VOICES OF EXPERIENCE

ARTIFICAL INTELLIGENCE

Demystify. Debunk. Deliver.

Advanced Grid Research
OFFICE OF ELECTRICITY
US DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy by National Renewable Energy Laboratory under contract No. DE-AC36-08G028308, Subtask OEPG.10294.03.01.09 in conjunction with the Utility Analytics Institute under Purchase Order 222037.

DOE/GO-102020-5439

How do we start? Now.

The concept of artificial intelligence (AI) is nearing a century in age[1] and many recent disruptions across industries are built on a backbone of AI. Yet, to many, the concept is still a relative black box. AI sounds like a transformational solution to challenges, but what happens on the inside to make it work?

Today, the complexity of the utility operating environment is expanding rapidly and utilities have no choice but to operate more efficiently and effectively to meet customer needs, adapt to the increasing need for resiliency, navigate the growing environmental challenges, and manage the changing energy mix. This market is a rife environment for AI to proliferate, and a growing number of utilities are on the precipice of transformation because of it, which is the topic of this *Voices of Experience* guide.

Voices of Experience is an initiative sponsored by the U.S. Department of Energy Office of Electricity's Advanced Grid Research group designed to bring utilities together to share their knowledge, insights and lessons learned through implementing the emerging technology that is reshaping the electric power industry. You are encouraged to download the *Voices of Experience* series from SmartGrid.gov/voices and tap into a growing industry knowledge base built on utility experience. Each guide is intended to stand alone, but together they build a more complete understanding of some of the challenges of transforming this industry. Past topics include:

- **Smart Grid Customer Engagement**
- **Advanced Distribution Management Systems**
- **Integrating Intermittent Resources**
- **Leveraging AMI Data and Networks**

The purpose of ***Voices of Experience|Artificial Intelligence*** (VOE|AI) is to better prepare utilities to use AI to address the complexity and challenges of their new reality. One of the initial insights from this project is—regardless of size, location and level of maturity—utilities across the country are interested in and passionate about this topic because they understand its importance to our electric systems. Utilities want to discuss their challenges and successes and learn from others so they can better prepare for the future. And their main message: The time is now. **The speed of the market and new customer expectations demand innovation that AI can enable.**

About This Guide

The information in this guide came directly from utility staff who are working through the challenges of building internal skill sets, investing in a new set of technologies to access and manage more data than ever before, and demonstrating its value to their organizations.

This effort started with a kickoff workshop hosted by Salt River Project (SRP) and Arizona Public Service in Phoenix, Arizona, at Utility Analytics Week 2019, followed by a second workshop hosted by CPS Energy in San Antonio, Texas, and a series of interviews with individuals at utilities on specific aspects of AI.

In total, approximately 61 people representing 40 utilities participated in this initiative through the in-person workshops and phone interviews. Wherever possible, this guide preserves the voices of the participants from the peer-to-peer discussions. However, the themes and common ideas that emerged have been summarized and edited into a single narrative with insights and advice without attribution to any one person or organization.

Contents at A Glance

Utilities engaged in developing AI capabilities have learned lessons and gained insights along the way that can be applied to other utilities preparing for the data-driven future. The goal of this guide is to provide information that might not be accessible elsewhere— the kind you might get from talking to a colleague at a neighboring utility.

A few things to note:

- **All utilities are different and have unique strategies, advantages, limitations and requirements. This document is not a road map that must be followed. It is a compilation of advice and insights that other utilities have learned through their own experience.**

- **Much of the advice and insights are not attributed to a single source because they are summaries from group discussions. Examples attributed to specific utilities are included with permission from the source of the information.**

- **Along the way, the working group identified helpful resources, including documents produced by the Electric Power Research Institute (EPRI), industry stakeholder groups and the national labs. These are not intended to be comprehensive but rather to point readers to additional information on the topic.**

And finally, this guide is not a how-to manual or technical report that must be read from cover to cover. It is simply meant to demystify AI, debunk some of the fear around it, and help utilities to deliver the value it promises by sharing what utilities are learning as they explore and leverage AI in their organizations.

[1] History of artificial intelligence (2020 May 4) Wikipedia: https://en.wikipedia.org/wiki/History_of_artificial_intelligence

Key Takeaways and Insights

Start Simple, Start Now

AI is not simple by definition, but utilities can start taking small steps and leading with simple examples to demonstrate its value to the organization. You don't need to start with a heavily engineered AI product with multiple layers and lots of features; you just need some data and a data scientist that understands how to build basic machine-learning models that can deliver or indicate value. And you need to do it now.

> ***"Just do it. Try it, even if it means starting local in your machines. Demonstrate how it works and how it is interconnected, because it helps you to really communicate and articulate the problem with your business area."***
> **– Rolando Vega, Sr. Manager Advanced Analytics, CPS Energy**

AI Must Be a Core Competency

Data analytics in general has not been a core competency historically for utilities, but it will become one. Regardless of their maturity level in analytics and AI, most utilities are amidst a digital transformation journey that eventually will lead to developing the capabilities to manage, analyze and leverage their data for AI. So, encourage your staff to take advantage of modern certifications that support AI and advanced analytics. Their jobs are not going away but may look a little different in the new environment.

> ***"If you outsource it entirely, you lose the opportunity to learn and adapt. If you want to incubate a new capability or discipline into your organization...you need to actually incubate it."***
> **– Joe Tellez, CTO, Tacoma Public Utilities**

Failure is Part of Innovation

Utilities as an industry justifiably are risk averse; however, the ones working in AI today have found they can generate acceptance of failure by reducing the overall cost of failure through streamlining resources, time boxing efforts and generally adopting Agile development frameworks in projects. The speed of the market and new customer expectations require innovation even if it means failing along the way. So, reduce the cost of failure and speed up the process!

> ***"The whole idea is to reduce the cost of failure so that it's OK to fail. If you have an idea and want to try it out, and you try it and it fails, great. Go onto the next one. Once you have reduced that cost, then you actually start seeing a lot of innovation. Rather than just automating, you actually start innovating."***
> **– Gary Gauthier, Manager of Operational Technologies, DTE Energy**

People Come First

At a high level, people need to know that the purpose of AI is not to replace jobs but to help people do their jobs better. Yes, a successful AI project will have at least some impact on work processes—and, in most cases, a large impact—changing how employees perform their day-to-day work. Change management, partnership and communication are essential for employees to adapt to new working methods. Get users involved in developing the solution to their problems up front, and ensure they are part of the change.

"There needs to be an ongoing feedback loop built into the project cycle."
– Rolando Vega, CPS Energy

Education is Imperative

AI projects involve people with diverse roles and responsibilities across the organization. From the initial stages of attaining buy-in from leadership to delivery of an AI solution, any number of employees will have a hand in how a project is developed, with only a few of them having formal education in data science or AI. As a result, utility leadership must support internal education and awareness of AI from the top down. This includes becoming familiar with common terminology and developing a basic understanding of algorithms, tools and data.

"Education is imperative; there needs to be some level of baseline knowledge to effectively deploy the tools developed to solve problems. Analytics leaders must ensure executive sponsorship and must put the training and awareness programs in place to develop readiness in the organization. This is a lesson I learned the hard way."
– Amelia Badders, Director Enterprise Advanced Analytics, CPS Energy

Foster Collaboration

AI projects can be transformational when partners across the AI life cycle—from concept to production to day-to-day use—collaborate effectively. This includes business leaders, subject matter experts (SMEs), data scientists, data engineers and solution engineers… not to mention the folks from operations, engineering, marketing, customer service, finance, etc. that will be using the solution. Developing AI solutions is as much a cultural exercise as it is a technical one. Teams need to understand the different needs, processes and limitations their collaborators face, so they can engineer solutions to those problems.

"Creating smart partnership between the business and the teams developing the AI ensures everyone is part of the change, which is essential for success."
– Vivian Bouet, CIO, CPS Energy

SALT RIVER PROJECT STANDS UP ANALYTICS DEVELOPMENT PROGRAM

Salt River Project (SRP) is standing up what they are calling an "Analytics Development Program," as part of their Analytics Community, to help increase data literacy in the workforce.

"Our main focus—at least initially—is on creating a common "base-level" of analytics knowledge so that students from multiple disciplines and business areas come out with a shared vocabulary and understanding of analytics, data, and methods," said Mark Drewes, Manager, Analytics Center of Excellence, SRP. "Our hope is that this will lead to greater use of analytics, and increased collaboration between analysts over time."

The current curriculum starts with data scientists from the Analytics Community leading an in-person introduction course, and then augmenting that with online learning and regular meetups over the course of the program.

"In the future, we are also planning on creating a more-compact and higher-level version of this curriculum to deliver to leaders throughout the organization," said Drewes.

What is Artificial Intelligence?

One finding over the course of this effort was that the term "artificial intelligence" means vastly different things to different people. One of the biggest confusion factors is all the different terms used as synonyms for AI, such as machine learning, deep learning and cognitive computing to name a few. These terms are not interchangeable, but they often are used that way.

> *"One crucial part of getting your utility to move forward in AI is to establish common definitions; otherwise, you will end up overworked and chasing your tail."*
> **– Norv Clontz, Director, Data Science Innovation, Duke Energy**

The English Oxford Living Dictionary defines AI as, "The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making and translation between languages."

AI is defined as machines mimicking human intelligence and decision-making. Thus, AI is the operationalization of data science and, in that respect, the deepening of decision science into business operations.

> *"The computer can say, 'That's a cat,' but then you have to say, 'What do you do about the cat?' That is where you move from performing (advanced analytics) to enabling AI."*
> **– Norv Clontz, Duke Energy**

Since its inception in the mid-20th century[2], the field of AI research has been in a constant and increasingly rapid state of evolution. Today, applied AI does not replace the efforts of humans or replicate the human mind, but instead replicates elements of human reasoning in analyzing data and solving problems.

