

Applying Artificial Intelligence & the Environmental Industry

Presented by :

Andrew Kirkman, bp Remediation Management

Collaborators:

Irene Montero, bp Remediation Management

Pramod Neelappa, bp Remediation Management

Emily Moore, bp

Leslie Steele, Kleinfelder

Nathan Stevens, Kleinfelder

Andy Chopra, Independent



The views and opinions expressed in this presentation are solely those of the author (presenter) and do not necessarily represent the views or opinions of bp.

Presentation Overview

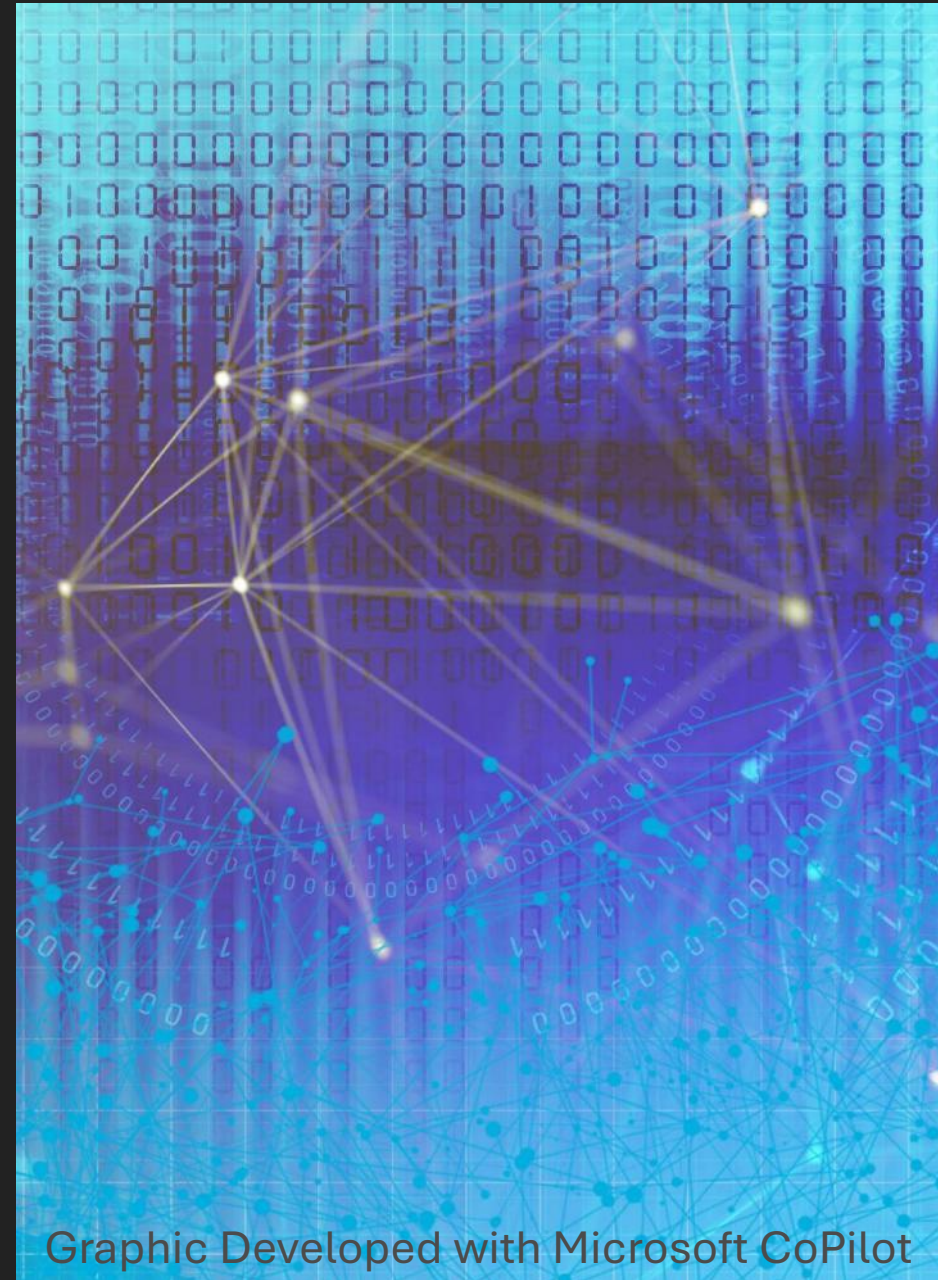
Introduction to Artificial Intelligence (AI),
Machine Learning (ML) and relationship to
existing disciplines

Review of applications of AI to our Industry

Discuss resulting themes of best practices

Note: This presentation was aided using Generative
AI in picture content and framing talking points
around best practice themes

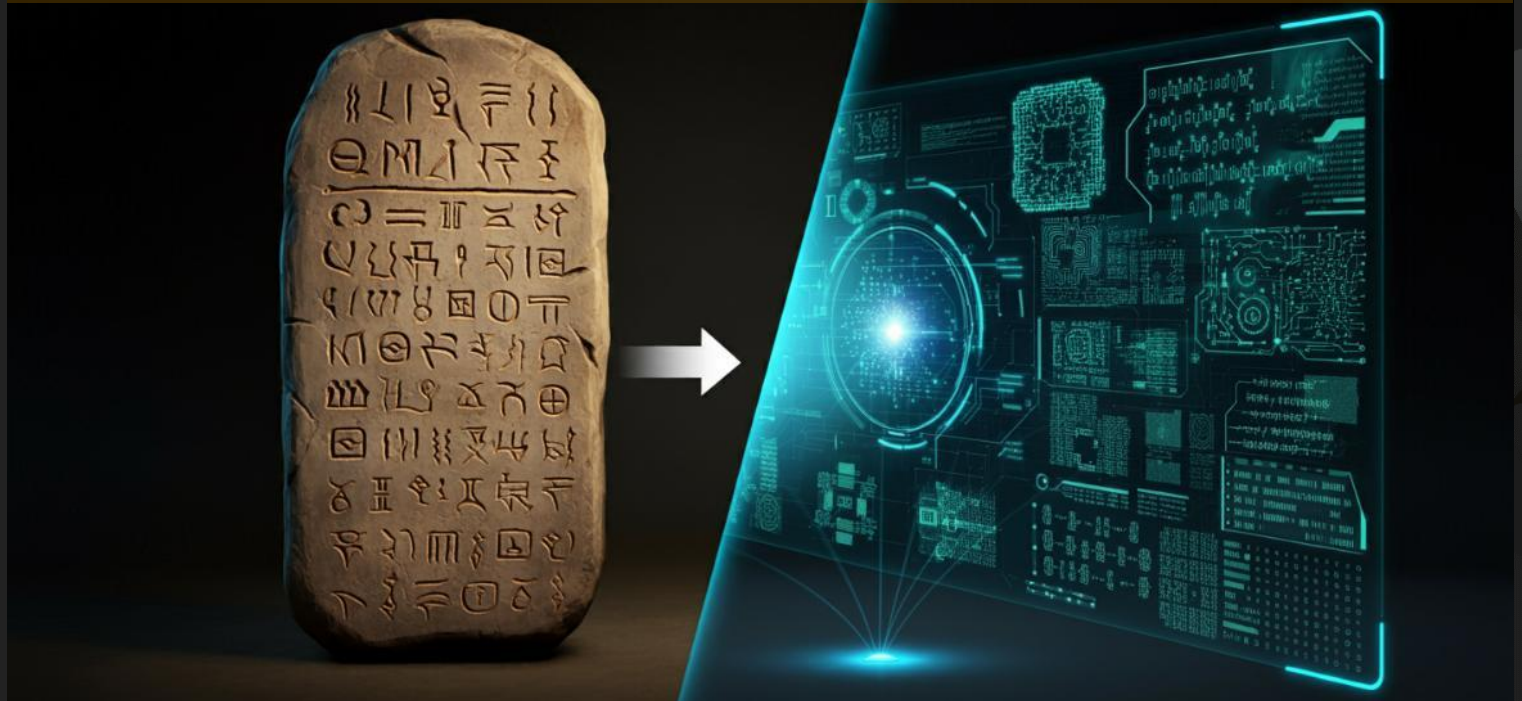
All slides were human led; some were AI aided



