Introduction to CPMAI

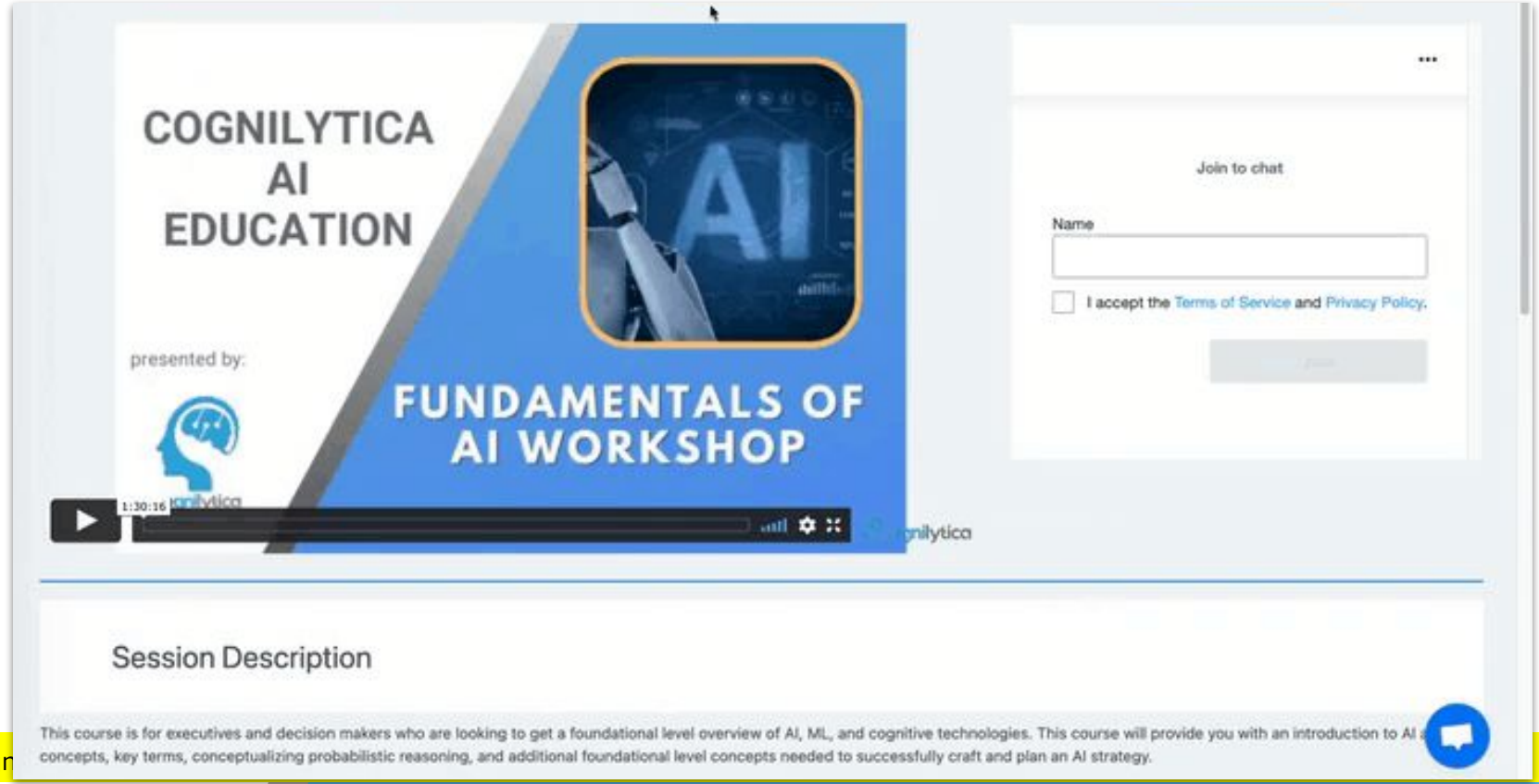
Best Practices and Methodology for doing AI right

Presented by Cognilytica Analysts:

Kathleen Walch
Ronald Schmelzer

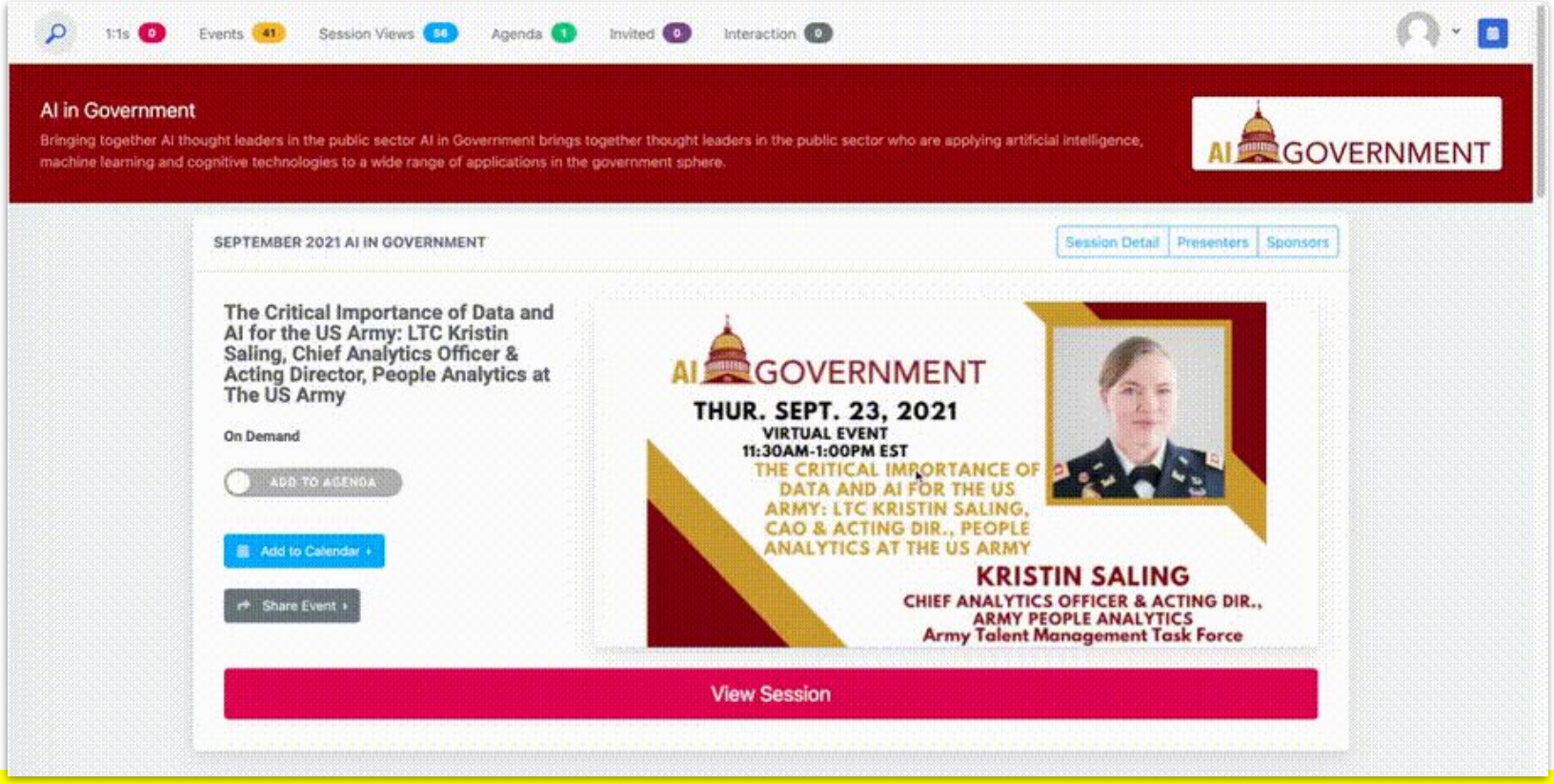
Cartoon illustrations by Timo Elliott (timoelliott.com)

