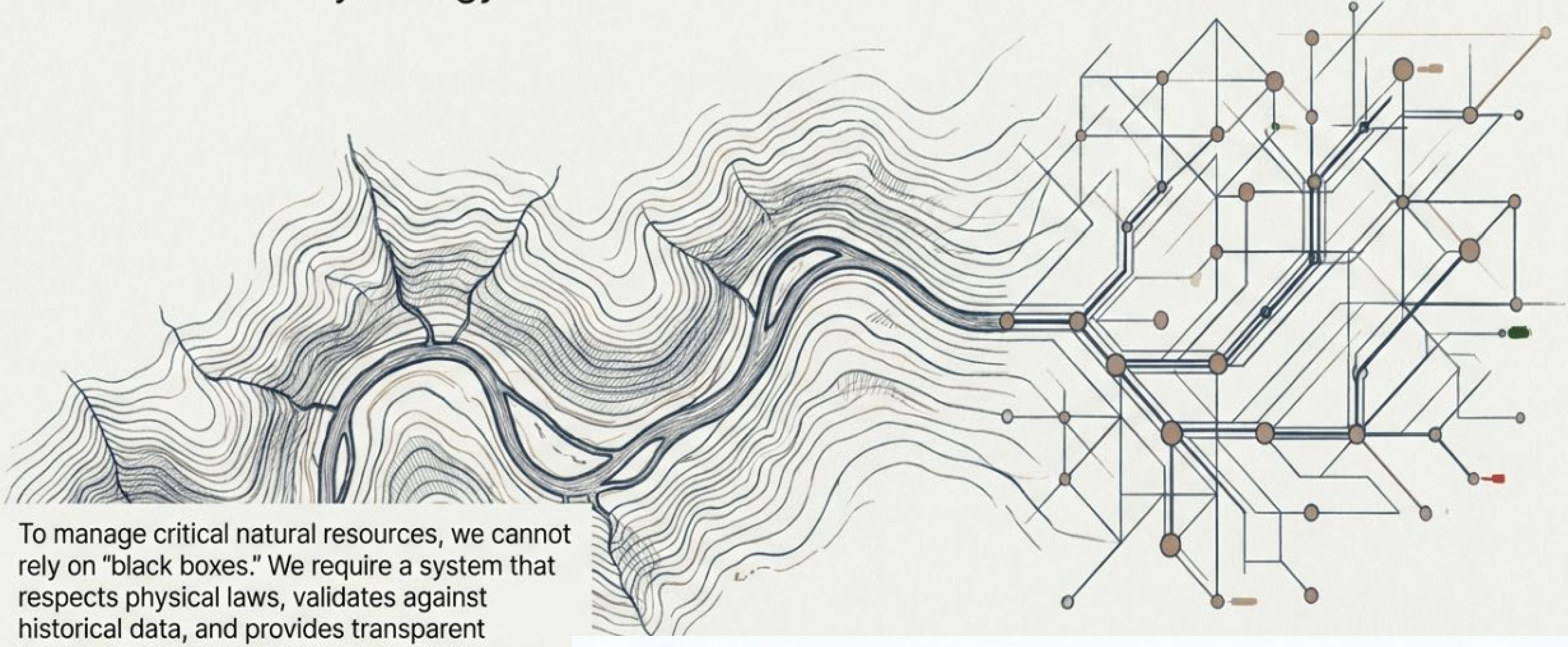
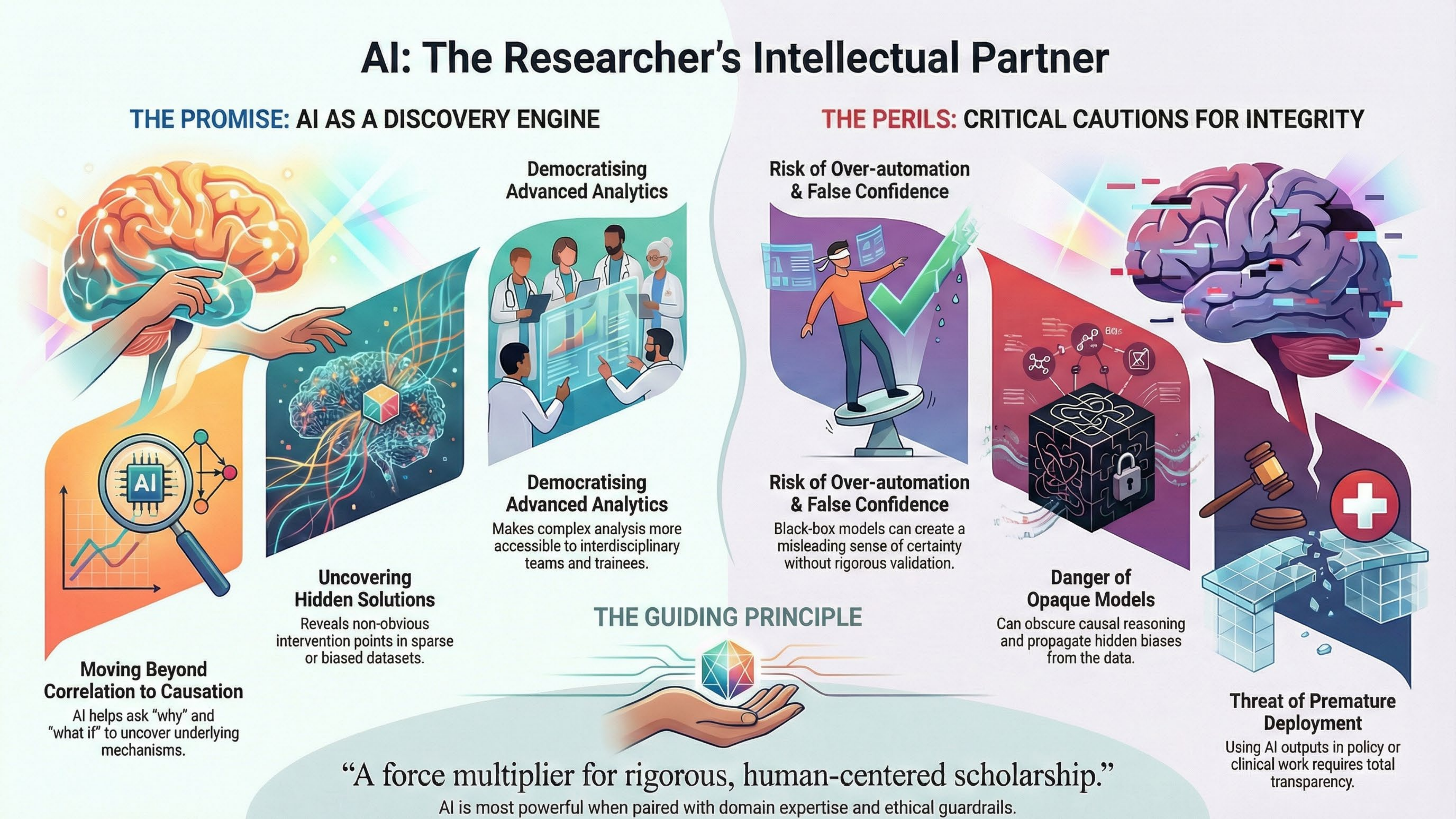
The Pillars of Trustworthy AI

Bridging Physics, Data, and Decision-Making in Climate and Hydrology.