Graphic Developed with Microsoft CoPilot

What Can AI Do For the Environmental Remediation Industry?

Can it take the same data set and find better conclusions?



Is time to delivery reduced?

Does Data or AI Drive the Ultimate Insight?

Recent Proposal was to develop AI model that would review sites documents and select the remedy using historical remedy selection documents.

Sounds Great.... What data will it be calibrated too? Data Amount, Quality and Boundaries

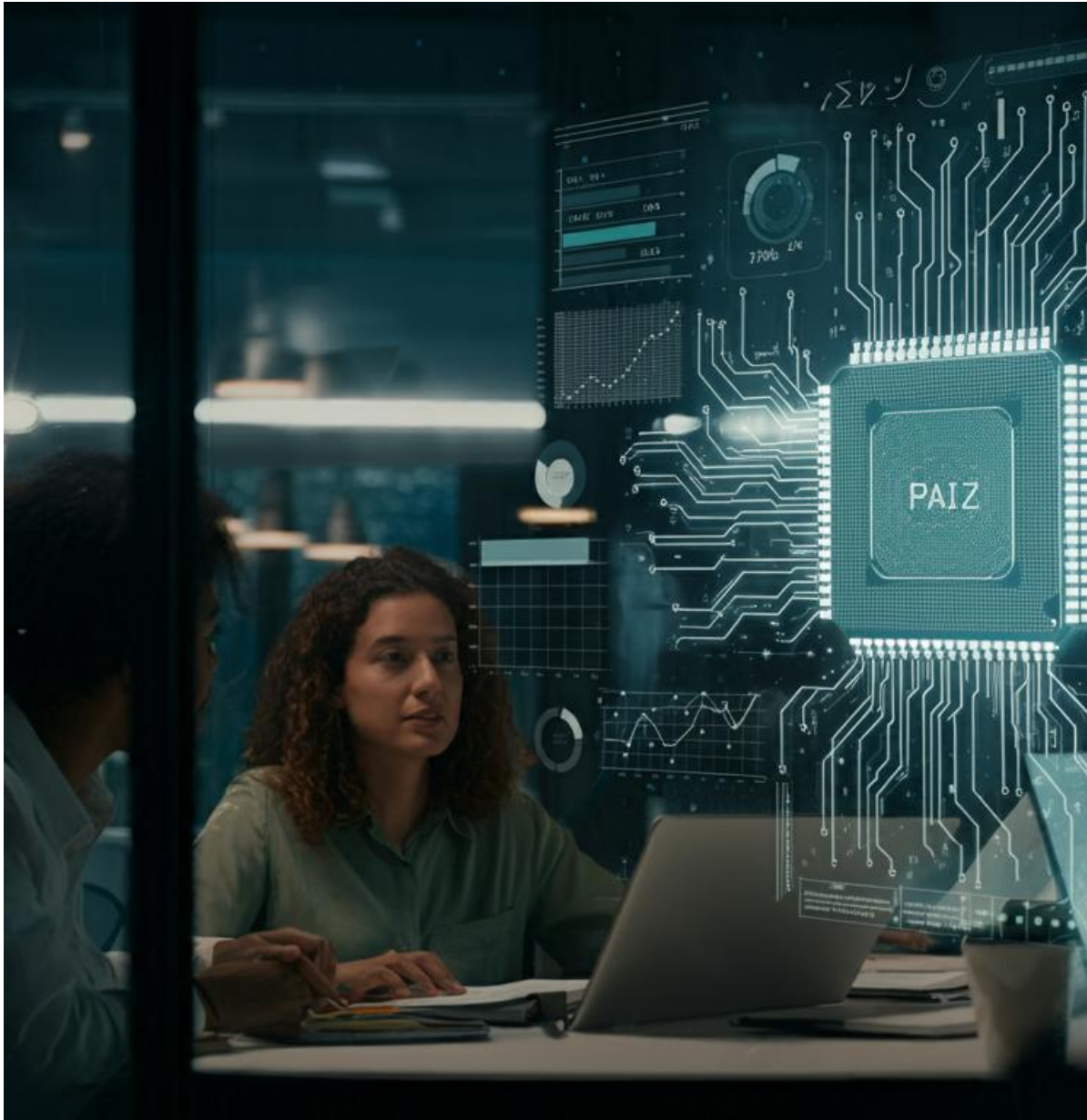
1. Wide data set – lots of variables relative to records
2. Will the historical data set to be used for calibration biased away from updated understanding
3. Will data set include performance monitoring of past remedies
4. How does this improve over the knowledge compiled in simple screening tables in recent remedy selection guidance documents

Does compiled and organized data provide the clarity?



Is AI taking our data and providing the insight





Where was the Human in that Last Proposal?

What Skills are needed to leverage AI?

Coding alone?

Data management?

Field experience in area of application?

Is AI replacing Humans or integrating with them?

