computing hypothesis can be expressed as a logical inference, it may be tested using mechanical theorem proving algorithms.** Typically, however, cognitive computing applications solve problems in domains with supporting data that are not so neatly structured. These domains, like medicine and finance, have rich bodies of supporting data that are better suited to statistical methods like those used in scientific experimental design.

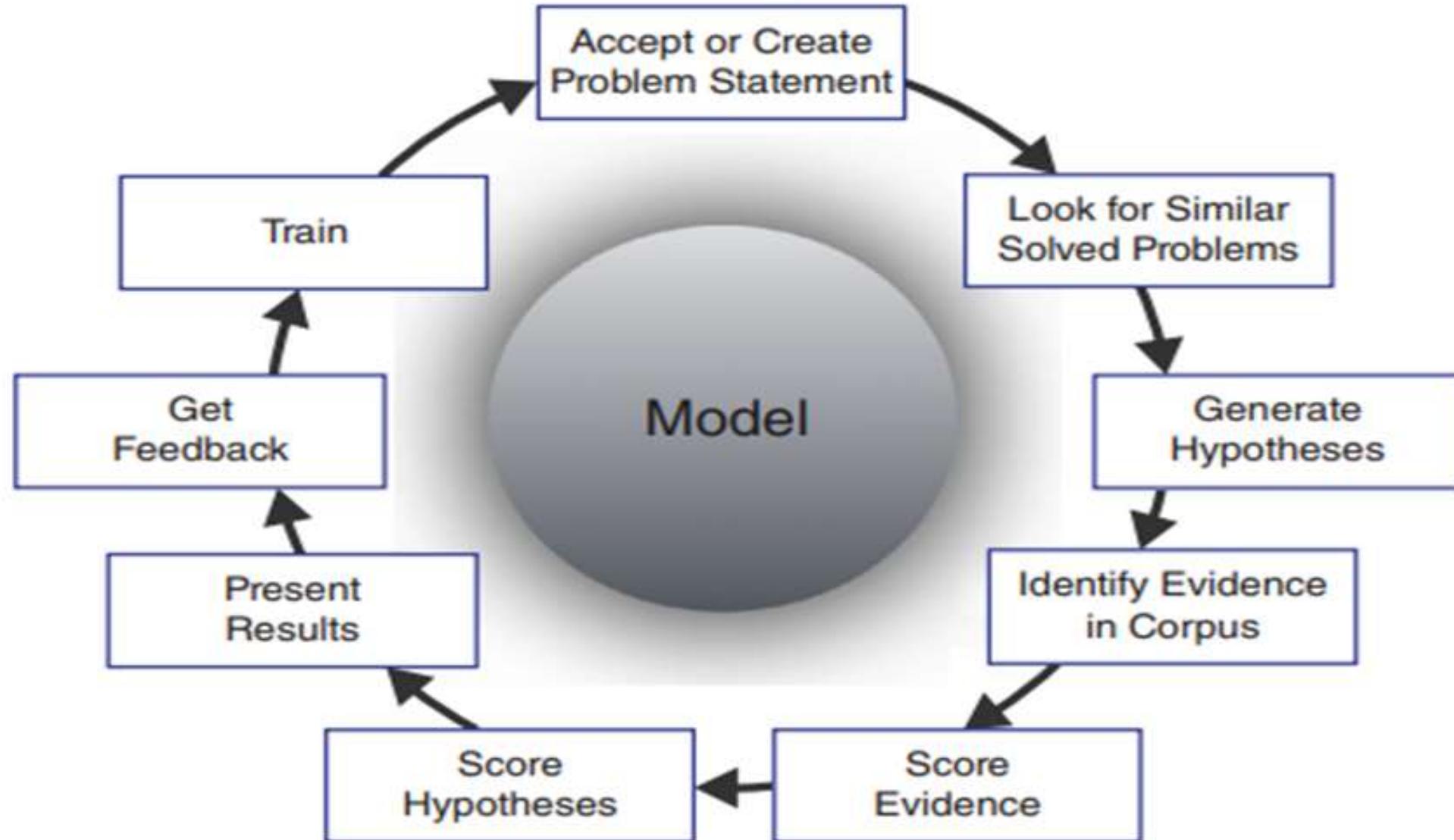


Figure The continuous machine learning process

- **Hypothesis Generation**
- The discussion about the scientific method said that a hypothesis is formulated to answer a question about a phenomenon based on some evidence that made it plausible.
- The experimental process is designed to test whether the hypothesis applies in the general case, not just with the evidence that was used to develop the hypothesis.
- there are two key ways a hypothesis may be generated. The first is in response to an explicit question from the user, such as “What might cause my fever of 102 and sore throat?” it may recognize that there are too many answers to be useful and request more information from the user to refine the set of likely causes.
- This approach to hypothesis generation is frequently used when the goal is to detect a relationship between cause and effect in a domain in which there is a known set of causes and a known set of effects, but there are so many combinations that the mapping of all causes to all effects is an intractable problem for humans to solve.

- The second type of **hypothesis generation** does not depend on a user asking a specific question.
- Instead, the system constantly looks for anomalous data patterns that may indicate threats or opportunities.
- Detecting a new pattern creates a hypothesis based on the nature of the data.
- For example, if the system is monitoring network sensors to detect threats, a new pattern may create a hypothesis that this pattern is a threat, and the system must either find evidence to support or refute that hypothesis.
- If the system is monitoring real-time stock transactions, a new pattern of buying behavior may indicate an opportunity.
- In these systems, the type of hypotheses that will be generated depends on assumptions of the system designers rather than on the actions of the users.
- Both types of application have the system generate one or more hypotheses based on an event, but in the first case, the event is a user question, and in the second it is driven by a change in the data itself.

Hypothesis Scoring:

- cognitive computing systems build a corpus of relevant data for a problem domain. Then, in response to a user question or change in the data, the system generates one or more hypotheses to answer a user's question or explain a new data pattern.
- The next step is to evaluate or score these hypotheses based on the evidence in the corpus, and then update the corpus and report the findings to the user or another external system.
- Hypothesis scoring is a process in which the representation of the hypothesis is compared with data in the corpus to see what evidence exists to support the hypothesis, and what may actually refute it (or rule it out as a valid possible explanation).
- scoring or evaluating a **hypothesis** is a process of applying statistical methods to the hypothesis-evidence pairs to assign a confidence level to the hypothesis. The **actual weights** that are assigned to each piece of supporting evidence can be adjusted based on experience with the system and feedback during training and during the operational phase.
- If none of the hypotheses score above a predetermined threshold, the system may ask for **more evidence** (a new diagnostic blood test) if that information could change the confidence in the hypothesis.
- Techniques for measuring the proximity of two data elements or structures for pattern matching, such as the fit between two hypothesis-evidence pairs, generally rely on a binary vector representation (such as a sparse distributed representation [SDR]) that can be manipulated using matrix algebra with readily available tools.
- The generation/scoring loop may be set to continue until a user is satisfied with the answer or until the system has evaluated all options

Presentation and Visualization Services

- As a cognitive computing system cycles through the hypothesis generation and scoring cycle, it may produce new answers or candidate answers for a user.
- In some situations, the user may need to provide additional information. How the system presents these findings or questions will have a big impact on the usability of the system in two ways.
- First, when presenting data supporting a hypothesis such as a medical diagnosis or recommended vacation plan, the system should present the finding in a way that conveys the most meaning with the least effort on the part of the user and support the finding with relevant evidence.
- Second, when the system requires additional details to improve its confidence in one or more hypotheses, the user must present that data in a concise and unambiguous way.

- The general advantage for visualization tools is their capability to graphically depict relationships between data elements in ways that focus attention on trends and abstraction rather than forcing the user to find these patterns in the raw data
- Following are three main types of services available to accomplish these goals.
 - Narrative solutions, which use natural language generation techniques to tell a story about the data or summarize findings in natural language. This is appropriate for reporting findings or explanations about the evidence used to arrive at a conclusion or question.
 - Visualization services present data in nontext forms, including:
 - Graphics, ranging from simple charts and graphs to multidimensional representations of relationships between data.
 - Images, selected from the data to be presented or generated from an underlying representation. (For example, if feature extraction detects a “face” object, a visualization service could generate a “face” or pictograph from a standard features library.)
 - Gestures or animation of data designed to convey meaning or emotion.
 - Reporting services refers to functions that produce structured output, such as database records, that may be suitable for humans or machines

- **Infrastructure**
- The infrastructure/deployment modalities layer represents the hardware, networking, and storage foundation for the cognitive computing application.
- As noted in the discussion on emerging neuromorphic hardware architectures, “Future Applications for Cognitive Computing,” most cognitive computing systems built in the next decade will primarily use conventional hardware.
- The two major design considerations for cognitive computing infrastructure decisions are:

The two major design considerations for cognitive computing infrastructure decisions are:

■ **Distributed Data Management**—For all but the smallest applications, cognitive computing systems can benefit from tools to leverage distributed external data resources and to distribute their operational workloads.

- Managing the ongoing ingestion of data from a variety of external sources requires a robust infrastructure that can efficiently import large quantities of data.
- Based on the domain, this may be a combination of structured and unstructured data available for batch or streaming ingestion.
- Today, a cloud-first approach to data management is recommended to provide maximum flexibility and scalability.

■ **Parallelism**—

- The fundamental cognitive computing cycle of hypothesis generation and scoring can benefit enormously from a software architecture that supports parallel generation/scoring of multiple hypotheses, but performance ultimately depends on the right hardware.
- Allocating each independent hypothesis to a separate hardware thread or core is a requirement in most cases for acceptable performance as the corpus scales up and the number of hypotheses increases.
- Although performance improvements should be seen within the system as it learns, the rate of data expansion in the corpus generally outpaces this performance improvement.
- That argues strongly for the selection of hardware architecture that supports relatively seamless expansion with additional processors

Design Principles for Cognitive Systems

- In a cognitive computing system, the *model* refers to the corpus and the set of assumptions and algorithms that generate and score hypotheses to answer questions, solve problems, or discover new insights.
- How you model the world determines what kind of predictions you can make, patterns and anomalies you can detect, and actions you can take.
- The initial model is developed by the designers of the system, but the cognitive system will update the model and use the model to answer questions or provide insights.
- The *corpus* is the body of knowledge that machine learning algorithms use to continuously update that model based on its experience, which may include user feedback.

- **Design Principles for Cognitive Systems**

- A cognitive system is designed to use a model of a domain to predict potential outcomes.
- Designing a cognitive system involves multiple steps. It requires an understanding of the available data, the types of questions that need to be asked, and the creation of a corpus comprehensive enough to support the generation of hypotheses about the domain based on observed facts.
- Therefore, a cognitive system is designed to create hypotheses from data, analyze alternative hypotheses, and determine the availability of supporting evidence to solve problems.
- By leveraging machine learning algorithms, question analysis, and advanced analytics on relevant data, which may be structured or unstructured, a cognitive system can provide end users with a powerful approach to learning

COGNITIVE DESIGN PRINCIPLES

- A cognitive system should be able to understand any kind of data and communicate with multiple data sources to process structured and unstructured data and generate meaningful insights or recommendations.
- Natural language processing, text to speech conversion, speech to text conversions are used widely in cognitive systems. The process involved in designing a cognitive system is iterative and should have 7 basic steps as shown in Figure.
- These are not sequential steps, and hence in the implementation process, we can go from one step to other depending on the results at each step

