Applying Large Language Models in Hydraulic and Hydrologic Engineering

William Katzenmeyer, P.E., C.F.M

Water Environment Federation
the water quality people®

Why Large Language Models

Large Language Models are at the frontier of AI/ML research and capability, with widely available consumer interfaces

Hundreds of Billions of continued capital investment is underway to continue improving capabilites.

Test Time Compute and Reinforcement Learning on Reasoning Tasks push the frontier for science reasoning and coding

Water Environment Federation
the water quality people®

What Am I Doing with LLM's?

Building HEC-RAS Automation Tools

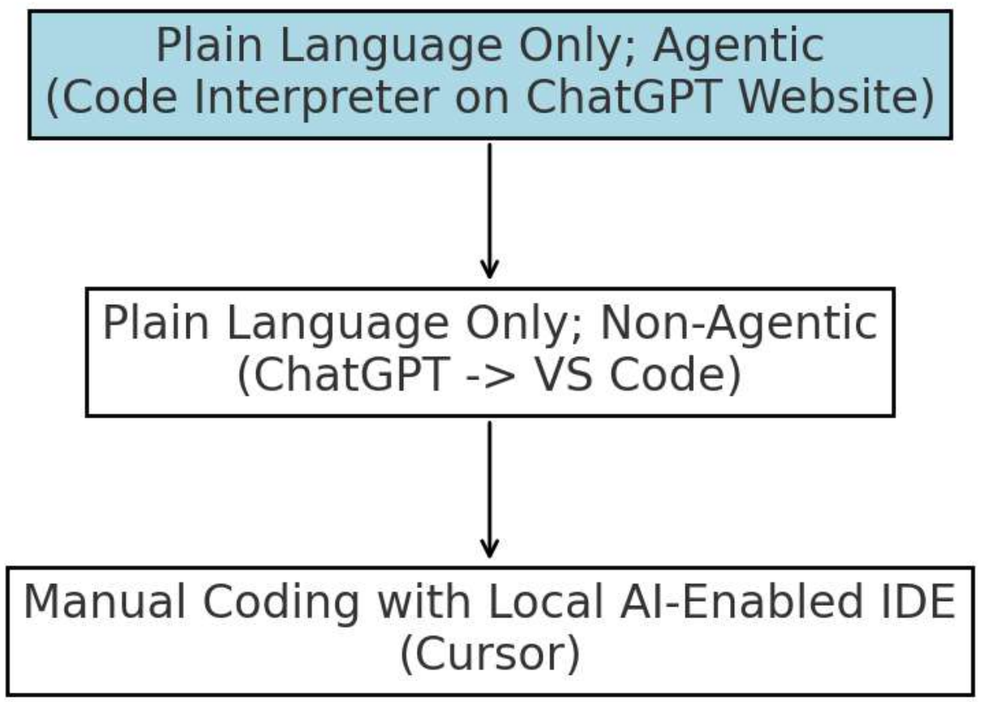
- HEC-Commander Tools
- RAS-Commander Library
- Lots More!

Daily Driver Tasks

- Writing Scripts
- Web Research
- Software Assistance
- Report Editing and Revision

Water Environment Federation
the water quality people®

3 Levels of Interaction for Leveraging Code with LLM's



- Writing and Research Assistant
  - Distilling and Summarizing
  - Addressing Grammar, Voice and Structure
  - Drafting and Editing Reports
  - Create QR Codes and Perform Research
  - Data Analysis and Figures

- Software Documentation Expert
  - User Guide with Active Reasoning
  - "How do I do this in Software Y"

- Fully Code-Enabled Assistant
  - Code Interpreter (Generic Assistant)
    - Small Helper Scripts
    - Converting GUI Workflows to Python
    - Agentic Task Automation
  - Running Simple Scripts Locally
  - Coding in your local Development Environment

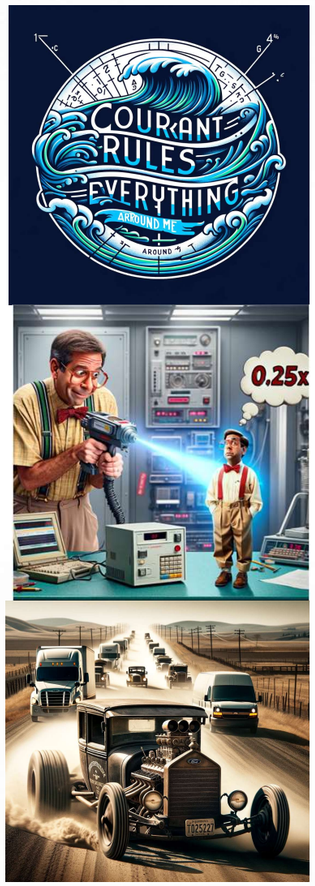
Water Environment Federation
the water quality people®