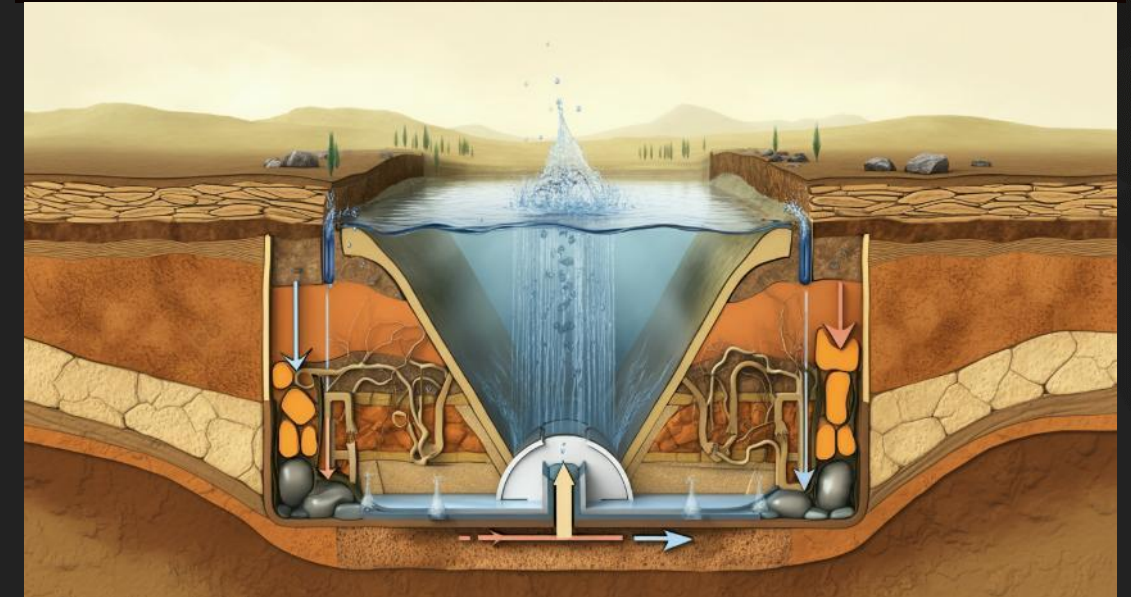
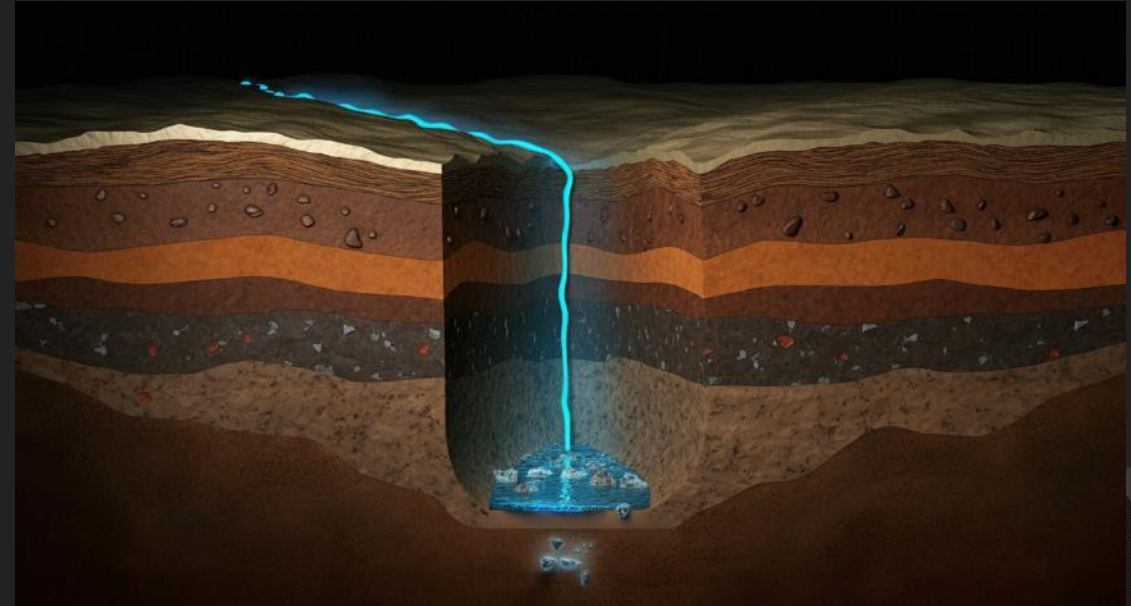
Does use of AI improve the outcome or replace the human at the cost of quality/confidence but at a lower price/time frame?



Is AI Reliable?

and if not, how do we leverage it?

E.g. Input for Image Generator –
Cross-Section of an Aquifer with
Groundwater Extraction



To answer some of these questions, consider what has occurred in this talk up to this point.

Who decided this was a good topic to present?

Where did question originate from?

Has AI enhanced the presenter or replaced?

Who Selected which image best represented groundwater?





Artificial Intelligence?