How to Ask a Question



- Sign

How to Download Slides



About Cognilytica

- Cognilytica is an ***AI-focused analyst and advisory firm***.
- Market research, advisory & guidance on ***AI, ML, & Cognitive Technology***
- Provides ***role-specific education*** on AI, ML, and emerging technology
- Focused on ***enterprise and public sector adoption*** of AI, ML, and Cognitive Technology
- Kathleen Walch and Ron Schmelzer are ***Principal Analysts and Managing Partners*** of Cognilytica
- Produce the popular ***AI Today podcast*** and ***AI communities***
- Contributing writers to ***Forbes***

80%+ of AI Projects Fail

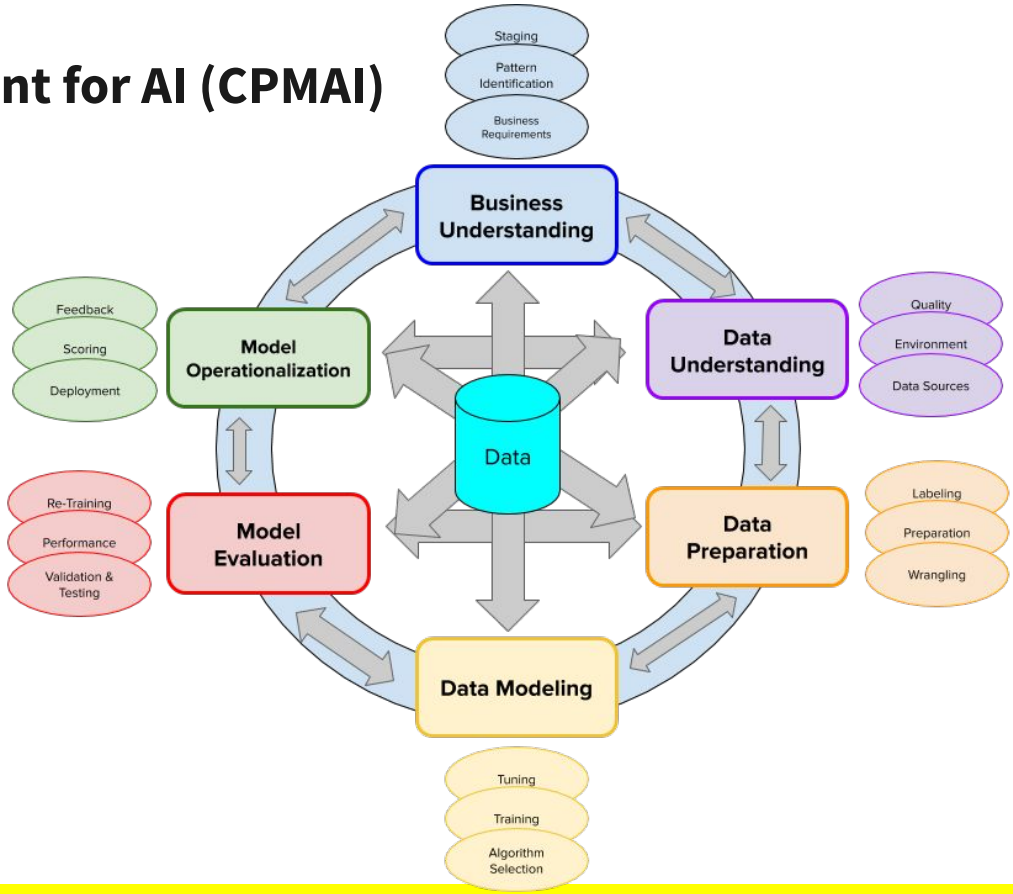
- According to TechRepublic, 85% of AI projects eventually fail to bring their intended results to the business.
- According to Gartner, 85% of Machine Learning (ML) projects fail.
- According to IDC, most organization polled reported failures among their AI projects, with a quarter of them reporting up to a 50% failure rate
- According to a report from the Project Management Institute, only about 70% of all projects completed in 2017 met original goals and business intent.

It doesn't have to be this way - follow a proven, best practices methodology for doing AI right

Don’t Be Another AI Failure statistic

- You need to adopt a proven, successful data focused project management methodology
  - Think of AI projects as you would Big Data projects, not Enterprise software development projects
- You need to move beyond methodologies such as the Scientific Method or Agile to run data projects
- You need to develop an intelligent data forward mindset

Cognitive Project Management for AI (CPMAI)



Reason
Why AI Projects Fail

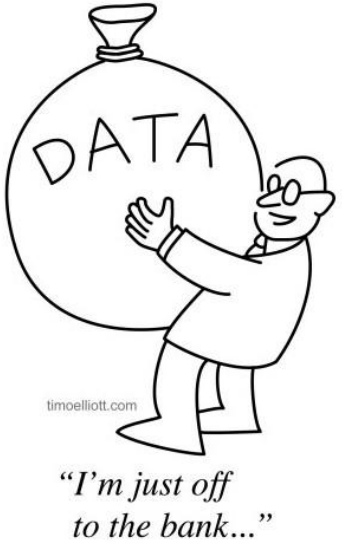
AI projects are NOT like traditional software development projects

AI is Not About Application Development

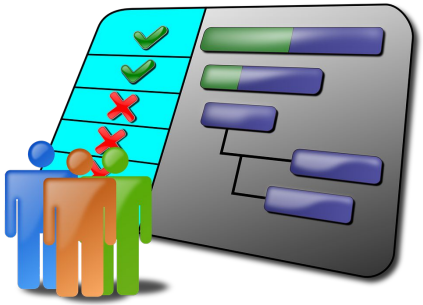
AI is not about application development! It's about data

- The code is a small part of making AI work. ***And not even the most important part***

If you run your AI Projects like you run your Application Development Projects you're going to find out the hard way that it won't work.