Figure 1. AI Evolution Over Time

A FRAMEWORK OF THE EVOLUTION OF AI

TASK AUTOMATION	DECISION AUTOMATION	INTELLIGENT AUTOMATION	GENERAL INTELLIGENCE
1950's	1990's	2000 – 2010's	Who Knows?
▪ Defined tasks ▪ Process automation ▪ Robotic Process Automation	▪ Business rules ▪ Machine learning	▪ Focused on narrow task ▪ Interpreting sensory information ▪ Pattern recognition ▪ Deep learning ▪ Neural networks	▪ Self-learning ▪ Reasons through novel situations ▪ Cognitive models of the world and how it works

Source: Duke Energy

AI has several different subcategories that characterize the degree to which it can replicate human thought and behavior. Oftentimes, the term "machine learning" (ML) is used interchangeably with AI; however, ML technically is a subset of AI and the basis for much of AI being developed and applied in industry today. The standalone definition of ML is, "The ability of a machine or algorithm to progressively improve upon a set task."[3]

Following are the three types of machine learning:

- **Supervised learning.** This is the simplest and most applied form of ML, where an algorithm is fed examples of labeled data to help it learn and eventually predict outcomes within a given task.
- **Unsupervised learning.** For this form of ML, an algorithm is fed unlabeled data and can independently group and classify the data under a given task.
- **Reinforcement learning.** This ML is when an algorithm includes a feedback loop to reinforce positive and negative outcomes, ideally becoming autonomous over time.

Following are some of the most commonly applied ML techniques:

- **Classification.** Observations are classified into categories.
- **Regression.** A value is predicated from a continuous data set.
- **Clustering.** Observations are assigned into categories, or clusters.
- **Recommendation systems.** Past observations are used to make recommendations.
- **Anomaly detection.** Observations that are outliers to a data set or an expected pattern are detected.
- **Dimensionality reduction.** Less important variables are minimized to determine the most significant ones.

Deep learning (DL) is a subset of machine learning where multiple layers of neural network are constructed to progressively extract information and achieve higher accuracy of learning result. For example, a neural network can include models created using any of the aforementioned ML techniques, itself managing the outcomes of each to deliver an eventual solution or finding.

> *"With machine learning you not only predict or decide what will happen, but you use that feedback to evaluate whether it will be a true case or not, and then make your next iteration more reliable. And eventually, at certain configurations or locations or setups, we can make the system autonomous."*
>
> **– Yingchen Zhang, Manager, Sensing, Measurement and Forecasting Group, NREL**

Figure 2. AI Relationships
DEFINITIONS HELP US DISCUSS THE TOPIC WITH CLARITY

ARTIFICIAL INTELLIGENCE

Any technique that enables computers to ***mimic human intelligence***, using logic, if-then rules, decision trees and machine learning (including deep learning)

MACHNE LEARNING

A subset of AI that includes abstruse statistical techniques that enable machines to ***improve at task with experience***

DEEP LEARNING

The subset of machine learning composed of algorithms that permit software to ***train itself*** to perform tasks, e.g., speech and image recognition, by exposing ***multilayered neural networks*** to vast amounts of data

Source: Duke Energy

[2] History of artificial intelligence (2020, May 4) Wikipedia: https://en.wikipedia.org/wiki/History_of_artificial_intelligence
[3] Heidenreich H (2018, December 4) What are the types of machine Learning? Towards Data Science: https://towardsdatascience.com/what-are-the-types-of-machine-learning-e2b9e5d1756f

The Digital Transformation

Data analytics in general has not been a core competency historically for utilities, and those at the forefront of this evolution say developing the skills and capabilities is a necessary journey. Regardless of their maturity level in analytics and AI, most utilities are amidst a digital transformation that eventually will lead to developing internal AI capabilities.

Many utilities are experiencing growing pains along this journey. Managing data, establishing acceptance, driving cultural change and executing on projects all present roadblocks and obstacles.

Unlocking Value from Data

While advanced metering infrastructure (AMI) provided unprecedented amounts of data to utilities, it also awakened many to what existing data they had but were not benefiting from yet.

Prior to smart meters, most meters were read 12 times per year. According to a report published by IEEE, 15-minute read intervals record 96 meter reads per day and approximately 2880 reads per month—a 287,900% increase in just kilowatt-hour (kWh) data. For utilities collecting additional data such as voltage, current and meter health, that number is substantially higher.

Utilities have managed sets of big data for many years—from meters, supervisory control and data acquisition (SCADA) networks, enterprise resource planning systems, financial systems and customer information systems—but that data has not been the basis of much decision-making until recently. Today's onslaught of AMI, new sensing technologies and unstructured image and customer data is catalyzing a profound shift in how utilities think about leveraging data to drive decisions.

> ***"We all want better data, but nobody really has been thinking about unstructured data... It's really the gold mine. There's so much there in the text, there's so much there in the pictures that...[can] augment your traditional data, your structured data. It might not be the Holy Grail, but it's close."***
> **– Workshop participant in Phoenix.**

Before AMI, the grid was managed, operated and maintained based on primary circuit models and engineering analysis. Years of experience provided veteran employees with the knowledge, expertise and intuition to ensure reliability. Today, with the amount of data at hand and the increasing complexity of the grid, decisions can be made based on data instead of intuition. AI will be the key to using data to optimize operations, resources and customer service.

The amount of data coming into a utility in a single hour is far beyond the processing power of humans. AI helps to make use of that data through programmed rules designed to sort, prepare and move data around as needed.

Once data is sorted and prepared, AI at the application level can be leveraged on a case-study basis to inform business decisions at a speed and level of accuracy never experienced before. This is when transformation at the business level begins.

AI Use Cases Across the Organization

High-value opportunities to employ AI are arising across different parts of the organization, and some do not require cohesion of data from across the enterprise but can be performed on isolated sets of data.

The following table contains examples of some of the more common use cases shared by participants in VOE|AI discussions.

Table 1: Examples of AI Use Cases Across the Organization

Document Automation	Analyze and optimize purchasing of supplies based on underutilized incentives in contracts and sales agreements.
Document Classification	Use ML models to scan through documents and, based on content or the structure of a document, predict what record series they each fall into.
Work Order Optimization and Prioritization	Create document classification algorithms to review and classify work orders as well as optimize critical work orders in generation facilities. The algorithms can also use textual analysis on human-generated work orders.
Load Forecasting Distributed Energy Resources (DER)	Information on type of equipment and weather conditions can be used to inform production of residential solar. Pair this with propensity modeling to deliver short- and long-term forecasts.
Curtailment Estimation	Use weather and meter production data to estimate curtailment needs for wind and solar units.
Electric Vehicle (EV) Projections	Self-reported EV adoption data from the customer relationship management (CRM) system can be used in combination with department of motor vehicles data and quarterly and forecast data from EPRI to forecast demand for EV charging as well as inform planning and construction of charging stations.
Optimize DER	Use distribution system data to optimize storage, solar and other DER at the residential level.
Identifying Faulty Equipment	Use a supervised ML model to identify faulty equipment based on excessive energy use.
Predictive Maintenance	Use phasor measurement unit, AMI and geographic information system (GIS) data to try to predict asset failure and prioritize maintenance on power delivery assets.
Vegetation Management	Analysis of images and light detection and ranging (LiDAR) data can be used to optimize vegetation management.
Outage Management	Build a comprehensive regression-based tool, using AMI data from past outages, to predict the potential magnitude of outages prior to events.
Modeling and Validating Connectivity	Use AMI, SCADA and GIS data to model and validate connectivity at the household level.
Call Center Staffing	Using ML, predict call center volume from 1 day to 14 days in advance to inform scheduling.
Call Center Response Optimization	Use an ML model to predict times and conditions of customer complaints down to the individual customer and address them proactively with automated notifications.
Call Center Automation	A web-based bot tool using natural language process can help to answer customer questions as they type.

UNSTRUCTURED DATA

Utility data is usually thought of as meter readings and other telemetry models—readings from devices on the system that are transmitted to a central collection point as structured data that is then dumped into a database. While this process is handy for monitoring the system, what is currently lacking is the paring of the structured data with unstructured data that are public and instantaneous—pictures, voices and text messages for example.

Consider that an outage can be identified using only unstructured data like news feeds and pictures that are taken and posted on Twitter. Or smoke that can be detected using image recognition. Paired with the system data, the utility can not only detect the outage, but also identify safe areas that are still powered. Translated into outage management, the utility can now dispatch their crews quickly and safely without waiting for a call from a customer.

There are many other examples of where the use of unstructured data can make a huge impact on utility operations. One is short-term solar forecasting, which no numerical weather model can predict accurately because cloud motion in short duration is very hard to track using these computationally intensive models. What technology companies and national labs are working on is using pictures of the sky and deep neural networks to predict the motion of the clouds to give very accurate 5, 10 and 15-minute solar forecasts specific to the regions captured in the cloud images. The next step for the utility would be to schedule flexible resources in those regions according to the forecasts. That might include, for example, dispatch ramping and reserve generators. Locally, with the advancement in demand side management, flexible loads can be better utilized using the short-term solar forecasting. For example, if a solar ramping down at one location is forecasted, a building management system can tell a connected thermostat to pre-cool or pre-heat the building in order to mitigate the solar ramping.

While utilities see the value in unstructured data, many have barely scratched the surface of extracting the value in their structured data that they already collect. There is also pressure to do more with the existing data and make their existing models more accurate. And, there is fear that adding unstructured data (that may be inaccurate) to the mix will make their existing models less precise rather than more.

This, however, may just be a step in the data evolution. Other industries with “inaccurate” or “incomplete” data are using AI and ML to complete the information. (That is how Netflix makes movie recommendations specific to their customers and online retailers make product suggestions to shoppers.) The utility industry is likely to adopt these methods as well, making them the new normal for utility operations.

Additional Resources:
An Introduction to AI, its Use Cases, and Requirements for the Electric Power Industry (EPRI); Natural Language Processing and its Application in the Utility Industry (EPRI); Voices of Experience|Leveraging AMI Networks and Data (DOE); Solar Irradiance Capturing in Cloudy Sky Days–A Convolutional Neural Network Based Image Regression Approach (IEEE); Matrix Completion for Low-Observability Voltage Estimation (IEEE)

Be People Centric

As if utilities did not already have enough on their hands trying to acquire talent to replace a generation of SMEs who are approaching retirement, the introduction of AI is posing new questions as to what the future workforce needs to look like. As these unknowns become clear, the demand for talent is likely to escalate.

An upsurge of new job roles will be needed to develop AI solutions, formulate and lead new AI business strategies, and manage AI projects. Although this will pose a challenge for utilities, there is a big upside: AI will enable systems and equipment to do a lot more; it will empower the utility workforce to do a lot more as well.