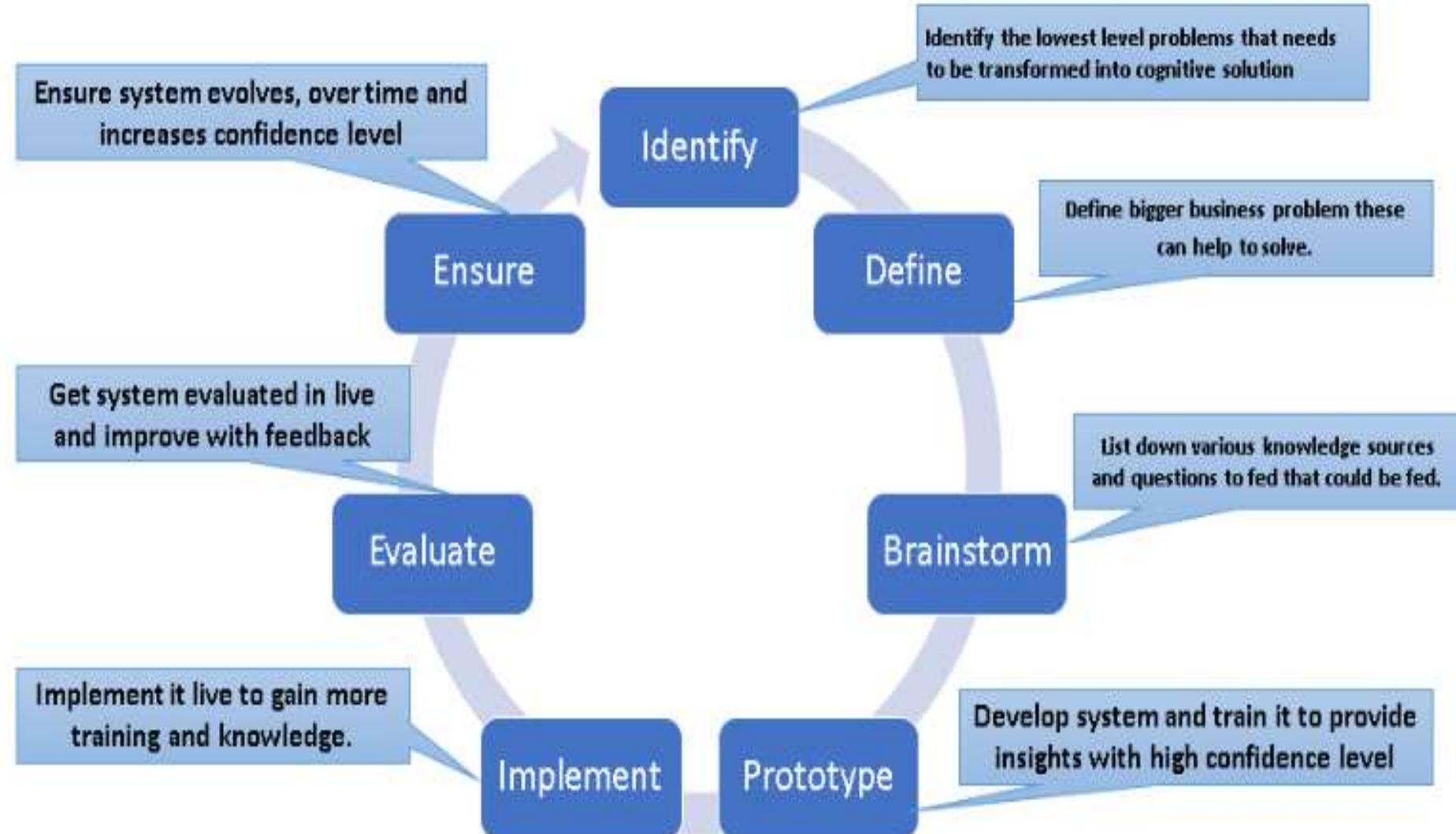
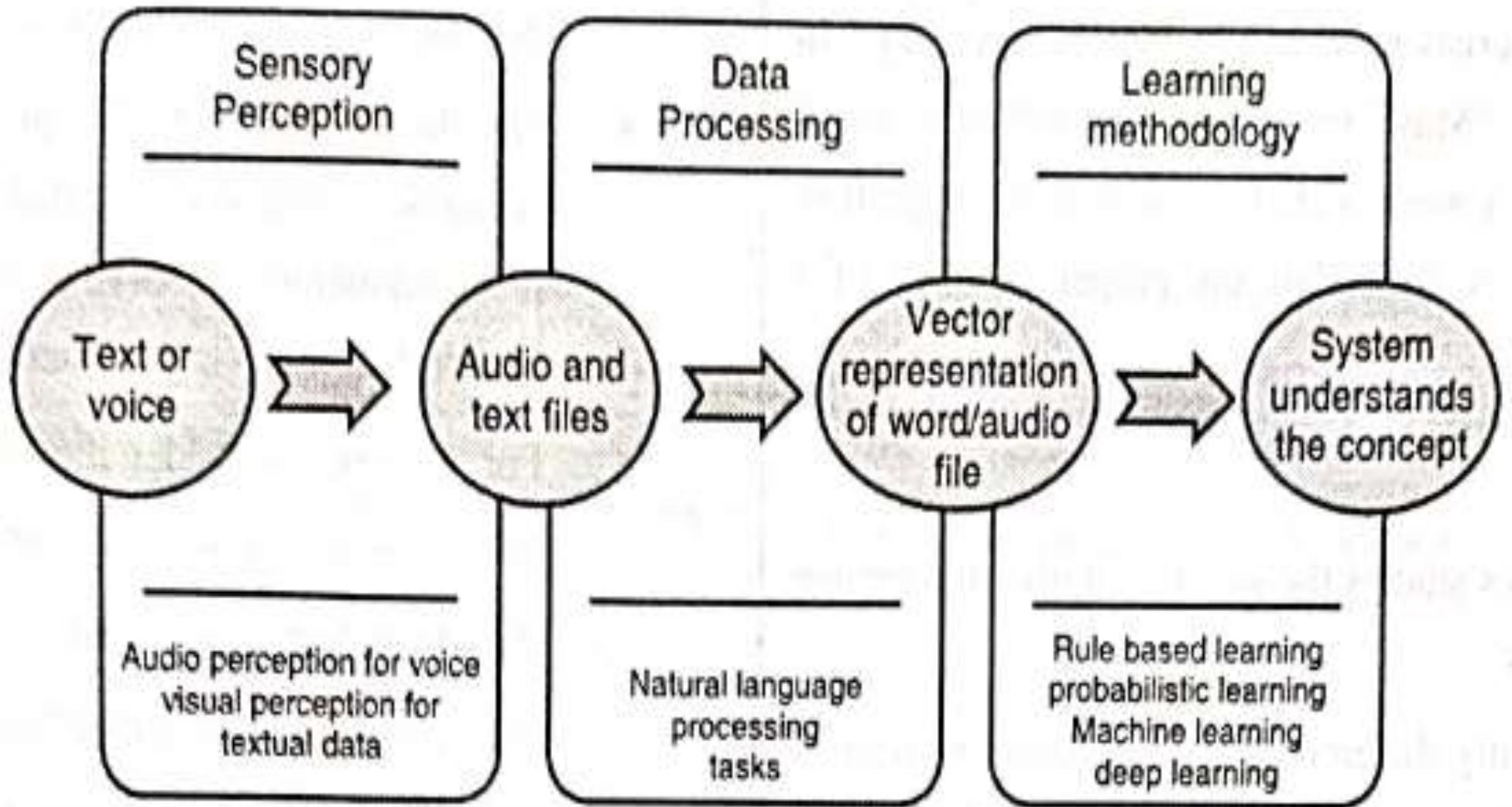


Figure 1: Cognitive design principles

- The **maturity** of a cognitive model is measured by the **confidence** level of the recommendations or the actions taken by a cognitive process.
- Hence, improving the confidence level of a system plays a vital role in making the cognitive process successful.
- There are multiple factors which can influence the confidence level depending on the data it has access to.
- In a normal scenario, the **confidence level** can be improved based on the **human actions taken to the recommendations** suggested by the cognitive system.
- These are self-evolving and require human intervention for updating the context and improving the confidence level.

The Role of NLP in a Cognitive System

- NLP is a set of techniques that extract meaning from text. These techniques determine the meaning of a word, phrase, sentence, or document by recognizing the grammatical rules—the predictable patterns within a language.
- They rely, as people do, on dictionaries, repeated patterns of co-occurring words, and other contextual clues to determine what the meaning might be.
- NLP applies the same known rules and patterns to make inferences about meaning in a text document.
- Further, these techniques can identify and extract elements of meaning, such as proper names, locations, actions, or events to find the relationships among them, even across documents.
- These techniques can also be applied to the text within a database and have been used for more than a decade to find duplicate names and addresses or analyze a comment or reason field, for instance, in large customer databases.



Cognitive Approach to NLP

The Importance of Context

- Translating unstructured content from a corpus of information into a meaningful knowledge base is the task of NLP.
- Linguistic analysis breaks down the text to provide meaning. The text has to be transformed so that the user can ask questions and get meaningful answers from the knowledge base.
- Any system, whether it is a structured database, a query engine, or a knowledge base, requires techniques and tools that enable the user to interpret the data.
- The key to getting from data to understanding is the quality of the information. With NLP it is possible to interpret data and the relationships between words.
- It is important to determine what information to keep and how to look for patterns in the structure of that information to distill meaning and context.
- NLP enables cognitive systems to extract meaning from text, Phrases, sentences, or complex full documents provide context so that you can understand the meaning of a word or term.

- This context is critical to assessing the true meaning of text-based data.
- Patterns and relationships between words and phrases in the text need to be identified to begin to understand the meaning and actual intent of communications.
- When humans read or listen to natural language text, they automatically find these patterns and make associations between words to determine meaning and understand sentiment.
- There is a great deal of ambiguity in language, and many words can have multiple meanings depending on the subject matter being discussed or how one word is combined with other words in a phrase, sentence, or paragraph.
- When humans communicate information there is an assumption of context.
- There are many layers to the process of understanding meaning in context. Various techniques are used such as building a feature vector from any information that can be extracted from the document. Statistical tools help with information retrieval and extraction. These tools can help to annotate and label the text with appropriate references (that is, assigning a name to an important person in the text). When you have a sufficient amount of annotated text, machine learning algorithms can ensure that new documents are automatically assigned the right annotations.

Connecting Words for Meaning

- The nature of human communications is complicated. Humans are always transforming the way language is used to convey information.
- Two individuals can use the same words and even the same sentences and mean different things. We stretch the truth and manipulate words to interpret meaning.
- Therefore, it is almost impossible to have absolute rules for what words mean on their own and what they mean within sentences.
- To understand language we have to understand the context of how words are used in individual sentences and what sentences and meanings come before and after those sentences.
- We are required to parse meaning so that understanding is clear. It is not an easy task to establish context so that those individuals asking questions and looking for answers gain insights that are meaningful.

Understanding Linguistics

NLP is an interdisciplinary field that applies statistical and rules-based modeling of natural languages to automate the capability to interpret the meaning of language. Therefore, the focus is on determining the underlying grammatical and semantic patterns that occur within a language or a sublanguage (related to a specific field or market). For instance, different expert domains such as medicine or laws use common words in specialized ways. Therefore, the context of a word is determined by knowing not just its meaning within a sentence, but sometimes by understanding whether it is being used within a particular domain. For example, in the travel industry the word “fall” refers to a season of the year. In a medical context it refers to a patient falling. NLP looks not just at the domain, but also at the levels of meaning that each of the following areas provide to our understanding.

- **Language Identification and Tokenization**

In any analysis of incoming text, the first process is to identify which language the text is written in and then to separate the string of characters into words (tokenization). Many languages do not separate words with spaces, so this initial step is necessary.

- NLP looks not just at the domain, but also at the levels of meaning that each of the following areas provide to our understanding.

- **Language Identification and Tokenization**

In any analysis of incoming text, the first process is to identify which language the text is written in and then to separate the string of characters into words (tokenization). Many languages do not separate words with spaces, so this initial step is necessary.

Phonology

- Phonology is the study of the physical sounds of a language and how those sounds are uttered in a particular language.
- This area is important for speech recognition and speech synthesis but is not important for interpreting written text.
- However, to understand, for instance, the soundtrack of a video, or the recording of a call center call, not only is the pronunciation of the words important (regional accents such as British English or Southern United States), but the intonation patterns.
- A person who is angry may use the same words as a person who is confused; however, differences in intonation will convey differences in emotion.
- When using speech recognition in a cognitive system, it is important to understand the nuances of how words are said and the meaning that articulation or emphasis conveys.

- **Morphology**
- Morphology refers to the structure of a word.
- Morphology gives us the stem of a word and its additional elements of meaning. Is it singular or plural? Are the verbs first person, future tense, or conditional? This requires that words be partitioned into segments known as morphemes that help bring understanding to the meaning of terms.
- This is especially important in cognitive computing, since human language rather than computing language is the technique for determining answers to questions.
- Elements in this context are identified and arranged into classes. There are elements including prefixes, suffixes, infixes, and circumfixes. For example, if a word begins with “non-” it has a specific reference to a negative. There is a huge difference in meaning if someone uses the verb “come” versus the verb “came.” Combinations of prefixes and suffixes can be combined to form brand new words with very different meanings.

- Morphology is also used widely in speech and language translation as well as the interpretation of images.
- Although many dictionaries have been created to provide explanations of different constructions of words in various languages, it is impossible for these explanations to ever be complete (each human language has its own context and nuances that are unique).
- In languages such as English, rules are often violated. There are new words and expressions created every day.
- This process of interpreting meaning is aided by the inclusion of a lexicon or repository of words and rules based on the grammar of a specific language.
- For example, through a technique called parts of speech tagging or tokenization, it is possible to encapsulate certain words that have definitive meaning.
- This may be especially important in specific industries or disciplines. For example, in medicine the term “blood pressure” has a specific meaning,
- However, the words blood and pressure when used independently can have a variety of meanings. Likewise, if you look at the elements of a human face, each component may independently not provide the required information

- **Lexical Analysis**
- Lexical analysis within the context of language processing is a technique that connects each word with its corresponding dictionary meaning.
- However, this is complicated by the fact that many words have multiple meanings. The process of analyzing a stream of characters from a natural language requires a sequence of tokens (a string of text, categorized according to the rules as a symbol such as a number or comma).
- Specialized taggers are important in lexical analysis. For example, an n-gram tagger uses a simple statistical algorithm to determine the tag that most frequently occurs in a reference corpus.
- The analyzer (sometimes called a lexer) categorizes the characters according to the type of character string.
- When this categorization is done, the lexer is combined with a parser that analyzes the syntax of the language so that the overall meaning can be understood.
- The lexical syntax is usually a regular language whose alphabet consists of the individual characters of the source code text.
- The phrase syntax is usually a context-free language whose alphabet consists of the tokens produced by the lexer.
- Lexical analysis is useful in predicting the function of grammatical words that initially could not be identified. For example, there might be a word like “run” that has multiple meanings and can be a verb or a noun

- **Syntax and Syntactic Analysis**

- Syntax applies to the rules and techniques that govern the sentence structure in languages.
- The capability to process the syntax and semantics of natural language is critical to a cognitive system because it is important to deduct inferences about what language means based on the topic it is being applied to.
- Therefore, although words may have a general meaning when used in conversation or written documents, the meaning may be entirely different when used in context of a specific industry.
- For example, the word “tissue” has different definitions and understanding based on the context of its use. For example, in biology, tissue is a group of biological cells that perform a specific function. However, a tissue can also be used to wrap a present or wipe a runny nose. Even within a domain context, there can still be word-sense ambiguity. In a medical context, “tissue” can be used with skin or a runny nose.

- Syntactical analysis helps the system understand the meaning in context with how the term is used in a sentence.
- This syntactic analysis or parsing is the overall process for analyzing the string of symbols in a natural language based on conforming to a set of grammar rules.
- Within computational linguistics, parsing refers to the technique used by the system to analyze strings of words based on the relationship of those words to each other, in context with their use.
- Syntactical analysis is important in the question-answering process. For example, suppose you want to ask, “Which books were written by British women authors before the year 1800?” The parsing can make a huge difference in the accuracy of the answer. In this case, the subject of the question is books. Therefore, the answer would be a list of books. If, however, the parser assumed that “British woman authors” was the topic, the answer would instead be a list of authors and not the books they wrote

- **Construction Grammars**
- Although there are many different approaches to grammar in linguistics, construction grammar has emerged as an important approach for cognitive systems.
- When approaching syntactical analysis, the results are often represented in a grammar that is often written in text.
- Therefore, interpretation requires a grammar model that understands text and its semantics. Construction grammar has its roots in cognitive-oriented linguistic analysis.
- It seeks to find the optimal way to represent relationships between structure and meaning. Therefore, construction grammar assumes that knowledge of a language is based on a collection of “form and function pairings.”
- The “function” side covers what is commonly understood as meaning, content, or intent; it usually extends over both conventional fields of semantics and pragmatics.
- Construction grammar was one of the first approaches that set out to search for a semantically defined deep structure and how it is manifested in linguistic structures.
- Therefore, each construction is associated with the principle building blocks of linguistic analysis, including phonology, morphology, syntactic, semantics, pragmatics, discourse, and prosodic characteristics.

