



Treasures Untold: The Value of Legacy Data in the Age of AI

- What is Artificial Intelligence (Artificial Intelligence!) and where did it originate?
- Where do these programs get their data? How reliable is it?
- General Uses of AI
- Positioning your organization's data strategy for AI, with a focus on legacy data
- Preparing for the future of AI in Healthcare

Arrgh! What's Your Pirate Name?



- Front row = Sugar-Tongued Fishmonger Sparrow
- Second row = Matey McWinky Tuna Breath
- Third row = Royal Salty O'Patches Jackson
- Fourth row = Captain Jack Swimmin' O'Jelly
- Fifth row and beyond= Shipmate Sharkbait

What is Artificial Intelligence (AI)?

Computers that are being programmed to perform tasks that typically take more human thinking:

Learning

Reasoning

Problem-
Solving

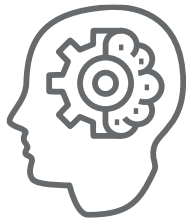
Perception

Decision-
Making

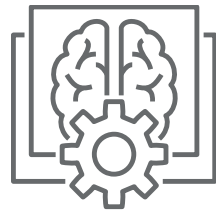
Where did it originate? The term “**Artificial Intelligence**” was coined in 1956 at a conference hosted by Dartmouth.

How is AI accomplished?

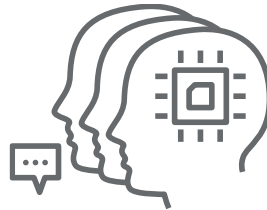
Many different techniques are being used to enable machines to demonstrate intelligent behavior:



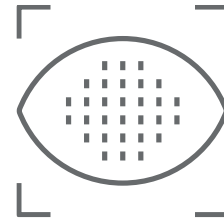
Machine Learning



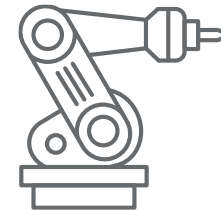
Deep Learning



Natural Language Processing



Computer Vision



Robotics



Expert Systems

The *amount* and the *quality* of the data is important.

Data is the keystone to “teaching” AI

The Learning Process includes:

- **Data collection** with a review of quality, diversity and dataset size
- **Data preprocessing**—cleaning and transforming
- **Choosing a learning algorithm**—depends on the nature of the data and what is available.
- **Training the model**—application of the learning algorithm,
- **Model evaluation and validation**—accuracy, precision, and recall are some metrics that are applied.
- **Deployment and inference**—making it available for use and will continue to “learn” as it encounters new data

Sources of data for “teaching” AI that are specific to the learning objective:

- Public datasets
- Private datasets
- Crowdsourcing
- Self-generated data

Best data formats for AI

- Structured and organized
- Numerically represented, if possible
- Highly reliable and clean
- Temporal or sequential
- Labeled and annotated



A Little Pirate Humor...

- Q: How do pirates know that they are pirates?
- A: They think, therefore they *ARRRRR!!!!*
- Q: How much did the pirate pay for his piercings?
- A: A buck-an-ear.
- Q: What do ye call a pirate with two eyes and two legs?
- A: A rookie.
- Q. What's a pirate's favorite type of exercise?
- A. The plank!

General Uses of AI

Non-Healthcare

- **Finance** – development of new trading algorithms, market interaction automation, and fraud detection and prevention.
- **Manufacturing** – production automation, supply chain management.
- **Customer Service** – chatbots for FAQ's, issue resolution, and recommendations.

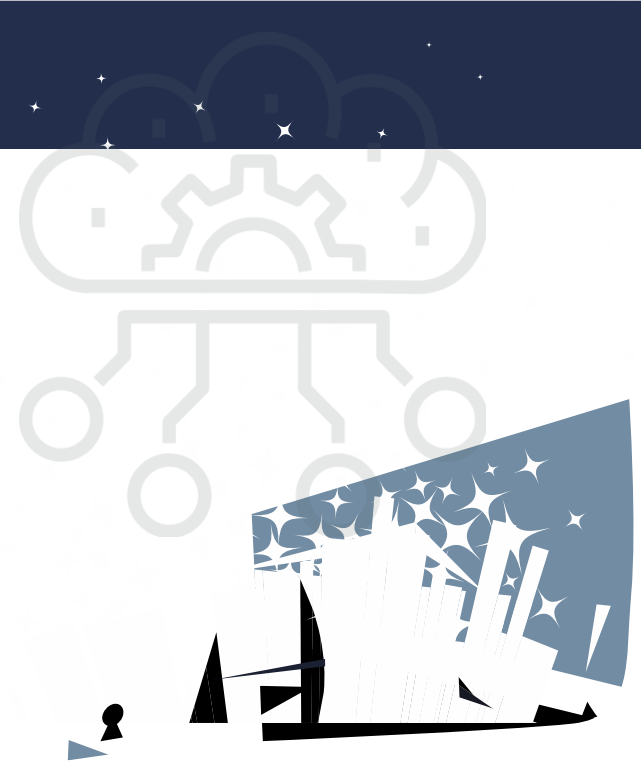
Healthcare

- **Risk stratification** – analyze a patient's likelihood of payment using demographics, historical payment behavior, income level, etc.
- **Diagnostic support** – develop new diagnostic tools and personalized treatment plans.
- **Proactive measures** around interventions to prevent payment challenges
- **Resource allocation and management assistance** – forecasting of demand around workload and resource needs across billing and collection workflow
- **Automated coding and documentation**
- **Claims processing review** to prevent denials or insurance payment delays
- **Charge Capture optimization**
- **Fraud detection** and revenue integrity auditing
- **Process automation** of manual, repetitive tasks

Positioning to maximize the value of current and legacy data

Your organization's data strategy for AI

- Include current and legacy system data for as far back as retention policies allow.
- Consider de-identified data for older data sets (beyond the legal retention period)



Legacy data has yet another use case for providing value to your organization

Preparing legacy data for AI teaching

Your archiving strategy matters:

- Data accessibility and centralization
- Improved data quality
- Increased sample size
- Longitudinal analysis for decision support or research

- Discrete data that is active or static
 - Preferred format as it is most ready for AI machine learning.
 - Uniformed format allows for the same learning algorithm to be used.
 - Best chance of being in the format AI learning requires.
- Image data
 - Clinical images with meta data is most helpful
- PDFs of clinical documents
 - Least desirable
 - OCR technology required – reliability is low, particularly if the document is hand-written
 - Requires many different learning algorithms to teach the AI how to learn from the different document formats it might encounter.

Preparing for the future

AI in Healthcare

- **Data Governance**
 - Legacy and production data sets
- **Data Retention**
 - To support identified and de-identified use cases
- **Data Cleanliness, Normalization, and Classification**
 - Legacy and production data sets
 - Go forward processes to keep data clean, organized and QA'ed
- **Discrete Data Archiving Strategy**

Before our Q&A, I leave you with this...

Yo-Ho! It's a Pirate's Life

May your anchor be tight,
Your cork be loose,
Your rum be spiced, and
Your compass be true!

WORK LIKE A
CAPTAIN
⚓
PLAY LIKE A
PIRATE

Arrgh! It's Q&A Time!

It wouldn't be a good pirate presentation without some treasure!



Shelly Disser, DBA
404-210-7175
sdisser@mediquant.com
www.mediquant.com

