AI Data Analytics in Environmental Laboratories

Reliable Data for Sound Decision Making

Matthew Cauthen

Environmental Consultant at Clinisys

Environmental Measurement Symposium
August 2024

clinisys

01 August 2024

Matthew Cauthen

Environmental Consultant at Clinisys

Mount Juliet, Tennessee

Bachelor of Science in Microbiology with a minor in Chemistry

Experience

26 Years in environmental laboratory space

- **10 Years bench chemist**
- **5 Years LIMS Admin**
- **11 Years LIMS Provider**

clinisys

AI Data Analytics in Environmental Laboratories

Topics

- Introduction to AI Data Analytics in Environmental Labs

- Automating & Enhancing Data Analysis

- Optimizing Experiments & Enabling Predictive Insights

- The Future of Environmental Laboratories with AI

- Words of Caution

- Conclusions

clinisys

Definitions

Analytics

The systematic computational analysis of data or statistics.

Data Analytics

A multidisciplinary field where there is extensive use of computer skills, mathematics, statistics, the use of descriptive techniques and predictive models to gain valuable knowledge from data through analytics.

Predictive Analytics

A form of analytics applying machine learning to generate predictive models.

Machine Learning

The ability of a machine to learn and then mimic human behavior that requires intelligence. This is accomplished through artificial intelligence, algorithms, and models.

Predictive Modeling

A statistical technique used to predict future behavior. The models can analyze both historical and current data and generate a model to predict potential future outcomes.

Why AI is a Must-Have Tool for Environmental Data Analytics?

- The ever-increasing volume, **complexity,** and pace at which environmental data is generated, pose **significant challenges** for traditional data analysis methods.

- Manual data cleaning, analysis, and visualization are **time-consuming**, prone to omission and human error.

- Identifying **hidden patterns** and trends within massive datasets, including **unstructured data**, requires **advanced** data mining techniques.

That's exactly where AI can help!

Image generated using Copilot via DALL-E

clinisys

Enabling Predictive Analytics & Insights

Predictive Modeling for Experiment Outcomes
- Analyze **historical** environmental **data** and **current conditions**.
- Predict experiment outcomes with **greater accuracy**.
- Allow researchers to adjust variables for better results.

Real-time Data Analysis & Visualization
- Visualize and **analyze** environmental **data in real-time.**
- Provide **immediate insights** into ongoing experiments.
- Enable **informed decision-making** throughout the research process.