HEC-Commander Repository

Open-Source Notebooks

HEC-Commander Repository (GitHub)

Blogs



Chat GPT

Learning Assistant

Water Environment Federation
the water quality people®

HMS-Commander

AI-Coded Jupyter Notebook Supporting

- Subbasin Parameter Editing
- DSS Output File Renaming
- Impervious Grid Scaling > 1.0
- Calibration Regions by shapefile
- CSV File Input
- Enables Linked HMS>RAS Calibration Workflows
- Modular Script Ready for AI-Editing for Bespoke Applications

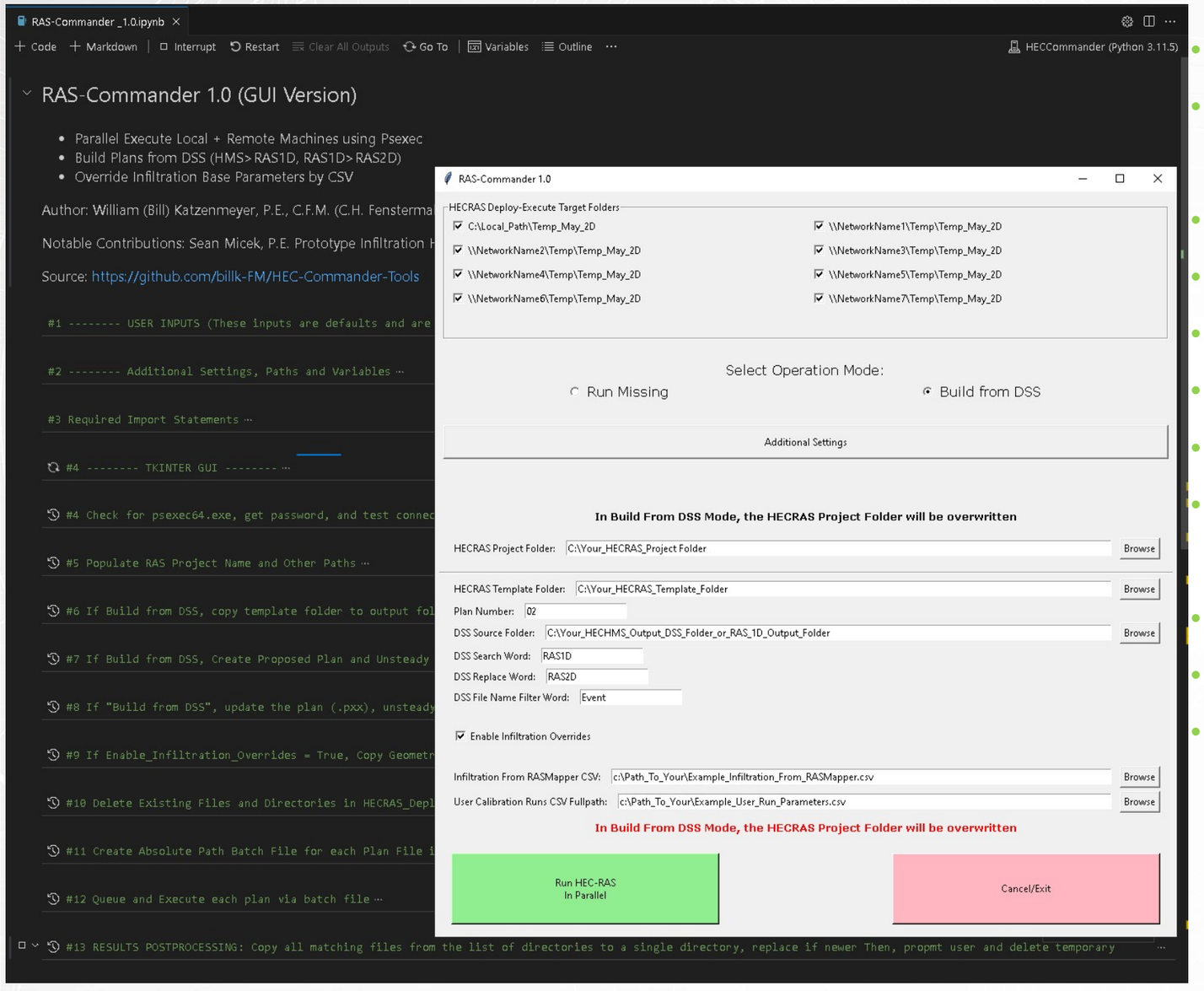
Water Environment Federation
the water quality people®