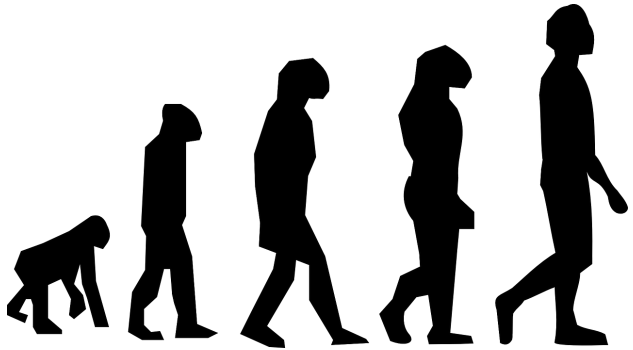
AI Projects as Data Management and Processing Projects



- *Use Data-Centric Methodologies*
- *Use Data-Centric Technologies*
- *Build Data-Centric Talent*
- *Still need Business-Centric ROI*
- *Leverage Existing Systems and Agile Methodologies*
- *Find ways to continuously update skills and methods*

Agile is NOT Good Enough (By Itself)

- *What does an "iteration" mean in a specific AI Project?*
- *The functionality stays the same while the data changes…*
- *Data Prep and Understanding go … where?*
- *Is this Cognitive Tech for internal use or external use?*
- *What to do about model training and re-training*
- *How do you test / manage highly individualized data sets?*

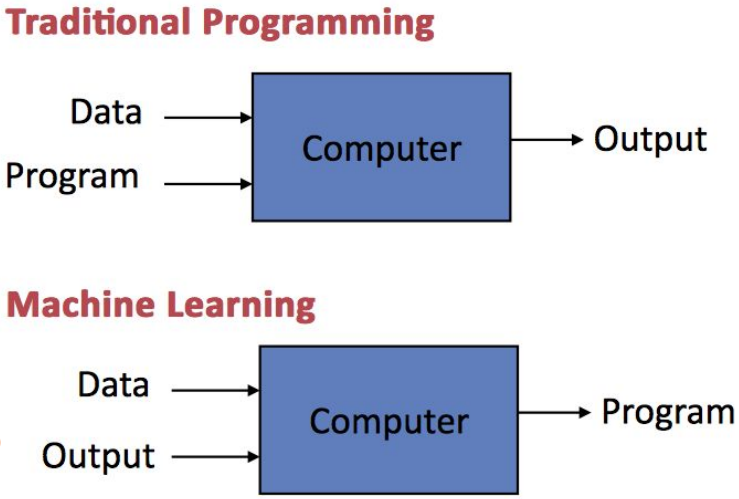


We need to evolve Agile Methodologies from a Data-First Perspective

AI and Machine Learning require a Data-First Approach

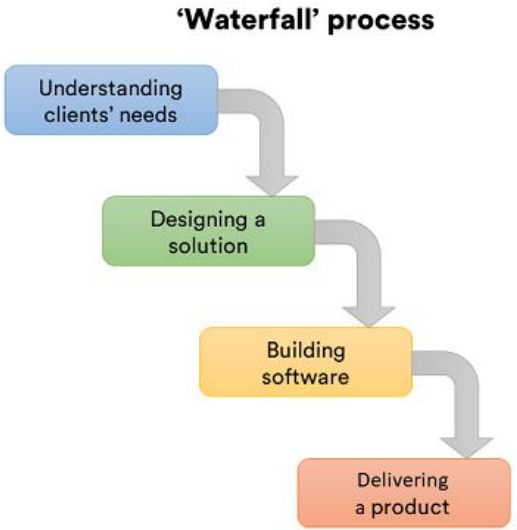
"The complexity in traditional computer programming is in the code (programs that people write). In machine learning, algorithms (programs) are in principle simple and the complexity (structure) is in the data. Is there a way that we can automatically learn that structure? That is what is at the heart of machine learning."

-- Andrew Ng, AI luminary



Project Management Approaches to AI are Challenged

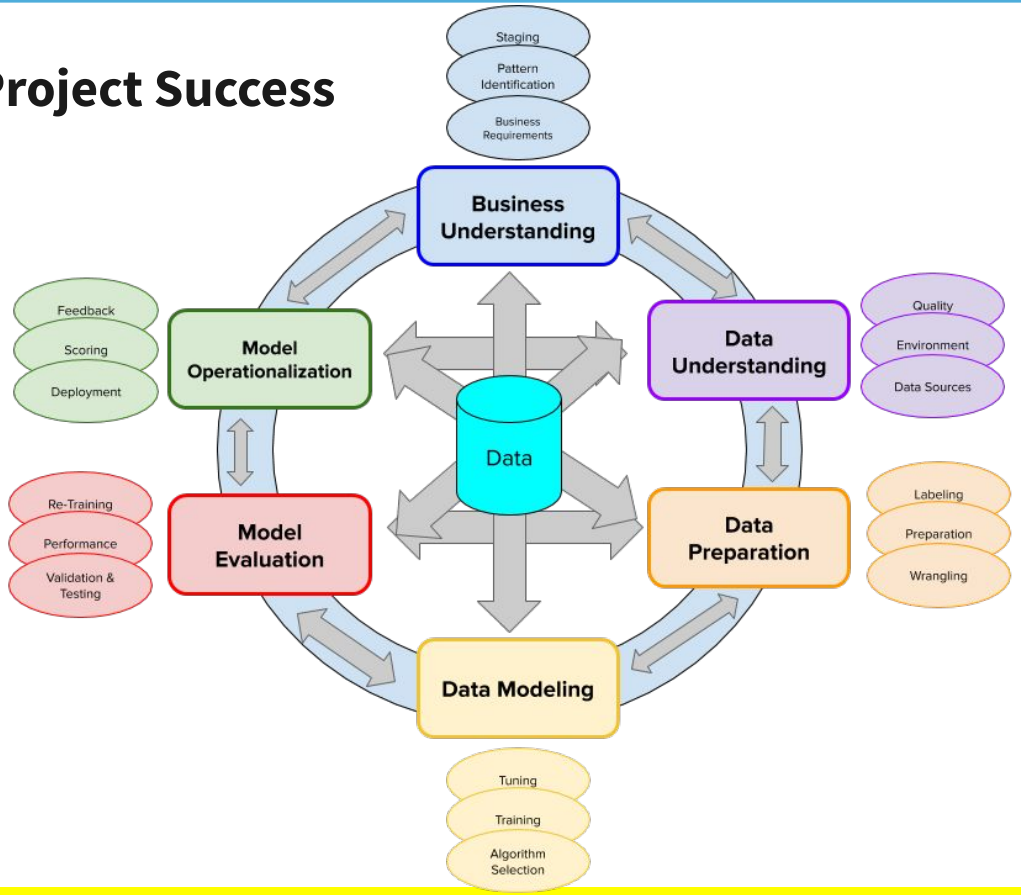
- ***"Waterfall" approaches definitely don't work***
  - Rigid and restrictive
  - Too lengthy
  - High risk of building the wrong model.
  - Requires teams to predict major obstacles
  - Unable to quickly respond to changing technology, requirements, needs, bias problems
- ***Agile approaches need some modifications***
  - Continuously learn and iterate
  - Opportunities to re-prioritize where necessary
  - Deliver value faster and more efficiently
  - ***Challenge: Simultaneous continuous model tweaking and deployment***



Update Agile Approaches with Data Science / Data Management Methodologies

CPMAI: Combining the Best for AI Project Success

- *The data-centricity of CRISP-DM*
- *Specific details for AI / cognitive technology projects*
- *Leverages the Seven Patterns to provide guidance*
- *All steps are iterative with each other*



Reason
Why AI Projects Fail

The ROI is not justified by the project

The AI Go/No-Go Decision

Is your machine learning project a go or no-go?