Hiring and training are not the only part of equipping a utility for AI. In fact, much of the important legwork happens before utilities get to the step of acquiring and developing new talent.

A successful AI project will have at least some impact on work processes—and, in most cases, a large impact—changing how employees perform their day-to-day work. Acceptance of AI project elements is complex, and it requires dedicated, intentional and ongoing effort.

Building Skills, Community and Buy-In Core Principles

AI holds boundless potential to improve the business. For a utility to get started, leadership and staff first need to understand how AI can help them, so they can support it. The groundwork needs to be laid in terms of demystifying AI, engaging employees, setting expectations, adapting workflows and thinking about creating people-centric solutions. Failure to do so will impact buy-in and confidence negatively at all levels.

> ***"Get people from the business involved up front in the concept development stage…so that they understand that you're not just handing them something saying, 'Here, this is going to help you make decisions about your job…' These team members know their data really well, and more importantly, know their challenges and pain points that are essential parts of designing a solution that will help them do their job better."***
> **– Rolando Vega, Ph.D., P.E., CPS Energy**

ADVICE FOR ENGAGING THE ORGANIZATION

Demystify AI

The first step to building support at all levels for AI is to demystifying it. This means providing education as to what AI is and can do—especially in the context of bringing value to the utility—and countering negative assumptions and bias. Simple examples and isolated proof-of-concept efforts are ways to demonstrate the inherent value of AI. Simple, small use cases also can be built into the workflow with less consequence and less risk, which can temper cultural resistance.

Engage Stakeholders at the Onset

At a high level, people need to know AI will not eliminate jobs. (Debunk this idea now!) Rather, it will enable them to do their existing jobs better. However, it is likely to bring changes to job roles and workflows. Change management, partnership and communication are essential for employees to adapt to these. Business process experts and SMEs are just as crucial to success as data engineers and data scientists, so engage them at the onset.

Create Community

Build a mechanism for people with interest and skills to work together. Even if AI skill sets exist in a utility, they might not be organized in dedicated advanced analytics teams—and that is fine. Especially for utilities just getting started, creating opportunities for people with skills and interest to meet, learn and collaborate will go a long way.

Accept Failure and be Willing to Pivot

The traditional project trajectory for most utilities involves a set of outcomes or completion points and is highly schedule oriented. Participants in the VOE|AI project emphasized how important it is for utilities to adopt a much more flexible, experimental attitude toward AI projects.

Enable the People Using the Solution to Co-Create it

To get people to accept changes into the workflow, there needs to be a constant feedback loop in the product delivery and development process. For utilities, this requires a paradigm shift to enable coordination between traditionally siloed teams across IT, AI developers (data scientists and AI data and solution engineers) and the people in the business units who will use the solution.

This recommendation goes hand in hand with engaging the business at the onset of a project and understanding that even when a tool has been developed, creating a valuable solution might take more work and iteration.

CASE STUDY: SALT RIVER BUILDS AN AI COMMUNITY

To grow and educate its analytics workforce, SRP started by holding analytics-themed events and brown bag lunches in 2017, where teams presented projects they were working on to get guidance, insights and knowledge from the larger community. Since then, the community has evolved and cross-functional teams have formed to tackle data projects. Some of these projects, such as forecasting short-term solar capacity, have resulted in significant dollar benefits.

“One of the best things that we have going on right now is our analytics community,” said Mark Drewes, Manager, Analytics Center of Excellence at SRP. “From a perspective of getting the word out of what all of these new tools and technologies are throughout our organization, I think that has been a great conduit for that. We’re getting analysts of all levels to come to those events.”

“Even if they’re Excel-based analysts, they’re at least hearing about these other technologies,” Drewes continued. “And they’re thinking about them and asking questions about them later on. With the analytics community also comes our analytics leadership team, which helps provide the guidance on which projects tie back to our strategic objectives from a company perspective. This determines what specific things we should be working on or what things the analytics community can start to train or educate on.

“We have been talking about getting value out of data and analytics for a long time. And now, thanks to the structure that the community provides, we are really seeing a maturity in the types of data projects being proposed and the skills being leveraged by SRP analysts. A few years ago, most data requests were operational-related, reports and dashboards focusing on what had already happened. Now, analysts are increasingly leveraging advanced analytic capabilities to help SRP make better decisions in the future.”

According to Drewes, “Our community efforts have accomplished two really important things. One is to get people to think about data in new ways and how they can use it, but also to get them to think, ‘OK. I see other groups in SRP doing these things, and these are the results they’re getting. I want to be part of that too and leverage some of this data that their groups are collecting, plus their kind of knowledge and skills to do the same thing.’”

AI Job Roles and Personas

To develop, deliver and deploy an AI product, there needs to be more than a data scientist doing analysis in a dark room somewhere; there needs to be people to manage the data coming in and the data coming out, as well as people to ensure the data scientist's analysis is delivered in a format that can be acted on by the business.

As utilities gain more sophistication in their analysis of data and transition from descriptive analytics and dashboards to predictive analytics built on AI, the level of skill and expertise also needs to evolve from basic tools like Excel and Power BI to programming interfaces like R and Python. To do this, utilities need an AI engineering team that can put those models into production.

That brings up another point: **successful AI projects depend on skill sets far beyond that of a data scientist.** AI is not just smart algorithms that replicate human-thought processes, it is the operationalization of those algorithms.

Operationalizing data science requires cross-functional teams with skills in project management, deep business knowledge, data science, data engineering and application development. More and more utilities are spinning out analytics communities and even formal organizations that act as a shared service across the enterprise. Though specific organizational models vary greatly in terms of formality, governance, reporting, roles and responsibilities, they share a common purpose to push the AI maturity of their organization forward.

> ***"End users don't want a black box solution. They want to explore alternative solutions, zoom in on certain ones, change inputs and participate in finding a solution. If your model result reinforces the decision they were going to make anyway, then they don't question it. But if it doesn't, they rightfully want to understand how it could be telling them something different than their 20 years of experience has."***
> **– Eina Ooka, Senior Quantitative Analyst, The Energy Authority Inc.**

Figure 3: CPS Energy's Analytics Personas

DIGITAL TRANSFORMATION (JAVA) SME

Data Warehouse	Works with Advanced Analysis and DBAs to identify valuable CUSTOMER-FACING data assets/products
API	▪ Dev/test/deploy new CUSTOMER-FACING API's ▪ Maintains Customer/Public API ecosystem
Data Catalog	▪ Develop API specification (RAML) ▪ Provide comprehensive testing cases and examples ▪ Publish API documentation to Exchange
Website	▪ Design ▪ Maintain

APPLICATION INTEGRATION (MIDDLEWARE) SME

Data Warehouse	Works with Advanced Analysis and DBAs to identify valuable EMPLOYEE- or 3rd PARTY-FACING data assets/products
Data Extract & Load	▪ Ingest data into databases ▪ Maintain workflows
API	▪ Dev/test/deploy new EMPLOYEE- or 3rd PARTY-FACING API's ▪ Maintains these API's
Data Catalog	▪ Develop API specification (RAML) ▪ Provide comprehensive testing cases and examples ▪ Publish API's documentation to Exchange

BUSINESS SME

Data Warehouse	▪ SQL queries ▪ Data engineering Biz dev ▪ Test data quality
Dashboards	▪ Develop/test low hanging fruit ideas ▪ Present ideas ▪ Drive business adoption
Data Catalogue	▪ Tag sensitive information ▪ Document data assets in Biz App ▪ Publish API documentation to Exchange

ADVANCED ANALYTICS SME

Data Warehouse	▪ Data engineering promotion and enterprise dev ▪ Propagate skills
API	▪ REST API calls from 3rd Party ▪ REST API calls from our System API's ▪ Propagate skills
Dashboard	▪ Develop/test complex high-value ideas ▪ Propagate skills
ML/AI	▪ Dev/test application ▪ ML automation ▪ Python scripting ▪ Propagate skills
Data Catalogue	▪ Tag sensitive SQL joins ▪ Document data assets in enterprise catalogue

Source: CPS Energy

Here are some of the common roles and personas (i.e., representative of a person) in AI projects:

Project Managers

Utilities have employed project managers across different business units for years. Formal education, certifications, time and experience have enabled project managers to refine their skills and processes. A growing number of Agile and Scrum concepts are being formally and informally integrated into project lifecycles and training is available. Unfortunately, AI projects are asking a lot more of project managers and no real education or certification exists for AI projects.

Core skills and tools:

- **Understand and manage cross-functional teams**
- **Book end the process with available data and eventual outcomes**
- **Balance the disproportionate effort of preparing data with other project elements**
- **Iteratively deliver the product with a minimum viable product (MVP) approach**
- **Gantt charts, Kanban boards, task lists, Excel-based tools and Microsoft Project**

Business SMEs

Employees embedded within business units are the most familiar with practice, process, questions, needs and pain points. Therefore, analysts within the business unit are crucial to provide unique representation on AI projects. In an ideal world, these people have data analyst and business intelligence skill sets that include deep knowledge of business process, business-specific data, and general knowledge of digital tools and process.

However, many utilities reported that, at present, individuals with such a comprehensive skill set are unique. As a result, it may take a combination of individuals to fulfill project needs until time, training and general analytics maturity expands.

Today, it is not common, but it is important to support business SMEs in the development of data knowledge and skills. These individuals play an increasingly crucial role in data governance by acting as intermediaries between the business, analytics teams and IT. They also are important data custodians who can share knowledge of data, document data, and create and enforce data governance processes that ensure quality, security and appropriate accessibility.

Core skills and tools:

- **Understand business processes, needs, challenges and pain points**
- **Knowledge of people and culture in the business**
- **Reporting and business intelligence (BI)**
- **Basic data management and governance skill set**
- **SAS, Tableau, SPSS and Power BI**

Data Scientist/Advanced Analytics Expert

Data scientists are highly trained individuals who create AI models based on their understanding of statistics, ML, optimization and other analytics. They also advise on how to apply the insights produced by models to solve problems and create value. This is their core value and should be their core function.

Today's data scientists are skilled data analysts with programming skills who can establish processing capabilities for large amounts of data and perform data mining and exploratory work on data. In large projects, this knowledge is important for them to collaborate with data engineers and application developers, but it is not their core responsibility.