HMS – Commander

Example CSV Input for HHS-commander & RAS-Commander 2D Infiltration Overrides

User run number From csv	Initial Deficit scale	Maximum Deficit scale	Percolation rate scale	Impervious area scale	Recession factor	Initial flow area ratio	Threshold Flow to Peak ratio	Time of concentration scale	Storage Coefficient scale
1	1.0	1	0.06	1	0.1	1	0.1	1	1
2	0.9	1	0.06	1	0.1	1	0.1	1	1

RAS-Commander
Parallelizing HEC-RAS across multiple networked machines via a Jupyter Notebook



AI-Coded Jupyter Notebook Supporting:

- Parallel HEC-RAS Execution
- Windows Native: Supports All Versions
- Leverage Multiple Workstations in Parallel
- Open Source, MIT License

Basic Components

- User Input and Settings
- New! Tkinter GUI
- File Deploy and Copy
- Batch File Creation
- Command Line Execution
- Results Collection

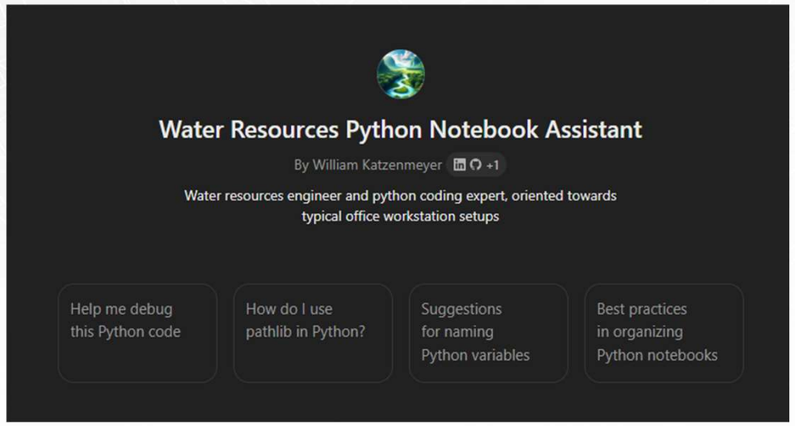
Flexible Operation:

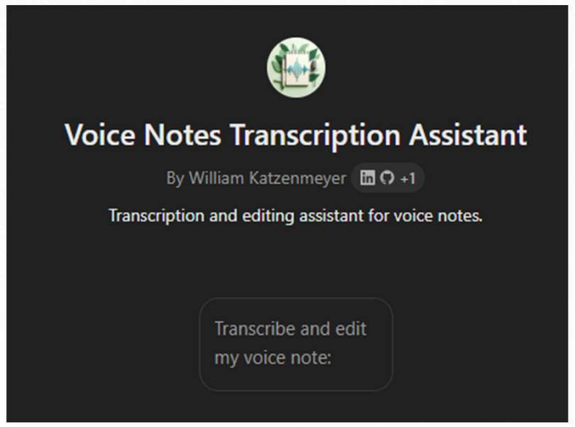
- Bring Your Own Project
- Create Plans from HMS DSS Input Files
- Optional 2D Infiltration Overrides

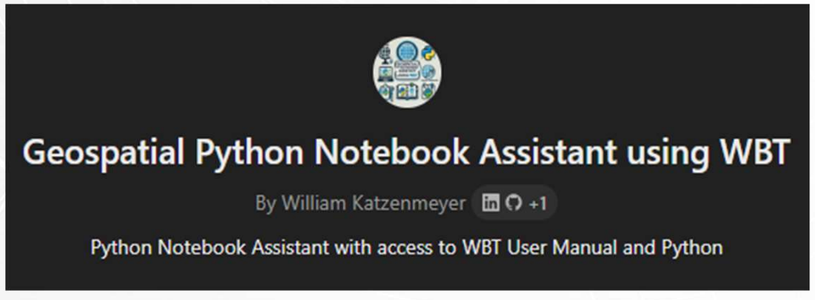
RAS-Commander is ready for AI-Assisted editing by end users to support their bespoke applications

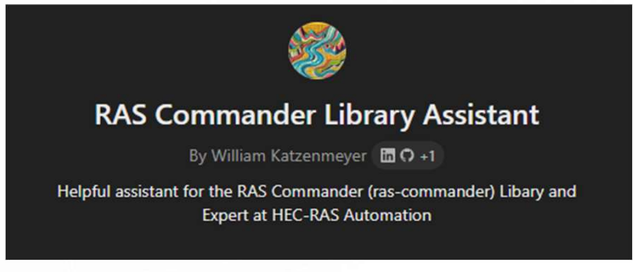
Water Environment Federation
the water quality people®