- **Discourse Analysis**
- One of the most difficult aspects of NLP is to have a model that brings together individual data in a corpus or other information source so that there is coherency.
- It is not enough to simply ingest vast amounts of data from important information sources if the meaning, structure, and intention cannot be understood.
- Certain assertions may be true or false depending on the context. For example, people eat animals, but people are animals, and in general don't eat each other. However, timing is important to understanding context. For example, during the 18th century, cigarette smoking was thought to be beneficial to the lungs. Therefore, if someone were ingesting an information source from that period of time, it would assume that smoking was a good thing. Without context there would be no way to know that premise of that data was incorrect.
- Discourse is quite important in cognitive computing because it helps deal with complex issues of context.
- When a verb is used, it is important to understand what that verb is associated with in terms of reference. Within domain-specific sources of data you need to understand the coherence of related information sources.
- For example, what is the relationship between diabetes and sugar intake? What about the relationship between diabetes and high blood pressure? The system needs to be modeled to look for these types of relationships and context.

- **Pragmatics**

- Pragmatics is the aspect of linguistics that tackles one of the fundamental requirements for cognitive computing: the ability to understand the context of how words are used.
- A document, an article, or a book is written with a bias or point of view.
- For example, the writer discussing the importance of horses in the 1800s will have a different point of view than the writer talking about the same topic in 2014.
- In politics, two documents might discuss the same topic and take opposite sides of the argument.
- Both writers could make compelling cases for their point of view based on a set of facts.
- Without understanding the background of the writer, it is impossible to gain insight into meaning.
- The field of pragmatics provides inference to distinguish the context and meaning about what is being said.
- Within pragmatics, the structure of the conversation within text is analyzed and interpreted.

- **Techniques for Resolving Structural Ambiguity**
- Disambiguation is a technique used within NLP for resolving ambiguity in language.
- Most of these techniques require the use of complex algorithms and machine learning techniques.
- Even with the use of these advanced techniques, there are no absolutes. Resolution of ambiguity must always deal with uncertainties.
- We can't have complete accuracy; instead, we rely on the probability of something being most likely to be true.
- This is true in human language and also in NLP. For example, the phrase, “The train ran late,” does not mean that the train could “run”; rather the train was expected to arrive at the station later than it was scheduled.
- There is little ambiguity in this statement because it is a commonly known phrase.
- However, other phrases are easily misunderstood. For example, examine the phrase, “The police officer caught the thief with a gun.” One might decide that it was the police officer that used a gun to arrest the thief. However, it may well have been the thief was using the gun to commit a crime. Sometimes, the truth of meaning can be hidden inside a complicated sentence.
- Because cognitive computing is a probabilistic rather than a deterministic approach to computing, it is not surprising that probabilistic parsing is one way of solving disambiguation. Probabilistic parsing approaches use dynamic programming algorithms to determine the most accurate explanation for a sentence or string of sentences.

- **Importance of Hidden Markov Models**
- One of the most important statistical models for processing both image and speech understanding are Markov Models . Increasingly, these models are fundamental to understanding the hidden information inside images, voice, and video.
- It is now clear that it is complicated to gain a clear understanding of the meaning that is often hidden within language.
- While the human brain automatically understands how to cope with the fact that the real meaning of a sentence may be indirect, “The cow jumped over the moon” may seem like an impossible task if the sentence were read literally.
- However, the sentence refers to a song for young children and is intended to be unrealistic and silly. The human mind calculates the probability that this sentence is intended to be a literal action between the cow and the moon.
- The human understands through the context of their environment, which may dictate a specific interpretation. The way systems interpret language requires a set of statistical models that are an evolution of a model developed by A.A. Markov in the early 1900s.
- Markov asserted that it was possible to determine the meaning of a sentence or even a book by looking at the frequency that words would occur in text and the statistical probability that an answer was correct. The most important evolution of Markov’s model for NLP and cognitive computing is Hidden Markov Models (HMMs)

- The premise behind HMMs is that the most recent data will tell you more about your subject than the distant past because the models are based on the foundations of probability.
- HMMs therefore help with predictions and filtering as well as smoothing of data. Hidden Markov Models (HMMs) are intended to interpret “noisy” sequences of words or phrases based on probabilistic states.
- In other words, the model takes a group of sentences or sentence fragments and determines the meaning. Using HMMs requires thinking about the sequence of data. HMMs are used in many different applications including speech recognition, weather patterns, or how to track the position of a robot in relationship to its target.
- Therefore, Markov models are very important for when you need to determine the exact position of data points when there is a very noisy data environment. Applying HMMs allows the user to model the data sequence supported by a probabilistic model.
- Within the model, an algorithm using supervised learning will look for repeating words or phrases that indicate the meaning and how various constructs affect the likelihood that a meaning is true.
- Markov models assume that the probability of a sequence of words will help us determine the meaning.
- There are a number of techniques used in HMMs to estimate the probability that a word sequence has a specific meaning.
- For example, there is a technique called **maximum likelihood estimation** that is determined by normalizing the content of a corpus.

- The value of HMMs is that they do the work of looking for the underlying state of sentences or sentence fragments. Therefore, as the models are trained on more and more data they abstract constructs and meaning.
- The capability to generate probabilities of meaning and the state transition are the foundation of HMM models and are important in cognitive understanding of unstructured data.
- The models become more efficient in their ability to learn and to analyze new data sources. Although HMMs are the most prevalent method in understanding the meaning of sentences, another technique called maximum entropy is designed to establish probability through the distribution of words.
- To create the model, labeled training data is used to constrain the model. This classifies the data. There are a number of approaches that are important in understanding a corpus in context with its use in a cognitive system.

- **Word-Sense Disambiguation (WSD)**
- Not only do you have to understand a term within an ontology, it is critical to understand the meaning of that word. This is especially complex when a single word may have multiple meanings depending on how it is used. Given this complexity, researchers have been using supervised machine learning techniques.
- A classifier is a machine learning approach that organizes elements or data items and places them into a particular class. There are different types of classifiers used depending on the purpose.
- For example, document classification is used to help identify where particular segments of text might belong in your taxonomy.
- Often classifiers are used for pattern recognition and therefore are instrumental in cognitive computing. When a set of training data is well understood, supervised learning algorithms are applied. However, in situations in which the data set is vast and it cannot easily be identified, **unsupervised learning techniques are often used to determine where clusters are occurring**.
- Scoring of results is important here because patterns have to be correlated with the problem being addressed. Other methods may rely on a variety of dictionaries or lexical knowledge bases. This is especially important when there is a clear body of knowledge—in health sciences, for example.
- There are many taxonomies and ontologies that define diseases, treatments, and the like. When these elements are predefined, it allows for interpretation of information into knowledge to support decision making. For example, there are well known characteristics of diabetes at the molecular level, the evolution of the disease, and well-tested, successful treatments.

- **Semantic Web**
- NLP by itself provides a sophisticated technique for discovering the meaning of unstructured words in context with their usage. However, to be truly cognitive requires context and semantics.
- Ontologies and taxonomies are approaches that are expressions of semantics. In fact, the capability to combine natural language processing and the semantic web enables companies to more easily combine structured and unstructured data in ways that are more complicated with traditional data analytic methods.
- The semantic web provides the Resource Description Framework (RDF), which is the foundational structure used in the World Wide Web for processing metadata (information about the origins, structure, and meaning of data).
- RDF is used as a way to more accurately find data than would be found in a typical search engine. It provides the ability to rate the available content. RDF also provides a syntax for encoding the particular metadata with standards such as XML (Extensible Markup Language) that supports interoperability between data sources.
- One of the benefits of schemas that are written in RDF is that it provides one set of properties and classes for describing the RDF generated schemas. The semantic web is instrumental in providing a cognitive approach to gaining insights from a combination of structured and unstructured sources of information in a consistent way

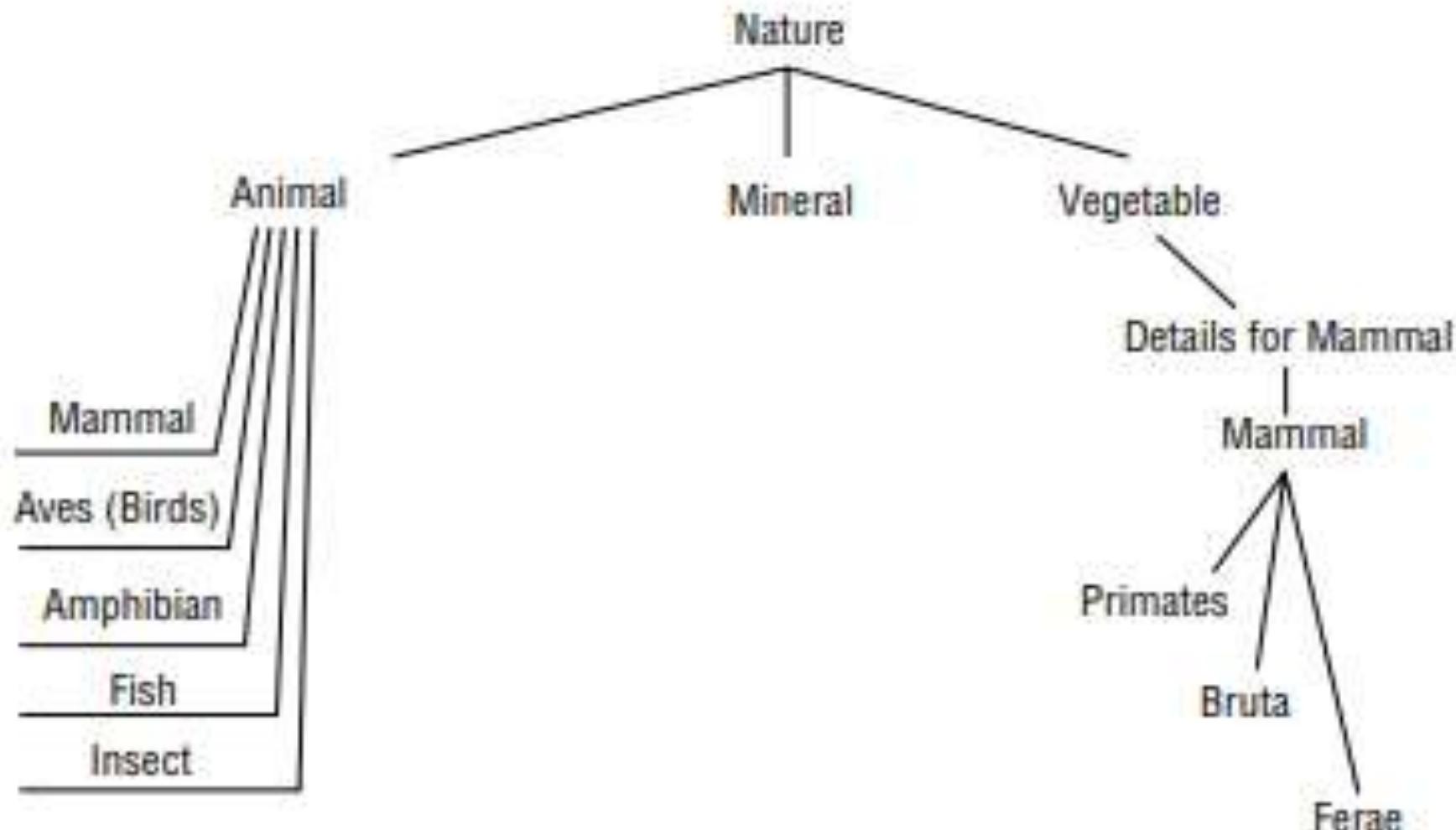
- **Applying Natural Language Technologies to Business Problems**
- Cognitive computing has found application in various real-world scenarios, enhancing human decision-making, automating complex tasks, and improving overall efficiency. Here are some examples of real-world use cases of cognitive computing
- **IBM Watson for Oncology** is used for analyzing medical literature, clinical trial data, and patient records to recommend personalized treatment options for cancer patients.
- Cognitive computing systems analyze vast amounts of financial data in real-time to detect patterns and anomalies, helping financial institutions identify potential fraudulent activities.
- Most Companies are utilizing cognitive computing to create intelligent virtual assistants and chatbots that can understand natural language, answer customer queries, and provide personalized assistance.
- Cognitive systems assist in automating the recruitment process by analyzing resumes, screening candidates, and even conducting initial interviews, streamlining the hiring process.
- **Enhancing the Shopping Experience**
- Leveraging the Connected World of Internet of Things

- **Leveraging the Connected World of Internet of Things**
- As more and more devices, from cars to highways and traffic lights, are equipped with sensors, there will be the ability to make decisions about what actions to take as conditions change.
- Traffic management is a complex problem in many large metropolitan areas. If city managers can interact with sensor-based systems combined with unstructured data about events taking place (rallies, concerts, and snow storms), alternative actions can be looked at.
- A traffic manager may want to ask questions about when to reroute traffic under certain circumstances. If that manager can use an NLP interface to a cognitive system, these questions can be answered in context with everything from weather data to density of traffic to the time when an event will start. Individual domains such as traffic routing and weather predictions will each have their own Hidden Markov Models.
- In a cognitive system it is possible to correlate this data across domains and models. Matching this data with an NLP engine that interprets textual data can result in significant results. The NLP question and answer interface can help the human interact with this complex data to recommend the next best action or actions

- **Voice of the Customer**
- The capability for companies to understand not only what their customers are saying but also how it will impact their relationship with that customer is increasingly important.
- One technique companies use to understand customer attitudes is sentiment analysis.
- Sentiment analysis combines text analytics, NLP, and computational linguistics in order to make sense of text-based comments provided by customers.
- For example, a company can analyze customer sentiment to predict sales for new product offerings from one of its divisions.
- However, a customer is not simply a customer for a single business unit.
- Many customers actually will do business with several different business units within the same company.
- Creating a corpus of customer data across business units can enable the customer service representative to understand all interactions with customers.
- Many of these interactions will be stored in notes in customer care systems.
- These same customers may add comments to social media sites and send e-mail messages directly to the company complaining about problems.
- There are subtleties in how customers use language that need to be understood to get a clear indication of customer intent

- **Fraud Detection**
- One of the most important applications for NLP and cognitive computing is the detection of fraud. Traditional fraud detection systems are intended to look for known patterns from internal and external threat databases. Determining risk before it causes major damage is the most important issue for companies dealing with everything from hackers to criminal gangs stealing intellectual property.
- Although companies leverage firewalls and all sorts of systems that put up a barrier to access, these are not always effective. Smart criminals often find subtle techniques that go under the radar of most fraud detection systems. Having the capability to look for hidden patterns and for anomalies is critical to preventing an event from happening.
- In addition, leveraging thousands of fraudulent claims documents, an insurance company can be better prepared to detect subtle indications of fraud.
- NLP-based cognitive approaches can enable the user to ask questions related to the corpus of data that has been designed based on a model of both acceptable and unacceptable behavior.
- This corpus can be fed with new information about detected schemes happening somewhere in the world.
- Understanding not only the words but also their context across many data sources can be applied to fraud prevention. Understanding word sense in complex documents and communications can be significant in preventing fraud.