Note that it is not uncommon to overextend the role of the data scientist; many utilities admitted to imposing extremely high and oftentimes unrealistic expectations on data scientists to fulfill larger roles across the life cycle of a project than they have bandwidth for or qualification to do so.

Core skills and tools:

- **Develop and tests AI/ML models**
- **Data engineering and management knowledge and skills (a support skill set)**
- **Application development and application programming interface (API) knowledge skills (a support skill set)**
- **ML and applied statistics, Python and R scripting, Scala, Apache Spark, Hadoop**
- **Jupyter and Rstudio**

Data Engineer

Data engineers support the AI process through designing, developing and maintaining data ecosystems. With large amounts of data coming from different sources, this role oversees the extract, transform, load (ETL) function. They also oversee big data warehouses that are the primary source for data being used by the data science team, often referred to as data lakes in the utility and many other industries.

Core skills and tools:

- **Support identification of essential data assets and planning**
- **Develop and manage ETL infrastructure and workflows**
- **Work with other parties in the publication and documentation of data as well as developing and instituting data governance policies and procedures**
- **Python scripting, Hadoop, SQL, NoSQL, MapReduce, Hive, Pig**
- **MongoDB and Cassandra tools**

Application Engineer/Full-Stack Developer

Application developers manage the development of infrastructure beyond the data layers overseen by data engineers. This role often overlaps with the data engineers, particularly on smaller teams. A key part of this role is API development, testing and management for customer- and noncustomer-facing applications, the latter of which overlaps with the data engineer role.

Core skills and tools:

- **Service management**
- **Full-stack programming**
- **Utility-specific IT infrastructure; SAP and Oracle**
- **Hadoop, Spark**
- **AWS and other cloud computing solutions**

ADVICE FOR BUILDING AI TEAMS

Start Small

Start with a small team containing the critical skills, give them some data and some hypotheses, and let them play with it. A simple project can go a long way in demonstrating the value of AI.

Establish Job Descriptions

Work on establishing correct job descriptions and classifications. If they do not fit naturally into the existing organizational chart, that is fine. Put them where it makes some sense and functionally allow them to operate how they need to.

Educate Staff

For existing employees—especially IT staff—take advantage of modern certifications that support AI and advanced analytics, such as Amazon Web Service (AWS) cloud certification, Agile project management certifications and DevOps. Their jobs are not going away but may look different in the new environment.

Consider Outside Partners

For organizations that lack some or all of the critical skills to move forward, consider partnering with a vendor or service provider, but do not put everything into the hands of the outside party. Use partners strategically to supplement and train your staff in the short term so you develop internal competency for the long term.

CASE STUDY: HOLISTICALLY BUILDING AI TEAMS AT DUKE ENERGY'S MADLAB

Duke Energy's MADLab initiative, where MAD stands for machine learning, artificial intelligence and deep learning, is an important driver of innovation culture at the utility. The group began as a collaboration between a small number of data scientists from marketing and engineers and architects from IT to deliver an image analytics product in late 2017. The two teams have since collaborated on AI projects with businesses across Duke Energy, and both teams have grown organically over time based on growing demand across the utility.

Today, Duke Energy's MADLab consists of a team of nearly 30 data scientists, data engineers, and developers from marketing and IT teams who are co-located and function as an integrated team. Projects span business units across Duke, from generation to customer.

The group works differently with teams across the business on each project, and structures each project team according to the type, size and nature of the project. "On smaller projects or more experimental proof-of-concept (POC) efforts, the Duke Energy MADLab team works more or less autonomously, assigning its own staff as project managers and Scrum masters to lead a project that might last just a few weeks," said Norv Clontz, cofounder of the Duke Energy MADLab. "On a larger effort or when rolling out a successful POC, we will work with budget allocated from the business and collaborate on a much deeper level with business SMEs. Those are the big-budget projects you hear so much about, but they all started small."

Since its founding, Duke Energy's MADLab has seen many successful AI use cases go into production across the enterprise, some of which are estimated to garner millions of dollar savings each year. It is now delivering projects in emerging areas that include deep learning for video and image, analytics from data collected from drones, neural networks, natural language processing and optical character recognition.

"What we have developed is special, but in theory, it's replicable at any scale…you just need a small team that is qualified to access and play with the data, and then the willingness to pursue areas of promise, supplementing skills from other teams and through contract partners."

Regardless of these successes, Clontz admits the Duke Energy MADLab team still needs to engage in awareness building and marketing of its capabilities to businesses that might otherwise seek solutions outside of Duke Energy.

"We try to inform business teams of opportunities we see while working on assigned projects, where we see a high potential for value or potential transformational opportunities. One thing we have been trying to do is to encourage more discovery work when we are on-site with other teams, but only in areas we are confident about, and over time becoming a trusted internal partner to the business."

AI Project Needs and Processes

AI projects are not linear in nature. They require discovery, iteration and a lot of adaptation from concept to finish. This can constitute a nightmare for planners, or it can be considered a framework for continuously identifying opportunities.

The life cycle of an AI project is more akin to a flowchart than a traditional plan-execute-control-close model. At all stages of a project, stakeholders must constantly reevaluate their next steps, determining where to allocate resources and when to pivot. Part of this process is constant delivery of products into the hands of customers, from which the team garners essential feedback to plan next steps.

Even small AI projects are complex. It is important to maintain communication between all parties involved. In the tech industry, it is common to hold daily stand-up meetings among teams that are 100% allocated on a project. It is less common for utility teams to be 100% focused on a single effort, but utilities that participated in the VOE|AI workshops stated their teams meet multiple times each week.

Similarly, coordinating information from across the enterprise, namely establishing transparency between operating and finance groups, will increase the visibility of benefits that can be delivered by AI.

> ***"It's important to understand all the people and processes associated with AI projects. It's not just data scientists creating models and then handing those off to end users. The effort might start with someone from the business and a data scientist collaborating to clearly define the problem, the solution today and solution tomorrow. However, as the project formalizes, these parties are just a small part of the overall work. There's data and feature engineering involved, database administration, middleware, API management and the design of customer or business experience...all these pieces of the puzzle that need to come together for the AI model to help us make more optimized decisions."***
> **– Rolando Vega, Ph.D, P.E., CPS Energy**

Selecting and Pursuing AI Projects

Even if a utility had all the resources in the world, the most sweeping AI efforts should start from humble beginnings. The first projects should be the result of isolated POC efforts that demonstrate the inherent value in the effort.

These were the two most common questions participants in the VOE|AI workshops asked about AI:

- **How should my organization get started?**
- **How do we identify and select projects that can deliver value?**

With the amount of data and new technology available, it can be difficult to figure out where to start. At the same time, utilities with constrained resources need to understand where their dollars can produce a tangible road forward. Should it be spent on staff? New technologies? Educating existing staff?

Internal stakeholders also can muddy the process of identifying projects strategically based on their own biases.

"We get new requests in on a daily basis that we can never get to, not even at the experimental level. We have had to find ways in which to optimize, and it's never that straightforward."
– Norv Clontz, Duke Energy

ADVICE FOR GETTING STARTED

Build Simple Examples to Attain Buy-In

Although AI can and should be transformational for utilities in the long run, utilities just getting started will need to lay a foundation first. Building simple examples, aside from being lower in cost and risk, will help to gain buy-in by demystifying AI and creating value quickly, even if it is just incremental value. Additionally, smaller use cases with a lower initial cost and impact to work processes will naturally integrate into the business more easily than a much larger transformational initiative.

Take an Enterprise-Wide Approach to Evaluating Metrics and Value

Utilities traditionally do not share data across the enterprise at a high level. Personnel often are disproportionately focused on solving their own business problems with their own data, investing in singular projects and pursuing duplicate efforts at times. With more coordination, whole organizations can benefit from economies of scale with respect to AI infrastructure, model and tool development.

Consider the Customer

Some utilities, such as CPS Energy and Austin Energy, shared they include the customer journey as a criteria to evaluate project potential. It is important to ask: How can a project improve a customer pain point or result in a meaningful interaction or experience?

Do Not Isolate AI Projects

One benefit of AI projects is that work completed for one effort or use case, particularly in the areas of data delivery and preparation, easily can be transferred over to other projects. Even at the discovery stage, several utilities reported it is common for analysts to uncover areas of value while trying to solve separate problems. As one member stated, "It's like using the same recipe and starter to bake another loaf of bread."

Do Not Chase Perfection, Chase Improvement

Plan and prioritize based on value to stakeholders. Within the sphere of ML, data scientists constantly are looking to improve the accuracy of their models. Typically the more data and detail in the model, the more accurate it is. The challenge for teams is knowing when a model is good enough.

> *"We can keep working to chase that .01% improvement, but is it worth it for what it will cost? Data Scientists love playing around with data and chasing perfection, but you need to put some kind of wrapper around these types of projects so they do not go on with no end. Get to level of accuracy that you need and publish the project, but keep ideas for potential improvement documented in case you ever revisit in the future."*
> **– Mark Drewes, SRP**

CASE STUDY: AVANGRID'S PROJECT SELECTION CRITERIA

As part of its wider innovation initiative, Avangrid Inc., a sustainable energy company based in Orange, Connecticut, holds regular meetings with employees from across the enterprise to share projects and ideas. One outcome of these meetings was the development of a holistic set of criteria to use for AI case study selection.

Avangrid considers the following criteria as well as overall strategic alignment to corporate strategy:

- **Sustainability.** Does the project address an environmental sustainability challenge?
- **Business value.** Are there both financial and nonfinancial business drivers?
- **Data.** What is the quality and availability of data needed?
- **Regulatory.** Is there a regulatory obligation associated with the project?
- **Business readiness.** Will the project begin to deliver value immediately?
- **Success.** How likely is the team to be successful?

"With these types of projects, the integration piece is always going to be challenging, so we let the amount of value that we think we will get really drive the decision of whether or not to move forward," said Ahad Esmaeilian, manager of smart grid innovation at Avangrid.

AI Project Management Principles

AI is too complex to be managed within a linear framework.

Utilities have long seen analytics project efforts fall flat under traditional waterfall-style project management practices. Some complained traditional frameworks often prioritize efficiency over value, often go over budget, drag on and risk delivery of obsolete solutions by the time the project ends. With AI projects, it is crucial to build in stages, reevaluate and sometimes pivot plans and strategy.