General AI

Handles any problem, creative thinking

Applied AI

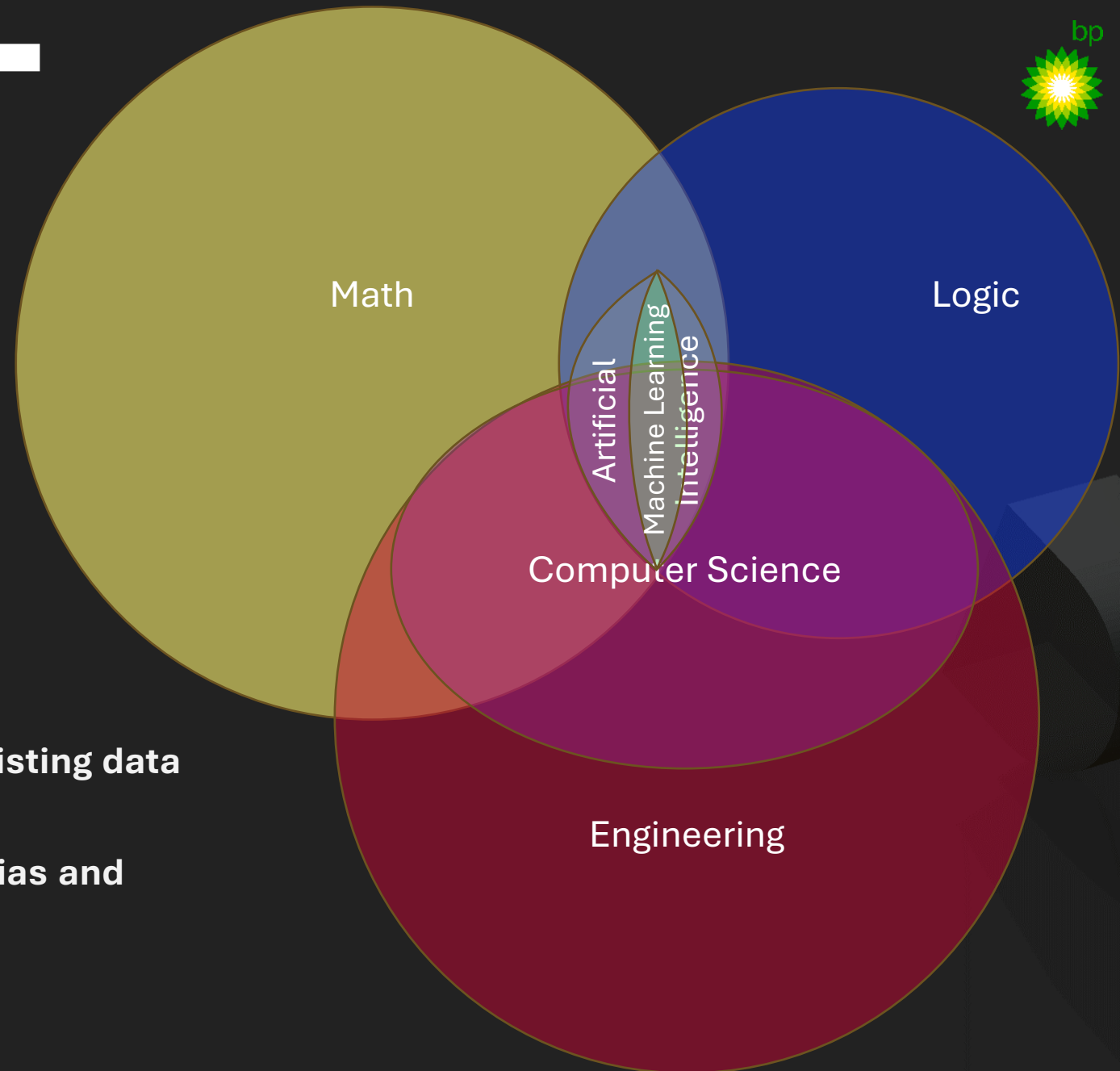
Problem specific

Generative AI

Generating new data based on patterns in existing data

Explainable AI

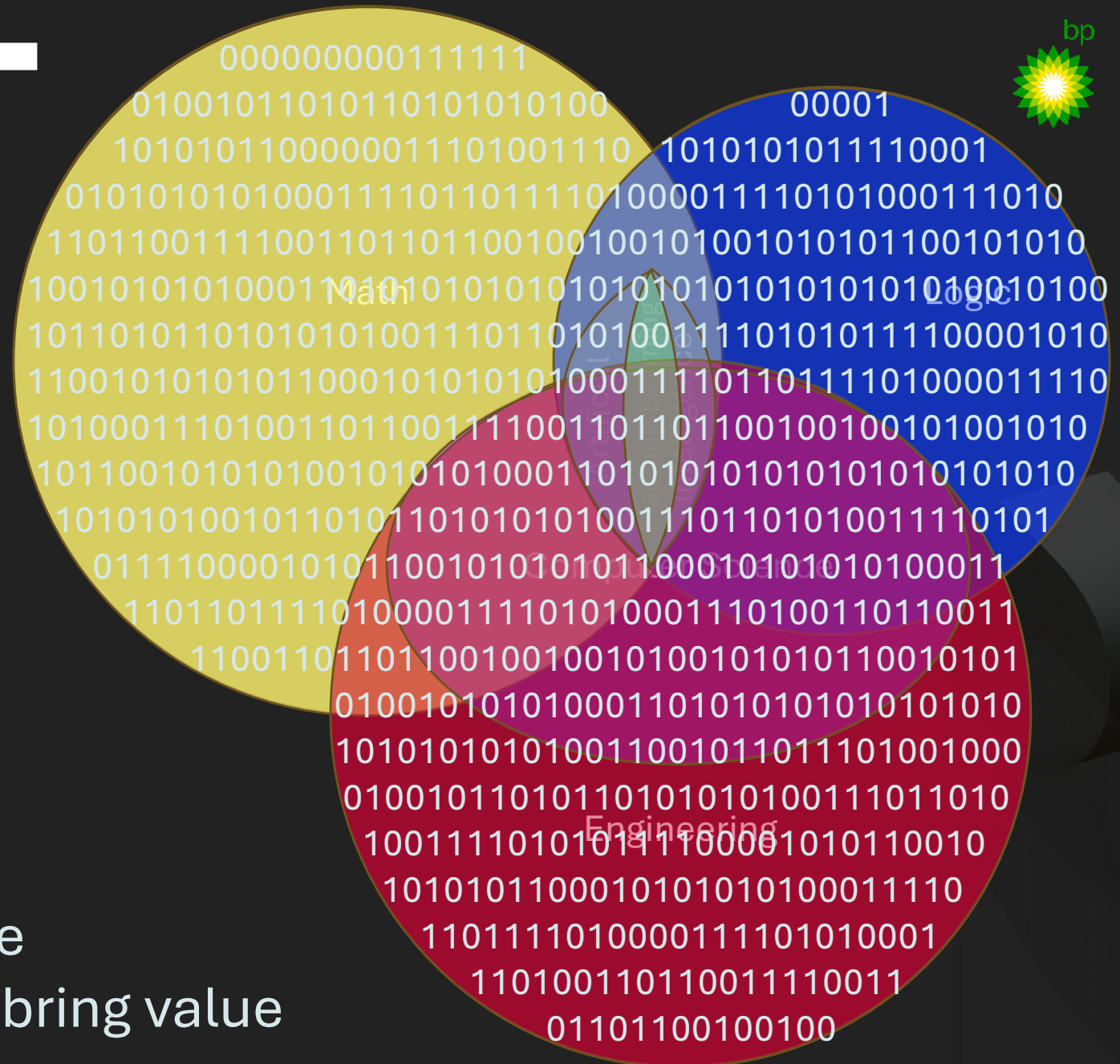
Transparency and understanding to reduce bias and promote accountability.





The Data set is more critical than the Technology

Invest in Data Quality and Quantity

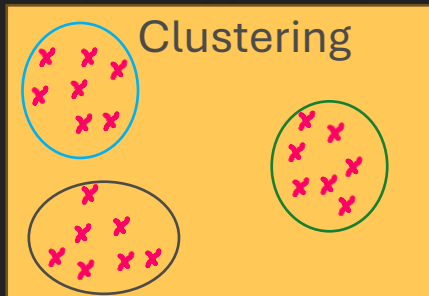


They all rely on data. Data is the most powerful aspect that will bring value

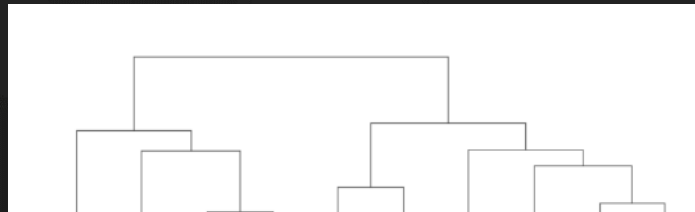


Diving Deeper into Artificial Intelligence

Machine Learning Examples



Dendrogram
Hierarchical Cluster



Machine Learning (ML) is an Applied AI which has been used in forensics for 20+ years

Supervised – regression algorithm or neural network trained on existing data

Unsupervised – Taking an existing data set and classifying based on each records values for multiple attributes, a vector → e.g., Principal Component Analysis or Clustering

There are 1000s of ML Approaches available

Consider whether AI/ ML for a given problem is:

1. A digital technology ready for application or;
2. A less proven application that requires the scientific method for each application (e.g., validation, uncertainty analysis)

A vendor selling a proprietary AI model as a technology doesn't often bear the risk to the outcome.