To manage critical natural resources, we cannot rely on "black boxes." We require a system that respects physical laws, validates against historical data, and provides transparent outputs for high-stakes decision-making.

Based on the research and frameworks of Dr. Debaditya Chakraborty

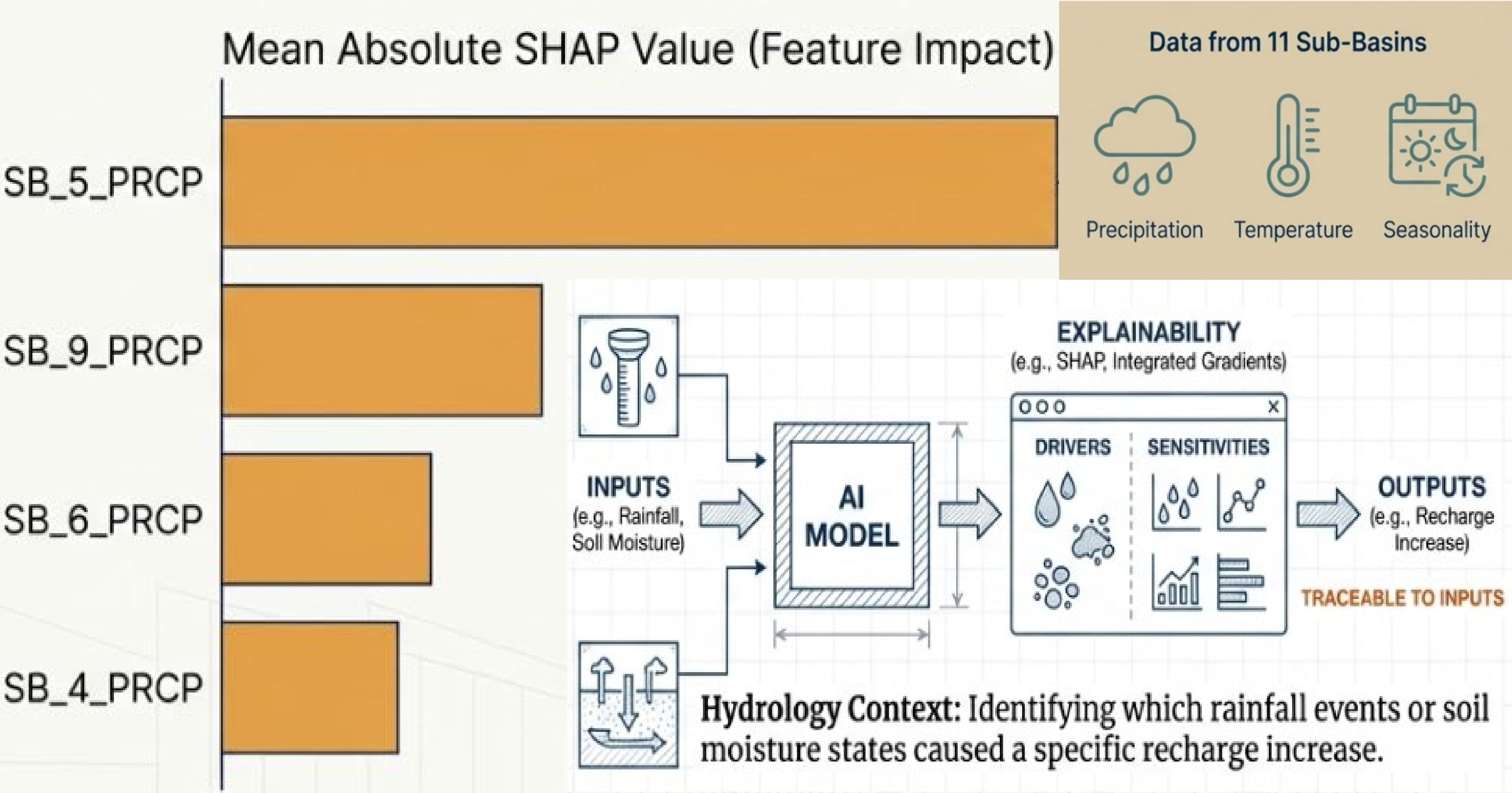

AI: The Researcher's Intellectual Partner
THE PROMISE: AI AS A DISCOVERY ENGINE
THE PERILS: CRITICAL CAUTIONS FOR INTEGRITY
Democratising Advanced Analytics
Risk of Over-automation & False Confidence
AI
Democratising Advanced Analytics
Makes complex analysis more accessible to interdisciplinary teams and trainees.
Risk of Over-automation & False Confidence
Black-box models can create a misleading sense of certainty without rigorous validation.
Uncovering Hidden Solutions
Reveals non-obvious intervention points in sparse or biased datasets.
Danger of Opaque Models
Can obscure causal reasoning and propagate hidden biases from the data.
THE GUIDING PRINCIPLE
Moving Beyond Correlation to Causation
AI helps ask "why" and "what if" to uncover underlying mechanisms.
Threat of Premature Deployment
Using AI outputs in policy or clinical work requires total transparency.
"A force multiplier for rigorous, human-centered scholarship."
AI is most powerful when paired with domain expertise and ethical guardrails.

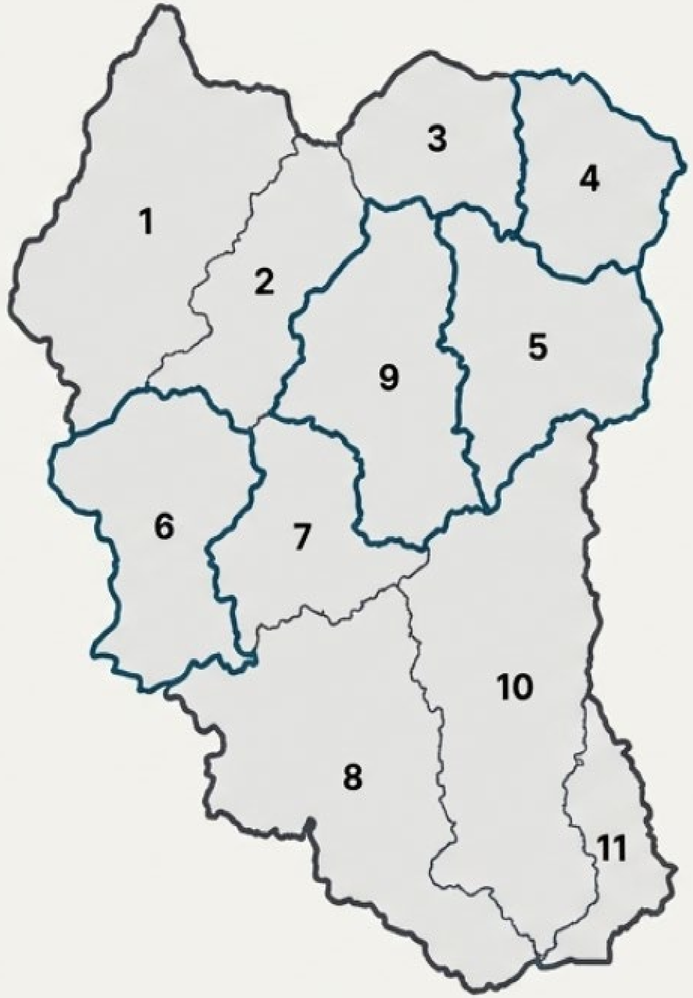
Core Question: Can Users Understand Why the AI Produced a Result?