Use this chart to gauge the feasibility of your AI project from a business, data and implementation standpoint.

Business feasibility	Data feasibility	Implementation feasibility
Is there a clear problem definition?	Do you have the required data that measures what you care about?	Do you have the required technology and skills?
Is the organization willing to invest and change?	Is there a sufficient quantity of data needed to train systems and do you have access to that data?	Can you execute the model as required in a timely manner?
Is there sufficient ROI or impact?	Is the data of sufficient quality?	Does it make sense to use the model where you plan to use it?

In order for an AI project to actually go forward, you need three things in alignment: the business feasibility, the data feasibility, and the technology / execution feasibility

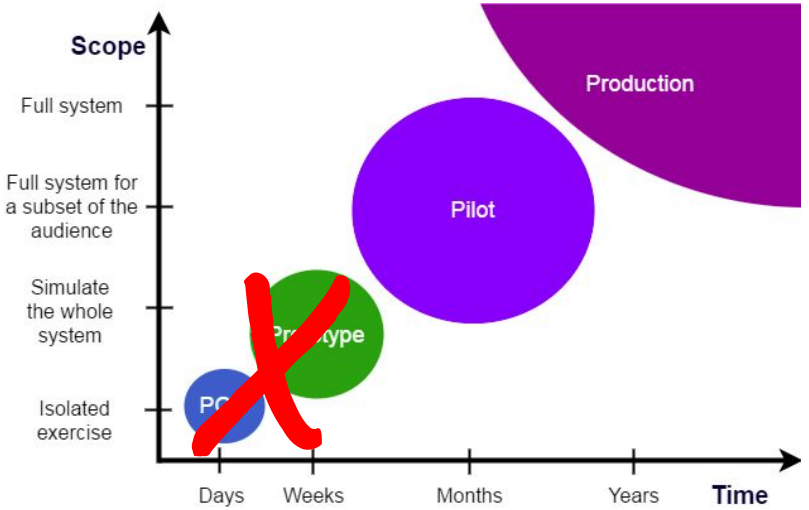
Reason
Why AI Projects Fail

The time between pilot to full production is too long. What happened to iteration?!

(and why are you doing Proof of Concepts?)

ML Project staging: The PoC vs. the Pilot - Go MVP!

- ***IMPORTANT: Distinguish between a Proof-of-Concept (PoC) and Pilot!***
- A Proof-of-Concept (PoC) is meant to illustrate a concept in an almost-but-not-really-real environment that never gets deployed into real-world production
- A Pilot is a real-world project that uses emerging technologies in a protected, safe environment but used in real-world production
- AI that's confined to a lab / experiment setting ***does NOT provide return and often fails!*** (ahem, autonomous vehicles much?)
- Much of the effort is upfront: algorithm selection, training set creation, model testing, validation, hyperparameter tuning… SO WHY POC or "Prototype"??!!!



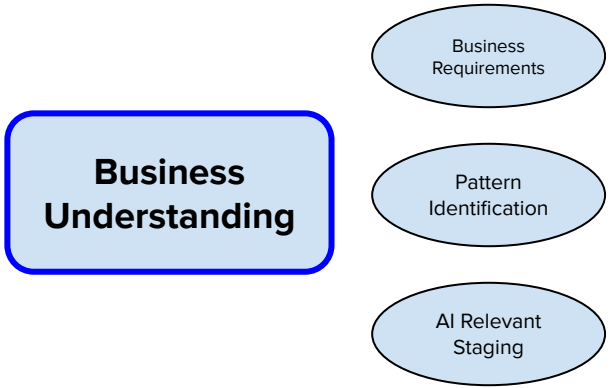
Avoid PoCs and go for Real-World Pilots!

Lean Agile Speak: Minimum Viable Product (MVP)

CPMAI Phase I: Business Understanding

Answer the fundamental questions and perform tasks:

- What problem are we attempting to solve?
- Should we solve this problem with AI / Cognitive Technology?
- What portions of the project require / do not require AI?
- What AI patterns are we implementing?
- What are the criteria for project success?
- What requirements are needed to complete the project?
- What other important considerations are there for the AI project?
- What skills are necessary for successful project completion?

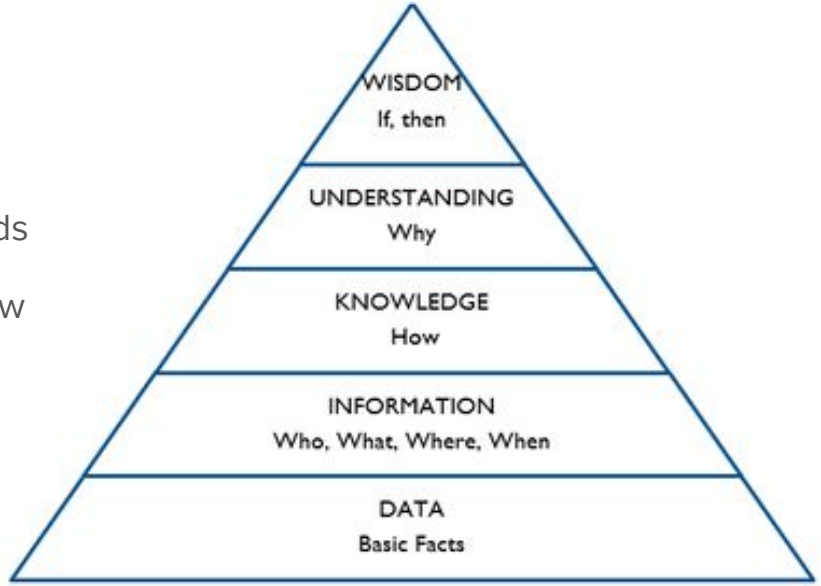


Reason
Why AI Projects Fail

Data Quantity Issues (and a lack of data understanding)