GPT Assistants



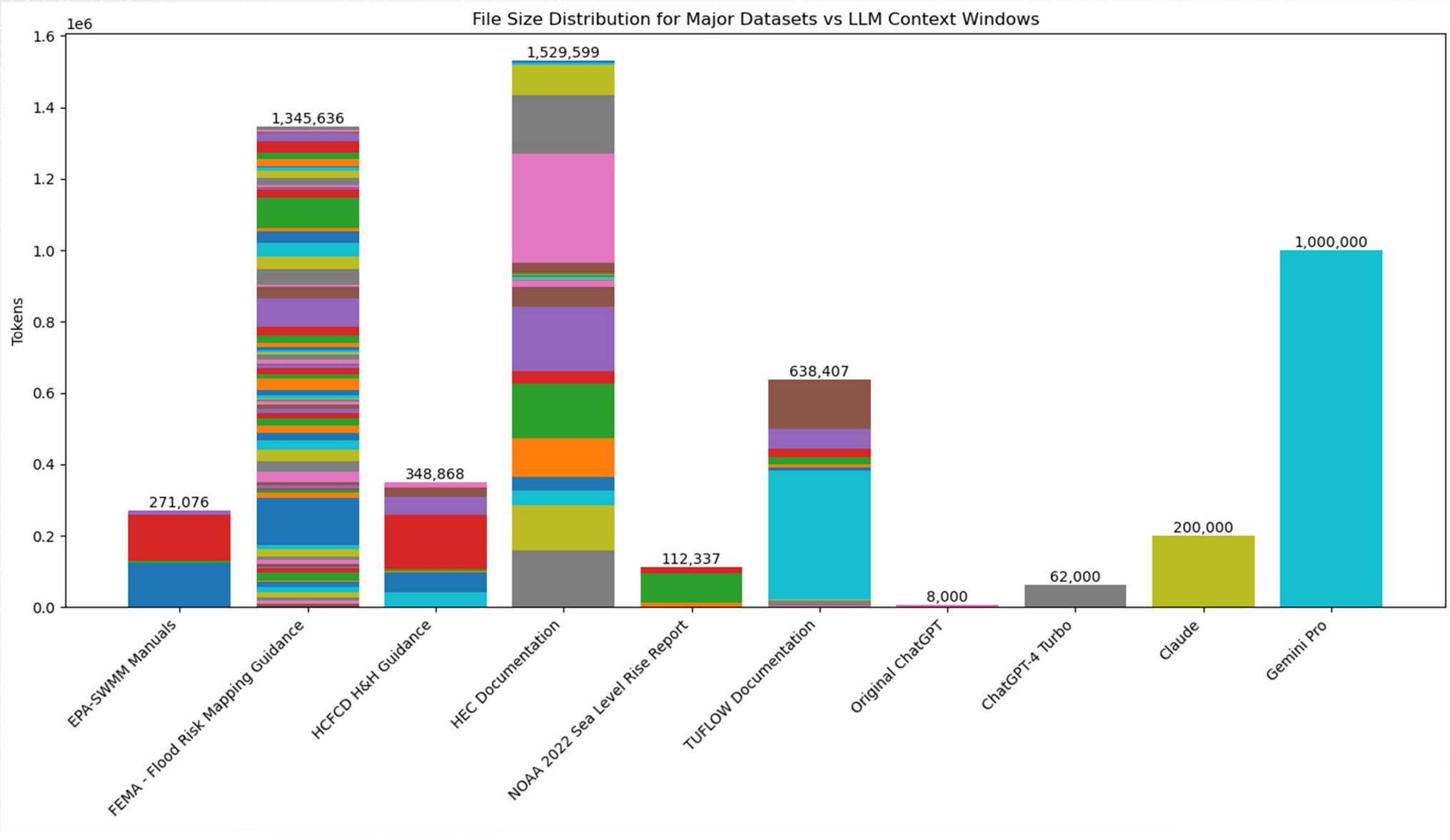






Water Environment Federation
the water quality people®

Context Window Limitations



Water Environment Federation
the water quality people®

Avoiding the Pitfalls and Limitations of LLM's

1

False Recall (Hallucinations)

2

Confident, Plausible Errors/ High Error Rates

3

Non-Deterministic Operation

4

Limited Input and Output

5

Rapidly Changing Technologies, Terminology and Platforms

6

Data Privacy and Confidentiality

Water Environment Federation
the water quality people®

Neural Networks are incredible at Compression

BUT Compression leads to Errors and Hallucinations

Neural Networks compress language data into latent representations, but it's a lossy compression. A model's weights are best relied upon for reasoning and not it's recall.

