AI-Supported Analysis and Research on Water Quality: Evaluation, Treatment, Prediction & Management

PARTICIPANTS' TRAINING GUIDE

Six Live Workshop Sessions — One Session Every Two Weeks

Live Session Length	2 hours per week, maximum
Total Duration	12 weeks / 12 live contact hours + short async prep
Prerequisites	None — no prior AI, ML, or programming experience required
Culminating Asset	One capstone project, built incrementally across Weeks 2–6
Final In-Person Hands-On Consolidation Workshop	3-Day Practical Training Workshop

1. Course Overview

Course Summary

This training guide supports a six-week, live virtual workshop that builds practical, sustainable AI capability for scientists and researchers working in chemistry, environmental science, biotechnology, nanotechnology, and water research. Each week centres on one 2-hour live session that blends short theory segments with hands-on coding and structured peer collaboration, applied throughout to real open water-quality assessment data. Participants build a single capstone project incrementally from Week 2 through Week 6, so that skills are practised on a real, personally meaningful problem rather than isolated exercises.

Overall Learning Objectives

- Build working AI/ML literacy — confidently distinguish AI, ML, and deep-learning approaches and choose the right one for a given water-research question
- Independently source, clean, and structure real data from USGS NWIS, EPA EDG, Kaggle, and GRDC/HydroSHEDS
- Build, evaluate, and interpret classical ML, deep-learning, and generative AI models for water-quality prediction, forecasting, and treatment optimisation
- Design a no-code/low-code automated monitoring and alerting workflow for ongoing lab use
- Apply an ethical, bias-aware lens to AI use in environmental monitoring and decision-making
- Deliver a complete, defensible capstone project and leave with a personal roadmap for embedding AI in their own lab

Prerequisite Skills

- A background in a relevant scientific domain (chemistry, environmental science, biotechnology, nanotechnology, water/hydrology, or related field)
- Basic computer literacy: comfortable using a web browser, uploading/downloading files, and following along with on-screen instructions
- A free Google account, used to access Google Colab (no local software installation required)
- No prior programming, statistics, or AI/ML background is required; all technical content starts from first principles
- Curiosity and willingness to experiment with code during live sessions

Time Commitment

Component	Time per Week
Live workshop session	2 hours (fixed weekly slot)
Asynchronous pre-session prep	≈ 20–30 minutes (short video/reading before each session)
Capstone project work (self-paced)	≈ 30–60 minutes recommended, between sessions

Materials Needed

- A laptop or desktop computer with a modern web browser and stable internet connection
- A free Google account for Google Colab
- Optional: a participant's own raw research dataset, for use in the capstone project from Week 2 onward

How This Guide Is Organised

Each of the following six week-sections follows a four-part structure: a weekly breakdown of theme, objectives, and key topics; a minute-by-minute session plan for the 2-hour live workshop; and an assessment and resources section with a short self-check quiz, the week's capstone milestone, and recommended further reading.

Six-Week Roadmap at a Glance

Week	Theme
1	Foundations of AI in STEM and Data Fundamentals
2	Classical Machine Learning for Water Quality Prediction
3	Advanced ML and Deep Learning for Treatment Optimization
4	Generative AI for Hydrological Modelling and Scientific Discovery
5	Agentic AI and No-Code Automation in Research
6	Domain-Specific Applications, Ethics, and Capstone Presentation

WEEK 1 · WORKSHOP

Foundations of AI in STEM & Data Fundamentals

1. Breakdown

Learning Objectives

- Distinguish AI, machine learning, and deep learning, and identify which water-research problems each is suited to
- Set up and navigate a Google Colab notebook without prior programming experience
- Write and run basic Python code (variables, loops, functions) using water-quality examples
- Programmatically retrieve a live dataset from USGS NWIS and a compliance dataset from the EPA EDG
- Clean, merge, and visualise real water-quality data into a single tidy table

Key Topics

- AI vs. ML vs. DL vs. data science: definitions and capability boundaries
- The applied-AI research lifecycle (problem → data → model → evaluation → responsible use)
- Google Colab environment: notebooks, cells, runtimes, Drive integration
- Python essentials: variables, data types, control flow, functions
- Pandas DataFrames and NumPy arrays
- Tidy-data principles and metadata handling
- USGS NWIS web services and EPA EDG dataset structure
- Data cleaning: missing values, outliers, timestamps

2. Session Plan (120 Minutes)

Time (mins)	Module	Format	Description
0:00–0:10	Welcome, Orientation & Icebreaker	Discussion	Introductions, workshop goals, overview of the 6-week arc and the capstone project participants will build.
0:10–0:25	The AI/ML/DL Landscape for Water Scientists	Lecture	Plain-language definitions, real examples from water research, and a myth-busting Q&A.
0:25–0:35	Live Demo: Touring Google Colab	Instructor Demo	Screen-shared walkthrough of notebooks, cells, runtimes, and saving work to Google Drive.
0:35–1:00	Hands-On: Python Basics via Water Examples	Hands-On Lab	Guided coding of variables, loops, and functions using pH/turbidity example values, run live alongside the instructor.
1:00–1:30	Hands-On: Retrieving & Cleaning Real Data	Hands-On Lab	Participants query a USGS NWIS station and an EPA EDG dataset, then clean and merge both into one tidy DataFrame.
1:30–1:45	Breakout: Data Quality Scavenger Hunt	Peer Breakout	Small groups compare datasets and catalogue data-quality issues found (gaps, units, outliers), reporting back one insight each.

Time (mins)	Module	Format	Description
1:45–2:00	Wrap-Up, Q&A & Capstone Scouting Assignment	Discussion	Recap, preview of Week 2, and an async assignment to scout a candidate dataset/question for the capstone project.

3. Assessment & Resources

Knowledge Check (Self-Assessment Quiz)

1. Which of the following best describes machine learning?

   A) A branch of AI where systems improve at a task from data rather than explicit rules B) Any code that runs automatically C) A synonym for deep learning D) A type of database

   Answer key: A

2. A tidy dataset is one where:

   A) Every column is numeric B) Each variable is a column, each observation is a row C) There are no missing values D) It is stored as a CSV

   Answer key: B

3. What is the main advantage of Google Colab for this workshop?

   A) It requires a paid subscription B) It runs Python in the browser with no local installation C) It only works offline D) It cannot access the internet

   Answer key: B

4. USGS NWIS is primarily a source of:

   A) Satellite imagery B) Real-time and historical water physical/chemical data C) Chemical patent filings D) Weather forecasts only

   Answer key: B

Practical Milestone

No formal milestone due yet — participants should have completed, reusable "Water Data Loader" notebook and begin scouting a dataset/question for their capstone project (due Week 2).

Recommended Resources

- USGS NWIS site and web-services documentation — waterdata.usgs.gov/nwis
- US EPA Environmental Dataset Gateway — edg.epa.gov
- Pandas "10 Minutes to pandas" official quick-start guide
- Google Colab official introduction notebook ("Welcome to Colab")