- **Taxonomies**
- A taxonomy is a hierarchical framework, or schema, for the organization of organisms, inanimate objects, events, and/or concepts. We see taxonomies daily as humans, and we don't give them much thought. Taxonomies are the facets, filters, and search suggestions commonly seen on modern websites.
- For example, books can be categorized as fiction and nonfiction at a high level. That may work in some instances, but in most cases, that is too high of a grouping level, so we further subdivide each high-level category until we are satisfied we have achieved an appropriate grouping level. Figure 1 shows an example of a taxonomy for books.
- Another taxonomy example is how you sort your documents on your computer. For example, some may choose to start with a subject and then sub-divide by year, while others may do the opposite.
- There are no absolute right and wrong with taxonomies, just degrees of appropriateness. The most important question to ask when creating a taxonomy is, “does this hierarchical grouping meet my needs?”



Taxonomy of nature

- **Taxonomies** A taxonomy is a representation of the formal structure of classes or types of objects within a domain.
- Taxonomies are generally hierarchical and provide names for each class in the domain.
- They may also capture the membership properties of each object in relation to the other objects. The rules of a specific taxonomy are used to classify or categorize any object in the domain, so they must be complete, consistent, and unambiguous.
- This rigor in specification should ensure that any newly discovered object must fit into one, and only one, category or object class.
- The concept of using a taxonomy to organize knowledge in science is well established. In fact, if you've ever started a guessing game by asking "is it animal, mineral, or vegetable?" you were using Linnaeus' 1737 taxonomy of nature.
- Linnaeus called those three categories the "kingdoms" in his taxonomy (Figure). Everything in nature had to fall into one of those categories, which he further divided into class, order, genus, species, and variety.
- At any level in a taxonomy, there can be no common elements between classes. If there are, a new, common higher-level category is required.

- Members of any class in the taxonomy inherit all the properties of their ancestor classes.
- For example, if you know that humans are mammals, you know that they are endothermic (warm-blooded) vertebrates with body hair and produce milk to feed their offspring.
- Of course, you also know that humans breathe, but you know that because everything in the class mammals belong to the phylum chordata, which are all animals, and animals respire.
- Inheritance simplifies representation in a taxonomy because common properties need be specified only at the highest level of commonality.
- In a cognitive computing system, the reference taxonomy may be represented as objects in an object-oriented programming language or in common data structures such as tables and trees.
- These taxonomies consist of rules and constructs that will not likely change over time.

- **Ontologies**
- According to Wikipedia, an [ontology](#) “encompasses a representation, formal naming, and definition of the categories, properties, and relations between the concepts, data, and entities that substantiate one, many, or all [domains of discourse](#).”
- In other words, ontologies allow us to organize the jargon of a subject area into a controlled vocabulary, thereby decreasing complexity and confusion.
- Without ontologies, you have no frame of reference, and understanding is lost.
- As Robert Engles states in his blog post [*On the role and the whatabouts of Ontology*](#), ontologies are “essential in modern architectural patterns to ensure data quality, governance, findability, interoperability, accessibility, and reusability.”
- For example, an ontology will allow one to associate the Book taxonomy with the Customer taxonomy via relationships.
- An ontology is more challenging to create than a taxonomy because it needs to capture the interrelationships between business objects/concepts by encapsulating the language and terminology of the business area you are modeling.
- A properly created ontology will expose the understanding of how the elements in the model relate to each other. Based on this understanding, one can infer intent via the relationships. A virtual assistant like Alexa uses these relationships phrases and synonyms of those phrases to define the user’s intention.

- An ontology provides more detail than a taxonomy, although the boundary between them in practice is somewhat fuzzy.
- An ontology should comprehensively capture the common understanding—vocabulary, definitions, and rules—of a community as it applies to a specific domain.
- The process of developing an ontology often reveals inconsistent assumptions, beliefs, and practices within the community.
- It is important in general that consensus is reached, or at least that areas of disagreement in emerging fields be surfaced for discussion.
- In many fields, professional associations codify their knowledge to facilitate communications and common understanding.
- These documents may be used as the basis of an ontology for a cognitive computing system.

- Other Methods of Knowledge Representation In addition to ontologies, there are other approaches to knowledge representation.
- Two examples are described in the following sections. Simple Trees
- A simple tree is a logical data structure that captures parent-child relationships. In a model where the relationships are rigid and formalized, a simple tree, implemented as a table with (element, parent) fields in every row, is an efficient way to represent knowledge. Simple trees are used frequently in data analytic tools and in catalogs.
- For example, a retailer's catalog may have 30 or 40 categories of products that it offers.
- Each category would have a series of elements that are members of that category.
- The Semantic Web Some members of the World Wide Web Consortium (W3C) are attempting to evolve the current web into a “semantic web” as described by Tim Berners-Lee, et al in Scientific American in 2001. In a semantic web, everything would be machine usable because all data would have semantic attributes, as described in the W3C's Resource Description Framework (RDF).The current web is basically a collection of structures or documents addressed by uniform resource locators (URLs). When you find something at an address, there is no required uniformity about how it is represented.
- By adding semantics, you would force structure into everything on the web. If we did have a semantic web, we could use more of what is on the web for a cognitive system without extensive preprocessing to uncover structural information

- **Applying Advanced Analytics to Cognitive Computing**

- Advanced analytics refers to a collection of techniques and algorithms for identifying patterns in large, complex, or high-velocity data sets with varying degrees of structure.
- It includes sophisticated statistical models, predictive analytics, machine learning, neural networks, text analytics, and other advanced data mining techniques.
- Some of the specific statistical techniques used in advanced analytics include decision tree analysis, linear and logistic regression analysis, social network analysis, and time series analysis.
- These analytical processes help discover patterns and anomalies in large volumes of data that can anticipate and predict business outcomes.
- Accordingly, advanced analytics is a critical element in creating long-term success with a cognitive system that can ask for the right answers to complex questions and predict outcomes.
- It explores the technologies behind advanced analytics and how they can be leveraged in a knowledge-driven cognitive environment.
- With the right level of advanced analytics, you can gain deeper insights and predict outcomes in a more accurate and insightful manner.

- Cognitive Analytics which will exploit the massive advances in High Performance Computing by combining advanced Artificial Intelligence and Machine Learning techniques with data analytics approaches.
- They add, “Cognitive Analytics applies human-like intelligence to certain tasks, and brings together a number of intelligent technologies, including semantics, artificial intelligence algorithms, deep learning and machine learning.
- Applying such techniques, a cognitive application can get smarter and more effective over time by learning from its interactions with data and with humans.”

- advantages that can be obtained from leveraging cognitive analytics. They include:
 - **Improved Customer Interaction.** Banik notes, “There are three areas where cognitive computing is useful for consumer interaction: enhanced client services; providing a tailored service; and guaranteeing a speedier response to consumer needs.”
 - **Enhanced Productivity.** Menon notes there are four areas of productivity where cognitive analytics can be beneficial: improved decision-making and better planning; significant cost reductions; better learning experiences; and better security and governance.
 - **Business Growth.** Banik writes, “Cognitive analytics promotes corporate success by: increasing sales in new markets [and helping] launch new goods and services.”

Table : Analytics Maturity Levels

ANALYTICS TYPE	DESCRIPTION	EXAMPLES OF QUESTIONS ANSWERED
Descriptive Analytics	Understand what happens when using analytic techniques on historical and current data.	Which product styles are selling better this quarter as compared to last quarter? Which regions are exhibiting the highest/lowest growth? What factors are impacting growth in different regions?
Predictive Analytics	Understand what might happen when using statistical predictive modeling capabilities, including data mining and machine learning. Predictive models use historical and current/real-time data to predict future outcomes. Models look for trends, clusters of behavior, and events. Models identify outliers.	What are the predictions for next quarter's sales by product and region? How does this impact raw material purchases, inventory management, and human resource management?

Prescriptive Analytics	<p>Use to create a framework for making a decision about what to do or not do in the future. The “predictive” element should be addressed in prescriptive analytics to help identify the relative consequences of your actions. Use an iterative process so that your model can learn from the relationship between actions and outcomes.</p>	<p>What is the best mix of products for each region? How will customers in each region react to advertising promotions and offers? What type of offer should be made to each customer to build loyalty and increase sales?</p>
Machine Learning and Cognitive Computing	<p>Collaboration between humans and machines to solve complex problems. Assimilate and analyze multiple sources of information to predict outcomes. Need depends on the problems you are trying to solve. Improve effectiveness of problem solving and reduce errors in predicting outcomes.</p>	<p>How secure is the city environment? Are there any alerts from the vast amount of information streaming from monitoring devices (video, audio, and sensing devices for smoke or poisonous gases)? Which combination of drugs will provide the best outcome for this cancer patient based on the specific characteristics of the tumor and genetic sequencing?</p>

- The next frontier, which comes with opportunities for enormous change, includes big data analytics and incorporates the technologies of machine learning and cognitive computing.
- As shown in Figure, there is a convergence of technologies cutting across analytics and artificial intelligence.
- One major push for this convergence is the change in the timing and immediacy of data.
- Today's applications often require planning and operational changes at a fast rate for businesses to remain competitive.
- Waiting 24 hours or longer for results of a predictive model is no longer acceptable.
- For example, a customer relationship management application may require an iterative analytics process that incorporates current information from customer interactions and provides outcomes to support split-second decision making, ensuring that customers are satisfied.
- In addition, data sources are more complex and diverse.
- Therefore, analytic models need to incorporate large data sets including structured, unstructured, and streaming data to improve predictive capabilities. The multitude of data sources that companies need to evaluate to improve model accuracy includes operational databases, social media, customer relationship systems, web logs, sensors, and videos.

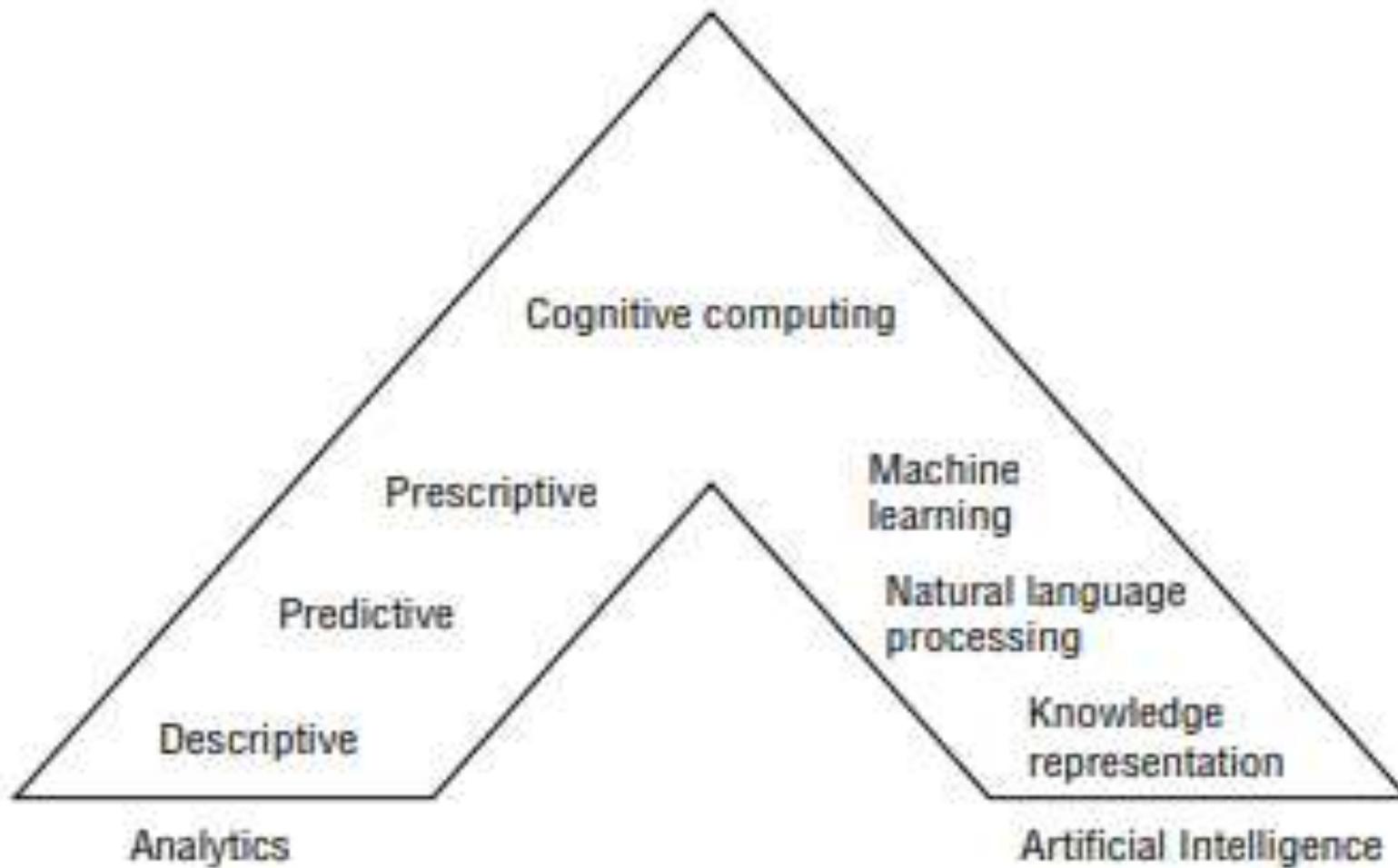


Figure : Converging technologies: analytics and artificial intelligence

- Increasingly, advanced analytics is deployed in high-risk situations such as patient health management, machine performance, and threat and theft management.
- In these use cases the ability to predict outcomes with a high degree of accuracy can mean lives are saved and major crises are averted.
- In addition, advanced analytics and machine learning are used in situations in which the large volume and fast speed of data that must be processed demands automation to provide a competitive advantage.
- Typically, human decision makers use the results of predictive models to support their decision-making capabilities and help them to take the right action

- There are situations, however, in which pattern recognition and analytic processes lead to action without any human intervention.
- For example, investment banks and institutional traders use electronic platforms for automated or algorithmic trading of stocks.
- Statistical algorithms are created to execute orders for trades based on pre-established policies without humans stepping in to approve or manage the trades.
- Automated trading platforms use machine learning algorithms that combine historical data and current data that may have an impact on the price of the stock.
- For example, a trading algorithm may be designed to automatically adjust based on social media news feeds. This approach can provide rapid insight as large volumes of current data are processed at incredibly fast speeds.
- Although taking action based on this early (and unverified) information may improve trading performance, the lack of human interaction can also lead to erroneous actions.
- For example, automated trading algorithms have responded to fake or misleading social media news feeds leading to a rapid fall in the stock market. A human would have hopefully taken the time to check the facts.