Verified Model Performance

0.7894

Test Data R^2





Methodology: Physically Meaningful Units (Sub-Basins) outperformed Arbitrary Geometric Grids.

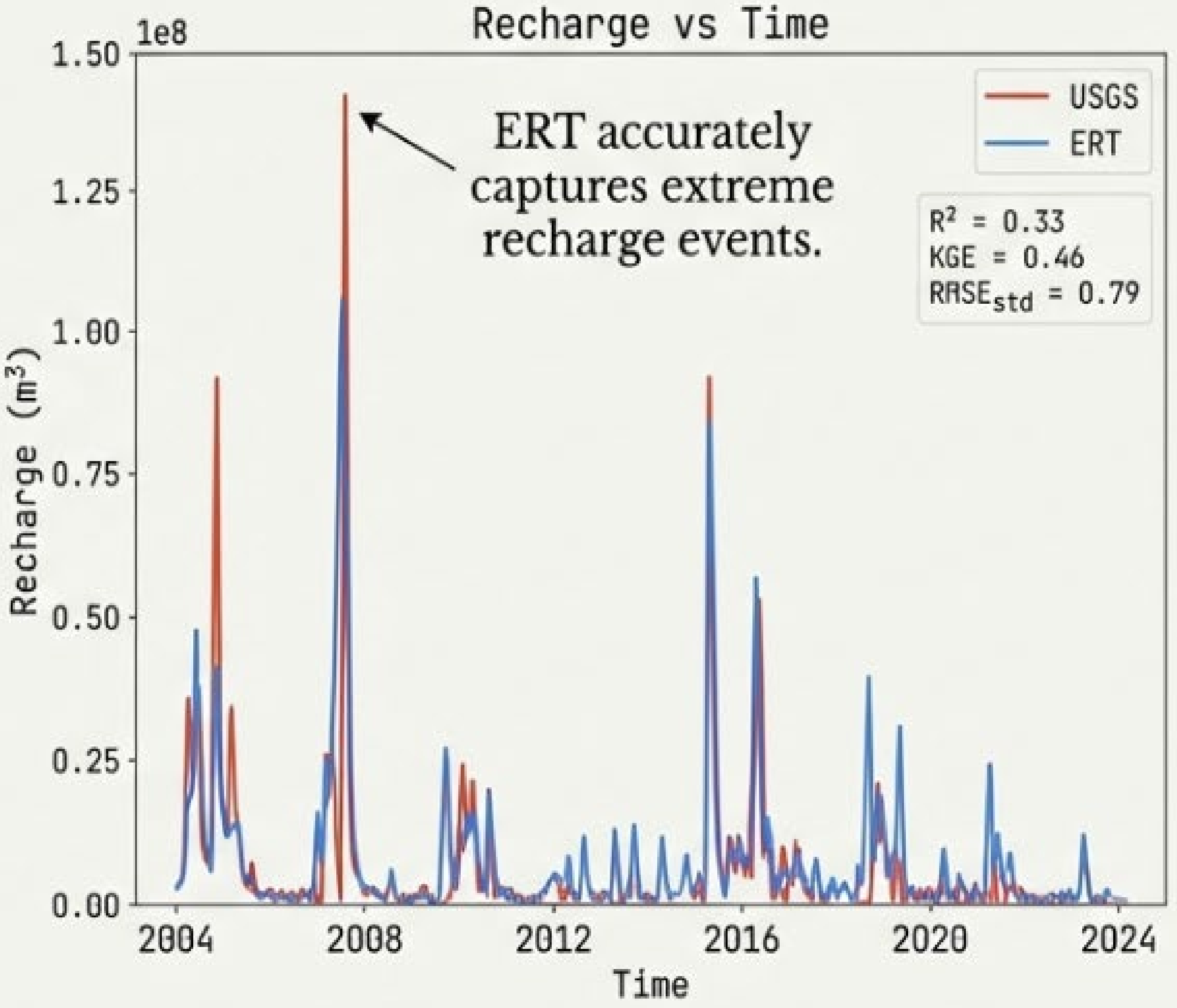
Model Selection: Identifying the Superior Architecture

The Contenders:

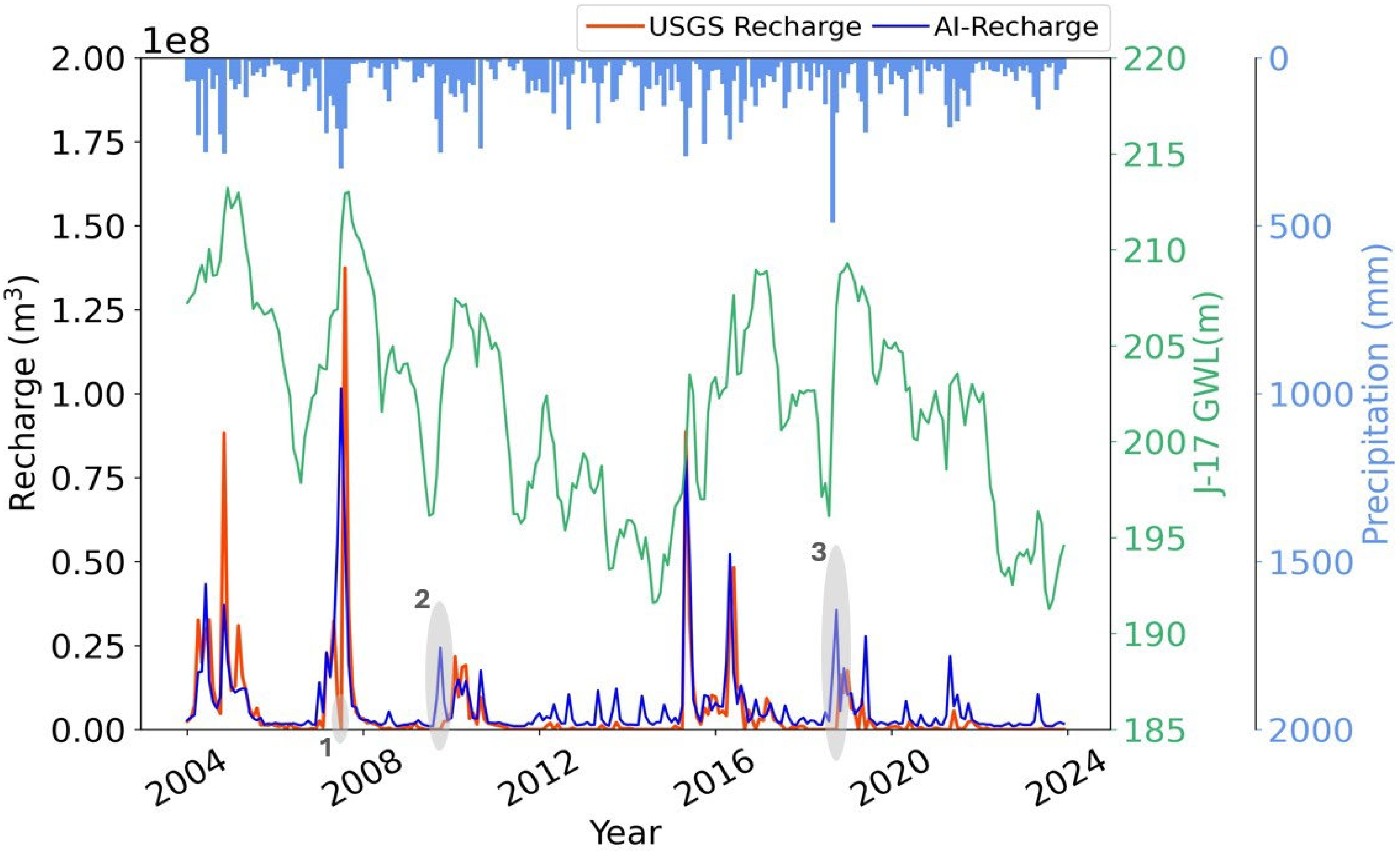
Five architectures tested: ERT, Random Forest (RF), XGBoost, HGBoost, and CNN.

The Winner: Extremely Randomized Trees (ERT)

- Peak Performance: consistently captured the timing and magnitude of peak recharge events (accounting for 70–80% of total recharge).
- The Failures: RF and CNN struggled with overfitting or timing shifts. Boosting models occasionally overestimated during droughts.



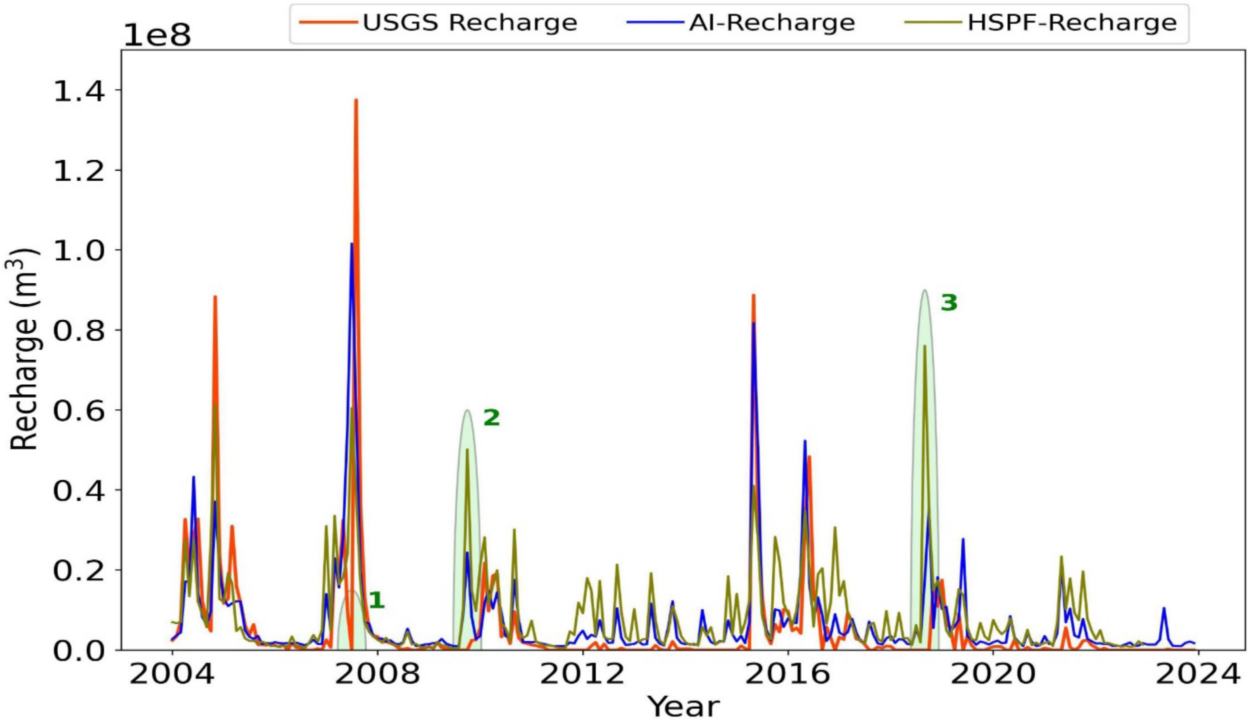
The Divergence: AI Detects “Hidden” Recharge Events



Conflict: AI, trained on physical models, predicts significant water entering the system when the underlying physical models assume zero recharge. Is AI **Hallucinating** or **Discovering**?