How Much Data is Needed?

Artificial intelligence, machine learning often touted as having less assumptions or non-parametric

This means more data is needed!

Traditional approaches with Mathematical representations of physical systems have assumptions but require much less data!

Ask if the data has ever been compiled and reviewed



Develop a
Conceptual Model
of the concern

Lack of Understanding
can be supplemented
with More Data

Combining non-
parametric with more
data can result in
improved understanding

Confirming the Behavior
reduces data needs

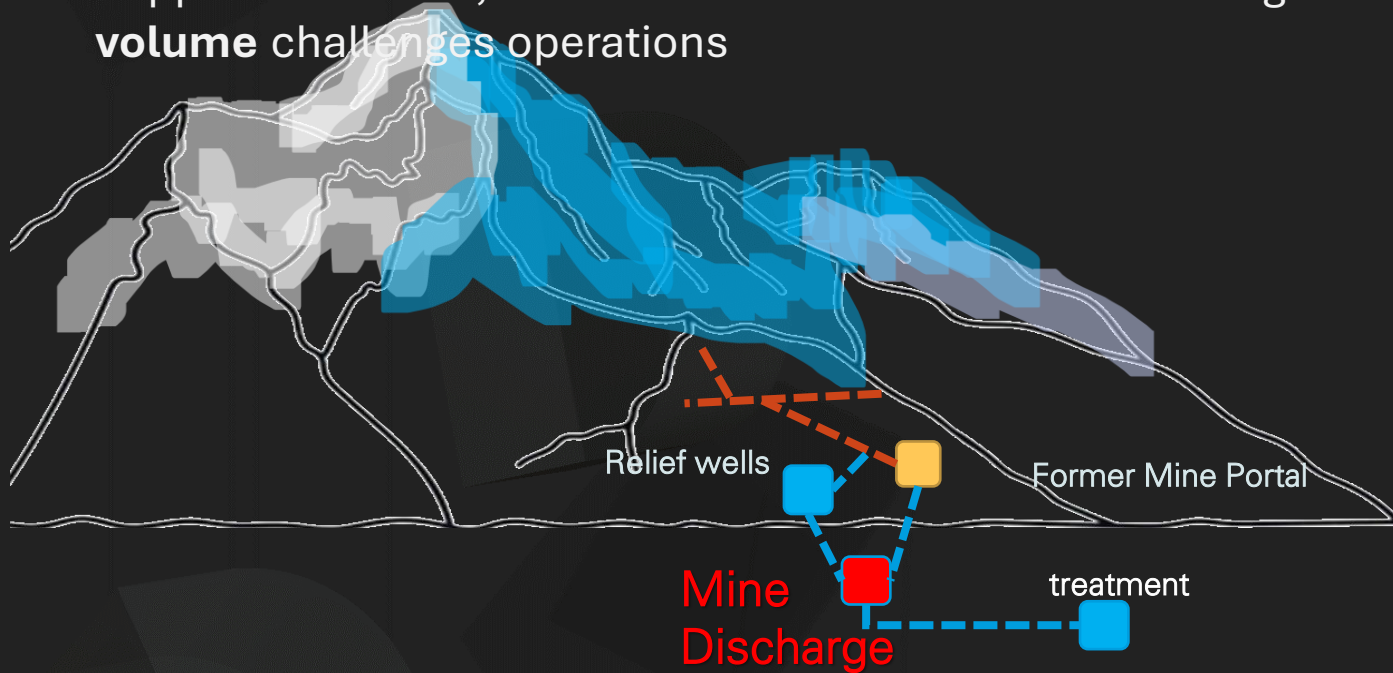
Assumptions based on
sound science, can be
combined with the right
data

Environmental
Characterization is
consistently data limited



Freshet — Snow Melt results in variable

discharge from former mine shaft. Treatment of mine water happens 24/7/365, but the variable nature of the discharge **volume** challenges operations



Data Sets

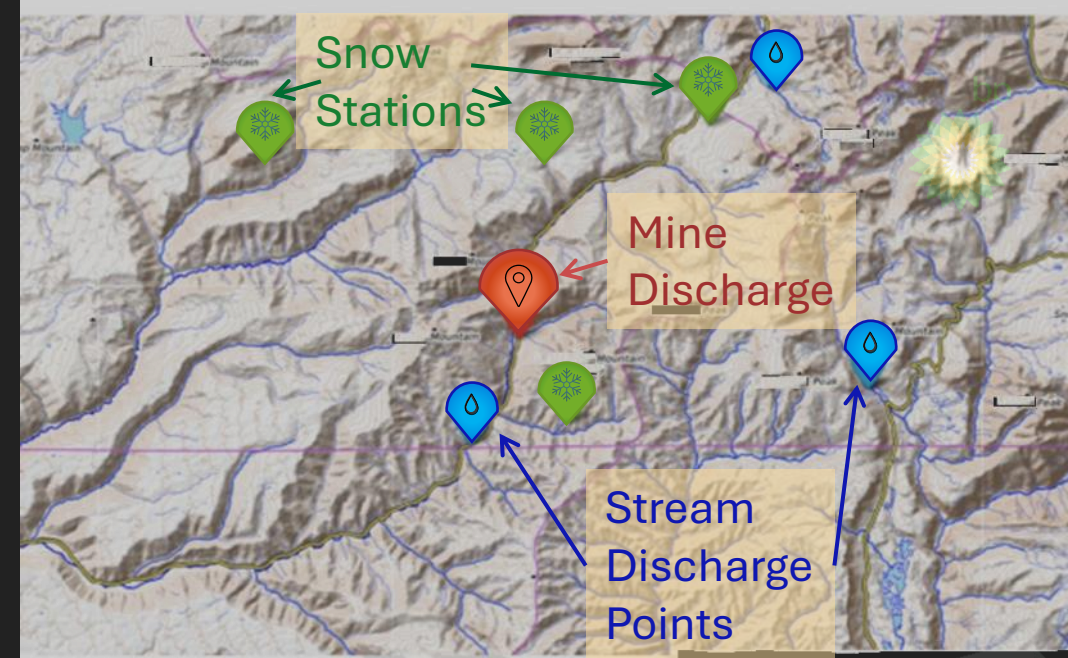
Snow Depth as SWE Air & soil Temperature