- The following two examples illustrate how companies are using machine learning and analytics to improve predictive capabilities and optimize business results.
- ■ Analytics and machine learning predict trending customer issues.
- The speed of social media can accelerate a small customer issue and grow it into a major complication before a company has time to react. Some companies decrease the time it takes to react to customer concerns by leveraging SaaS offerings that use machine learning to look for trends in social media conversations.
- The software compares the social media data to historical patterns and then continuously updates the results based on how the predicted pattern compares to actual results. This form of machine learning provides at least 72 hours of advanced warning on trending issues before they are picked up by mainstream media.
- As a result, marketing and public relations teams can take early action to protect the company's brand and mitigate customer concerns. Media buyers use the service to quickly identify changing customer purchasing trends, so they can determine where they should place their ads in mobile applications and web environments

- Analytics and machine learning speed the analysis of performance to improve service level agreements (SLA).
- Many companies find it hard to monitor IT performance at fast enough intervals to identify and fix small problems before they escalate and negatively impact SLAs.
- Using machine learning algorithms, companies can identify patterns of IT behavior and orchestrate its systems and operational processes to become more prescriptive.
- These systems can learn to adapt to changing customer expectations and requirements. For example, telecommunications companies need to anticipate and prevent network slowdowns or outages so that they can keep the network operating at the speeds required by their customers.
- However, it can be nearly impossible to identify and correct for interruptions in bandwidth if network monitoring is not done at a sufficiently granular level.
- For example, with a new machine learning solution offered by Hitachi, telecoms can analyze large streams of data in real time.
- Hitachi's customers can combine analysis of historical data and real-time analysis of social media data to identify patterns in the data and make corrections in the performance of the network.
- There are many situations in which this would be helpful to customers. For example, if a streaming video application shows a popular sporting event and the game goes into overtime, an adaptive system could automatically add an additional 15 minutes of bandwidth support, so end users are provided with consistent high-quality service. Machine learning can help the system adapt to a variety of changes and unusual occurrences to maintain quality of performance

- **Key Capabilities in Advanced Analytics**
- You can't develop a cognitive system without using some combination of predictive analytics, text analytics, or machine learning.
- It is through the application of components of advanced analytics that data scientists can identify and understand the meaning of patterns and anomalies in massive amounts of structured and unstructured data.
- These patterns are used to develop the models and algorithms that help determine the right course of action for decision makers. The analytics process helps you understand the relationships that exist among data elements and the context of the data.
- Machine learning is applied to improve the accuracy of the models and make better predictions. It is an essential technology for advanced analytics, particularly because of the need to analyze big data sources that are primarily unstructured in nature.

- **The Relationship Between Statistics, Data Mining, and Machine Learning**
- Statistics, data mining, and machine learning are all included in advanced analytics. Each of these disciplines has a role in understanding data, describing the characteristics of a data set, finding relationships and patterns in that data, building a model, and making predictions.
- There is a great deal of overlap in how the various techniques and tools are applied to solving business problems. Many of the widely used data mining and machine learning algorithms are rooted in classical statistical analysis.
- The following highlights how these capabilities relate to each other.
- **Statistics**
- It is the science of learning from data. Classical or conventional statistics is inferential in nature, meaning it is used to reach conclusions about the data (various parameters).
- Although statistical modeling can be used for making predictions, the focus is primarily on making inferences and understanding the characteristics of the variables.
- The practice of statistics requires that you test your theory or hypothesis by looking at the errors around the data structure. You test the model assumptions to understand what may have led to the errors with techniques such as normality, independence, and constant variances.
- The goal is to have constant variances around your model. In addition, statistics requires you to do estimation using confidence values and significance testing—test a null hypothesis and determine the significance of the results, called p-values.

- **Data mining**, which is based on the principles of statistics, is the process of exploring and analyzing large amounts of data to discover patterns in that data.
- Algorithms are used to find relationships and patterns in the data and then this information about the patterns is used to make forecasts and predictions.
- Data mining is used to solve a range of business problems such as fraud detection, market basket analysis, and customer churn analysis.
- Traditionally, organizations have used data mining tools on large volumes of structured data such as customer relationship management databases or aircraft parts inventories.
- Some analytics vendors provide software solutions that enable data mining of a combination of structured and unstructured data.
- Generally, the goal of data mining is to extract data from a larger data set for the purposes of classification or prediction.
- In classification, the idea is to sort data into groups. For example, a marketer might be interested in the characteristics of people who responded to a promotional offer versus those who didn't respond to the promotion.
- In this example, data mining would be used to extract the data according to the two different classes and analyze the characteristics of each class. A marketer might be interested in predicting those who will respond to a promotion. Data mining tools are intended to support the human decision-making process

- **Using Machine Learning in the Analytics Process**
- Machine learning is essential to improving the accuracy of predictive models in a cognitive environment. . One of the key differences in machine learning as compared to other mathematical approaches is the focus on using iterative methods to reduce the errors. Machine learning provides a way for systems to learn and thereby improve both the models and the results of those models. It is an automated approach that provides new ways of searching for data and enables many iterations of a model to occur, quickly improving accuracy
- These predictive models have a large number of attributes across many observations.
- The data sets are likely to be unstructured, massive in size, and subject to frequent change. Machine learning enables the models to learn from the data and enhance the knowledge base for a cognitive system.
- Hundreds or thousands of iterations of a model take place very quickly, leading to an improvement in the types of associations that are made between data elements.
- Due to their complexity and size, these patterns and associations could have easily been overlooked by human observation. In addition, complex algorithms can be automatically adjusted based on rapid changes in variables such as sensor data, time, weather data, and customer sentiment metrics. The improvements in accuracy are a result of the training process and automation. Machine learning refines the models and algorithms by continuously processing new data in real time and training the system to adapt to changing patterns and associations in the data

- Increasingly companies are incorporating machine learning to understand the context of various predictive attributes and how these variables relate to each other. This improved understanding of context leads to greater accuracy in the predictions.
- Companies are applying machine-learning technology to improve predictive analytics processes that have been in place for many years. For example, the telecommunications industry has used analytics to analyze historical customer information such as demographics, usage, trouble tickets, and products purchased to help predict and reduce churn.
- Over time the industry progressed from a focus on data mining of structured customer information to include text analytics of historical unstructured information such as comments on customer surveys and notes from call center interactions. Currently, an advanced analytic approach followed by some telecoms brings the unstructured and structured information together to develop a more complete profile of an individual customer. In addition, historical information can be combined with the most current information sourced through social media applications.
- Machine learning technology is used to train systems to quickly identify those customers that the company is most at risk of losing and develop a strategy to improve retention. Machine learning is applied in many industries including healthcare, robotics, telecommunications, retail, and manufacturing.

- Supervised and unsupervised machine learning algorithms are used in a variety of analytics applications.
- The machine learning algorithm chosen depends on the type of problem being solved and the type and volume of the data required to solve the problem.
- Typically, supervised learning techniques use labeled data to train a model, whereas unsupervised learning uses unlabeled data in the training process.
- Machine learning models that are trained on labeled data can then use this training to predict an accurate label for unlabeled data.
- Labeled data refers to the identification or tag that provides some information about the data. For example, unstructured data such as voice recordings could be “labeled” or “tagged” with a name of the speaker or some information about the topic on the recording.
- Humans often provide the labels to the data as part of the training.
- Unlabeled data does not include tags, other identifiers, or metadata. For example, unstructured data such as videos, social media data, voice recordings, or digital images would be considered unlabeled if it exists in its raw form without any preconceived human judgments about the data.

- **Supervised Learning**
- Supervised learning typically begins with an established set of data and a certain understanding of how that data is classified. Humans are involved to provide the training, and the analytical model is fit with data that is tagged or labeled.
- The algorithms are trained using preprocessed examples and then the performance of the algorithms is evaluated with test data.
- Occasionally, patterns that are identified in a subset of the data can't be detected in the larger population of data. If you fit the model to patterns that exist only in the training subset, then you create a problem called overfitting.
- To protect against overfitting, testing needs to be done against both labeled data and unlabeled data.
- Using unlabeled data for the test set can help to evaluate the accuracy of the model in predicting outcomes and results. Some applications of supervised learning include speech recognition, risk analysis, fraud detection, and recommendation systems. The following tools and techniques are often used to implement supervised learning algorithms.

- **Regression**—Regression models were developed in the statistical community. LASSO regression, Logistic regression, and Ridge regression can be used in machine learning.
- LASSO is a type of linear regression that minimizes the sum of squared errors.
- Logistic regression is a variant of standard regression but extends the concept to deal with classification. It measures the relationship between a categorical-dependent variable and one or many independent variables.
- Ridge regression is a technique used to analyze data with highly correlated independent variables (collinearity).
- Ridge regression introduces some bias to the estimates in order to reduce the standard errors or misleading variances that result from collinearity with least squares estimates.

■ Decision tree—

- A decision tree is a representation or data structure that captures the relationships among a set of categories.
 - Leaf or end nodes represent the categories, while all other nodes represent “decisions” or questions that refine the search through the tree.
 - A variety of machine learning algorithms are based on traversing decision trees. For example, gradient boosting and random forest algorithms assume that the underlying data is stored as a decision tree.
 - Gradient boosting is a technique for regression problems, which produces a prediction model in the form of an ensemble.
 - Random forest is an algorithm that organizes data using classification and regression trees to look for outliers, anomalies, and patterns in the data.
- The algorithm builds a model by initially selecting predictors at random and then continuously repeats the process to build hundreds of trees.
- Random forest is a bagging tool (ensembles of regression trees fit to bootstrap samples) that leads to more accurate models by relying on an iterative approach to many alternative analyses.
- After the trees are grown, it is possible to identify clusters or segments in data and rank the importance of variables used in the model. Leo Breiman, Statistics Department at the University of California, Berkeley, created this algorithm and described it in a paper published in 2001. Random forest algorithms are used extensively in risk analytics.

- **Neural networks—**
- Neural network algorithms are designed to emulate human/animal brains. The network consists of input nodes, hidden layers, and output nodes. Each of the units is assigned a weight. Using an iterative approach, the algorithm continuously adjusts the weights until it reaches a specific stopping point.
- Errors identified in training data output are used to make adjustments to the algorithm and improve the accuracy of the analytic model.
- Neural networks are used in speech recognition, object recognition, image retrieval, and fraud detection.
- Neural networks can be used in recommendation systems like the Amazon.com system that makes selections for the buyer based on previous purchases and searches. Deep neural networks can be used to build models on unlabeled data.
- The Google Brain project used 16,000 processors in a neural network to discover patterns in images that converged as it learned to recognize cats, without a predefined template for cat images. Microsoft's Project Adam can identify dog breeds from photographs, using asynchronous neural networks.
- Neural networks can be challenging to use in environments that require an audit trail or traceability, such as financial services trading systems. Because these systems are designed to learn autonomously, it could be cost-prohibitive to track changes made in a way that would satisfy the requirements of an outside examiner

- **Support Vector Machine (SVM)**—
- SVM is a machine learning algorithm that works with labeled training data and output results to an optimal hyperplane.
- A hyperplane is a subspace of the dimension minus one (that is, a line in a plane).
- SVM is usually used when there are a small number of input features.
- The features are expanded into higher dimension space.
- SVM is not scalable to billions of elements of training data.
- An alternative algorithm for situations with extremely large volumes of training data would be logistic regression.
- **k-Nearest Neighbor**

(k-NN)—k-NN is a supervised classification technique that identifies groups of similar records.

The k-Nearest Neighbor technique calculates the distances between the record and points in the historical (training) data.

It then assigns this record to the class of its nearest neighbor in a data set. k-NN is often selected when there is limited knowledge about the distribution of the data

- **Unsupervised Learning**
- Unsupervised learning algorithms can solve problems that require large volumes of unlabeled data. As in supervised learning, these algorithms look for patterns in the data that enable an analytics process.
- For example, in social analytics, you may need to look at large volumes of Twitter messages (tweets), Instagram photos, and Facebook messages to collect adequate information and develop insight into the problem you want to solve.
- This data is not tagged; and given the large volume, it would take too much time and other resources to attempt to tag all this unstructured data.
- As a result, unsupervised learning algorithms would be the most likely choice for social media analytics. Unsupervised learning means that the computer learns based on an iterative process of analyzing data without human intervention.
- Unsupervised learning algorithms segment data into groups of examples (clusters) or groups of features. The unlabeled data creates the parameter values and classification of the data.

- Unsupervised learning can determine the outcome or solution to a problem, or it can be used as the first step that is then passed on to a supervised learning process.
- The following tools and techniques are typically used in unsupervised learning.
 - ■ Clustering techniques are used to find clusters that exist in the data sample. Clustering categorizes the variables into groups based on certain criteria (all the variables with X or all the variables without X).
 - The K-means algorithm can estimate the unknown means based on the data. This is probably the most widely used unsupervised learning algorithm. It is a simple local optimization algorithm.
 - ■ The EM-Algorithm for clustering can maximize the mixture density given the data.