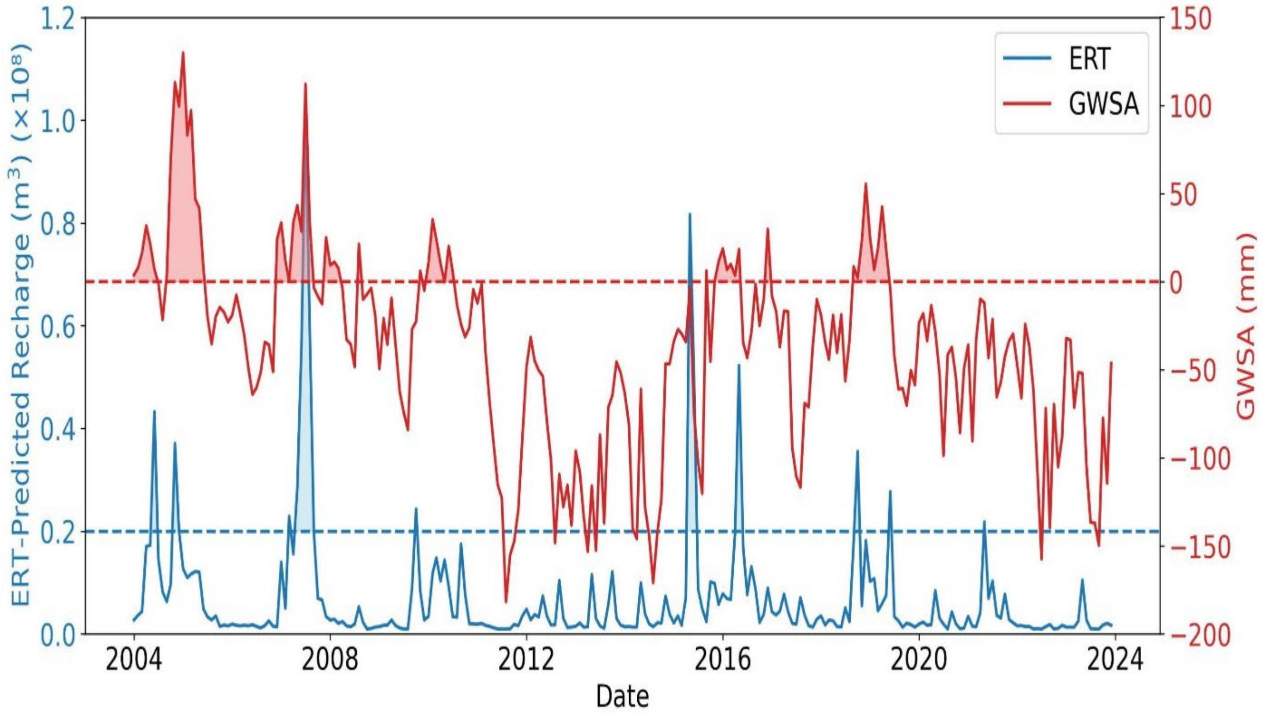
VERIFICATION FROM SPACE AND SIMULATION

Evidence A: HSPF Simulation



Simulation confirms the timing of the disputed recharge events.

Evidence B: GRACE Satellite Data



Satellite gravity data confirms positive storage anomalies during AI-predicted recharge events.

Forecasting Aquifer Health: Tree-Based AI Outperforms Deep Learning

Predicting water levels in the vital Edwards Aquifer is crucial.
A study compared two AI model families for 1-12 week spring flow and groundwater level forecasts.

XGBoost: The Most Accurate & Stable Performer

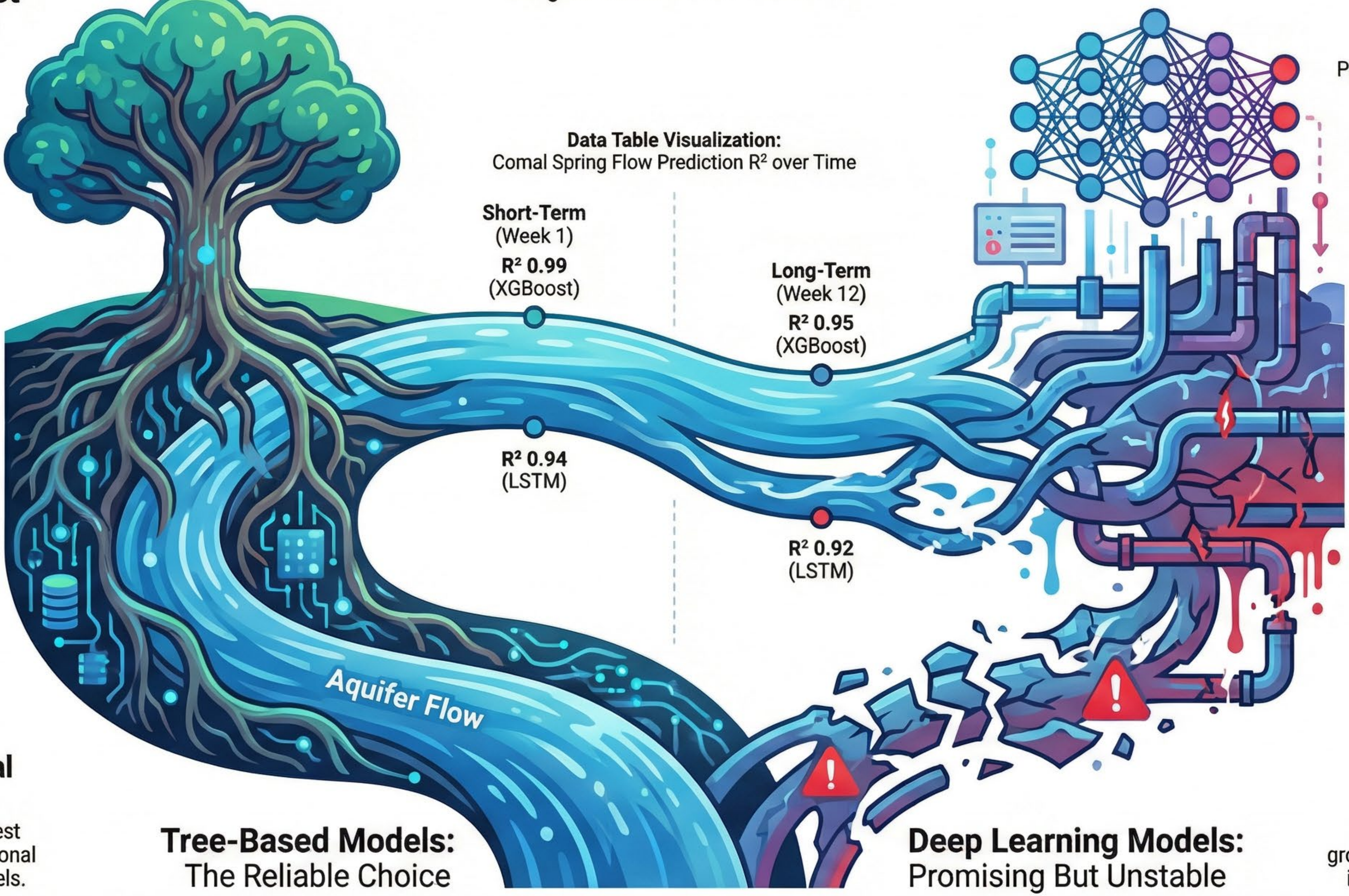
Consistently delivered the highest accuracy across all 1-12 week forecast periods.

Robust Even in Extreme Low-Flow Conditions

At week 8, XGBoost correctly predicted critical low groundwater levels (CS4) 50% of the time.

Efficient and Practical for Real-World Use

Achieved top results with modest data needs and less computational power than deep learning models.

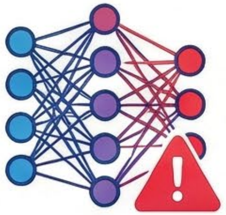


Tree-Based Models: The Reliable Choice

Deep Learning Models: Promising But Unstable

Strong Start, But Faltering Finish

Performed well in the short-term (1-4 weeks), but accuracy and stability declined significantly over longer periods.