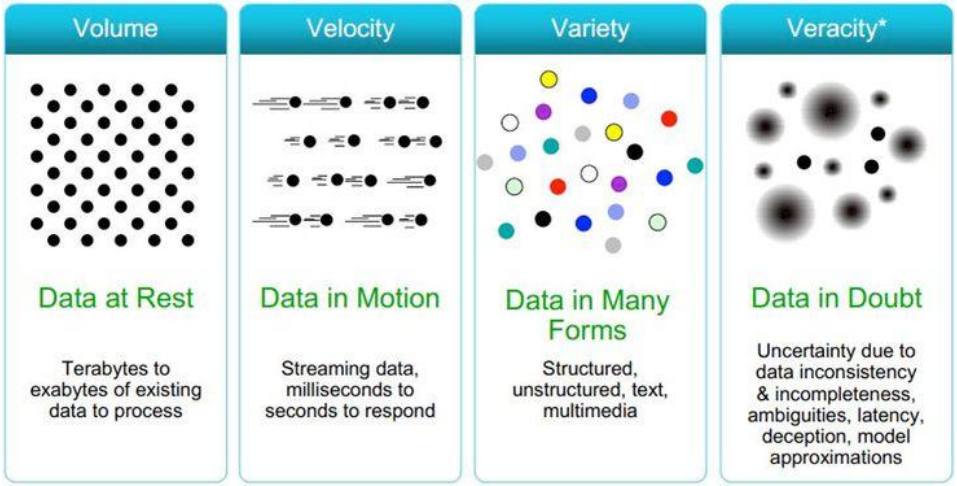
The DIKUW Pyramid

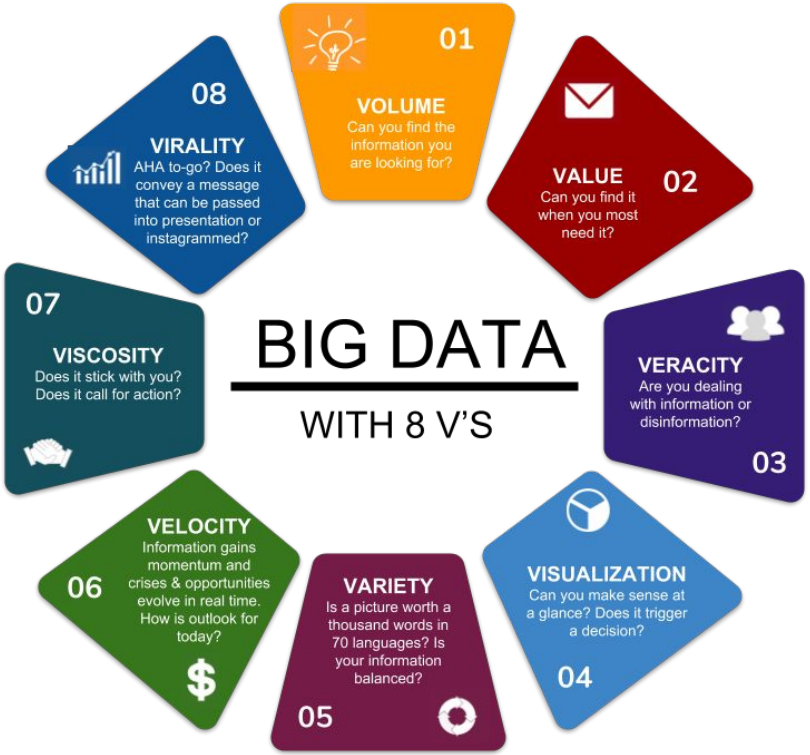
- Remember ***WHY*** we're doing this.
- Addressing higher-value information & operational needs
- Enabling machines to do more of what we are doing now
  - Taking the ***machine out of the human***
- Evolutions from Industry 1.0 - 4.0
  - ***Economic disruption or economic enablement?***



The DIKUW Pyramid provides hints to understanding how to plan, manage, and operationalize AI & ML Projects

What Big Data Taught Us

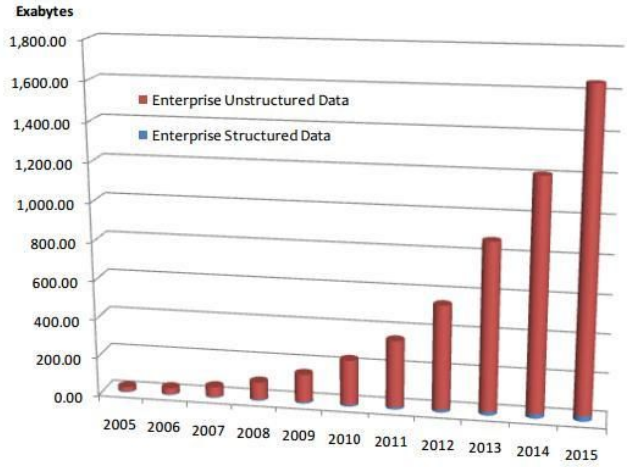




The big shift over the past decades is from collecting data to understanding it to turning it into knowledge, conclusions, and actions. Moving our way up the DIKUW Pyramid!

The Untapped Value of Unstructured Data

- Over ***80% of data in the enterprise is unstructured***
  - Email, documents, images, video, messages, voice mail
- Another 10% is semi-structured
  - Log files, JSON API messages, more
- So much value in that unstructured data, but you can't just query it or use standard analytics tools

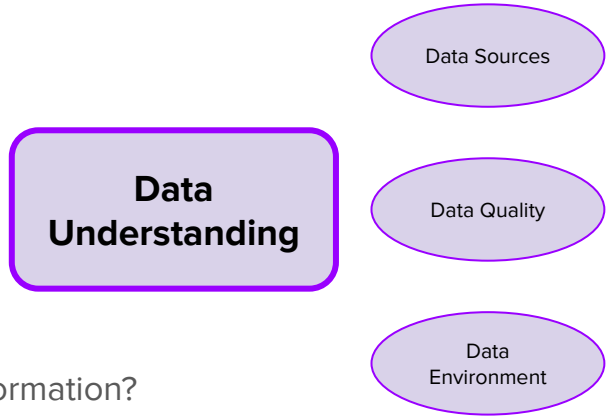


We want to do more to process and extract more value from Unstructured Big Data

CPMAI Phase II: Data Understanding

Answer the fundamental questions and perform tasks:

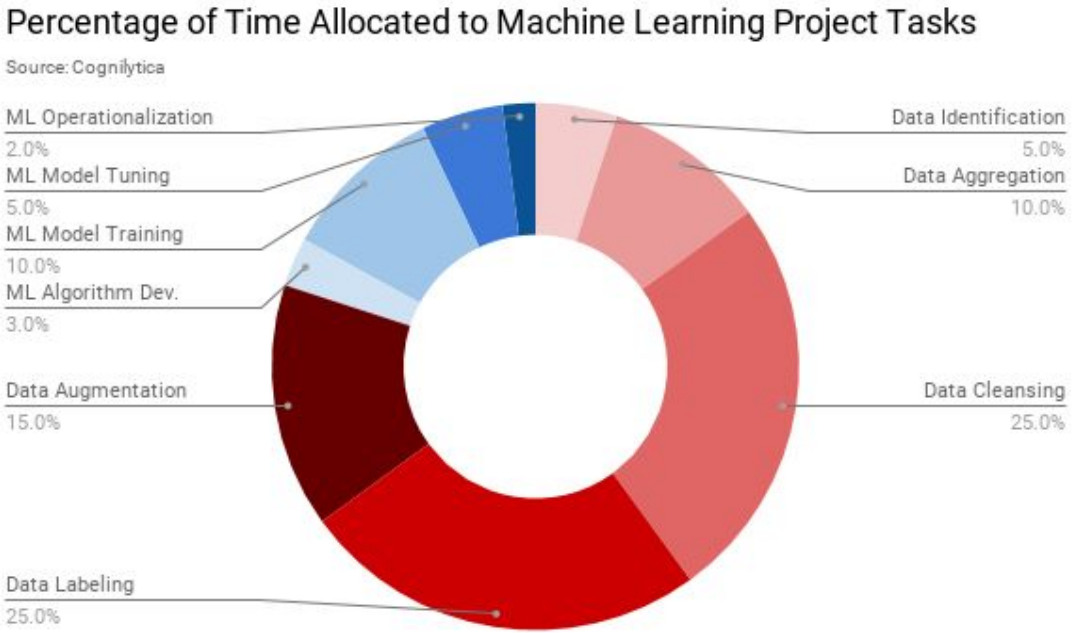
- What data are necessary to achieve the objectives?
- What is the quantity and quality of our data?
- What internal and external data is necessary?
- What requirements to augment existing data?
- What are the requirements for ongoing data gathering and preparation?
- What are the requirements for technology for data manipulation and transformation?
- What other important data-relevant considerations are there for the AI project?



Reason
Why AI Projects Fail

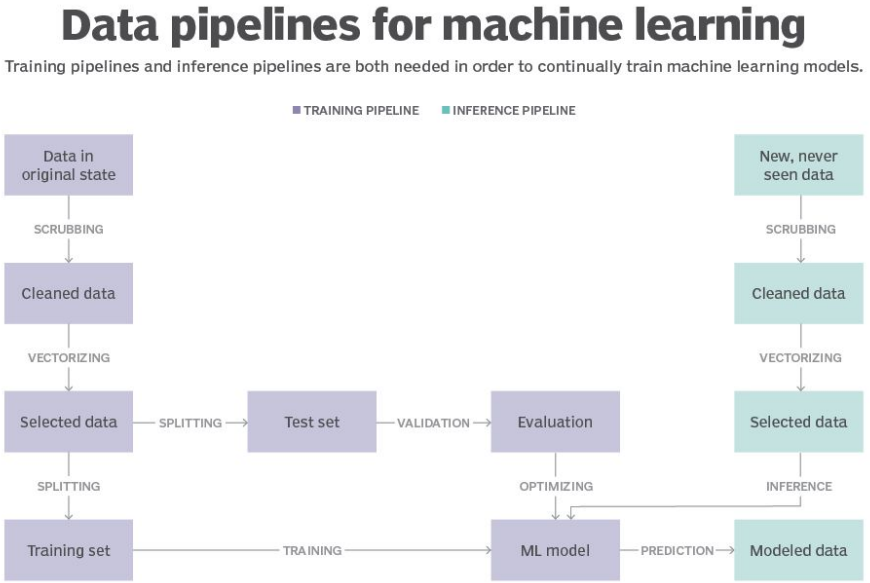
Data Quality Issues

80% of AI Projects are Data Engineering



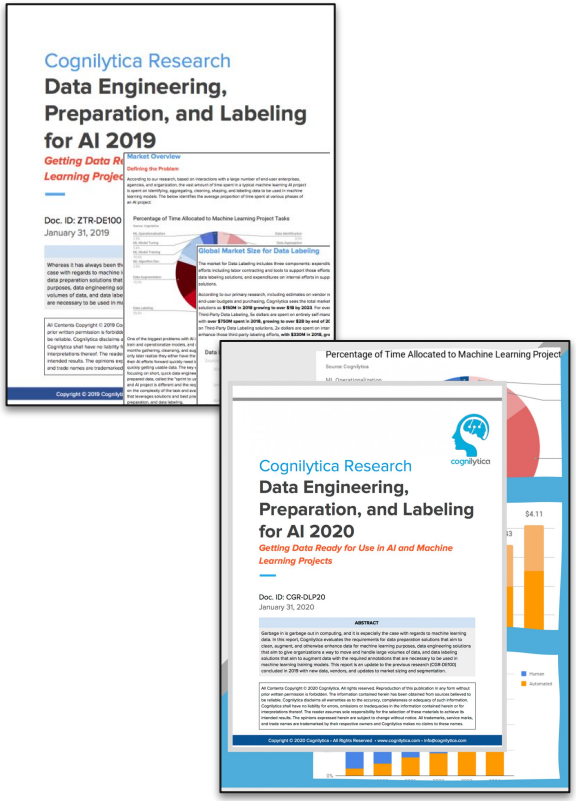
AI-Specific Needs of Data Preparation

- ***Data Acquisition / Ingest / Capture***
  - ETL
  - Cloud-based data
- ***Merging***
  - Combining data sources
- ***Cleaning***
  - Deduping, removing extraneous, bad data
- ***Enhancing***
  - Adding necessary data for models
- ***Filtering***
  - Eliminating bias
- ***Feature Engineering***
  - Assisting with enhancement (see future on multiplying classification data sets)
- ***Retraining Pipelines***
  - Creation of pipelines to deal with model iteration



Data Labeling: The Achilles Heel of AI

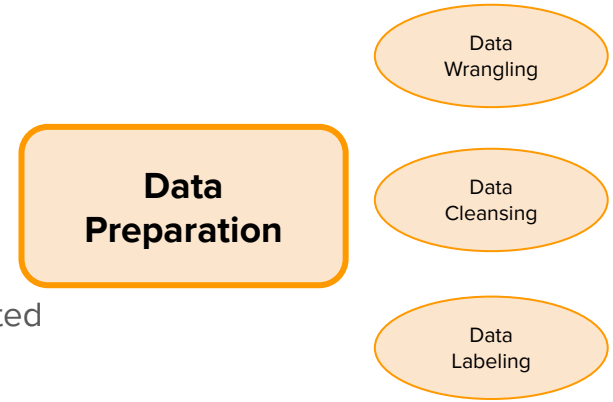
- In order for Supervised Learning approaches to work, they must be fed clean, ***well-labeled data that the system can use to learn from example***
- But how do you get Labeled Data?
- ***Do it yourself***
- ***Find a source of already labeled data***
- ***Get your Users to Do it***
- ***Hire a Contractor Workforce***
- ***Contract with Third Party Data Labeling Firms***



CPMAI Phase III: Data Preparation

Answer the fundamental questions and perform tasks:

- How must data be transformed to meet requirements?
- Implementation of data cleansing, transformation, and manipulation
- Iterations of the data engineering pipeline
- Means by which data quality can continuously be monitored and evaluated
- Use, extension, and modification of third-party data
- Human-involved data annotation and manipulation (“labeling”)
- Performance of additional data augmentation steps