- ■ **Kernel density estimation (KDE)** estimates the probability distribution or the density of a data set. It measures the relationship between random variables.
- KDE can smooth the data when inferences are made from a finite data sample.
- KDE is used in analytics for risk management and financial modeling.
- ■ **Nonnegative matrix factorization (NMF)** is useful in pattern recognition and to solve challenging machine learning problems in fields such as gene expression analysis and social network analysis. NMF factorizes a non-negative matrix into two non-negative matrixes of lower rank, and can be used as a clustering or classification tool. Used in one way, it is similar to K-means clustering. With another variation, NMF is similar to probabilistic latent semantic indexing—an unsupervised machine- learning approach for text analytics.
- ■ **Principal Components Analysis (PCA)** is used for visualization and feature selection. PCA defines a linear projection where each of the projected dimensions is a linear combination of the original.

- **Singular Value Decomposition (SVD)** can help to eliminate redundant data to improve the speed and overall performance of the algorithm.
- SVD can help decide which variables are most important and which ones can be eliminated.
- For example, assume you have two variables that are highly correlated, such as “humidity index and probability of rain” and, therefore, do not add value to the model when used together. SVD can be used to determine which variable should be kept in the model. SVD is often used in recommendation engines.
- **Self Organizing Map (SOM)** is an unsupervised neural network model that was developed in 1982 by Tuevo Kohonen.
- SOM is a pattern recognition process. The patterns are learned without any external influences. It is an abstract mathematical model of topographic mapping from the (visual) sensors to the cerebral cortex.
- It is used to understand how the brain recognizes and processes patterns. This understanding of how the brain works has been applied to machine learning pattern recognition. These techniques have been applied to manufacturing processes.

- **Predictive Analytics**

- Predictive analytics is a statistical or data mining solution consisting of algorithms and techniques that can predict future outcomes.
- Data mining, text mining, and machine learning can find hidden patterns, clusters, and outliers in both structured and unstructured data. These patterns form the basis of the answers and predictions made with a cognitive system.
- Predictive modeling uses the independent variables that were identified through data mining and other techniques to determine what is likely to occur under various future circumstances. Organizations use predictive analytics in many ways, including prediction, optimization, forecasting, and simulation.
- Predictive analytics can be applied to structured, unstructured, and semi-structured data. In predictive analytics, the algorithm you use applies some sort of objective function.
- For example, Amazon.com uses an algorithm that learns about your buying behavior and makes predictions regarding your interest in making additional purchases. The focus on analyzing unstructured data represents a change for the use of predictive analytics. Traditionally, statistics and data mining technology have been applied to large databases of structured data. The internal operational systems of record at an organization are typically stored as structured data.
- However, it is the wide range of data types in unstructured data that represents the majority of data required to form a knowledge base for a cognitive system. These unstructured data sources include e-mails, log files, customer call center notes, social media, web content, video, and literature.
- Until recently, it has been more difficult for companies to extract, explore, and leverage unstructured data for decision making.
- Technology advancements such as Hadoop have improved the speed and performance of statistical analysis of unstructured data. The ability to analyze these unstructured data sources is key to the development of cognitive systems

Table 6-2: Predictive Analytics Use Cases

USE CASE	EXAMPLE	HOW PREDICTIVE ANALYTICS CHANGES OUTCOMES
Predicting consumer behavior	A manufacturer can identify patterns in consumer preferences that it could not recognize using traditional analysis of the data. Use of predictive analytics has improved supply chain management and the ability to react to consumer demand. This manufacturer can now predict customer orders 4 months in advance with an accuracy rate of approximately 98 percent.	The company deployed a real-time data warehouse to ensure that multiple sources of data could be well integrated and available at the right time for analytics. The company is building more accurate models using timely data and diverse data types. The models are designed to identify hidden patterns and create accurate forecasts.
Sales and inventory forecasting	A large multistore retailer uses advanced analytics to develop models at a faster pace than in the past using larger volumes of data. This company benefited by improving the accuracy of its sales forecasting models and reducing inventories. The company achieved 82 percent accuracy in its forecasting, a major improvement compared to traditional approaches.	This retailer implemented an analytics platform that standardizes and automates a portion of the predictive analytics process. Using this platform, the company can build 500 predictive models per month as compared to 1 model using traditional methods. The increased granularity in its models is yielding greater accuracy.

Predicting failures in machinery	A medical equipment manufacturer embeds sensors in its equipment to monitor performance. The recorded data is constantly streamed and analyzed to predict potential failures with enough lead time to make adjustments and avoid harm to patients.	Advanced analytics is used to build sophisticated algorithms that can uncover hidden patterns of failure and monitor sensitive equipment more accurately than traditional methods. The volume of data that needs to be analyzed is large and streaming.
Predicting and reducing fraud	An insurance company used advanced analytics to transform its approach to claims processing and improve fraud detection. The company improved its success rate in pursuing fraudulent claims from 50 percent to 90 percent and saved millions of dollars.	Predictive analytics is used to look at the whole claims process differently. Patterns of fraud are analyzed and used to rate the likelihood that each new claim may be fraudulent. Text mining is incorporated into the system to gain insight from analyzing the content of police reports and medical records.

- **Text Analytics** Given the business value of text-based unstructured sources, text analytics is a critical element of cognitive systems.
- Text analytics is the process of categorizing unstructured text, extracting relevant information, transforming it into structured information, and analyzing it in various ways.
- The analysis and extraction processes used in text analytics leverage techniques that originate in computational linguistics, natural language processing, statistics, and other computer science disciplines.
- The text can be extracted and transformed, and then analyzed iteratively to identify patterns or clusters and determine relationships and trends.
- In addition, the transformed information from text analytics can be combined with structured data and analyzed using various business intelligence or predictive and automated discovery techniques.

- The need for businesses to make decisions based on real-time information makes text analytics an increasingly important capability.
- For example, a telecom provider wanting to understand which customers might be most likely to switch to the competition unless they receive the right incentive needs real-time data on customer sentiment.
- The accuracy of predictive models that rely on customer sentiment data requires rapid analysis of large volumes of unstructured data.
- Sentiment scoring and natural language processing engines can build more accurate models and improve the speed of analysis.
- Machine learning can improve the models' capability to react to sentiment data from social media as it is fed back into the model.
- Text analytics is widely used to help organizations increase customer satisfaction, build customer loyalty, and predict changes in customer behavior.
- Text analytics can also improve search capabilities in areas such as faceted navigation

Table : Text Analytics Use Cases

Marketing	Churn analysis, voice of the customer, sentiment analysis, customer survey analysis, social media analysis, market research
Operations	Voice of the employee, document categorization, competitive intelligence
Legal/Risk and Compliance	Document categorization, risk analysis, fraud detection, warranty analysis, e-discovery

- **Image Analytics**
- The sources used to develop knowledge corpora for a cognitive system are likely to include videos, photos, or medical images.
- There has been an enormous increase in the volume of images created and managed by governments, organizations, and individuals.
- As a result, image analytics capabilities are important in cognitive computing. The ability to quickly identify clusters and patterns in these images can have a major impact on IT and physical security, healthcare, transportation logistics, and many other areas.
- For example, facial recognition technology is used to both verify and identify individuals as a means to help prevent fraud and solve crimes.
- Governments use facial and image analytics to anticipate and prevent terrorist activities. Facial recognition can be used in video indexing to label faces in the video and identify the speakers in the video.
- Although facial recognition is a significant part of image analytics, cognitive systems will demand the capability to identify content in many different types of images.
- Image analytics can index and search video events by classifying objects into different categories such as people, animals, and cars, or to look for anomalies in a medical digital image such as an X-ray or CT scan.
- Facial recognition was one of the earliest research areas in the field of image analytics. For example, Facebook's research into facial recognition has resulted in software called DeepFace that is based on an advanced machine learning neural network.

- **Speech Analytics**
- Text, image, and speech analytics can be used in a cognitive system to provide the right context to answer a question correctly or make an accurate prediction.
- Although text analytics is used to gain insight into sentiment, many emotions and attitudes can be easily masked in text. Images and speech can provide many more clues to a person's emotions and anticipated actions.
- Speech analytics is the process of analyzing recorded speech to extract information about the person speaking or the content of his words.
- Identifying the patterns of words and phrases that are good indicators of emotion and intent to act in a certain way can lead to improved accuracy of predictive models.
- Speech analytics has been applied to call center processes for many years. There was significant research in the field of automatic speech recognition (ASR) as early as the 1950s.
- ASR systems needed to provide accurate information for people with vastly different speech patterns and regional accents without the need for training.
- Statistical models can create speech-clustering algorithms for different word and sound reference patterns.
- Although various statistical modeling techniques were applied to solving the problem of ASR, the hidden Markov model (HMM) and the stochastic language model became the most widely used techniques in the 1980s.
- As call centers developed in the 1990s as an important way to create a more efficient and cost-effective way to properly route calls, companies such as AT&T began using automatic speech recognition technology as part of the call center process.

- **Using Advanced Analytics to Create Value**
- Ultimately, the goal of deploying advanced analytics processes and cognitive computing is to improve decision making. Companies are using analytics to differentiate from the competition by doing a better job of listening to customers, anticipating their needs, and making highly targeted offers. Government agencies use analytics to differentiate their cities by making them safer, more responsive to citizens' needs, and more ecologically sound. Healthcare organizations use analytics to improve physician training, eliminate unnecessary hospitalizations, and improve overall quality of care. Building the analytics models and cognitive computing environments that support these improvements in decision making requires more data, more accurate data, more refined data, and the ability to manage and interpret data from all input streams at fast speeds.
- The ability to make faster and more execution-oriented decisions depends on reducing the “degree of difficulty” of interpreting the right data from all input streams. You need to manage the volume and complexity of the data.
- At the same time, you need to sample high-velocity data in a meaningful way. Raw data as captured by systems and sensors has potential value, but needs to be processed and analyzed to build business value.
- Business value increases as volume, complexity, and speed are managed during the analytics process. The value to an organization comes from acting on the analysis of the data.

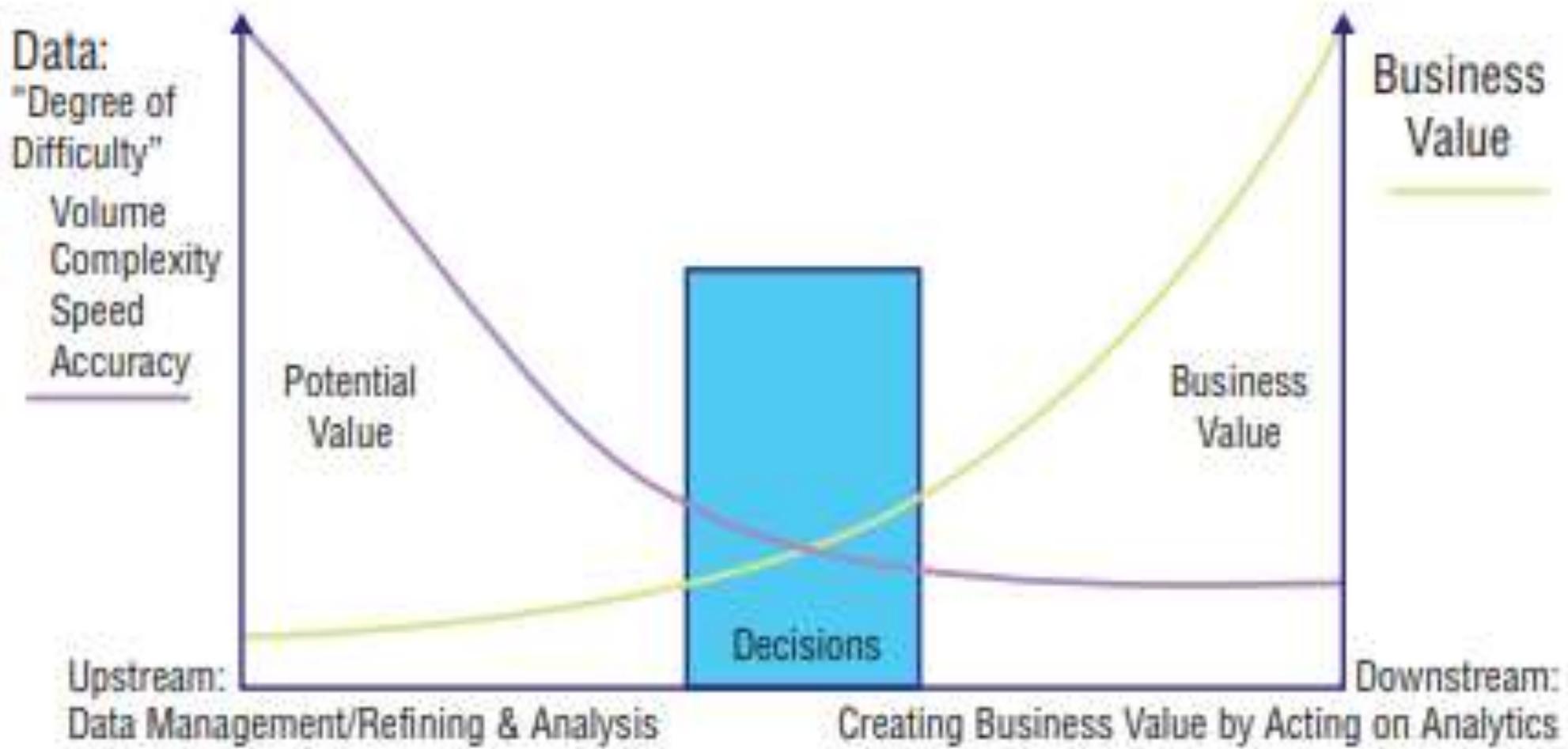


Figure : Refining raw data to create business value

- **Building Value with In-memory Capabilities**
- Managing the dimensions of speed, complexity, and volume can be best addressed by an optimization strategy that includes both software and hardware.
- To scale appropriately to support big data analysis, many companies opt for hardware that is pre-integrated and optimized to run advanced analytics workloads.
- To achieve business value, the analytical models and predictions need to be fully integrated into operational business processes.
- All applications need access to the necessary data at the right time.
- Capabilities such as columnar alignments, graph databases, and in-memory computing can help support advanced analytics.
- To enhance the speed, shared memory, shared disk, high-speed networks, and optimized storage are valuable techniques.