Prone to Catastrophic Failure

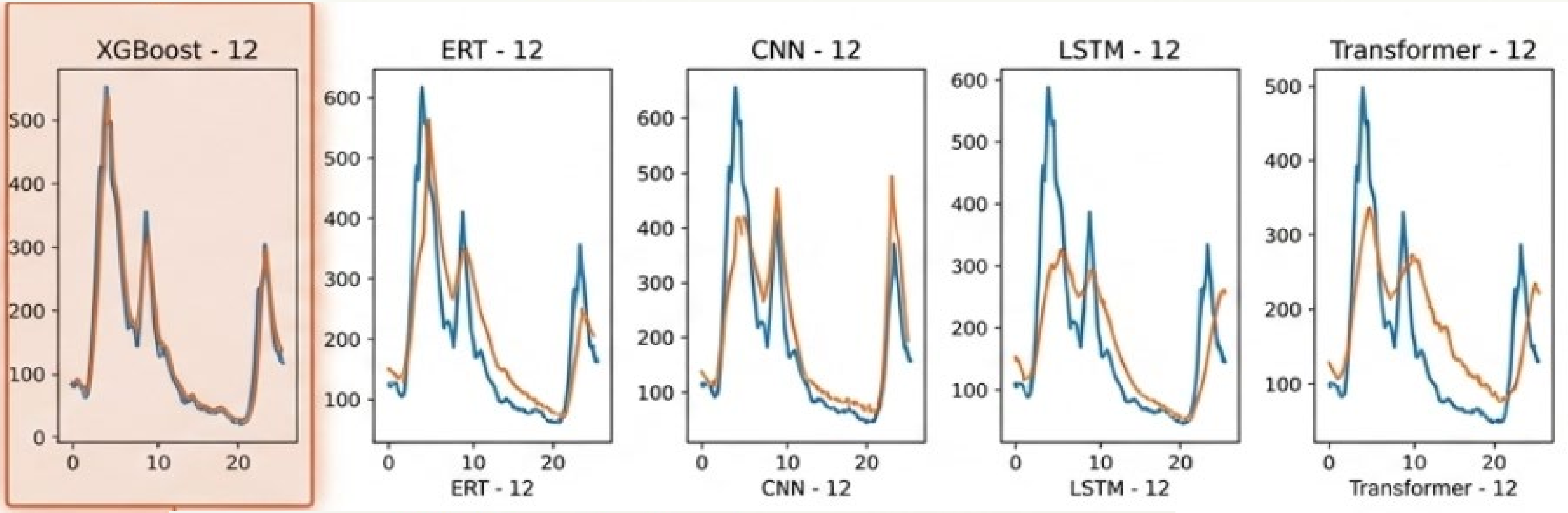
The CNN model completely broke down during week 7 (spring flow) and week 9 (groundwater) forecasts.

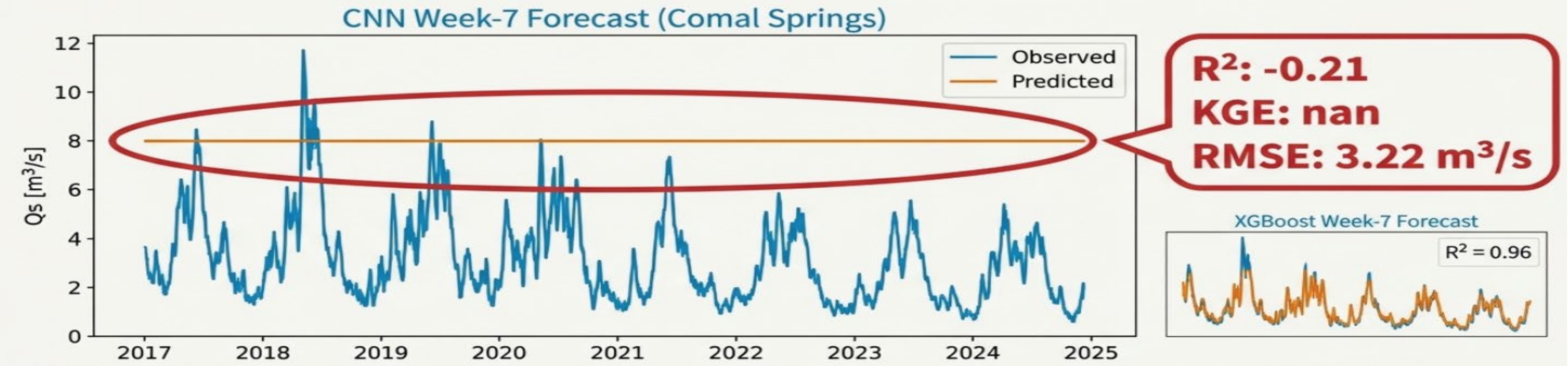
Struggled with Extreme Conditions

LSTM accuracy for critical low groundwater levels (CS4) fell to 0% in some medium-term forecasts.

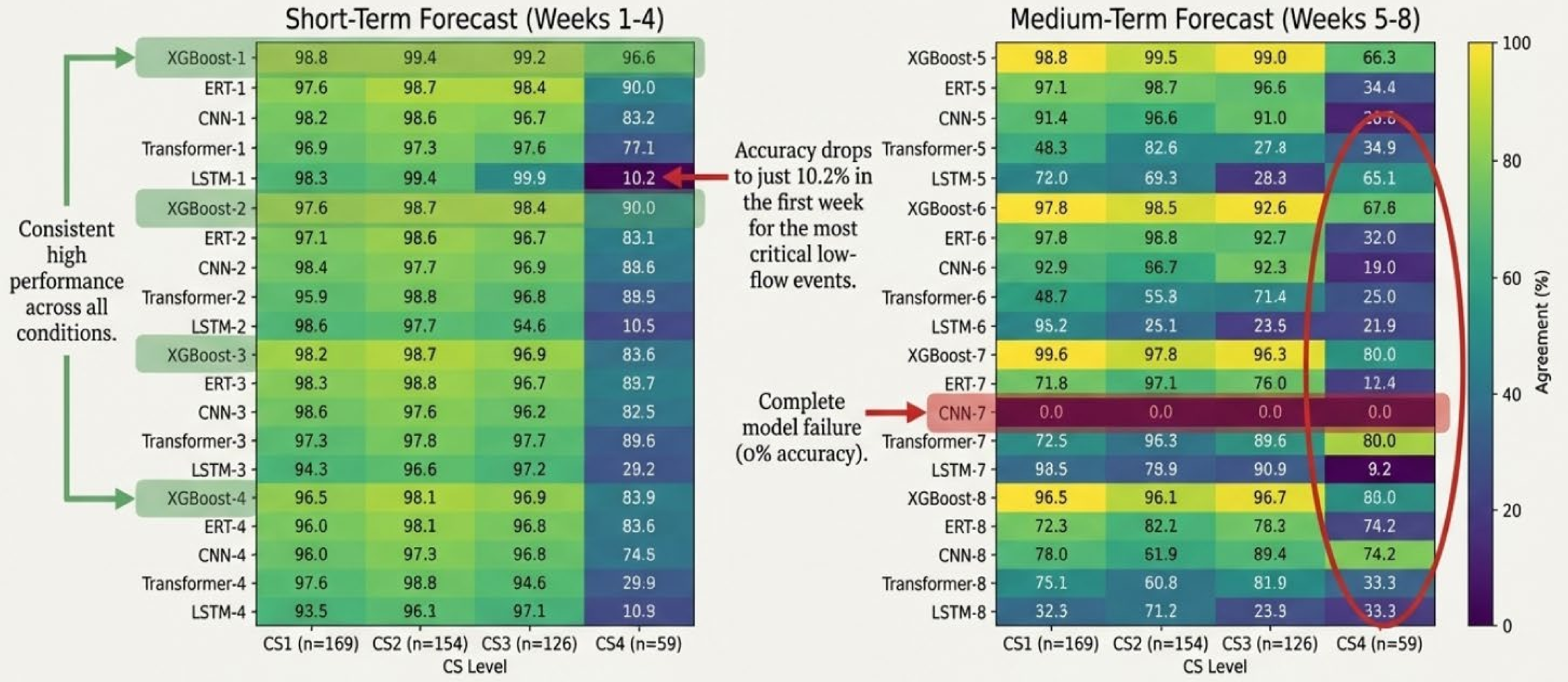
Visualising the Collapse: Stability vs. Instability