Identifying Trends and Patterns
- Analyze large environmental datasets to **discover hidden patterns**.
- **Identify correlations** between different environmental factors.
- Lead to **new scientific understanding** and knowledge generation.
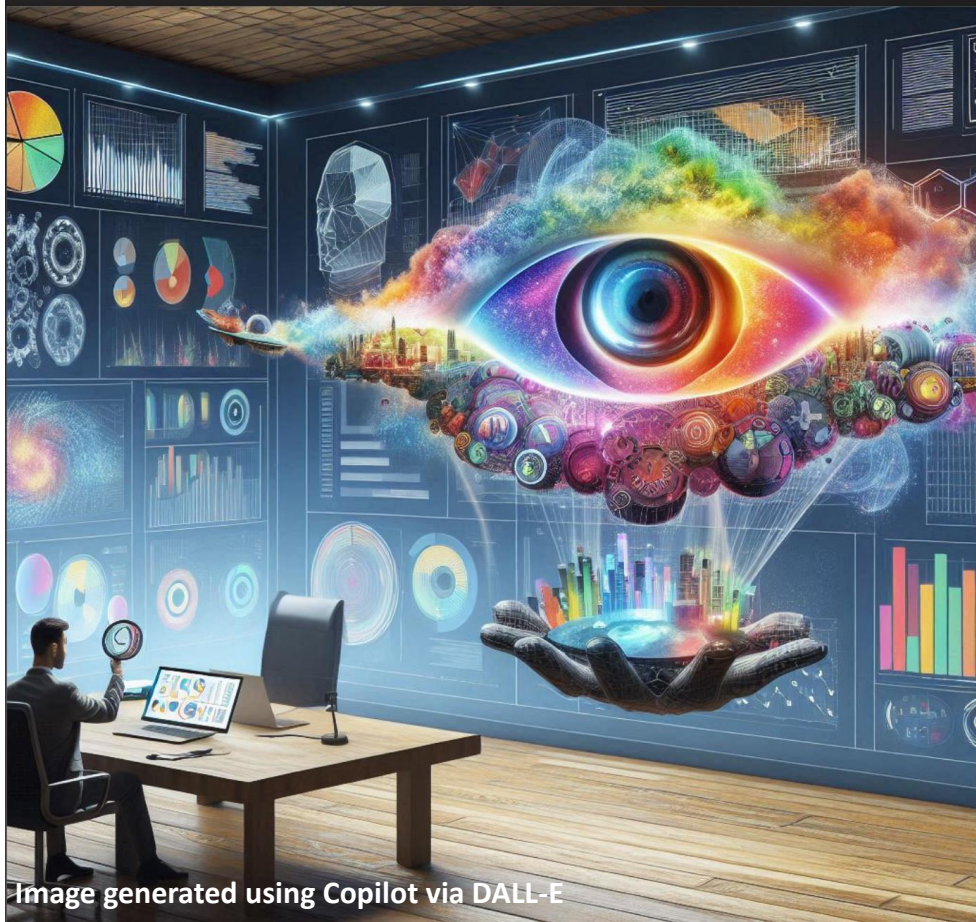
Generating Interpretable Insights & Visualizations

Image generated using Copilot via DALL-E

Explainable AI (XAI)

- AI models can **explain the reasoning** behind their predictions and insights derived from environmental data.
- This transparency **builds trust** in AI-driven results and allows researchers to understand the model's decision-making process.

Automated Report Generation

- AI can automatically generate comprehensive reports **summarizing key findings** from environmental data analysis.
- This saves researchers/professionals time and ensures consistency in reporting, facilitating predictable and clear delivery of results.

Advanced Data Visualization

- AI can create **interactive visualizations** of environmental data, allowing researchers to explore and **manipulate data in real-time**.
- This empowers researchers to identify patterns, test hypotheses, and gain **deeper insights** from highly complex datasets.