Concepts of Agile project management and development frameworks have been working their way steadily into utilities in all business units—not just in IT and analytics teams. Each concept provides structure to address complex and adaptive problems while ensuring delivery of solutions.

Most of the utilities engaged in the VOE|AI project reported they employed Scrum Alliance-certified project managers, but several other frameworks exist, including PMI-ACP, IC Agile, SAFe, APM (Agile project management), EXIN Agile and Lean/Kanban. Each of these are tailored differently to individuals, depending on their role and level of experience.

> *"When we launched our data science program, learning how to draw time boxes around projects was something that we realized we had to do. Perfection isn't possible, and trying to get there is time-consuming, expensive, and usually more than what the business needs."*
> **– Mark Drewes, SRP**

Integrating AI into the Business

AI is the operationalization of data science—and this last step involves both technical and cultural challenges.

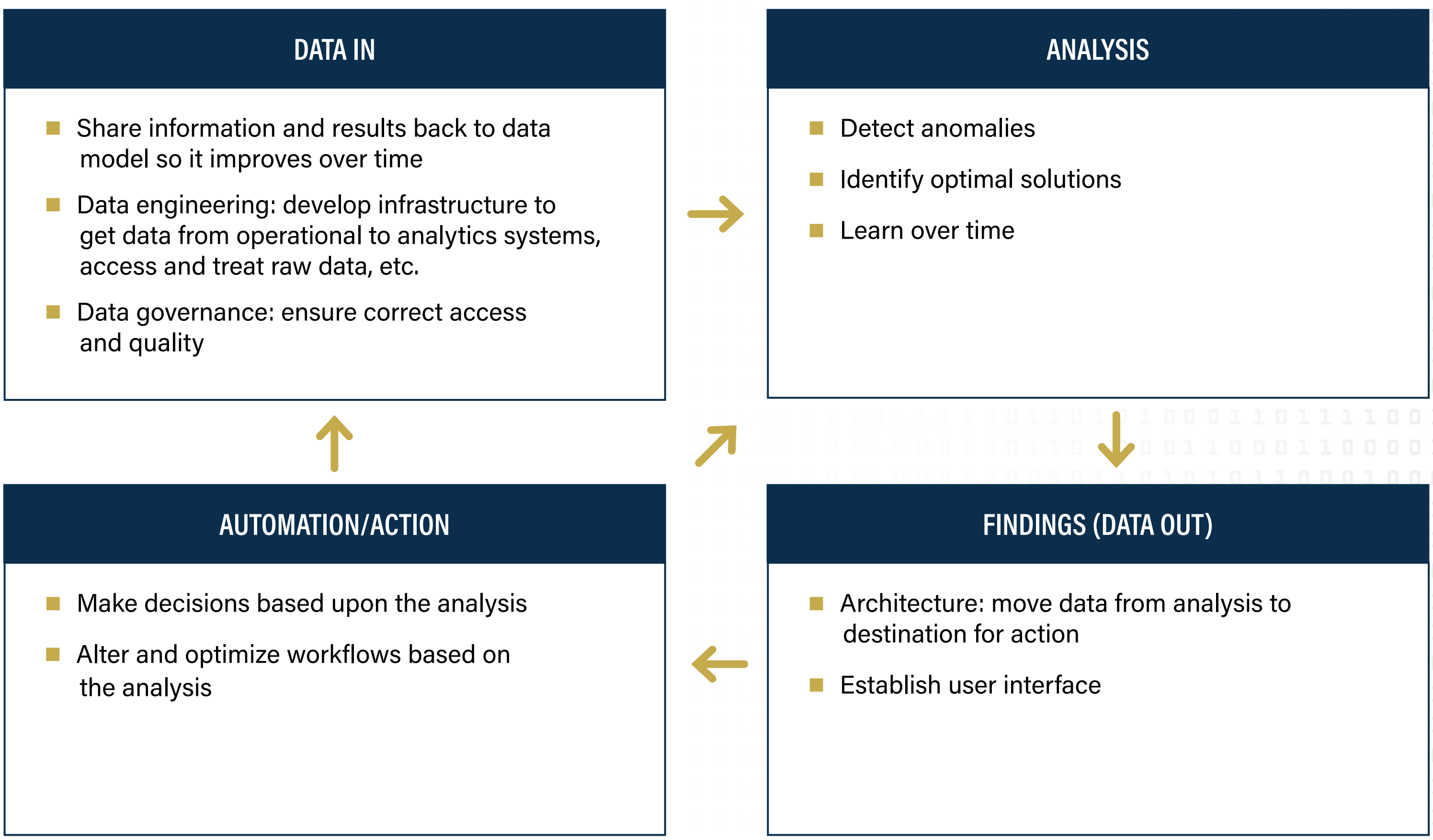
The term data science has been used interchangeably with AI. However, that is incorrect. Data science must be operationalized—that is, moving from POC to being embedded into business process—for it to fall into the category of AI.

What does this mean? Certainly, data scientists can perform analysis on data and develop known rules that help to identify a problem—such as a broken asset, an inefficient process or how customers behave—but is that improving on what the workforce already knows and can do today?

The ability to draw insights from data, no matter how complex, is just the starting point. Embedding these decisions into workflows to create improvement—like making more decisions in the same amount of time, processing information faster and improving the accuracy of decision making—is what leads to transformation.

Participants in the VOE|AI discussions consistently stated that operationalizing AI was one of the most difficult aspects of projects. This process requires the development of a system that provides data to models and pushes the model findings out to workflows. The figure below illustrates this process.

Figure 4: Operationalizing Data Science



Source: Utility Analytics Institute

ADVICE FOR MANAGING AI PROJECTS

Enable Collaboration

Create an environment that fosters collaboration, such as building in regular meetings between cross-functional project teams and locating teams together rather than segregated within business units or departments. This will help teams who develop the solutions and those who manage system operations to work together more effectively and efficiently, ultimately providing solutions to the business more rapidly.

Use Agile Development Principles

Keep teams focused and moving with Agile project management frameworks that enforce delivery of findings and products in short periods of time. Deliver solutions and iterations rapidly, seek feedback and then revise. After a model is first published (once it works), there is still the element of developing a user interface that people find useful and intuitive.

Use the Right Tools

Today's developers need to work with an extensive network of outside parties and administrators that manage different components in the AI life cycle. Look for technologies that can streamline this process. Technology providers are spinning out tools like this at a rapid rate, such as Cloudera Data Science Workbench, Databricks and MuleSoft, that help with efforts at all project stages. Investigate the use of APIs and API management tools to speed up the integration.

CASE STUDY: MOVING FROM EXPERIMENT TO PRODUCT

***"Getting models off desktops and into production is a big challenge."* – Mark Drewes, SRP**

Mark Drewes, Manager, Analytics Center of Excellence, SRP, provided some insight into some of his team's challenges to operationalize AI. "We can't have our data scientists working on these models every day to keep them up to date. The process of getting the models off their desktops (where they use Python with isolated data) and into an application where it runs automatically every day, with up-to-date and sometimes streaming data, is anything but straightforward.

"What this really means is moving what someone is doing in isolation on their computer into a distributed computing environment, and that doesn't even take into account the human side of things," Drewes noted.

Moving into production, models often need to be tweaked and altered to function properly in the new environment (that is, off the desktop). Once into production, there likely will be subsequent updates that need to be made to the product, based on building out originally planned features or user feedback.

"At the end of the day, this requires extensive back and forth between the data science team, AI engineers/architects and [the system administrators] in IT," Drewes said, "and there needs to be some order to that chaos."

"When you try to solve a problem, you may be successful and solve that problem, but you may not. That is a significant challenge that our team is facing; we do not always solve the original problem, but we usually solve three others that create a lot of value. Our role includes communicating these new solutions to our business partners. It's a whole new way of thinking about things."
– Amelia Badders, CPS Energy

WHAT IS DEVOPS?

DevOps is the combination of cultural philosophies, practices and tools that increases an organization's ability to deliver applications and services at high velocity, evolving and improving products at a faster pace than organizations using traditional software development and infrastructure management processes. This speed enables organizations to better serve their customers and compete more effectively in the market. Engineers work across the entire application life cycle—from development and test to deployment and operations—and develop a range of skills not limited to a single function.

Additional Resources:
https://devops.com/; https://aws.amazon.com/devops/; https://www.scrumalliance.org/; https://www.pmi.org/certifications/types/agile-acp
https://www.apm.org.uk/resources/find-a-resource/agile-project-management/

Data: The Base of the Iceberg

Data is the foundation for AI, and the more data available to train an algorithm, the better the outcome. There is a high level of effort required to get data to a state where it can be leveraged effectively by AI at the application level.

At the same time, opening access to data sets that traditionally have been siloed has caused cultural conflicts around how to properly manage and share data. Not only are there fears in terms of security, but other forms of anxiety and resistance toward AI in general also impede the willingness to increase access to data.

Utilities report that data management and preparation consist of the bulk of the work for AI projects. Several experienced VOE|AI participants cited the 80/20 rule as a standard expectation, where 80% of the effort in an AI project is spent assembling and preparing data and the other 20% is dedicated to developing and deploying models.

> *"The loudest pain point isn't even in advanced analytics; it's how manually intensive it is to both access the data and do something meaningful with it."*
> **– Gary Gauthier, DTE Energy**

Why so much work? At the point of collection, raw data must be ingested and flow through a highly engineered pipeline that ensures it goes to the right place and is formatted properly for use. With thousands of sources of data coming into a complex network of utility systems, this process can be debilitatingly complicated. As a result, many utilities report their data oftentimes is underutilized.

In the context of managing a complex system of data, human intervention is critical in terms of developing standards and rules for data. For utilities with mission-critical data, it is crucial to impart systems of governance and stewardship upon data.

When utilities can start to work through that complexity, they begin to unlock the value.