Soil Moisture Stream Discharge

Air & soil Temperature

Stream Discharge

Mine Discharge



After Montero et al., 2023

1. DATA EXPLORATION

2. MACHINE LEARNING APPROACHES

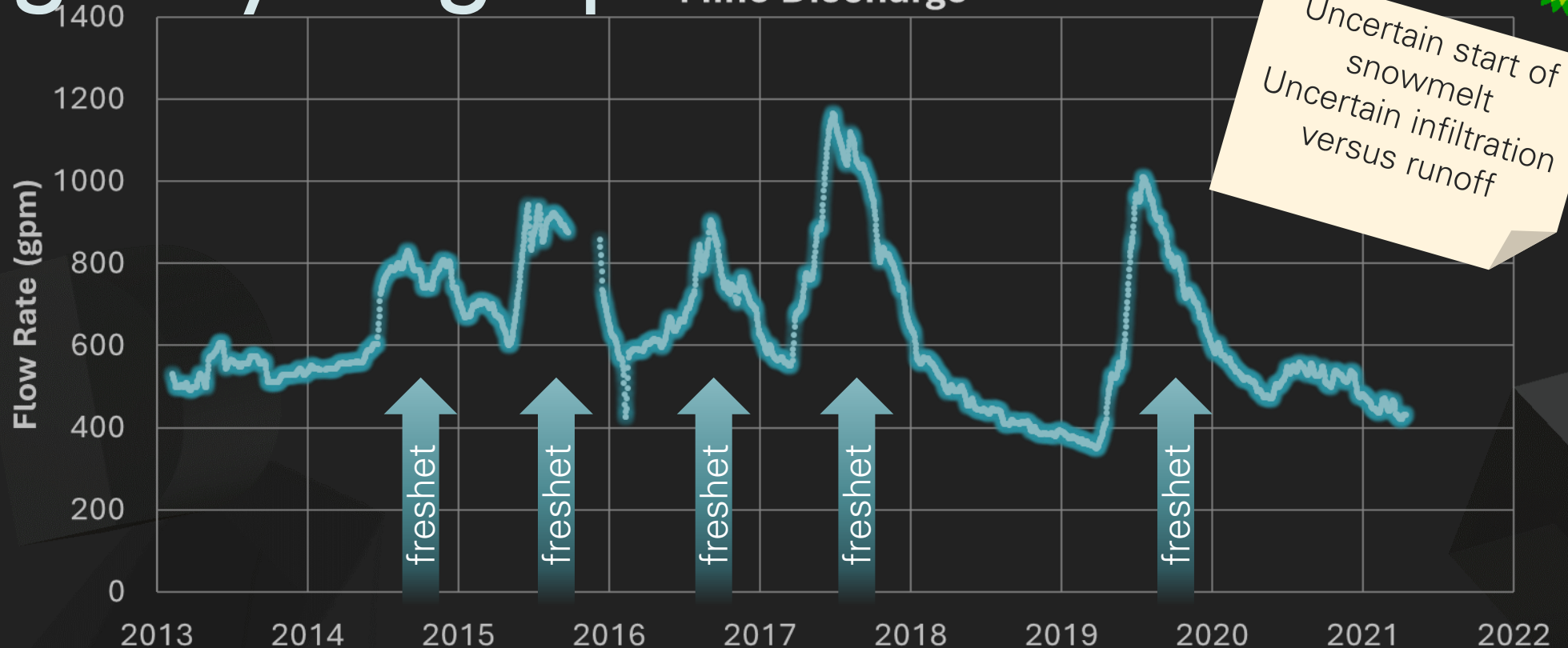
Multivariate Bayesian Ridge (https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.BayesianRidge.html)

Prophet - [Prophet | Forecasting at scale. \(facebook.github.io\)](https://facebook.github.io/prophet/)

3. VALIDATION AND RE-TRAINING

Montero, I., Neelappa, P., Chopra, A., Ismayilov, T., 2023, *Building a Machine-Learning Predictive Tool for Management of Water Treatment Resources*, Association of Environmental Health Sciences West Coast Conference, March 2023.

Discharge Hydrograph



Discharge Measurement consists of:

- Discharge from former mine portal
- Managed flow from relief wells. These wells manage the elevation of water inside the tunnel within a narrow range
- Does each data point above represent an event or are multiple data points needed to represent the nature of a given event?

After Montero et al., 2023



Key Learnings

Model version: 4.1.0.2022

Selected model:

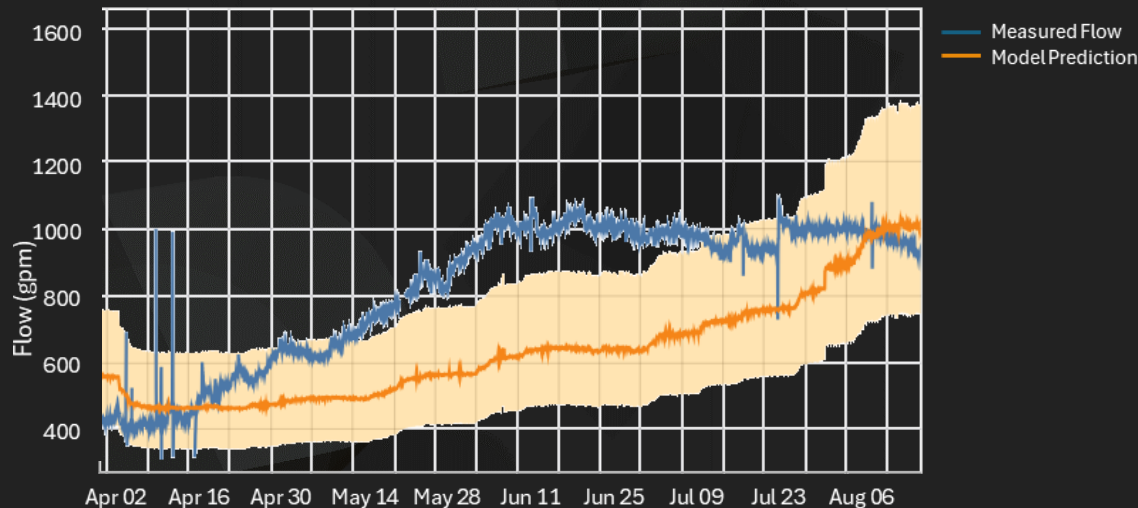
20_weeks

Forecast start date:

2023/04/01

Forecast end date:

2023/08/19



Investment:

Low Investment
Compilation of Data
Investment, online
and structured

9-12 Months Data
Engineer/Scientist

Results/Learnings

CSM supported model
optimization and
identification of
erroneous data

Poor Data Quality
Limited Outcome

Soil Sensors In
non-ideal
locations

Measurements prior
to 2016 were
inaccurate due to
sensor issues.

Model could
identify correct
Magnitude or
Timing, not both

Data allowed for
both standard and
AI approaches

Data Engineers were
reluctant at times to
quantify uncertainty

Did the model result in
a mass balance?