This lossy recall manifests as made-up references, misquotes and other incorrect information

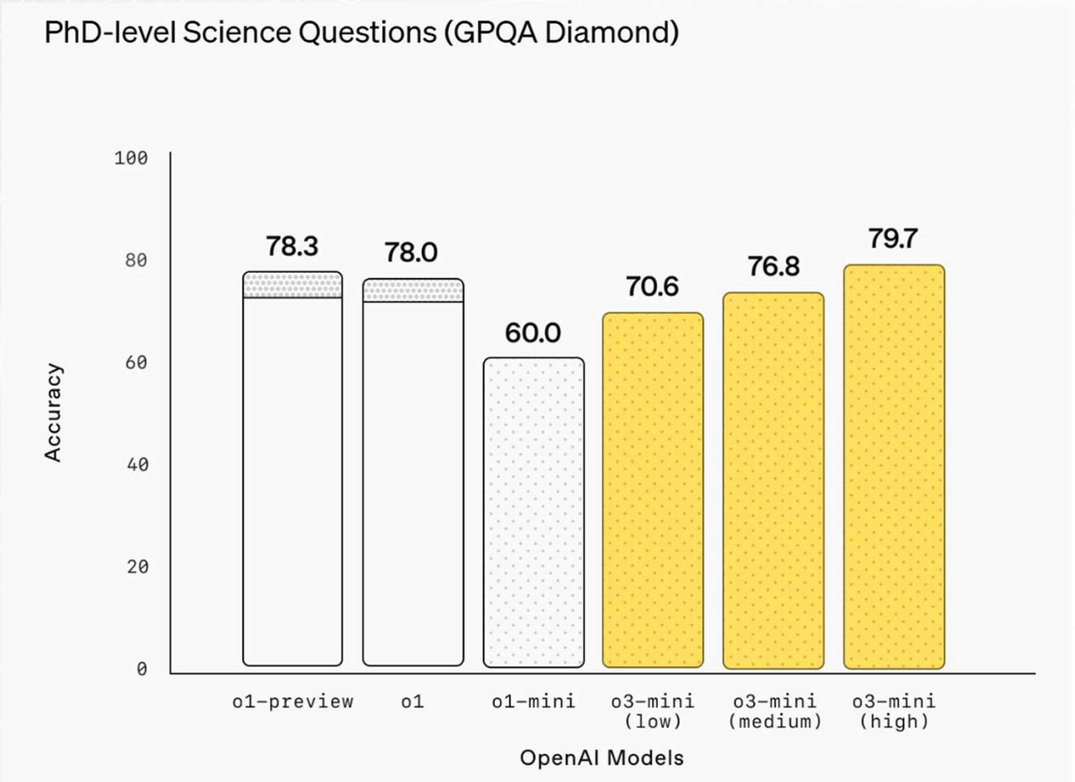
The best LLM platforms provide the models with specific context and valid links for the model to formulate a response, such as:

- **Perplexity**
- **OpenAI and Google "Deep Research" Tools**

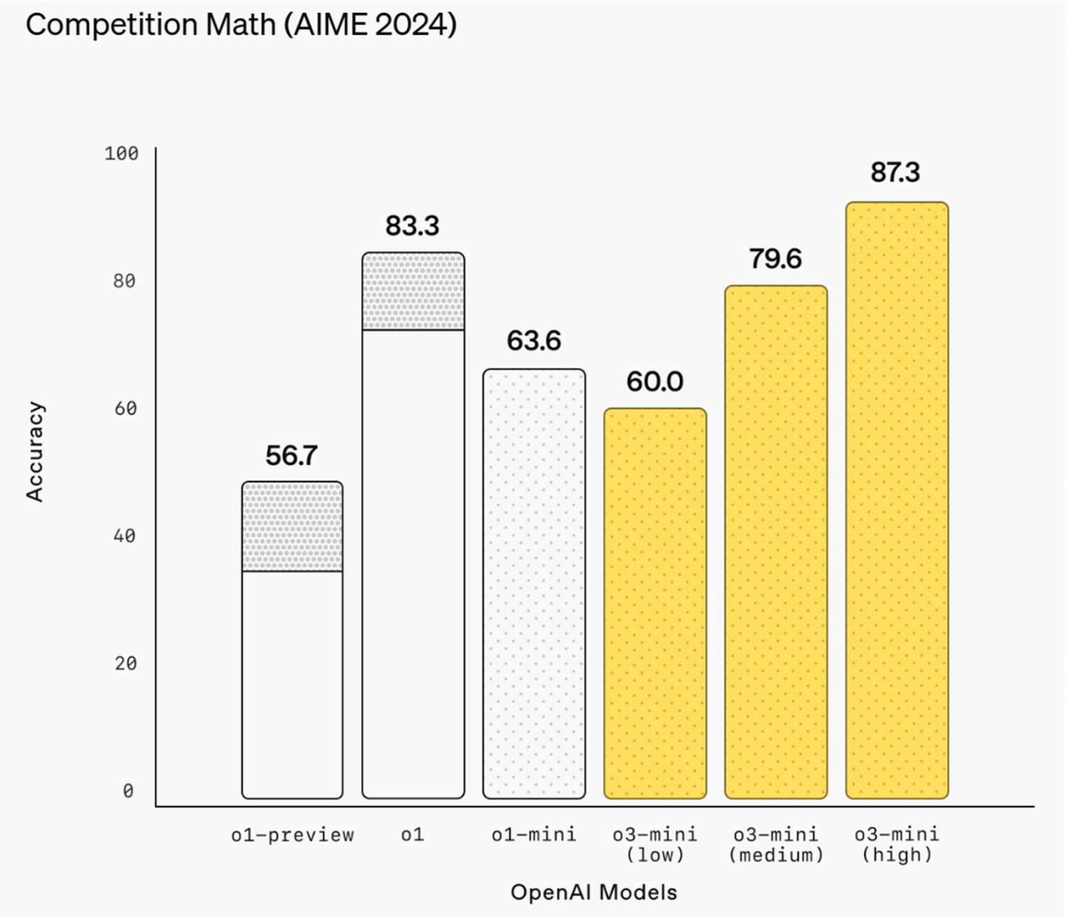
To reduce errors and hallucinations, providing accurate context is key!

Water Environment Federation
the water quality people®

Tempering Expectations



PhD-level science: *On PhD-level biology, chemistry, and physics questions, with low reasoning effort, OpenAI o3-mini achieves performance above OpenAI o1-mini. With high effort, o3-mini achieves comparable performance with o1.*