> *"While legacy systems can be challenging when customer expectations are being driven by the "Digital Dragons" (Google, Amazon, etc.) who have invested heavily in state-of-the-art data and analytics systems, start with where you are. Get value from the data you have as you move towards more advanced AI development. When you release the data from even your legacy systems, and you give the business access to it , people are very, very grateful."*
> **– Vivian Bouet, CPS Energy**

Figure 5: Key Data Management Activities

Ingest/Acquire Data	Build Data Pipeline	Establish Governance
Charactierize Data Files	Raw File Store	Security and Permissions
Files	Determine Destination	Quality, Accuracy, and Safety Protocols
Database	Cleanse and Format Raw Data	Monitor Usage, Destinations, and Outcomes
Streaming Data	Process Data for Destination	
Sensors/IoT	Catalogue and Describe Data	
Subscriptions	Publish Data	

Source: Utility Analytics Institute, 2020

Data Infrastructure and Engineering

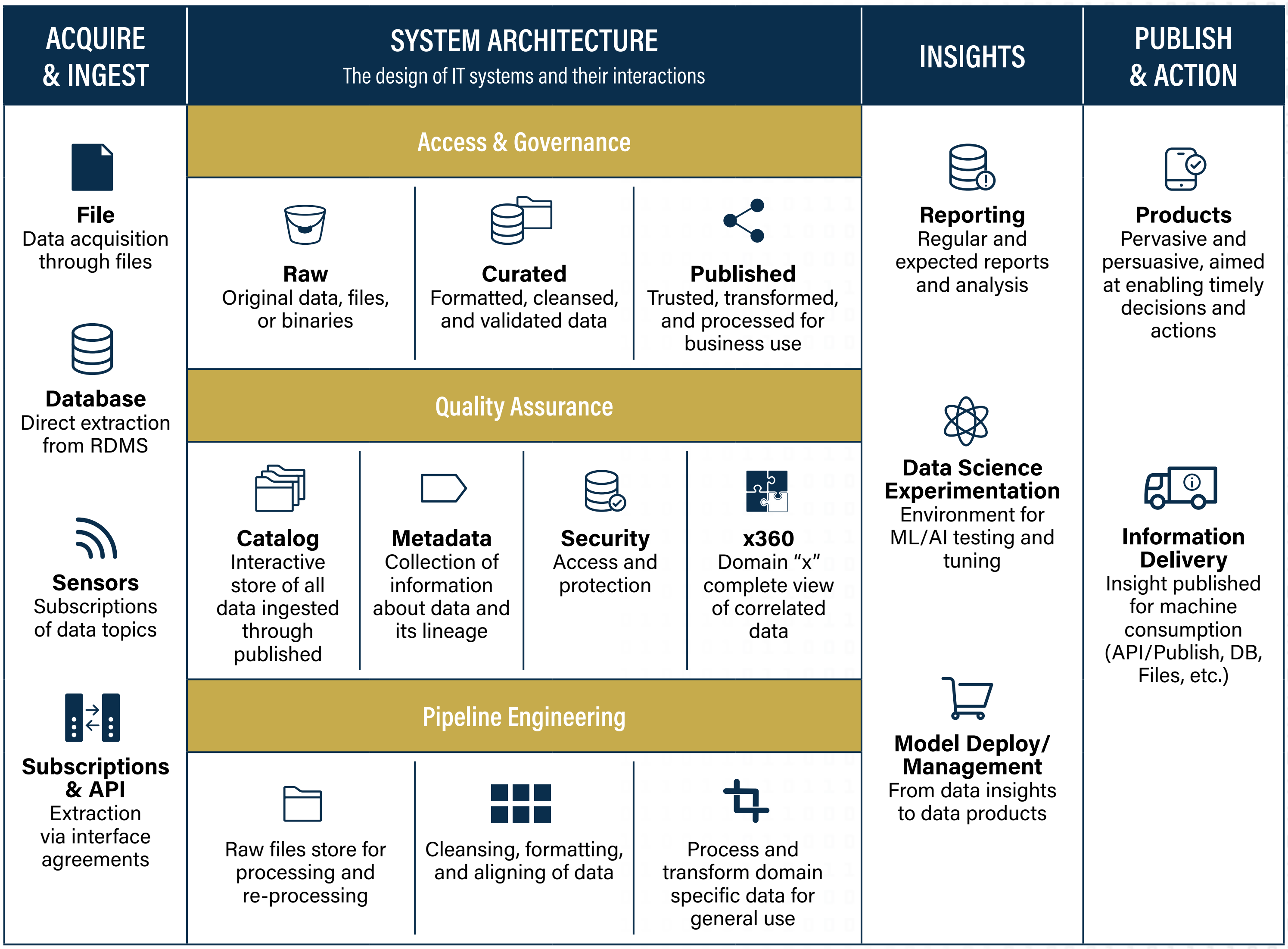
Data engineering for AI projects should remove complexity, not increase it.

Data engineering is an activity that requires collaboration between many stakeholders across groups, including IT, analytics and the business. Mapping out how data moves through an organization is akin to designing the customer journey. It can be done carelessly, resulting in gaps and bottlenecks, or it can be done intentionally based on the final desired outcome.

At the core of all data activities, data engineers (who often report into IT groups) have the technical skills to plan and develop systems architecture to support the movement of data throughout the organization, as well as the knowledge and skills to inform proper governance over data.

It is crucial for business partners to participate in the design and inform the work of data engineers. The figure below illustrates the data engineering value chain and the steps that need to take place to get data into the hands of business analysts and data scientists.

Figure 6: Data Engineering Value Chain



Source: JP Dolphin, PG&E

Table 2: Data Engineering Key Terms and Definitions

Acquire and Ingest	The first step is to absorb or collect data, whether it is a file or streaming service. The data engineering team will determine how to process the data.
System Architecture	What is the plan for how IT systems interact? Should a standard database structure or vendor platform be used? Thoughtful planning, in this case called system architecture, can help to ensure IT systems are as future proof as possible.
Data Pipeline	The data pipeline takes the raw data and puts it in an accessible format and a location where it can be stored efficiently.
Data Quality	This step includes cataloging the data and metadata, which is information about the data. For example, which meter did this read come from? How accurate is it? Synchronizing the time stamps of different data sets, such as SCADA and AMI data, is another important step for analytics.
Data Access and Governance	Understanding who has access to what data is not only important for cybersecurity concerns but also for determining who has read access vs. write access and who is responsible for improving data quality or creating calculated fields.
Data Lab or Data Lake	This is the environment where analysts and data scientists can play with and manipulate the data. Without this space, they will be doing things on their own local machine (the opposite of a best practice), which does not allow the analysis to be scaled or shared across teams.
Tools and Databases	The last step in the chain is the development of any systems, tools or databases—or maybe a dashboard—for sharing and publishing the data.

Source: Voices of Experience|AMI 2019

A WORD ABOUT THE DATA PIPELINE

The data pipeline is a crucial element of the production process. In the experimental stage, data scientists typically perform a request to access data and work with it in a closed environment, or sandbox. In the production stage, the process of moving the data from the source to the model needs to be automated. With complex, heavily customized system architectures, IT teams rightfully are hesitant about engaging in this process at a rapid speed, which creates a conflict with AI project principles.

CASE STUDY: CPS ENERGY MOVES DATA THROUGH REST APIS

As with many utilities, CPS Energy's Advanced Analytics team relied on data requests and individual middleware projects to establish ongoing data flows for specific projects. "Middleware governs all of the possibility for AI, and right now it's the biggest bottleneck to speedy innovation," said Dr. Rolando Vega, senior manager advanced analytics at CPS Energy.

The utility turned to the development of a representative state transfer (REST) API to solve this problem. A REST API allows data to be transferred from its source system to a client application through a web-based protocol.

According to Colt Allen, machine learning engineer at CPS Energy, "Through the use of our REST API, we can now securely retrieve and manipulate time-series data from our PI Historian, without human intervention from IT." In the past, CPS Energy would have to request that data from PI be transferred to our enterprise data warehouse. "Since so many of our projects rely heavily on PI data, establishing a direct connection to that system has provided benefits in terms of speed and efficiency," Allen added.

Allen shared some of the specific benefits of the REST API:

- **Easy access.** Open source clients like python can be used to access PI Web API.
- **Reliable.** Data is consumed directly from its source.
- **Faster analytics.** Query response from API is faster than querying from a database
- **Saves storage.** Avoids data duplication.
- **Reduces process execution time.** Eliminates the middleware process to load data into tables.
- **Flexibility.** Provides the ability to query only the necessary attributes and fields, just like SQL.

Data Governance

Do not overlook data governance.

Data governance provides rules and guidelines about how data can be accessed and who can have read vs. write access. Strong data governance will save time and money over the long term by reducing demands on IT staff and protecting the integrity of data and systems.

Many utilities focused early efforts and investments on more basic data integration activities at the engineering level. This included meeting demands for ingestion and storage of data, breaking down physical and cultural data silos and establishing infrastructure. The difficulty in creating access to data locked in legacy systems while building bandwidth and infrastructure to accommodate new data cannot be understated.

Components of Data Governance

- **Assigning access and restrictions over data**
- **Development of quality, preparation, documentation and maintenance policies**
- **Delegating roles and responsibilities to ensure accountability**
- **Data decision processes, including data integrity and legal obligations**
- **Partnering with parties responsible for architecture and system integration**
- **Introduction and management of tools for quality, workflow and cataloging**

ADVICE FOR DATA MANAGEMENT

Start Simple

Do not get buried in an overly ambitious data road map or large complex project. Start with useful data on hand, choosing sources of external and unstructured data as it adds value to models.

Break Down Data Silos

Break down physical and cultural data silos. Much of the data in utilities today is heavily siloed, limiting the value that can be derived from it. Similarly, some utilities report cultural attitudes toward data ownership limit its availability to other teams.

Clarify Roles and Responsibilities

Data engineering and governance are shared roles across IT, analytics teams and the business units. Assign these roles with intention and clarity and create systems for accountability.

Enable Data Access

Compile, centralize and then democratize. A modern data environment is becoming less and less of an optional consideration for utilities. Not having an enterprise-wide data warehouse or data lake will sorely limit the ability to perform data science.

Create Rules

Do not build data swamps. Data lakes that store data specifically for the purpose of data science and AI are crucial, but these easily evolve into data swamps when there are no strict rules, protocols and cataloging.

Invest in Tools

Invest in tools and processes that streamline access to data. As data moves through different environments, it constantly is reorganized and transformed. Minimizing the number of transformations has both efficiency and quality benefits.