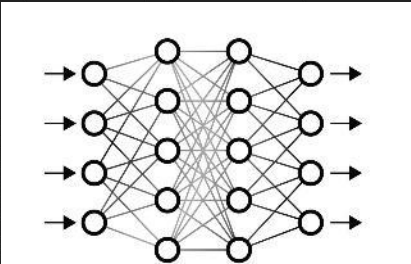


Hybrid models

Use rules (analytical derivations, or logical relationships) to put ‘guardrails’ on ML models.

NVIDIA Talk to bp this year described transforming the error function in neural networks with analytical solutions representative of the target physical system.

Improved efficiency, accuracy, and cohesion.



+

$$\frac{\partial c}{\partial t} = D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} + D_z \frac{\partial^2 C}{\partial z^2} - V \frac{\partial C}{\partial x} - \lambda C$$

Machine Learning

- High prediction accuracy
- Fast/data efficient (once trained)
- Transparency challenged

Mathematical Model

- Requires knowledge of behavior
- Requires numerical solution for behavior
- High prediction accuracy

Rules (Expert Systems)

- Heuristics based on experience
- Explicit connections
- Logical underpinning
- Slow/data intensive



Pretrained Models

Microsoft Custom Vision

customvision.ai

Base Model – Specific AI approach with an initial set of training data

In this case → Photographs with general categories of labels for those pictures by Microsoft

bp progressed training to classify photographs for traffic safety

Less data required than starting from Zero

The purpose was to photograph Work Zones near traffic and ensure the barriers met traffic safety plan requirements.

Traffic App



Results: Traffic App is useful but does not replace the Human

Investment:

MS Power App to Collect Photo's of work zones



Human Classification of Photographs



CustomVision AI
800 training classified photographs

Results:

90% Success during Training

84% Success during Pilot



4 False Positives

12 False Negatives

7 Poor Photo Quality

107 Total Photographs during Pilot





Traffic App Incorrectly Indicated Compliant Work Zone



Who/What Failed??

Review of the Training photographs revealed

No Active Street work had been used in the training data set

Training Data set was at retail stations

The scenario was outside of the bounds of the original training data

Prior photographs did not account for Spotter Position, only set-up of traffic barriers

Perhaps the application did not fail rather it wasn't trained, and the training data set was incomplete

Traffic app is worthwhile for what it was trained on but its important to understand the bounds of the training data set.

Continuing Applications of Photograph based AI

Real Time Traffic Safety Vehicle Detection



All art here is AI generated with Google Image fx



Remote Site Wildlife Detection
Real Time





AI & Active Retail Station Maintenance Data

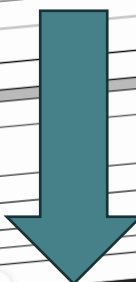
CLIENT / CUSTOMER: [REDACTED]
 SITE NAME: [REDACTED]
 FCL DEPT: [REDACTED]
 TECHNICIAN NAME: [REDACTED]

TASK NO: [REDACTED]
 DATE / TIME: [REDACTED]
 SIGNATURE: [REDACTED]

Expectations of using this PM survey form:
 1. All urgent non contract works to be communicated directly with the Business Support Coordinator
 2. All contract based price remedial work identified to be completed during PM where equipment / parts available.
 3. All contract based price remedial work to be communicated to the Business Support Coordinator
 4. All non contract remedial work to be communicated to the Business Support Coordinator
 5. All non contract remedial work to be completed fully and comments clearly written for ease of reading by others.
 6. All sections are to be completed fully and comments clearly written for ease of reading by others.
 7. Photographs are to be taken of any remedial work completed / to be completed and attached to the photo pages (8 & 9) on this form.

Category	Check Complete			Comments/Remedial Repairs Required	Remedial Work Completed on Site	To be reviewed by Service Manager
	Yes	No	N/A			
CSD CHECKS - 6 MONTHLY - See Content Locker App - CSD Test Procedures Folder for full details on Process						
a. Shaver Valves - Have all these units passed. (Close and test pour, no fuel should pass)	✓					
b. Breakaway / Staircase - Have all these units passed. (Free Movement / visual inspection/Expiry Date) (Free Movement / visual inspection/Expiry Date) (Free Movement / visual inspection/Expiry Date) (Free Movement / visual inspection/Expiry Date)	✓					
c. Caltex Truck stop OPW Breakaway Expiry Date: See Content Locker for full details on the OPW Breakaway Process.	✓					
d. Nozzle Auto Shut Off - Have all these units passed (Test pour, nozzle shuts off)	✓					
e. PLIDI M.I.D. - Has the Leak Detectors passed the test as per the instructions? (Independent/ Private Customer only)	✓					
f. Emergency Stop - Has the unit passed the test as per the instructions? (Independent/ Private Customer only)	✓					
PUMPS & DISPENSERS						
1. Are all hoses in good condition i.e. No cracking, abrasion, bulging etc?	✓					
2. Test hose continually for all hoses. Are all hoses < 1 m length?	✓					
3. Record any hoses by pump number & grade with a resistance of > 1 meg ohm.	✓					
4. Are all hose fittings/breakaways tight & leak free & good condition?	✓					
5. Are all auto clips & poststrips functional where fitted?	✓					
6. Are nozzle holsters in good condition?	✓					
7. Are nozzle spouts and splash guards in place?	✓					
8. Are all hoses up off ground i.e. not lying on the ground and in danger of being run over?	✓					

Unstructured Data



Structured Database

Data Aggregation

Data Aggregation Using Human &



AI Optical Character Recognition (OCR)
from Azure Cognitive Library

Open AI Chat GPT 4.0 to query reports

300 Groundwater & Tank Pull Reports

Maintenance and Inspection Reports

Finance records



Human Validation of ChatGPT 50 reports

Note: OCR could read digitized pdfs & struggled with handwritten and symbols ✓



Human Led Statistics



Retail Data Science

	Sample Details and Analytical Results						
URS Sample Reference	SS43	SS44	SS45	SS46	SS47	SS51	SS52
Laboratory Sample Reference	730548.43	730548.44	730548.45	730548.46	730548.47	730548.51	730548.52
Date Sampled	24-Sep-09	24-Sep-09	24-Sep-09	24-Sep-09	24-Sep-09	24-Sep-09	24-Sep-09
Sample Location	Dispensers			Remedial excavation beneath dispensers			
	Beneath dispensing lines	Beneath Dispenser 1	Dispenser 2	South wall	West wall	South wall	North wall
Sample Depth (m below ground level)	1.1	2.3	2.5	1.2	2.4	0.5	2.4
Sample Soil Type	SILT	Sandy GRAVEL	Sandy GRAVEL	SILT	SILT	SILT	SILT
Guideline Soil Type ⁴	Sandy SILT	SAND	SAND	Sandy SILT	Sandy SILT	Sandy SILT	Sandy SILT
Sample of soil remaining or removed	Remaining	Removed	Remaining	Removed	Removed	Removed	Removed
Total Petroleum Hydrocarbons (TPH)							
C ₇ -C ₉	< 8.0	360 ^m	780 ^m	< 8.1	1600 ^m	< 8.9	600 ^m
C ₁₀ -C ₁₄	< 20	410	790	< 20	2,000	< 20	400
C ₁₅ -C ₃₆	< 30	1,100	4,500	< 30	1,300	36	9,000
Total hydrocarbons (C ₇ - C ₃₆)	< 60	1,800	6,100	< 60	4,900	< 60	10,000
BTEX Compounds							
Benzene	< 0.050	4.4 ^m	3.8 ^m	0.13	32 ^y	< 0.057	<u>5.00</u>
Toluene	< 0.050	9.6	2.9	< 0.050	55	< 0.057	0.27
Ethylbenzene	< 0.050	24	12	0.18	<u>120</u>	< 0.057	1.7
Total Xylenes	< 0.15	<u>127</u>	<u>148.2</u>	2.35	640 ^y	< 0.177	<u>200.39</u>