clinisys

# Automating Data Ingestion & Preprocessing

Data Wrangling

- AI can **automate the cleaning, review and formatting** of raw environmental data, addressing issues like missing values, outliers, and inconsistencies.
- This **ensures data quality** and consistency, a critical foundation for reliable analysis.
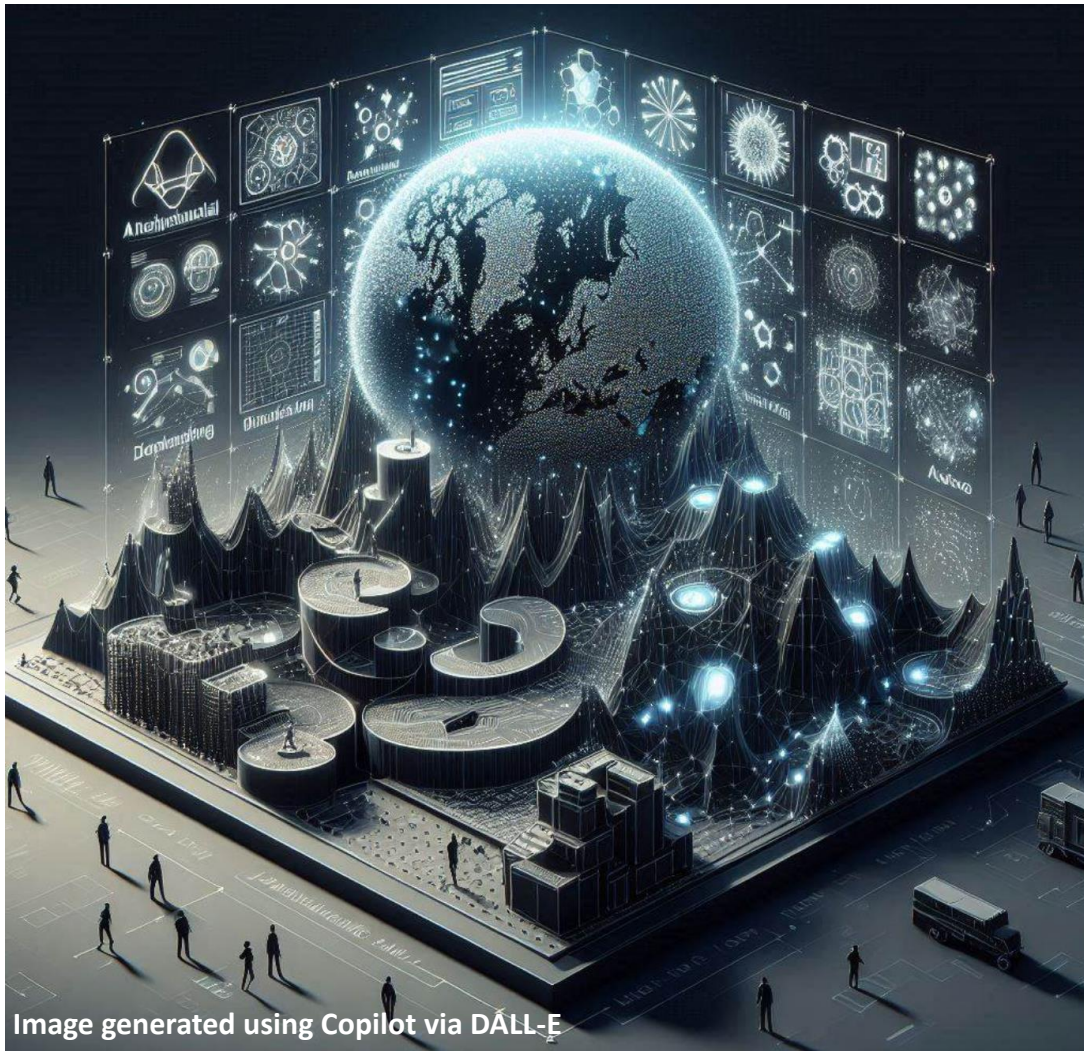
Feature Engineering

- AI can automatically **extract relevant features** from complex environmental data, saving researchers considerable time and effort.
- Feature engineering involves identifying and selecting the **most informative data points for analysis and review**, leading to more robust models and accurate results. Data Review Automation is like laboratory Nirvana.

Anomaly Detection

- AI can **identify unusual patterns** and potential errors in environmental data, alerting lab professionals and SMEs to potential issues.
- This **proactive approach** increases the efficiency of the data review act, as well as reducing the level of risk.