Observed (blue) and Predicted (orange) values illustrate dramatic differences in model performance.





Deep Dive: Accuracy Across All Critical Stages



The full performance matrix reveals a clear pattern: XGBoost maintains high accuracy across a wide range of conditions, while deep learning models show significant and unpredictable weaknesses, particularly as flow conditions become more critical.

Complexity Does Not Equal Accuracy

Despite the hype surrounding Deep Learning, simpler Tree-Based models (XGBoost) demonstrated superior performance, stability, and operational value compared to complex architectures like Transformers and LSTMs.

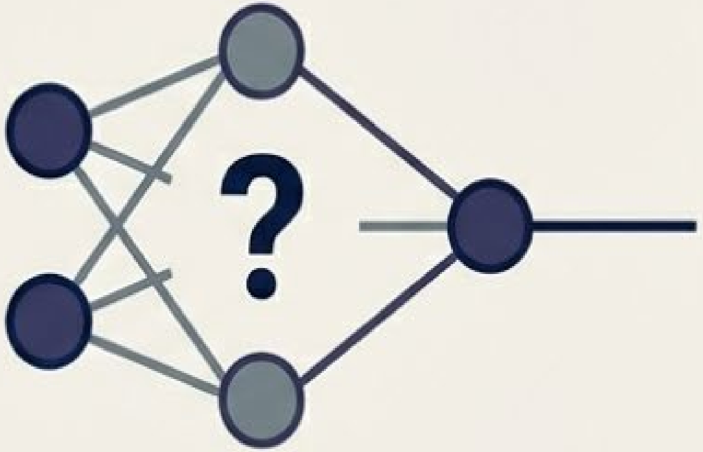
Data Scarcity

Temporally rich but spatially limited data prevents complex models from generalizing.

Over-Smoothing

Transformers and CNNs dampen volatility, missing the peaks and extremes essential for drought triggers.

The 'Black Box' Risk

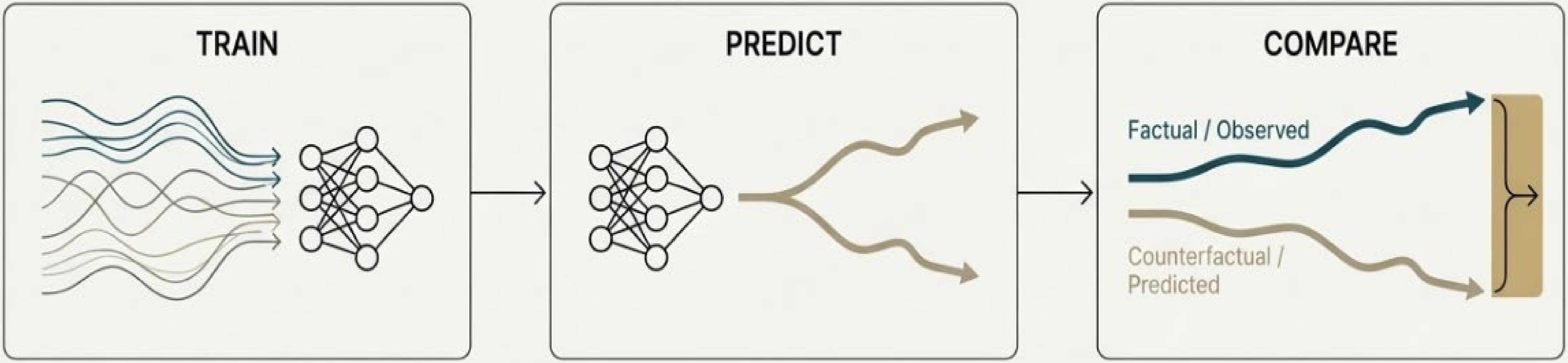


Occasional complete breakdowns (e.g., NaN errors) suggest models are learning learxing noise or 'hallucinating'.

To Prove the Impact, We Must See the Path Not Taken

How can we definitively measure the success of mitigation strategies implemented in 2006?
We need to know what would have happened without them.

Our Approach: Counterfactual AI (CAI)