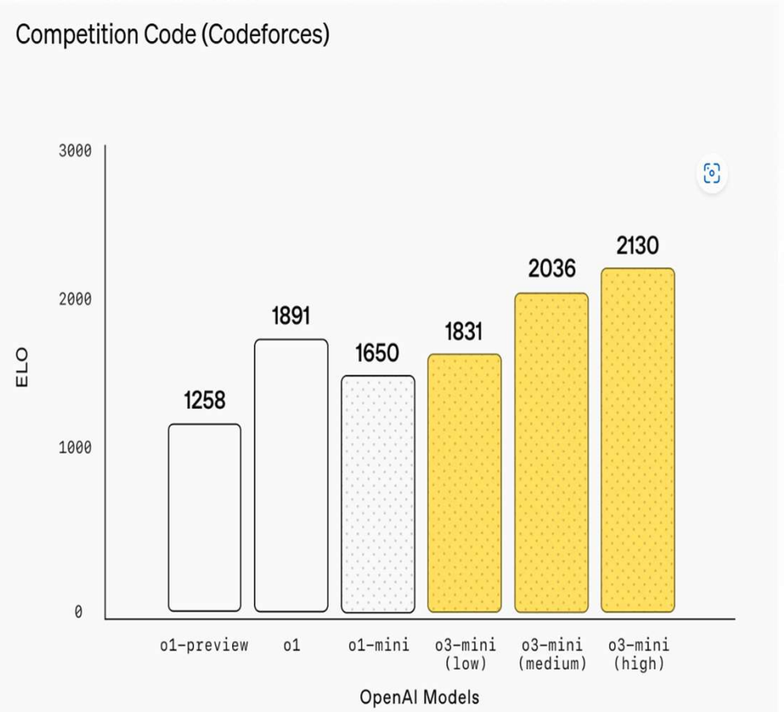


Water Environment Federation
the water quality people®

Coding Performance Scores

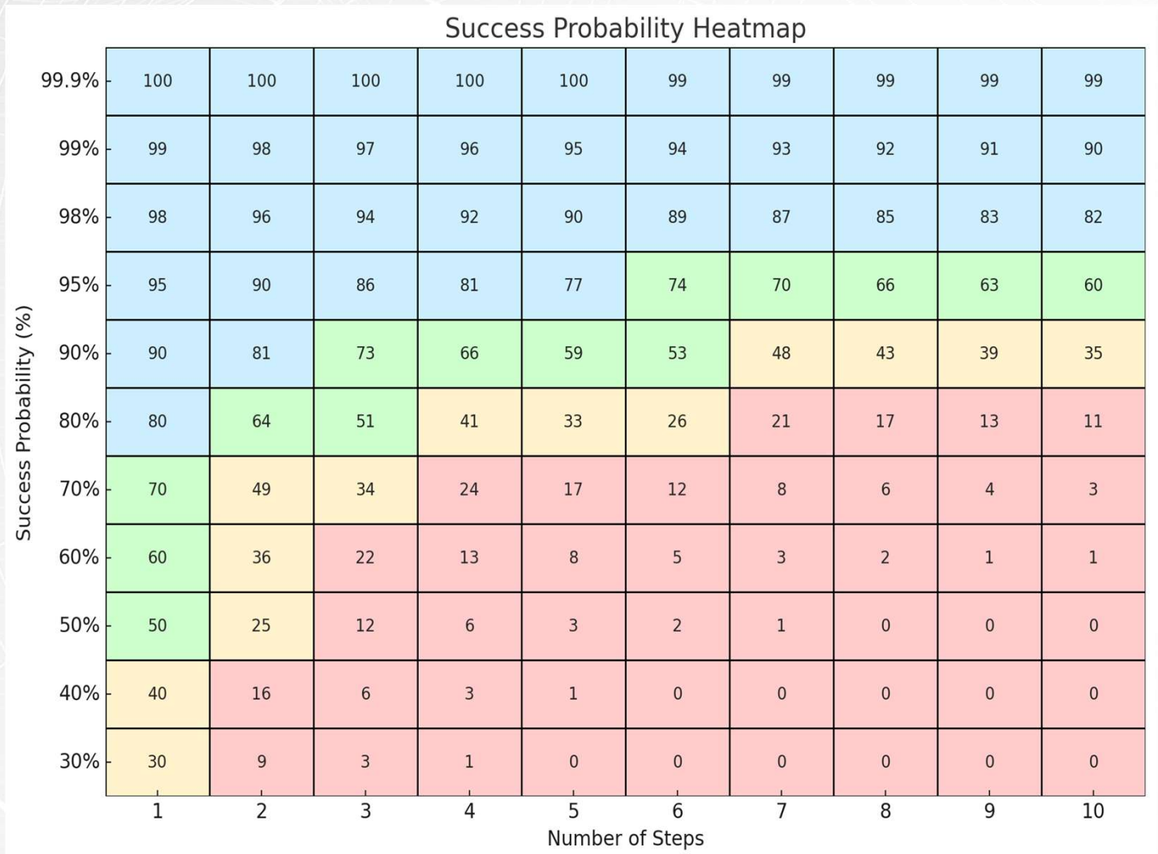
Benchmark (Metric)	DeepSeek V3
Architecture	MoE
# Activated Params	37B
# Total Params	671B
MMLU (EM)	88.5
MMLU-Redux (EM)	**89.1**
MMLU-Pro (EM)	75.9
DROP (3-shot F1)	**91.6**
IF-Eval (Prompt Strict)	86.1
GPQA-Diamond (Pass@1)	59.1
SimpleQA (Correct)	24.9
FRAMES (Acc.)	73.3
LongBench v2 (Acc.)	**48.7**

HumanEval-Mul (Pass@1)	**82.6**
LiveCodeBench (Pass@1-COT)	**40.5**
LiveCodeBench (Pass@1)	**37.6**
Codeforces (Percentile)	**51.6**
SWE Verified (Resolved)	42.0
Aider-Edit (Acc.)	79.7
Aider-Polyglot (Acc.)	**49.6**
AIME 2024 (Pass@1)	**39.2**
MATH-500 (EM)	**90.2**
CNMO 2024 (Pass@1)	**43.2**
CLUEWSC (EM)	90.9
C-Eval (EM)	**86.5**
C-SimpleQA (Correct)	**64.1**



***Competition coding**: On Codeforces competitive programming, OpenAI o3-mini achieves progressively higher Elo scores with increased reasoning effort, all outperforming o1-mini. With medium reasoning effort, it matches o1's performance.*

Agentic Operation and the March of 9s



Benchmarks are starting to represent the difficulty that we encounter in practice and are beginning to approach success rates that allow multi-step agentic behavior.