Image generated using Copilot via DALL-E

clinisys
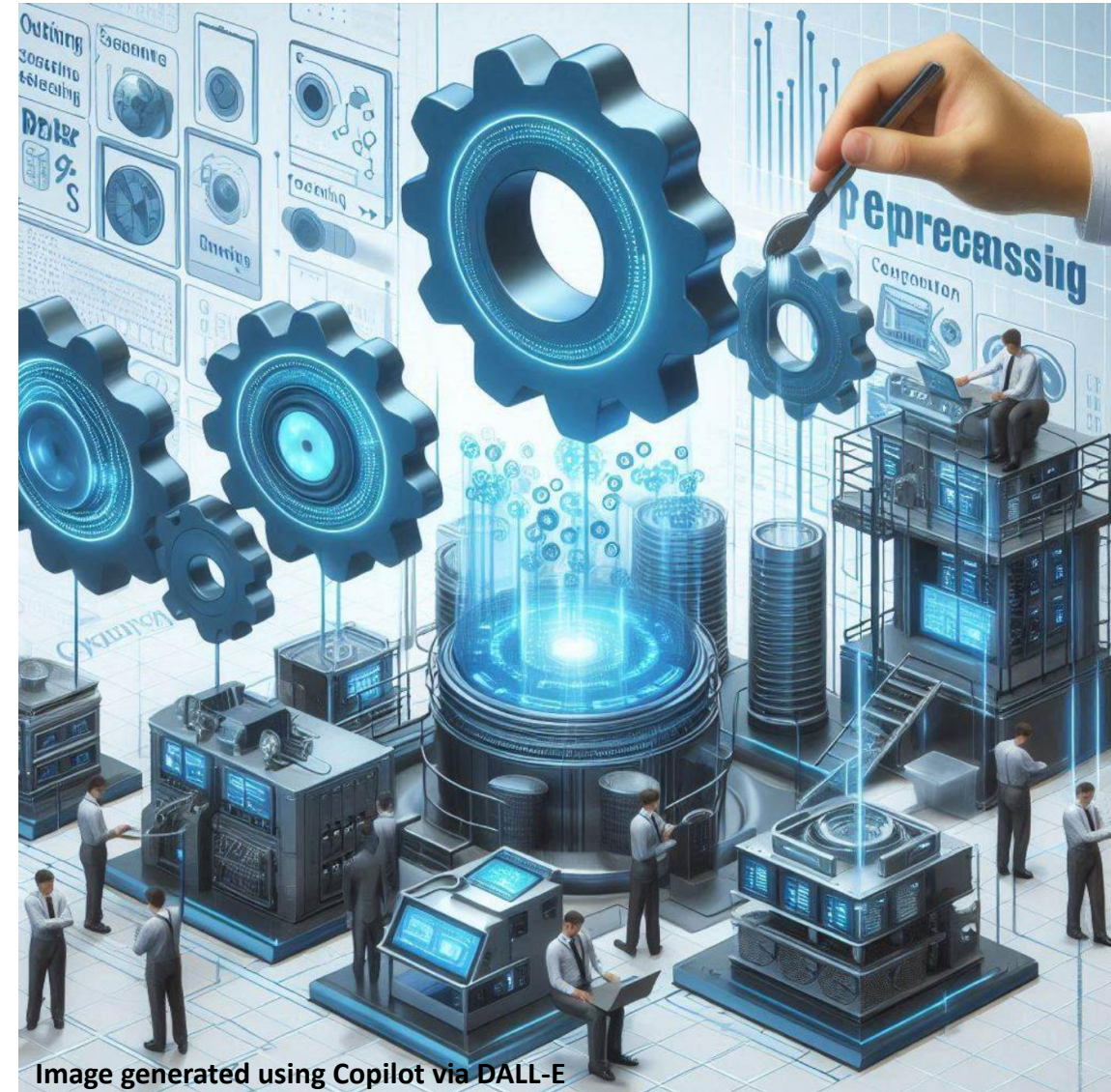
Analyzing Large & Complex Datasets

Image generated using Copilot via DALL-E

Machine Learning & Deep Learning
- AI models can analyze massive environmental datasets to **identify hidden patterns**, predict outcomes, and build informative models.
- These models can learn from vast amounts of data and uncover complex relationships between environmental variables.

Dimensionality Reduction
- AI can reduce the complexity of high-dimensional environmental data sets, making them **easier to visualize** and analyze.
- Dimensionality reduction techniques condense data sets into a smaller number of relevant dimensions while preserving essential information.

Clustering & Classification
- AI can **group similar environmental data points** together (clustering) and identify different types of patterns (classification).
- This helps researchers **understand the structure** and variability within their data, leading to more targeted analysis and informed decision-making.

clinisys

Streamlining & Optimizing Experiments

Predictive Modeling

- AI enables **prediction of experimental outcomes** based on existing environmental data, helping professionals design more efficient and effective experiments.
- This allows researchers to **optimize experimental variables** and resource allocation.