CASE STUDY: DTE DEMOCRATIZES DATA AND INCREASES INNOVATION

According to Gary Gauthier, manager of operational technologies at DTE Energy, "Innovation happens at all levels, from very basic to very advanced. Sometimes we can justify the value of an advanced project by performing very basic analysis."

As a result, his team has strived to democratize access to data and analytics tools, but not without having to work through a lot of complexity. "Because we want to provide SMEs with data to discover value, we needed to establish a sound IT environment and data governance process that would ensure safety, security and quality of our data. We also did not want to force employees who have always had access to data through a certain workflow to suddenly adapt to a new one," Gauthier explained.

For DTE, that means allowing data to be duplicated across systems, and supporting a variety of tools and applications to analyze data. On an ongoing basis, this requires a lot of planning around how to synchronize data across systems, and the establishment of rules for data as it moves across the enterprise. "We want to encourage people to use data, so we do what we can to increase access and open doors. It has required an immense cultural shift within IT, but it has paid off so far," said Gauthier.

Additional Resources:
Data Management Body of Knowledge (DMBOK2)
Data Management Maturity Model (DMM)
https://www.scrumalliance.org/
https://www.pmi.org/certifications/types/agile-acp
https://www.apm.org.uk/resources/find-a-resource/agile-project-management/

The Changing Technology Landscape

AI is not just about buying a set of new tools, it is a different way of thinking about technology processes, from purchasing to managing to retiring.

The technology landscape for AI is evolving rapidly. What is available today is much more than what was available yesterday, and drastically less than what will be available tomorrow. In the utility industry, IT investments traditionally have moved at a slower pace, where it is standard protocol to invest in on-premises systems expected to last for up to 20 years or more. The AI tool set today is almost entirely based in the cloud, and it is common for solution life spans to last no more than a couple of years (with lots of updates along the way).

Building the AI Architecture

For some utilities, their AI architecture might just be a slight extension to their existing system architecture. Others might build as much of their architecture as possible in the cloud on platforms like AWS and Microsoft Azure. Regardless, those who have taken steps to work out this challenge are at the forefront of the AI transformation and learning a lot along the way.

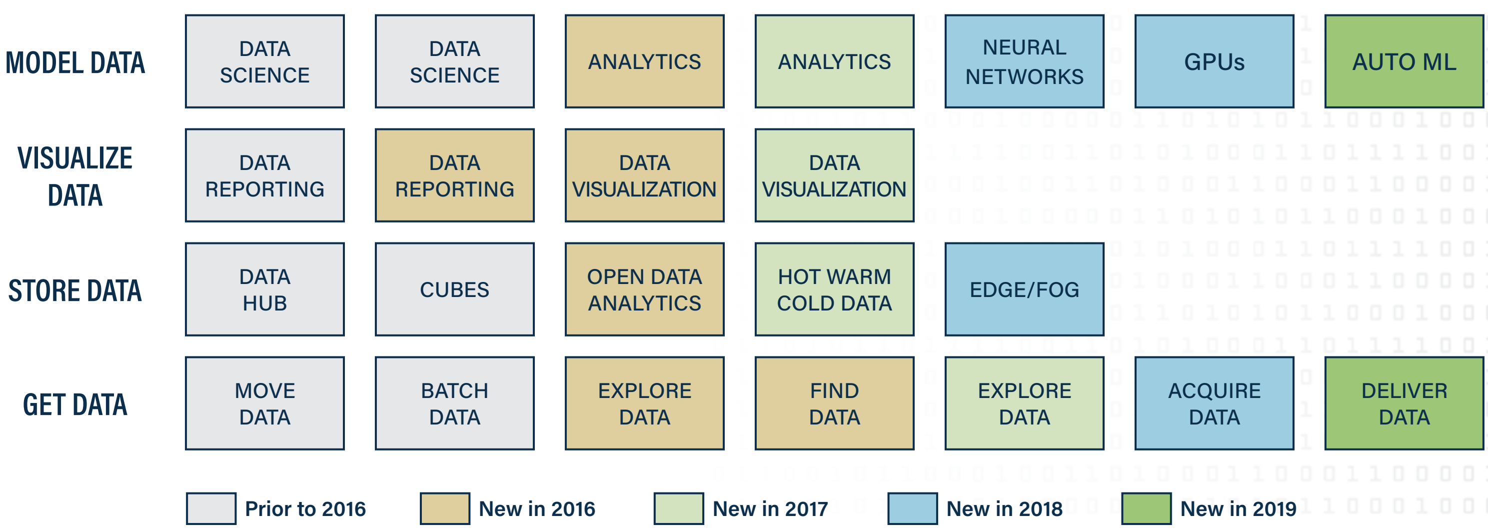
> ***"If you want to incubate a new capability or discipline into your organization...you need to actually incubate it."***
> **– Joe Tellez, Tacoma Public Utilities**

The complexity of the architecture will vary based on the type of AI, the data being used and the AI product. The Process section of this report discussed the challenges in operationalizing AI, of which technology constitutes a large portion.

The figure below illustrates how one utility built its AI architecture as of late 2019—each year it is integrating new tools to help its teams deliver value.

Figure 7: An AI Toolset

WHAT TOOLS DO DATA SCIENTISTS & AI ENGINEERS USE?



Source: UAI

So, what is required to establish a robust architecture for AI? According to utilities like Duke Energy, CPS Energy, Salt River Project, and Tacoma Public Utilities, whose teams are aggressively building out their technology capabilities, there is no finite answer to that question today—but it's important to start working through the process as soon as possible.

CASE STUDY: THE KEY TO AI TECHNOLOGY ADOPTION FOR TACOMA PUBLIC UTILITIES

The cloud is critical for utilities to keep up with the pace of technology innovation and to de-risk the process of designing and implementing an AI architecture.

According to Andrew Braeger, Solutions Architect at Tacoma Public Utilities, "The cloud provides you with a lot of different services that help you to innovate quickly and more easily than if you build all of this into your existing on-premises systems infrastructure. CIS, SCADA, Historian systems are evolving, but not at the same speed as data analytics and AI technology."

"If you go with a large tool like AWS, there is a whole ecosystem of tools that already integrate into that platform, managed blockchain, IoT, ML, etc.," he says. "It takes a lot of load off our internal IT teams to not have to figure out how to configure every new product we want to use."

Of course, there are challenges to moving significant portions of your architecture to the cloud—accounting procedures, cultural resistance, and in some cases security—but as the benefits become better understood, utilities are simultaneously finding ways around these.

Bottom line: there's a robust toolset for the full AI lifecycle, if you are able and willing to adopt it.

Purchasing AI Technology

Regardless of the specific approach to acquiring technologies, utilities need to consider how what they buy will impact the development of AI skills and capabilities.

There are a lot of definitions and offerings available for AI platforms and applications, and different ways to purchase them. We identified the following three product and purchasing categories, but in reality, most utilities are adopting solutions in all three categories: modular data science tools, custom platforms built in the cloud and outsource application development.

Modular Data Science Tools

There are numerous plug-and-play tools on the market that empower the citizen data scientist to develop models with drag-and-drop capabilities. What is great about these tools is they can open up the world of data science and its benefits to a wider group of people and start building buy-in with small projects. At the same time, these tools are not flexible for trained data scientists who are used to programming models in tools like R and Python. As well, some utilities pointed out they do not trust the stability of companies offering these solutions, many being start-ups and subject to bankruptcy or acquisition.

Custom Platforms Built in the Cloud

More utilities are using modern environments such as AWS and Microsoft Azure to build out their architecture and tool set, or at least components of it. This approach has a lot of benefits in terms of being able to move quickly and adopt and retire components alongside technology improvements, as opposed to being stuck with a product until it depreciates. However, utilities need to approach the cloud strategically, and processes and procedures need to evolve alongside adoption. As well, the process of data retrieval and integration with legacy systems will pose technical challenges.

Outsource Application Development

Vendors and services providers have existing robust analytics and AI capabilities. Utilities can leverage this expertise to get projects off the ground and rapidly deploy applications and products. However, some drawbacks exist, including the delivery of black box solutions and less flexibility to improve upon or pivot use cases. As well, if a utility is too dependent on outsiders to develop their AI capabilities, they lose the ability to develop their internal expertise, which can have long-term consequences.

> *"A best solution always is the best solution at the time. It doesn't mean that you're going to buy one thing and that it will last forever."*
> **– Darlene Berghammer, Contract Administrator, Austin Energy**

ADVICE FOR PURCHASING AI TOOLS

Do Not Wait for the Perfect Solution

The best solution always is the best solution at the time and does not mean it is going to last forever. Think critically about when and how to buy AI solutions. Product life cycles, fees and contract terms affect purchasing decisions.

Customize Your Purchase

Do not just go buy what other people buy. Consider a custom solution that will work for your employees, architecture and cloud strategy. Whatever is purchased should not only address a staff person's needs but also a business program, a performance need and even the customer journey.

Consider the Cloud

The cloud is going to be essential in this environment. Start working on breaking down barriers to adoption, such as accounting, security, cultural resistance, education and a lack of skill sets. It generally is understood the cloud is very secure, but utilities still need to train staff on secure cloud processes, which vary greatly from standard on-premises procedures. Also, work with accounting teams and across business groups to help characterize and justify subscription costs for cloud-based solutions.

Educate External Stakeholders

The degree to which accounting procedures are regulated varies from region to region. Collaborate with local peers to share knowledge and insights, and work together to educate regulators, board members and other decision makers on the value and customer benefits of technology that enables AI.

Appendix A

Voices of Experience | Artificial Intelligence Participants

This list includes companies that participated in this project by attending a workshop or participating in an interview. The DOE\UAI leadership team would like to thank everyone who supported this initiative, especially those who shared their experience with the readers of this guide.

American Electric Power
Ameren
American Transmission Company
Austin Energy
Avongrid, Inc.
Black Hills Corporation
Chelan County PUD
Colorado Springs Utilities
CPS Energy
Dominion Energy
DTE Energy
Duke Energy
EPRI
Evergy
Exelon Utilities
FirstEnergy
Grant County Public Utility District
Green Mountain Power
JEA
Los Angeles Department of Water and Power
Midcontinent Independent System Operator
National Renewable Energy Laboratory
New Jersey Natural Gas
NorthWestern Energy
Oncor Electric Delivery
Ontario Power Generation
Pepco Holdings Inc.
PPL Electric Utilties
Rappahannock Electric Cooperative
Salt River Project
Southern Company
Tacoma Public Utilities
Tennessee Valley Authority
Tucson Electric Power

Appendix B

Glossary of Terms

This glossary includes terms appearing in the *Voices of Intelligence|Artificial Intelligence* guide. It is not intended to be a comprehensive glossary of technical terms related to artificial intelligence or data science. Where possible, we used a primary source to define the term, but recognize there may be other definitions used in data science communities.