- *X Axis:*

Number of Sequential Steps

- *Y Axis:*

Success Probability for Single Response

- *Heatmap shows probability of success of multiple steps @ similar difficulty.*

Water Environment Federation
the water quality people®

Prompt Engineering

- The most impactful tips and tricks for improving your prompts generally revolve around providing the AI custom instructions and context

When prompting an LLM, focus on

- Providing clear, well-structured directions
- Use Delineators to Separate Instructions from Context

- **Understand the Limitations**
  - Limited Context Windows
  - Limited Retrieval from Large Documents
  - Probabilistic Operation, not Deterministic
  - File size and library limitations in Code Interpreter
  - No internet access (blame the AI safety patrol)

Basic "prompt engineering" is typically:

- Role (Persona)
- Constraints
- Contextual Data
- Instructions
- Desired Output
- Examples

- Prompts can also be structured as code

Your AI assistant does not know what it is doing here, unless you tell it.

Water Environment Federation
the water quality people®

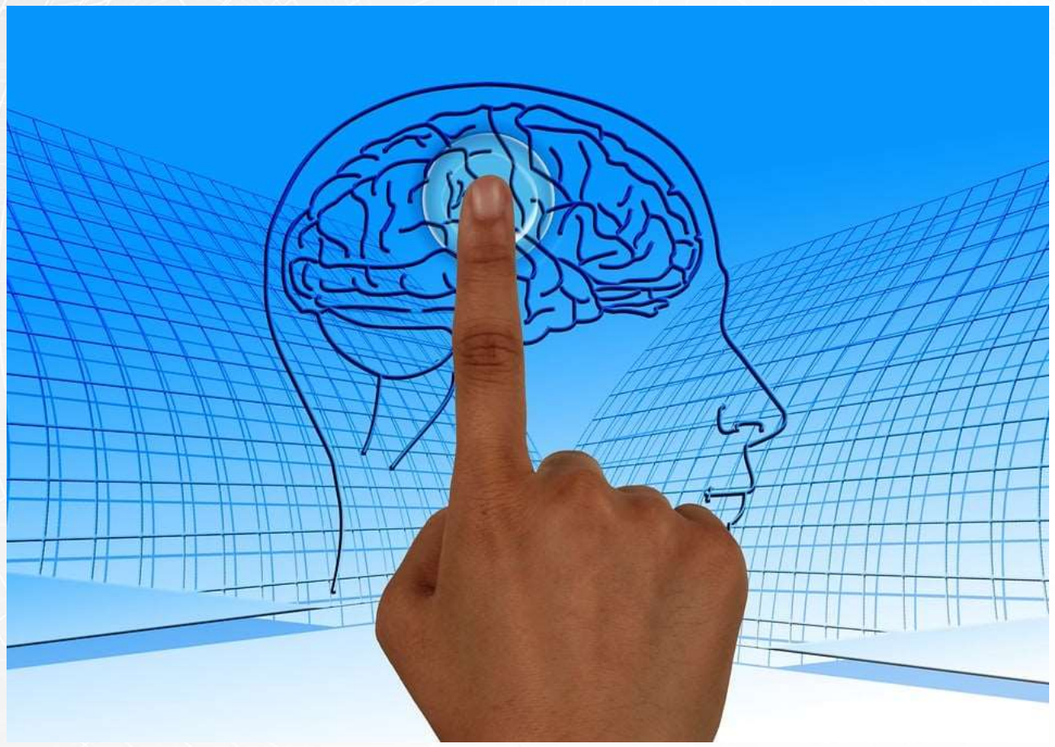
Maintaining Responsible Charge of Your Work

None of our tools are perfect. As a licensed professional, you are solely responsible for maintaining direct control and responsible charge of all aspects of your work product. Human review of all outputs will remain a steadfast requirement regardless of the advancements in Large Language Model capabilities.

As LLM's become more and more reliable, spotting the errors will require additional diligence. With better tools at our disposal, we get to spend more time on higher value tasks by directing agents to perform repetitive, verifiable tasks, data analysis, and bespoke software development.

Water Environment Federation
the water quality people®

Learn more



HEC-Commander Tools Repository on Github

RAS-Commander Library on GitHub (Coming Soon)

Full Momentum Vodcast

Full Momentum Episode 33: The Future of leveraging HEC-RAS with Automation and AI

Australian Water School Free Webinars on Youtube

AI Tools for Modeling Innovation

Applying AI to HEC-RAS Modeling Workflows

Australian Water School Paid Webinars and Courses

Short Course: AI Applications to HEC-RAS modelling workflows

Full Course: AI-assisted coding for water modellers

Youtube Demonstrations

https://www.youtube.com/@GPT_Commander

Water Environment Federation
the water quality people

Q&A

Discussion

Water Environment Federation
the water quality people®