Real-time Feedback

- Having **real-time data analysis** during experiments or research, provides opportunities for adjustments and optimized data collection.
- This allows researchers to **adapt their experiments dynamically** based on real-time insights, leading to more accurate data and improved efficiency.

Automated Lab Tasks

- AI-powered robots can **automate routine lab tasks** such as sample triage, labeling, batching and testing, freeing lab professionals to focus on more complex work, data review, and any others high-value tasks.
- Lab automation streamlines workflows, **reduces human error**, increases throughput, retains compliance, while allowing for competitive turnaround times.

Image generated using Copilot via DALL-E

clinisys

01 August 2024

Predictive Analytics at Higher Scale

Image generated using Google gemini

Predictive Modeling for Environmental Outcomes
- AI can go well beyond the lab, as it empowers **environmental analysis with predictive capabilities** at a totally different scale.
- This capabilities allows researchers and environment management organizations/agencies to **anticipate potential environmental challenges** and develop preventative measures.

Real-time Environmental Monitoring & Visualization
- Besides pure lab-generated data, AI can analyze complementary data from environmental sensors in real-time, providing **continuous monitoring and visualization** of environmental conditions.
- This empowers stakeholders and interested parties to identify and **respond to environmental changes promptly.**

Identifying Emerging Environmental Threats
- AI can analyze large and diverse environmental datasets to **identify emerging threats** and potential environmental hazards.
- By detecting early warning signs, researchers, policymakers and regulators can **take proactive steps to mitigate environmental risks.**

Words of Caution

The AI & Analytics - Only as Good as the Data Entered

- Garbage in, garbage out.

- Ideally, meaningful analytics comes from multiple sources. Who do you trust data from? Will results from Lab X skew data because their results were better? Worse? Made up?

Risks

- Discrimination of environmental data based on region or wealth of one area versus another or historical use of an area or political influence on environmental affairs or…

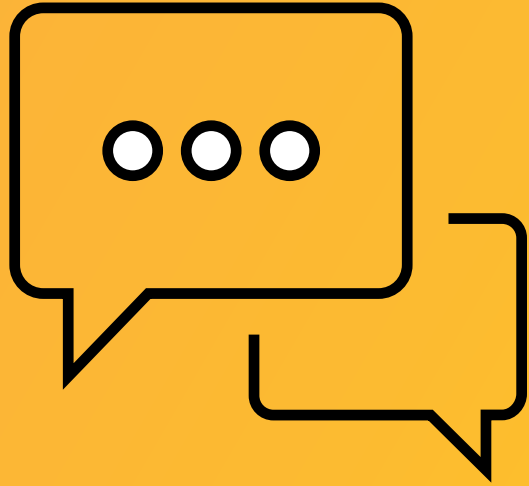## Acceptance of Analytics & AI

- While AI and analytics are incredible and powerful tools, they are new tools that will take time to fully understand and use effectively.

- In the past, many laboratories "simply" turned over the data to engineers, geologist, hydrologist, town councils, etc. to "massage" and make use of.

- For many, labs the use of analytics will be new, but there is an observed trend where more and more labs are showing interest in this powerful capability.

Conclusions

- AI can and will revolutionize environmental laboratories, automate tasks, transform the approach to data analysis, and **enable groundbreaking research**.

- As AI capabilities continue to evolve, we can expect even more transformative applications:

  - ✓ **Self-optimizing laboratories**: AI-powered systems that **autonomously manage workflows**, resource allocation, and experiment design.

  - ✓ **Hyper-personalized environmental monitoring**: AI-driven systems that **tailor environmental monitoring to specific locations** and potential risks.

  - ✓ **Democratization of environmental data**: AI facilitating **seamless data sharing** and collaboration among researchers and stakeholders.

Clinisys roadmap includes Gen AI-powered LIMS To be at the forefront scientific innovation

clinisys