Reason
Why AI Projects Fail

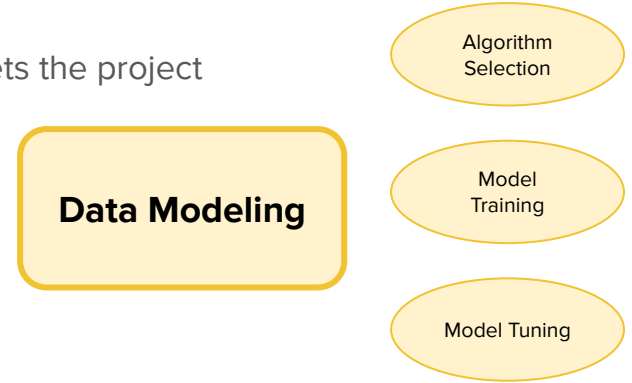
Trusting the Vendors Too Much

(Product mismatch, Overhype, and Oversell)

CPMAI Phase IV: Model Development

Answer the fundamental questions and perform tasks:

- How can the data be transformed into a machine learning model that meets the project requirements?
- Performance of model training activities
- Performance of model optimization activities
- Determination of appropriate algorithms, settings, and hyperparameters
- Creation of ensemble models
- Use of third-party models or extensions of models
- Model development as appropriate for selected machine learning technique



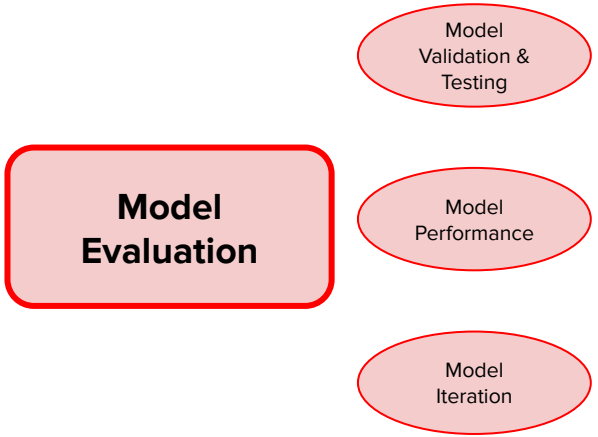
Reason
Why AI Projects Fail

Building a model before you know how it will be used in the real world

CPMAI Phase V: Model Evaluation

Answer the fundamental questions and perform tasks:

- Does the model meet requirements for accuracy, precision, and other metrics?
- Determining and evaluating concerns on overfit and underfit of models
- Evaluation of training, validation, and test curves for overall acceptability
- Evaluation of models against business Key Performance Indicators (KPIs)
- Determination of model suitability with regards to operationalization approach
- Determination of means for model monitoring
- Determination of means for model iteration and versioning

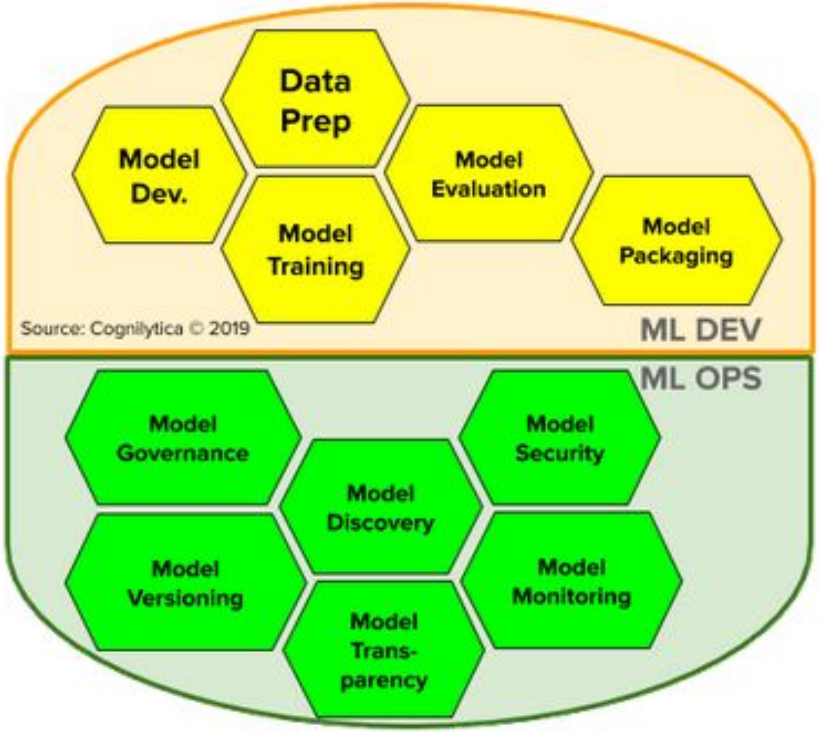


Reason
Why AI Projects Fail

AI Project lifecycles are continuous -- don't forget retraining (and model / data drift)

The Rise of Machine Learning Operation (MLOps)

- **DevOps for ML**
  - Model development lifecycle
  - Model deployment
  - Model versioning
  - IT Ops considerations for ML
- **Model-specific Opps**
  - Model drift
  - Data drift
  - Data provenance
  - Model Governance
  - Model discovery

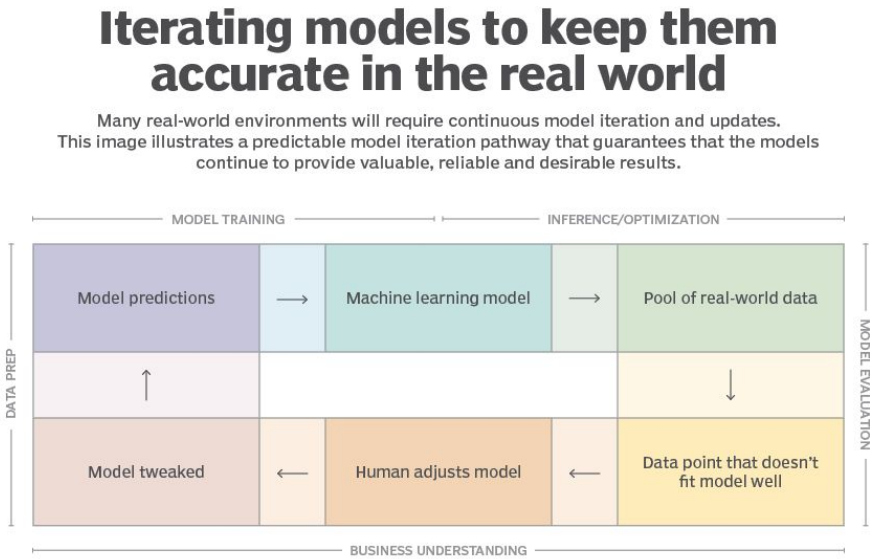


ML Ops is a Thing You Do. Not a Thing you Buy

Model Iteration

- If the real-world environment doesn't change much, then some models don't need frequent iteration
- Many real-world environments will require continuous model iteration and update
- CPMAI Phase V Model Evaluation requires ***continuous model evaluation, retraining, and operationalization***