Needed to experiment with prompts for ChatGPT 4.0

Initial questions too complex for ChatGPT 4.0 → human was replaced

Revised to ask more directly for data result where we would reason

50% to 80% accuracy depending on which question

```
"report_prepared_by": "NA",
"site_name": "NA",
"total_no_of_samples_tested": "NA",
"groundwater_depth": "NA",
"distance_to_town_water_supply_intake": "NA",
"groundwater_use_in_vicinity": "NA",
"samples": [
```

```
"sample_name": "SS14",
"contaminants_detected": ["TPH C7-C9", "TPH C10-C14", "TPH C15-C36", "Benzene", "Toluene", "Ethylbenzene", "Xylenes"],
"exceeds_industry_or_regulatory_guidelines": ["no", "no", "no", "no", "no", "no", "no"],
"contaminant_concentration": ["< 8.4", "< 20", "< 30", "< 0.050", "< 0.050", "< 0.050", "< 0.15"],
"sample_depth": "2.7"
```




Retail Results

AI Tools supported Data Aggregation from unstructured pdfs to structured format (e.g., SQL or comma delimited text)

Data Aggregation with Statistical analysis in MS Excel provided justification for recommendations on future management of retail sites

Financial Data recognized as incomplete record

- Stations with financial provision may be false positives,
- Those without may be false negatives.
- **This could introduce bias.**

A more robust assessment of environmental liability could rely on groundwater impacts as a metric to compare vs maintenance records

Investment:

12 Months Data Engineer/Scientist Aggregating Data even with AI tools

<40 hours to perform needed Data Analysis

Results

Pass and Fail from maintenance checks did not have high correlation to groundwater impacts

Inconclusive test results did have a high correlation to impacts – Why?...Theories below:

If it is Pass it works

If it fails it gets fixed quickly

If Inconclusive – how often did we look to fix vs retest?

ChatGPT was more reliable identifying groundwater results than soil results → likely due to Bias in reports



Resulting Thoughts – Why is 50-80% accuracy ok?

Ancient cultures- often came up with a system they knew was imperfect, made initial guesses and they used the errors to calculate more accurate answers.

We still use this method today, consider cost estimates, we take an informed guess at what the cost will be and factor in some buffers to manage uncertainty and as the project progress we refine expectations.

If Errors are known, communicated, & planned for they can be managed

Given we have predictions that are 80 percent accurate in characterizing if a given site is impacted:

It's difficult to rely on ChatGPT 4.0 solely to determine where impacts do and do not exist

The results could support prioritizing site reviews

Its less risky using ChatGPT 4.0 to identify trends across a portfolio of reports/sites

e.g., maintenance records vs likelihood of impact

If a given application of AI doesn't provide uncertainty, how do we know when to apply it?



Consider These Challenges

Communicate to an audience that biodegradation of oil is routinely occurring. The oil is not too toxic for microbes and in fact they attach on to it and develop enzymes to enhance extraction.

Can AI help explain or conceptualize this?

One needs to alert a team to the dangers of rolling out too many new digital products too quickly with lack of support to their company.

Can AI help message this?

Improving Communication with Generative Artificial Intelligence

AI art Generated with Google Image FX Studios

Communication Applications

- Continuing Education
- Stakeholder Discussion
- Sales
- Teamwork



9/19/2025





Conclusions

The value is generally coming from aggregating data over the improved data insights from AI

AI does not appear practical for routine use with subsurface characterization due to high variability and low data quantity

- AI doesn't decrease the amount of data needed
- Actually, it likely increases the amount needed
- AI can decrease analysis time

Whether it's viewed as a Best Ethical or Technical practice it is a Good Theme:

Use AI to enhance the job rather than replace the human

This ensures quality, understanding of uncertainty and enhances the employee

Application of AI benefits from understanding uncertainty

How the results are used is dependent on the uncertainty

- Broad trend reviews can reduce risk
- Prioritization of record review
- Be cautious of complete reliance in use on the results



A Framework for Evaluating AI Projects

Consider an AI Register:

Goal: What is the Goal of the Project

Is there an understanding of the expected behavior and its relationship to the variables

What is the value to be obtained? efficiency, insight

What is the value metric?

Data: Quality, Amount and Boundaries of Data

Is data amount supplemented by a precalibrated AI application?

Approach: Precalibrated, Build your own Supervised, or Unsupervised

- Precalibrated Models - e.g., MS Customvision.ai, Google Imagefx or natural Language models OpenAI Chat GPT
- Single AI approach or are multiple approaches being used to reduce the data set and then model behavior

Risks: Document the known limitations of the data, approaches, understanding

Identify the Impacts these could have and likelihood

Identify mitigations taken and how these can be validated

Stakeholder agreement: How will this be used in decisions

Uncertainty: How will uncertainty be quantified and managed in decisions?

Is it an entirely new application of AI?

- Using non-parametric or parametric analysis?
- Certain machine learning algorithms have been routinely forensics (e.g., clustering & principal component analysis)

Future Unknown Risk Management: Health Industry Register – If a staff looks at a patient record, they document their review and activities

- Could AI models be more transparent, document automatically what changes/decisions were made in case review is needed at later date
- Applies to AI managing remediation system operations perhaps



When should we Look to Artificial Intelligence for Solutions?

There are more reasons to improve data management than for AI alone.

Conceptualized after reading a digital systems article

[After Pratt, M.K., 2023, *Replacing vs. maintaining legacy systems* | online article TechTarget.com](#)

Reasons to use traditional methods

1. We've never fully implemented them
 - Data was never compiled and reviewed
2. Traditional methods still support organizational goals/insights
3. Large number of records do not exist 1k-10K per location/event/variable

Reasons to use AI methods

1. Traditional methods have been tried and are too resource intensive (e.g., reading 300 reports)
2. New Insights needed and traditional methods have not produced recently
3. Data too complex to identify dominant variables
4. Operational efficiency (e.g, ChatGPT 4.0 AI up to 80% correct, Graphic generation, seeding a presentation or report outline)