- Platforms that are designed for high-speed and volume analytics may rely on in-memory capabilities.
- Transaction processing, operational processing, analytics, and reporting and visualization can be integrated within one in- memory platform.
- It eliminates the time-consuming effort of data extraction and transformation.
- In-memory analytics provides a way to process large and complex analytic workloads quickly.
- It can improve application performance. These workloads can be chunked into smaller units and distributed across a parallel system. This approach is used primarily with structured data.
- In-memory analytics can help to overcome the challenges of trying to visualize and analyze big data at fast enough speeds.
- For example, with real-time streaming data, in-memory capabilities can ensure that computations are performed in RAM to increase processing times much faster than data access from disk.

- **Impact of Open Source Tools on Advanced Analytics**
- Open source analytics tools are having a major impact on the growth of predictive analytics at many organizations. The open source software environment and programming language, R, is fast becoming one of the primary tools for data scientists, statisticians, and other enterprise users.
- R, which is designed for computational statistics and data visualization, is the language of choice for graduate students doing research in advanced analytics and cognitive computing.
- Strong interest in R has led to a very active open source community. Members of the community share information on models, algorithms, and coding best practices.
- Users like the flexibility that a special-purpose programming language and environment offers for building custom applications. Some of the benefits of R include its flexibility and adaptability. R is actually an implementation of the statistical programming language S, developed at Bell Laboratories, as a higher-level alternative to using FORTRAN statistical subroutines

- Although R can be complicated to use unless you are an experienced data scientist or statistician, many vendors offer some sort of connection to R that makes it easier to use.
- Vendors are providing algorithms that are preset and ready to use in model development. The open source community has spawned and supports many projects that form the foundation for advanced analytics applications. For example, two important projects within the Apache Foundation are projects like Cassandra (distributed DBMS) and Spark (an analytics framework for cluster computing in the Hadoop space).

- **the seven key steps involved in designing a typical cognitive application:**
- 1. Defining the objective
- 2. Defining the domain
- 3. Understanding the intended users and their attributes
- 4. Defining questions and exploring insights
- 5. Acquiring the relevant data sources
- 6. Creating and refining the corpora
- 7. Training and testing

- **Defining the Objective**
- Creating a cognitive application has much in common with developing any other enterprise application. You need to understand what the objectives are for your application and how you will achieve those objectives.
- Therefore, the first step in developing a cognitive application is to understand the types of problems your cognitive application is going to solve. Your objective needs to consider the types of users you will be appealing to and if there will be multiple constituencies in your user base.
- What issues will your users be interested in, and what do they need to know? One of the unique differences between a cognitive application and traditional applications is that users should expect more than answers to queries.
- A cognitive application should provide answers to questions but also go deeper and explore context related to how and why something happened.
- Building traditional applications often begins with business process.
- In contrast, in a cognitive application you need to develop an objective based on knowledge and data.
- Therefore, in the design you need to set some parameters around the type of knowledge that is pivotal to your corpora. In other words, your objective should probably focus on a specific segment of an industry rather than attempt

- Several examples of objectives for cognitive healthcare applications follow:
 - ■ Provide personalized information and social support to help individuals optimize their health and wellness.
 - ■ Help health consumers take a more active role in managing their own health and the health of people they help care for.
 - ■ Help determine if the treatment plan selected for the patient is the best and most cost-effective.
 - ■ Provide additional knowledge to medical students to support what they learn on their subspecialty rotations.

- Cognitive applications are also a good solution in situations in which you would like to provide assistance to a customer service representative or salesperson.
- Consider a retail organization with a large and diverse group of sellers. This company has a small number of sellers with many years of experience and deep knowledge of the products the company sells.
- If a customer has specific requests or needs help comparing alternatives, these knowledgeable sellers can answer all their questions and make the sale.
- However, the company also has high turnover, and many sellers lack the knowledge to provide the right level of support to customers.
- The company decides to introduce a cognitive application designed to make all the sellers as smart as the most knowledgeable and experienced seller in the company.

- **Defining the Domain**
- The next step is to specify the domain or subject area for your cognitive application. Defining your domain is a prerequisite to identify and assess which data sources you will need for your applications.
- In addition, your domain definition will be useful in determining the subject matter experts that will be helpful in training the system. Objective is likely to help narrow the domain focus.
- For example, a cognitive application designed to train medical students would require medicine as the domain, whereas a cognitive application designed to help clinicians select the right treatment plan for breast cancer patients would require breast oncology as the domain. The medical domain will require comprehensive and broad-based medical taxonomies, ontologies, and catalogues, whereas the breast oncology domain will require segments of the medical ontology as well as additional data specific to the field.

Table Examples of Cognitive Application Domains

DOMAIN	SELECTED DATA REQUIREMENTS	SUBJECT MATTER EXPERTS
Medicine	International Classification of Diseases (ICD) codes, Electronic Medical Records (EMR), and research journals	Senior physicians in medicine and key specialties
Airplane Manufacturing and Maintenance	Complete parts list, maintenance records per airplane, and spare parts inventory	Mechanics who know how to anticipate failure and create effective repairs and experienced pilots
Retail	Customer and product data	Experienced sales associates

- Although the domain helps to define the data sources you might need, you may also include data sources that are not typically associated with solving problems in that domain.
- The inclusion of non-typical data sources is needed because cognitive systems support problem solving in a different way from traditional systems.
- A cognitive application is good at helping users to assimilate knowledge quickly and efficiently.
- Although some of this knowledge might be found in specific data sources, it may also incorporate information that is typically learned by experience.
- One of the advantages of a cognitive application is that it can provide every user with proven business practices and industry-specific knowledge that is well known to your most experienced domain experts.
- A cognitive system's greatest value comes from its ability to combine information from industry data sources with testing and refinement based on interactions with highly experienced experts.
- For example, when faced with an unusual problem, an airplane mechanic with 30 years of experience might remember similar situations that occurred in the past and recommend something like, “the problem is likely to be A or B and we should take these five steps to get the best result.”

- **Understanding the Intended Users and Defining their Attributes**
- You need to understand the types of users who will be accessing your cognitive application.
- Expectations for user and system interactions will have an impact on the development of the corpus, the design of the user interface, and how the system is trained.
- The level of accuracy required in a cognitive application will depend on the intended use case.
- For example, a scientist requires a much more precise level of accuracy than a customer service representative answering questions about replacement parts.
- However, it is unnecessary and unwise to attempt to anticipate all the questions your users will ask and all the different ways your cognitive application will be used.
- A cognitive application assumes that the data will grow and change as new data sources are discovered and added.
- In addition, the machine learning algorithms will refine the way questions are analyzed and answered. You need to build in flexibility so that your application can change as user requirements change.
- The learning process for a cognitive system is continuous, and as a result, your application will get smarter and deliver greater value to end users the more it is used.

- The following best practices can help to ensure that your cognitive application has the flexibility it needs to provide the right level of support for its users:
- ■ **Understand your users' level of understanding of the domain.** Is your cognitive application intended for consumers or domain experts? Will your users understand the meaning of industry-specific terms? Will your cognitive application be used to help train users in a particular domain?
- ■ **Plan for variations in types of questions and analysis required.** Will your application have users with varied backgrounds and levels of expertise? For example, if you are planning for consumers and domain experts, they are likely to ask questions using different words and language styles. Although these users may be looking for insights into a similar topic, they may have widely differing expectations for the level of insight required. The consumer may look for a definition, whereas the domain expert wants to compare alternative solutions to a complex problem.
- ■ **Keep the scope of your application broad enough to support different types of users.** If you are too specific or narrow in your definition of the domain, there may be subject areas that are not covered adequately in the corpus. It is better to err on the side of a little more coverage of your domain than less, as the learning process will successively refine the corpus toward the “right size” with increased usage.

- **Defining Questions and Exploring Insights** “The Foundation of Cognitive Computing,” a cognitive system delivers insight relative to a domain, a topic, a person, or an issue, based on training and observations from all varieties, volumes, and velocities of data.
- A cognitive system creates models to represent the domain and generates and scores hypotheses to answer questions or provide insight.
- To ensure your cognitive application delivers the insights your users are looking for, you need to begin by mapping out the types of questions they may ask.
- Users of a well-defined and trained cognitive application can benefit in many ways.
- One significant benefit is the capability to receive alternative answers to questions along with associated confidence levels. These benefits can be achieved only if the right set of data for the domain and the system is ingested into the corpus and then is properly trained and tested.
- However, before you even begin to train the system, you need to consider the types of questions your users will ask and the types of insight that your users will be looking for.
- Many of the early cognitive applications are of two main types: customer engagement, or discovery and exploration. Customer or user engagement applications typically leverage advanced Question-Answer Systems designed to answer questions as part of an ongoing dialogue with the user.
- Answers to questions may be provided as a set of alternatives with associated confidence levels. Discovery and exploration applications begin with data analysis rather than by asking questions. You may not know what to expect or exactly what questions to ask.
- Discovery applications are used in situations such as genomic exploration, security analysis, or threat prevention. Typically, in these situations, your cognitive application will begin by looking for patterns and anomalies in the data.

- The cognitive system can make associations between questions, answers, and content to help the user understand the subject matter at a deeper level. The questions users will ask can be placed in two general categories:
- ■ Question-Answer pairs—The answers to these questions can be found in a data source. There may be conflicting answers within the data sources, and the cognitive system will analyze the alternatives to provide multiple responses with associated confidence levels.
- ■ Anticipatory analytics—The user engages in a dialogue with the cognitive application. The user may ask some questions but not all the questions. The cognitive application will use predictive models to anticipate the user's next question or series of questions.

- **Typical Question-Answer Pairs** Developers of question-answer cognitive applications have found that they need to begin with approximately 1,000–2,000 pairs of questions and answers.
- You have already defined the users of your application, and you need to keep them in mind when creating the question-answer pairs.
- How will your representative group of users ask questions? Consider not only the content of the question, but also how it will be asked.
- The questions need to be in the voice of the end user. What style of language will they use?
- What technical terms are they likely to know?
- There are often many ways to ask the same question, and you need to consider the alternative styles of questioning when developing these initial questions.
- Although the answers need to use terms and a language style that will be understood by users, the content of the answer needs to be vetted by subject matter experts.

Table Question-Answer Pairs for Different Types of Users

QUESTION	ANSWER
Health consumer: What is a morcellator?	A morcellator is a device with a spinning blade that is used to shred a fibroid through an incision on a woman's abdomen. The force and speed of the device may cause cellular particles from the fibroid to become dispersed in the abdomen.
Gynecologist: What are the risks and benefits of using a morcellator for surgical treatment of fibroids?	Risks include potential spread of an occult uterine sarcoma. Benefits include smaller incisions for the patient, less bleeding, and quicker healing and recovery.

- You should define a sample set of questions prior to selecting the data sources needed to build the corpus.
- By choosing your information sources based on what is needed to answer a representative set of questions, your system can learn how to answer similar questions in the same domain.
- If you build the corpus first, you may make the mistake of tailoring your questions for training and testing to the information you already have at hand.
- When your cognitive application becomes operational, your users may have questions that cannot be answered by the system.
- It is expected that the corpus will need to be continuously updated during training and operation; however, you want to start out by including as many data sources as required to provide the right level of insight within your chosen domain.

- **Anticipatory Analytics**
- What if the user is not in a position to ask a specific question of the cognitive application?
- Anticipatory analytics can be used when there are many unknown factors making it difficult for a user to know what questions to ask.
- For example, in military or security analytics, you may not know when or where a future event will occur or even what event will occur. You need to observe the data and look for patterns without knowing what you are looking for.
- The data you need to observe and analyze may be unclean or subject to inconsistent definitions and inconsistencies in metrics or measurements around time and place.
- However, when this data is used for a cognitive security application, unclean data may provide valuable clues to anticipate events or actions. The anomalies or outliers in the data are used to build the models and anticipate changes that can identify security threats or military events in time to take corrective action.

- Anticipatory analytics is also used in cognitive applications that are designed to understand an individual's personal needs and help them to make good decisions.
- Because a user does not need to ask a question to be provided with a recommended action, the creators of the application need to focus on the different personal situations that might be best suited for assistance by a cognitive system.
- For example, a cognitive assistant could monitor a user's schedule and alert the user if there is a delay in a scheduled air flight or train.
- By monitoring personal medical devices and applications, a cognitive assistant could alert users they may be getting sick or help them keep on track with dietary goals.
- Users are increasingly sharing a lot of personal information on a variety of applications and devices—ranging from health monitoring devices to e-mail, travel, and calendar applications.
- A cognitive application can be trained to integrate this information to learn a lot about you.
- In addition, a cognitive application can be designed to be aware of what is happening in the world around you through geospatial, travel, health, and other applications.
- Therefore, a cognitive application that understands your location, your health and medical status, and the context of your questions can make personalized recommendations.
- An anticipatory cognitive application leverages data to make personal tasks easier and provide information you need before you ask for it

- **Acquiring the Relevant Data Sources**
- When developing a corpus you should determine the most relevant data sources. This is challenging because you cannot know with certainty what type of insights users might require as their needs change over time.
- However, taking the time to evaluate data sources you currently own and those you may want to acquire also offers great opportunity.
- You may discover that you have internal data resources that can provide new insight when leveraged by a cognitive system.
- Additionally, you may want to include social media data or other external sources.
- Cognitive systems provide an opportunity to leverage data sources in new ways. To start building the corpus, you need to understand your requirements for a variety of internal and external data sources.
- As you move through the testing process and your application becomes operational, you need to be prepared to add new sources as they become available and the scope of the application expands.

- ***The Importance of Leveraging Structured Data Sources***

- Much of the focus around cognitive computing has been data from unstructured data sources. However, cognitive solutions must gain insights based on the current state of customers or other constituents.
- Therefore, you need to know what internal data sources are going to be meaningful. For example, if the application is related to travel, the company needs internal data to relate to the details about customers or travel locations.
- A retail application needs data sources related to merchandise that has been ordered, what products have been sold, and who the customers are. A hospital-based healthcare application needs data on patient status, medical history, and hospital admissions.
- These data sources will most likely be stored as structured data in relational databases including customer data from a Customer Relationship Management system or patient data from an Electronic Medical Record for a healthcare application. Additionally, there could be streaming data sources that come from sensor networks.