Advanced Metering Infrastructure (AMI): AMI is an integrated system of smart meters, communication networks, and data management systems that enable automated two-way communication between a smart meter and a utility.

Agile: Agile software development refers to a group of software development methodologies based on iterative development, where requirements and solutions evolve through collaboration between self-organizing cross-functional teams.

Algorithm: In simple terms, an algorithm is a strict and logical sequence of explicit, step-by-step instructions to solve any problem. Algorithms are elementary building blocks that make up machine learning and artificial intelligence. Very often algorithm operates as a sequence of simple if → then statements or a sequence of the more complex construct of numbers and commands. The goal of an algorithm is to teach AI, neural network, or other machines on how to figure things out on their own.

Application Programming Interface (API): An API is a set of commands, functions, protocols, and objects that programmers can use to create software or interact with an external system. It provides developers with standard commands for performing common operations so they do not have to write the code from scratch.

Artificial Intelligence (AI): The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

Amazon Web Services (AWS) Cloud: A comprehensive, evolving **cloud** computing platform provided by Amazon that includes a mixture of infrastructure as a service (IaaS), platform as a service (PaaS) and packaged software as a service (SaaS) offerings.

Big Data: Large amounts of structured and unstructured data too complex to be handled by standard data-processing software.

Bots: A device or piece of software that can execute commands, reply to messages, or perform routine tasks, as online searches, either automatically or with minimal human intervention (often used in combination), such as intelligent infobots or shopping **bots** that help consumers find the best prices.

Business Intelligence: Refers to technologies, applications, and practices for the collection, integration, analysis, and presentation of business information. The purpose of **Business Intelligence** is to support better business decision-making.

Classification: In machine learning and statistics, classification is a supervised learning algorithm technique that allows machines to assign categories to data points (categorize data into a given number of classes). Classification (decision trees and neural network classifiers) can be used for text classification in marketing.

Clustering: A machine learning technique that involves the grouping of data points. Clustering is a common technique for statistical data analysis used in many fields including machine leanring. Clustering is used with applications like customer segmentation, fast search, and visualization.

Cognitive Computing: The use of computerized models to simulate the human thought process in complex situations where the answers may be ambiguous and uncertain.

Computer Vision: A field of AI used to obtain information from images. Computer vision uses AI technologies to solve complex problems such as object detection and image processing from image files (JPEGs) or camera feeds.

Confidence Interval: An interval about a point estimate that quantifies the statistical uncertainty in the true value being estimated due to variability.

Continuous Learning System (CLS): Systems that are inherently capable of learning from real-world data and are able to update themselves automatically over time while in public use.

Dashboard: A type of user interface that often provides at-a-glance views of key performance indicators (KPIs) relevant to a particular objective or business process. In other usage, "dashboard" is another name for "progress report" or "report."

Data Governance (DG): The process of managing the availability, usability, integrity and security of the data in enterprise systems, based on internal data standards and policies that also control data usage. Effective data governance ensures data is consistent and trustworthy and is not misused.

Data Lake: This the environment where analysts and data scientists can play with and manipulate the data. Without this space, they will be doing things on their own local machine (the opposite of best practices) because it doesn't allow the anlaysis to be scaled or shared across teams.

Data Mining: The process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

Data Pipeline: A set of operations connected in series, where the raw data is been processed into a format that's accessible by the down stream operations, and in a location where it can be stored efficiently.

Decision Tree: A decision support tool with a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. Decision tree is the most powerful and popular tool for classification and prediction.

Deep Learning: A subset of machine learning where multiple layers of neural networks are constructed to progressively extract information and achieve higher accuracy of learning result.

DevOps: Combines the words "development" and "operations." It encompasses developers and IT operations personnel within an organization.The goal of DevOps integration is to improve collaboration between development and operations teams.

Expert System: A form of AI that attempts to replicate a human's expertise in an area, such as medical diagnosis. It combines a knowledge base with a set of hand-coded rules for applying that knowledge. Machine learning techniques are increasingly replacing hand coding.

Feedback Loop: The process by which an AI model's predicted outputs are reused to train new versions of the model.

Gantt Chart: A Gantt chart, commonly used in project management, is one of the most popular and useful ways of showing activities (tasks or events) displayed against time. On the left of the chart is a list of the activities and along the top is a suitable time scale. Each activity is represented by a bar; the position and length of the bar reflects the start date, duration and end date of the activity.

Image Recognition: The ability of a system or software to identify objects, people, places, and actions in images. It uses machine vision technologies with artificial intelligence and trained algorithms to extract features and identify classes.

Kanban Board: An agile project management tool designed to help visualize work, limit work-in-progress, and maximize efficiency (or flow). Kanban boards use cards, columns, and continuous improvement to help technology and service teams commit to the right amount of work, and get it done.

Linear Framework: Writing all the steps in an action in a **linear** form. This approach is also called **linear** scripting. In this approach, the entire test case flow will be written in QTP in a **linear** fashion.

Machine Learning (Ml): The ability of a machine or algorithm to progressively improve upon a set task.

Metadata: Metadata describes other data. It provides information about a certain item's content.

Middleware: Middleware has two separate but related meanings. One is software that enables two separate programs to interact with each other. Another is a software layer inside a single application that allows different aspects of the program to work together.

Minimum Viable Product (MVP) Approach: A version of a **product** with just enough features to satisfy early customers and provide feedback for future **product** development.

Narrow Intelligence: AI that is programmed to perform a single task, whether that is checking the weather, being able to play chess, or analyzing raw data to write journalistic reports.

Natural Language Processing (NLP): A branch of artificial intelligence that deals with the interaction between computers and humans using the natural language. The ultimate objective of **NLP** is to read, decipher, understand, and make sense of the human languages in a manner that is valuable. Most **NLP** techniques rely on machine learning to derive meaning from human languages.

Neural Networks: A set of algorithms, modeled loosely after the human brain designed to recognize patterns. Neural networks interpret sensory data through a kind of machine perception, labeling or clustering raw input.

Optical Character Recognition (OCR): Optical character recognition or optical character reader is the conversion of images of typed, handwritten, or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene photo, or subtitle text superimposed on an image.

Pattern Recognition: The ability to detect arrangements of characteristics or data that yield information about a given system or data set. Pattern recognition is essential to many overlapping areas of IT, including big data analytics, biometric identification, security, and AI.

Personas: Fictional characters, which you create based upon your research in order to represent the different user types that might use your service, product, site, or brand in a similar way.

Phasor Measurement Unit (PMU): A device used to estimate the magnitude and phase angle of an electrical phasor quantity (such as voltage or current) in the electricity grid, using a common time source for synchronization. Time synchronization is usually provided by GPS and allows synchronized real-time measurements of multiple remote points on the grid.

PI: **PI** stands for Process Information. This application can efficiently record **data** from process control systems (ex. DCS, PLC) into a compressed time series database. Utilities use **PI** systems to record, analyze, and monitor real-time information.

Proof of Concept (POC): A realization of a certain method or idea to demonstrate its feasibility or a demonstration in principle with the aim of verifying that some concept or theory has practical potential.

Representative State Transfer (REST): An architectural style for providing standards between computer systems on the web, making it easier for systems to communicate with each other. REST-compliant systems, often called RESTful systems, are characterized by how they are stateless and separate the concerns of client and server.

Python: A high-level programming language designed to be easy to read and simple to implement. It is open source, which means it is free to use, even for commercial applications. Python can run on Mac, Windows, and Unix systems and has also been ported to Java and .NET virtual machines.

R: A widely used statistical programming language used in academia and industry. R works well with data, making it a great language for anyone interested in data analysis, data visualization, and data science.

Regression: A statistical approach that estimates the relationships among variables and predicts future outcomes or items in a continuous data set by solving for the pattern of past inputs, such as linear regression in statistics. Regression is foundational to machine learning and artificial intelligence.

Reinforcement Learning: A type of machine learning algorithms which allows software agents and machines to automatically determine the ideal behavior within a specific context, to maximize its performance. Reinforcement algorithms are not given explicit training labels or exact models; instead, they are forced to learn by correcting its actions to get closer to the observed reward with each iteration.

Robotic Proc Auto/Robotic process automation (RPA): Uses software with AI and ML capabilities to perform repetitive tasks once completed by humans.

Scrum: A widely used and popular agile framework that helps team to work together by selforginizing and learning from experience. The term agile describes a specific set of foundational principles and values for organizing and managing complex work.

Segmentation: The process of separating data into distinct groups. This is a core activity in most business problems. A well-defined **segment** is one in which the members of the **segment** are similar to each other and also are different from members of other **segments**.

Structured Data: Clearly defined data with easily searchable patterns.

Supervised Learning: Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output. Y = f(X) The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. The majority of practical machine learning uses supervised learning.

Telemetry Models: Model-driven **telemetry** is a new approach for network monitoring in which data is streamed from network devices continuously using a push **model** and provides near real-time access to operational statistics.

Textual Analysis: A methodology that involves understanding language, symbols, and/or pictures present in texts to gain information regarding how people make sense of and communicate life and life experiences. Visual, written, or spoken messages provide cues to ways through which communication may be understood.

Unstructured Data: Information that either does not have a predefined data model or is not organized in a predefined manner.

Unsupervised Learning: The training of an AI algorithm using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Unsupervised learning algorithms can perform more complex processing tasks than supervised learning systems.

Waterfall Model: A software development process introduced in 1970 that emphasizes that a logical progression of steps be taken throughout the software development life cycle (SDLC), much like the cascading steps down an incremental waterfall.

Voices of Experience | Artifical Intelligence

Go to SmartGrid.gov/Voices for more information on DOE's Voices of Experience initiative.

NREL Transforming ENERGY

UAI®

Advanced Grid Research
OFFICE OF ELECTRICITY
US DEPARTMENT OF ENERGY