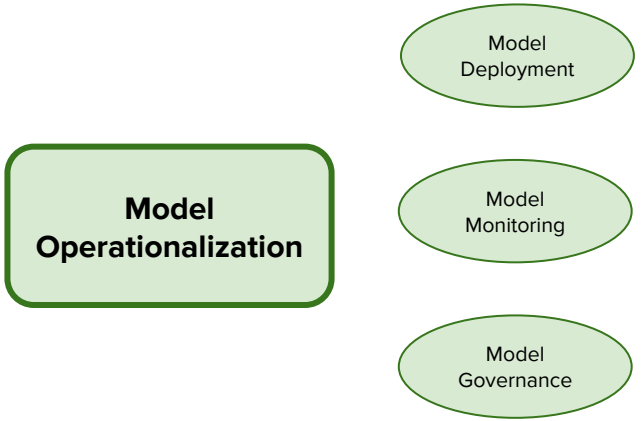
Successful AI projects have a predictable model iteration pathway that guarantees that the models continue to provide valuable, reliable, desirable results



CPMAI Phase VI: Model Operationalization

Answer the fundamental questions and perform tasks:

- How will this model be used in production / operational environments?
- What are the requirements for data flow for model to be useful?
- What are the requirements for performance?
- Operationalization of model in different environments
- Implementation of model monitoring
- Implementation of model versioning and governance
- Evaluation of business performance
- Determination of project success and iteration requirements



Number One Reason Why AI Projects Fail

Overpromising and Underdelivering

- Do you even know what problem you're trying to solve?
- Why are you tackling the hardest possible problem first?
- Why are you trying to tackle many AI Patterns at the same time?

Think Big. Start Small. ITERATE OFTEN.

This is the recipe for success.

CPMAI AI & Big Data Methodology Education & Certification

Vendor-Neutral self-paced online AI & ML Methodology Best Practice Education for:

- Technology Project Managers and Program owners
- Technology Implementers
- Contractors and SIs looking to certify their AI service offerings
- Professionals looking to advance their skills and careers
- ***Certification Included***

CPMAI: The Most Widely Adopted Best-Practices, Vendor-neutral Methodology for AI and Big Data Projects. Do AI the RIGHT Way.

More info at: courses.cognilytica.com

Some of the organizations that have completed Cognilytica Education:

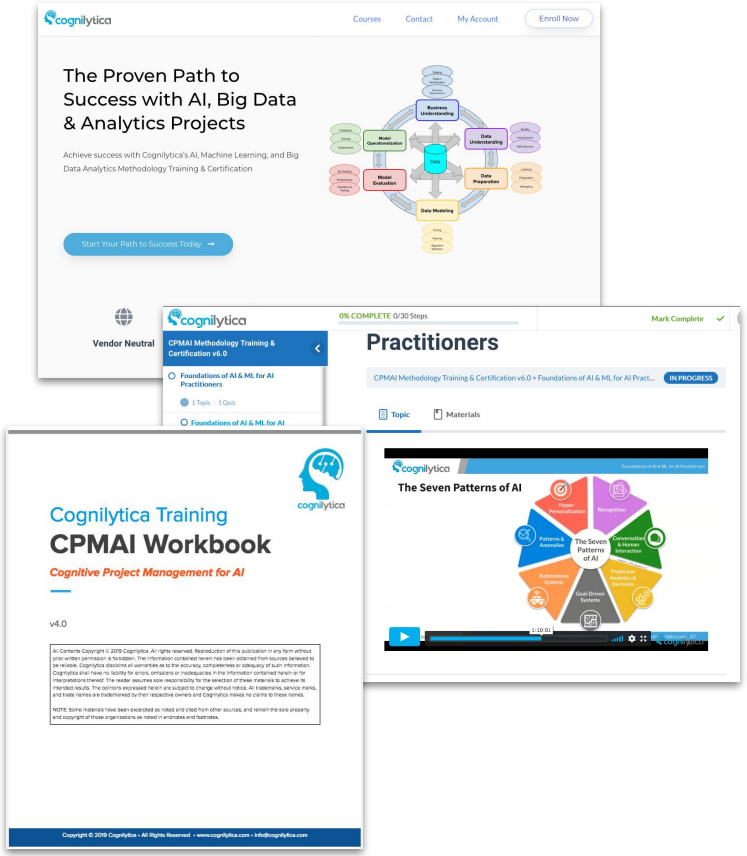
IRS, Department of the Army, Coca-Cola, Allegis Group, United States Postal Service, Department of Homeland Security, eia, epsilon, Partners Healthcare, Sub-Zero Wolf, USAA, Treasury Internal Revenue Service, National Geospatial-Intelligence Agency, GSA, Sandia National Laboratories, Atlassian, Department of Energy, motorola a lenovo company, Deloitte., MITRE, RioTinto, Department of Justice

- Sign up now for CPMAI training and certification @ **Courses.Cognilytica.com** • **EMAIL info@cognilytica.com for 10% discount** •

Strategic Support for Strategic AI

CPMAI Methodology and Certification

- ➔ **Course includes 15 modules and 24 hours of instruction**
- ➔ **Self-paced and supported virtual courses on AI topics:**
  - Foundations of AI & ML for AI Practitioners
  - Managing Data for AI
  - Applications of AI
  - Thinking and Acting like a Data Scientist
  - Ethical and Responsible AI
- ➔ **Gain understanding of key concepts in AI, machine learning, cognitive technologies, predictive analytics, Robotic Process Automation (RPA), and Big Data**
- ➔ **Certifications and exercises included**
- ➔ **Hands on workbook included to help you document project**
- ➔ **$2,495 per learner - Discounts for groups of 5 or more!**



• Sign up now for CPMAI training and certification @ **Courses.Cognilytica.com** • **EMAIL info@cognilytica.com for 10% discount** •

Ethical & Responsible AI Training and Certification

Vendor-Neutral online Ethical & Responsible AI Training & Certification allows you and your organization to:

- Gain a comprehensive understanding of boundaries for what is acceptable and not acceptable use of AI technology
- Understand methods, practices and processes by which AI systems can be implemented to fulfill governance, documentation, measurement, and mitigation
- Learn the necessary aspects to turn abstract ethical AI concepts into practical implementation actions.
- Be able to prove their knowledge of necessary elements of ethical and responsible AI systems
- ***Comprehensive Ethical Framework included***
- ***Certification Included***

BEING CERTIFIED ON ETHICAL AND RESPONSIBLE AI IS VITALLY IMPORTANT. Don't get left behind!

Thank You!

Presenters:

Kathleen Walch & Ronald Schmelzer

Cognilytica - http://www.cognilytica.com

info@cognilytica.com