- **Analyzing Dark Data**
- Dark data refers to data that has been stored over many years and sometimes decades. Much of this data has been stored but not previously analyzed.
- For example, dark data could be data about performance of a company's stock over a decade or data stored at the time of a security breach.
- With the cognitive system, the dark data can become the benchmark to analyze how things have changed over time. This data may provide new insights by using machine learning to look for patterns in data collected over many years.
- Given the advent of new analytics technologies, this dark data may now be an important internal data source depending on the domain

- **Leveraging External Data**
- What external data sources will support users? External data sources may include everything from industry-specific technical journals that are focused on new research findings to industry taxonomies and ontologies.
- In medical research there are results from clinical trials that might provide insights into drug interactions. Most industries have a wealth of third-party databases with both structured and unstructured data.
- Increasingly, there are stores of videos, images, and sounds that are of particular interest to either a specific industry or a technical discipline. Many industries have codified ontologies and taxonomies that are managed and updated by industry consortiums.
- These sources are critical in creating your corpus. However, you may find that you need to capture only a subset of the available data. These data sources often include the hierarchical classification of entities or concepts within a domain, which are important for determining context and meaning.
- You need to use caution when using these external data sources. For example, what is the origin of the data source? Who owns that data source and how and when was it created? More important, who is responsible for updating the data source on an ongoing basis? Equally important is the security and governance of the data sources.
- There are data sources that include private information that can be used under strict governance guidelines. If that data is misused, it can cause significant problems for an organization

Table Industry-Specific Taxonomies and Ontologies

INDUSTRY	TAXONOMY/ ONTOLOGY	PUBLISHER	DESCRIPTION
Healthcare	International Classification of Diseases (ICD)	World Health Organization	International codes for diseases, disease symptoms, and medical findings about diseases
Healthcare	Semantic taxonomy for the healthcare ecosystem	Developed by companies such as Healthline Corp.	Classifies healthcare information on the web and maps the relationship between consumer and clinical terminology
Construction	International Building Code (IBC)	International Conference of Building Officials	Standards and compliance regulations for international building codes
Finance	U.S. GAAP Financial Taxonomy	Financial Accounting Standards Board (FASB)	U.S.-based standards for financial accounting and reporting
Information Technology	NIST Cloud Computing Taxonomy	National Institute of Standards and Technology (NIST)	Companion to the NIST Cloud Computing Reference Architecture; goal to help communicate the offerings and components of cloud architecture

- **Creating and Refining the Corpora**
- Building a cognitive application requires extensive collaboration between the technology team and business experts.
- The initial steps in the development process include defining the objective and user expectations for the application.
- This stage requires substantial industry or domain expertise. The next series of steps in the application development process relies more heavily on the technology team.
- The actual creation of the corpus, model development, and training and testing of the system requires skills in areas such as software development, machine learning, and data mining.
- The creation of the corpus is not a one-time process. There is an initial effort to build a quality corpus (or corpora) that includes the selected data sources. However, there needs to be continuous re-evaluation of the data sources to determine if new sources need to be added or if enhancements to existing sources are required to improve outcomes from the cognitive application.
- You need to understand the life cycle for each of the data sources because many of these sources need to be updated at regular intervals. Therefore, you need to set a process in place to ensure that updates to data sources are made on a timely basis.

- **Preparing the Data**
- All data ingested into the corpus must first be validated to ensure that it is readable, searchable, and comprehensible. Structured, semistructured, and unstructured data are likely to be combined in various corpora included in a cognitive system. All data sources need to be evaluated to see if any transformations or enhancements are required prior to ingestion into a corpus.
- Are your text-based resources such as journal articles, textbooks, and research documents annotated with headings that provide queues to the cognitive system? Tagging should help the system identify and classify the content in specific articles. In addition, tagging can ensure that the cognitive system can quickly make appropriate associations between different data elements.
- The requirements to transform the structure of the data can vary based on the cognitive platform you use. The corpora of some early cognitive systems were ingested with primarily unstructured text-based content. As a result, complex structured data sources needed to be transformed into unstructured content prior to ingestion into the corpus. Initially, this transformation was time-consuming.
- However, services have been developed to help speed up the process of transforming data structures. Vendors are continuing to improve data preparation services for cognitive systems, making it possible for structured data to be automatically transformed within the system.
- These transformation and other data preparation services can have a positive impact on adoption rates for cognitive applications. Data from structured data sources such as Customer Relationship Management Systems or other database applications needs to be easily and quickly ingested into the system if business users are going to begin using the applications at a greater rate.
- It is not necessary that the complete database be ingested as is. Actually, it is quite common that only a segment of the existing data source is required to meet the requirements for the domain.

- **Ingesting the Data**
- Managing the data ingestion process efficiently is critical to the success of a cognitive application. Data ingestion is not something that happens just once during the development of the system.
- Existing data sources are subject to continuous updates and refinements to ensure they are accurate and up to date.
- The results of the training and testing of the models may indicate weak spots or limitations in the corpora that require the addition and revision of sources. In addition, changing user expectations are likely to result in new additions to the corpora. Delays in making the required updates to the corpora will decrease the effectiveness and accuracy of the system.
- Therefore, to maintain the viability of the cognitive system, data sources may need to be ingested in near real time. Typically, you will have access to a set of services that are designed to make the ingestion process fast, robust, and flexible.
- Although there may be some coding required, the ingestion services will include connectors and tools to make the process as seamless as possible. As in traditional data management efforts, you need to have controls and supports in place to maintain governance and anticipate and correct for errors.
- you must incorporate real-time traceability into the data ingestion process. If errors result in an unexpected halt in the ingestion process, you need to trace back to understand why the problem occurred and where you were in the ingestion process when it stopped.
- This is called checkpointing, and can use this information to restart the ingestion process in the right place. you may need to monitor the ingestion process to ensure that any records that are deleted or scrubbed to meet security requirements have been handled properly

- **Refining and Expanding the Corpora**
- a corpus needs to be continually refined to ensure that the cognitive application delivers accurate information and provides the right level of insight. Although you have completed extensive preparation for ingesting content needed to provide a good knowledge base for your cognitive application, it is hard to anticipate all data requirements at the outset.
- Early in the training process, you may find that the accuracy of the answer to a certain question is below your accepted threshold. By increasing the coverage (adding more data) for certain topic areas in your domain, you should improve accuracy.
- Plan for multiple iterations of this process of training, observing results, and then adding to the corpus.
- You need to establish an ongoing process of updating data requirements and adding to the corpus as you proceed through the testing process and after your application becomes operational.
- You can use expansion algorithms to determine which additional information would do the best job of filling in gaps and adding nuance to the information sources in the corpus. There will be situations in which you need to enrich data by providing lookups to additional sources that might have detailed information about customers or definitions of technical data.

- **Governance of Data**
- The corpora in your cognitive application will include a wide range of data sources. There may be personal data that is subject to the same data privacy rules that apply to data used in other systems in your organization.
- Therefore, you will need to comply with the same data privacy and security requirements of any system. There will be data that will be ingested into the corpus that might have restrictions on use based on governance requirements.
- In some situations there might be copyrighted images or content that is part of your corpus.
- Therefore, you want to make sure that you have a license for use of that content. In healthcare there are patient privacy rules that require that personal information be anonymized.
- In a retail system it will be important not to expose customers' credit card data.
- If a corpus includes social media data, you must be sure that you are not violating the privacy of users of those sites.
- For example, users might decide that they no longer want to allow access to location data. In some countries there are restrictions on where customer data can be stored.
- A cognitive system may require the highest level of governance and security because over time it will include sensitive data about competitive best practices. Therefore, in designing and operating a cognitive system, governance and security cannot be an afterthought.

- **Training and Testing**
- It is through an iterative process of model development, analysis, training, and testing that the cognitive system begins to learn.
- Deploying a scalable training and testing strategy can ensure your application works as intended when it becomes operational. You need to measure responses to determine what is the minimal level of accuracy that is acceptable.
- After this is established through the testing process, you can begin to establish the ground truth—a set of data that is the gold standard for accuracy of a model.
- It may require you to try additional data sets so that the information used for testing is objective.
- Initially, you create a ground truth that establishes what the system knows and understands.
- In Question-Answer based cognitive applications, you have a set of question-answer pairs that establish your ground truth. The questions represent the types of questions your users will ask. The answers to those questions are accurate, having been approved by domain experts.
- These question-answer pairs are developed in clusters around a topic to help with the machine learning process.
- Algorithms help the system to understand context by looking for associations and patterns in the clusters of question-answer pairs.
- Your training and testing strategy needs to compare new analysis against the ground truth and add to the base level of truth when needed to improve the accuracy of the system

- This is often an iterative process; each time the data is trained, the accuracy of the application improves.
- Cognitive systems are designed to learn from failure and improve through feedback. Your cognitive application may assign a high confidence level to answers that are obviously wrong.
- As part of the training process, you need to analyze why the system got the answer wrong. Although training the system should be an automated process, there are some aspects that involve manual intervention, particularly by subject matter experts.
- The steps that help to analyze the reason for the error and what corrective action to take to improve accuracy in the future. These errors are measured against key measures for monitoring cognitive system performance including recall, precision, and accuracy.

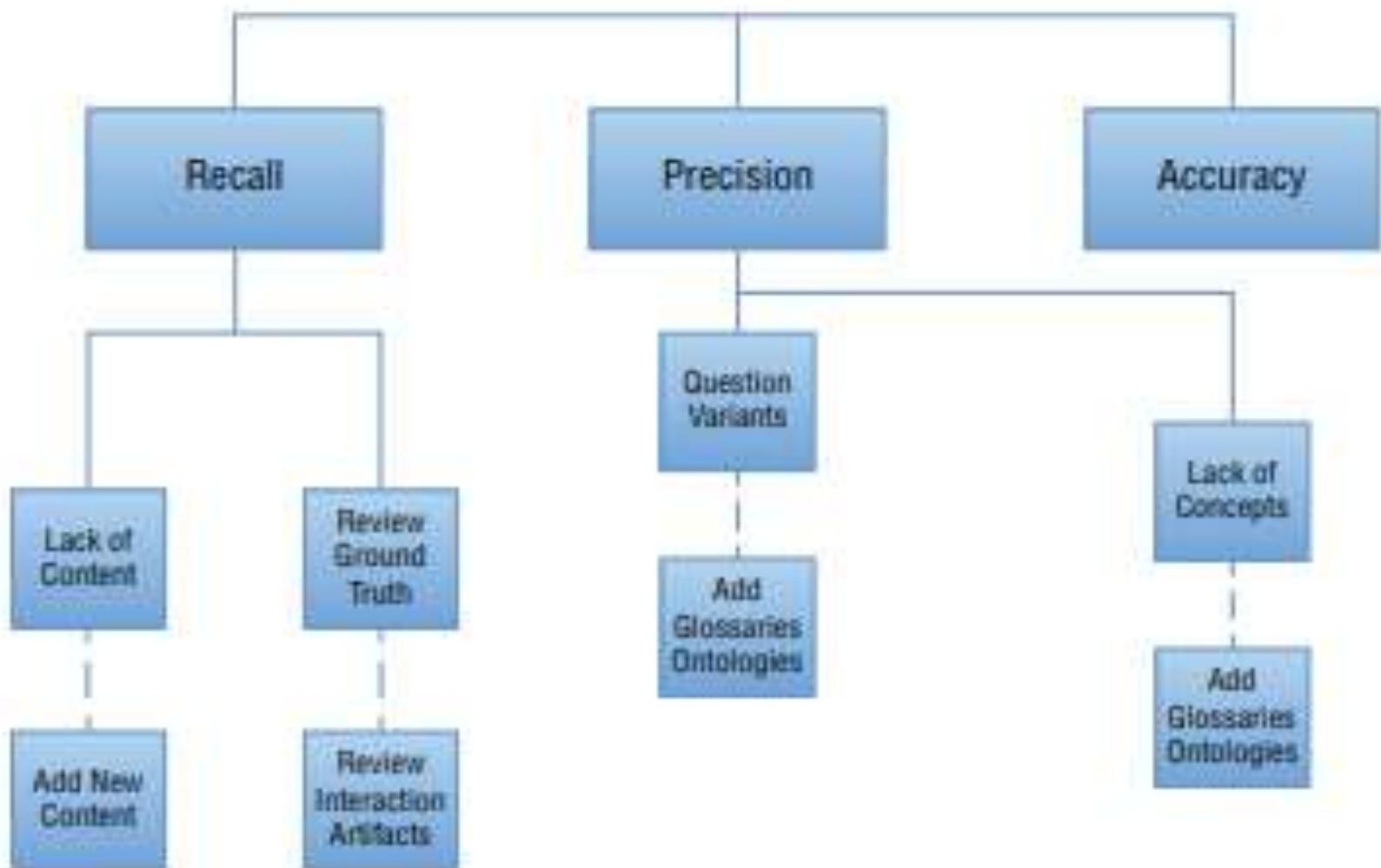


Figure : Improving accuracy of the models

- lack of data is not the only reason why an application may provide incorrect answers.
- Subject matter experts may need to review the ground truth and make adjustments to the answers provided to the system.
- Other errors may occur because the models cannot capture some of the relationships and nuances between similar data sources.
- One approach to improve this is by adding glossaries and ontologies that provide the system with cues to learn more about key concepts.
- Training and testing data can be one of the most time-consuming parts of the process of creating a cognitive system. The smaller the domain, the easier it will be to both create the corpus and find training data to ensure that the information can answer questions and learn over time.
- In this situation, you can select a sample data set that is representative of the type of questions and type of problems you are addressing

- If the domain is larger and more complex, it requires a larger set of sample data.
- In many situations, you can select sample data that is directly applicable to the problem.
- For example, the data regarding consumer questions about treatment of diabetes are well understood. However, you may have a situation in which there isn't a lot of certainty regarding outcomes.
- For example, if you want to understand data from traffic management in a large metropolitan area, you might need a vast amount of sensor data. You may not select the right set of data that is representative of the patterns you want to identify. As you can see, training and then testing results can be complicated by scope and scale issues.
- The most important part of the training process is to have enough data so that you are in a position to test your hypothesis.
- Often the first pass at training provides mixed results. This means that you either might need to refine your hypothesis or provide more data.
- This process is not unlike learning any new discipline where you start with your assumptions based on incomplete knowledge. As you learn more, you can determine that you need more data from more sources. As you gain more insights from the data, your assumptions will change.
- At this point you are ready to test your understanding of the domain to see if you have the right amount of knowledge or if you are still required to collect more data and learn more. This is precisely what happens in an automated fashion when you design a cognitive system.