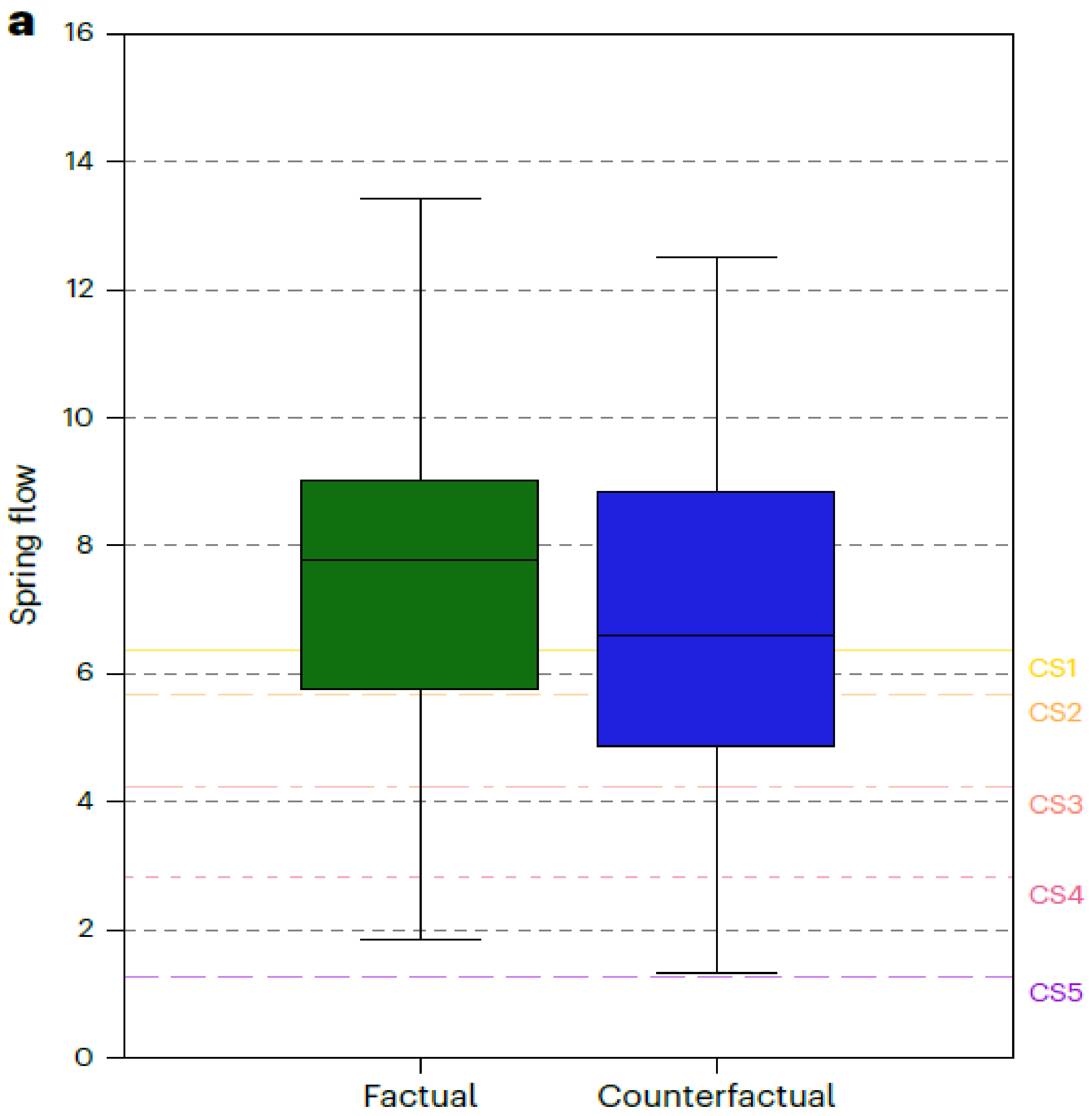


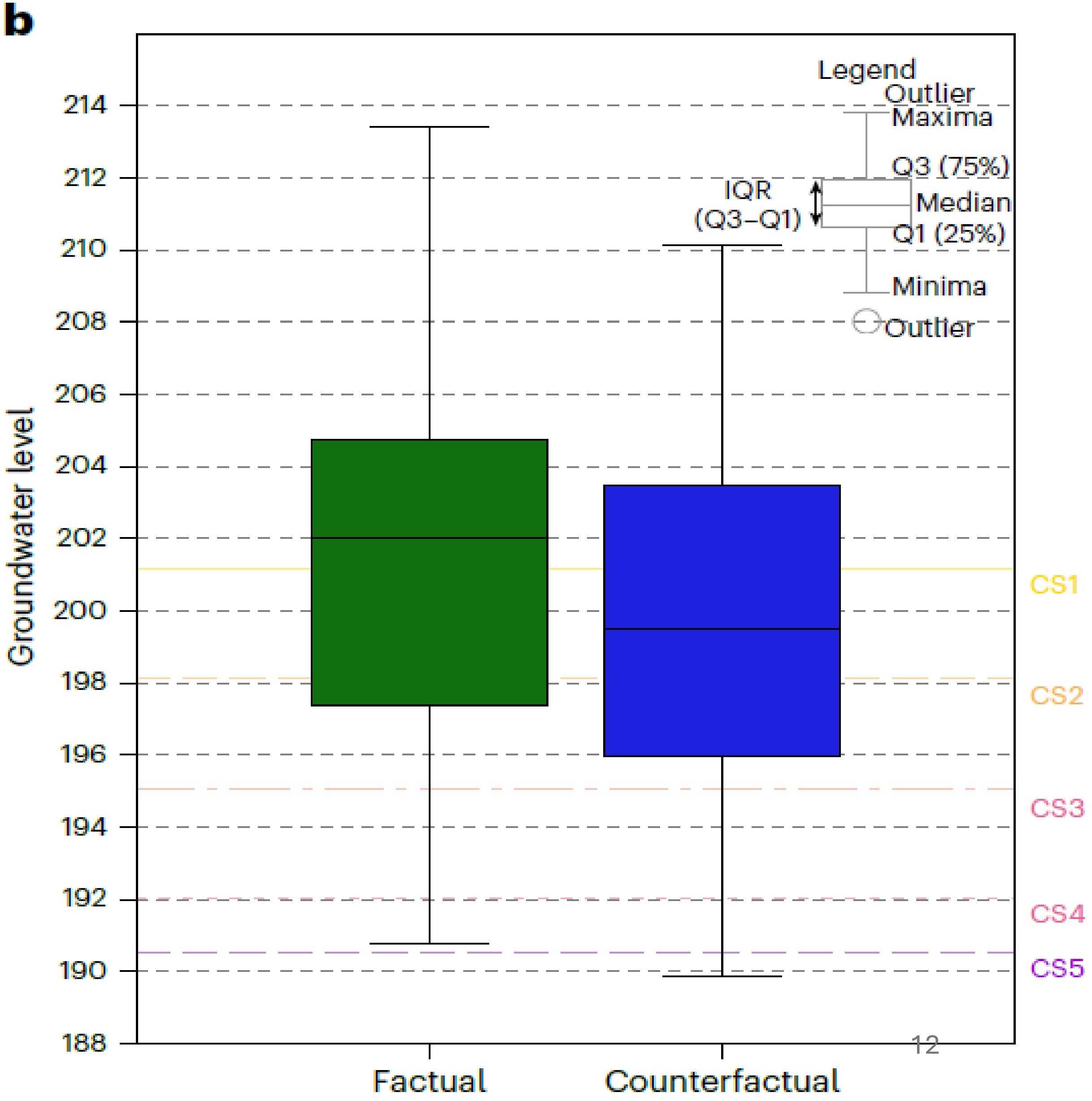
We built a tree-based ensemble AI model using extensive hydroclimatic data from a 60-year period *before* mitigation measures were in place (1946-2005).

The AI model then generated a "counterfactual" forecast for the 2006-2022 period, simulating a reality where no mitigation was ever implemented.

We compare this counterfactual prediction to the "factual" observed data from the same period. The difference between the two reveals the true impact of the conservation efforts.

Mitigation Measures Sustained Critical Spring Flows and Aquifer Levels Remained Higher & Stable





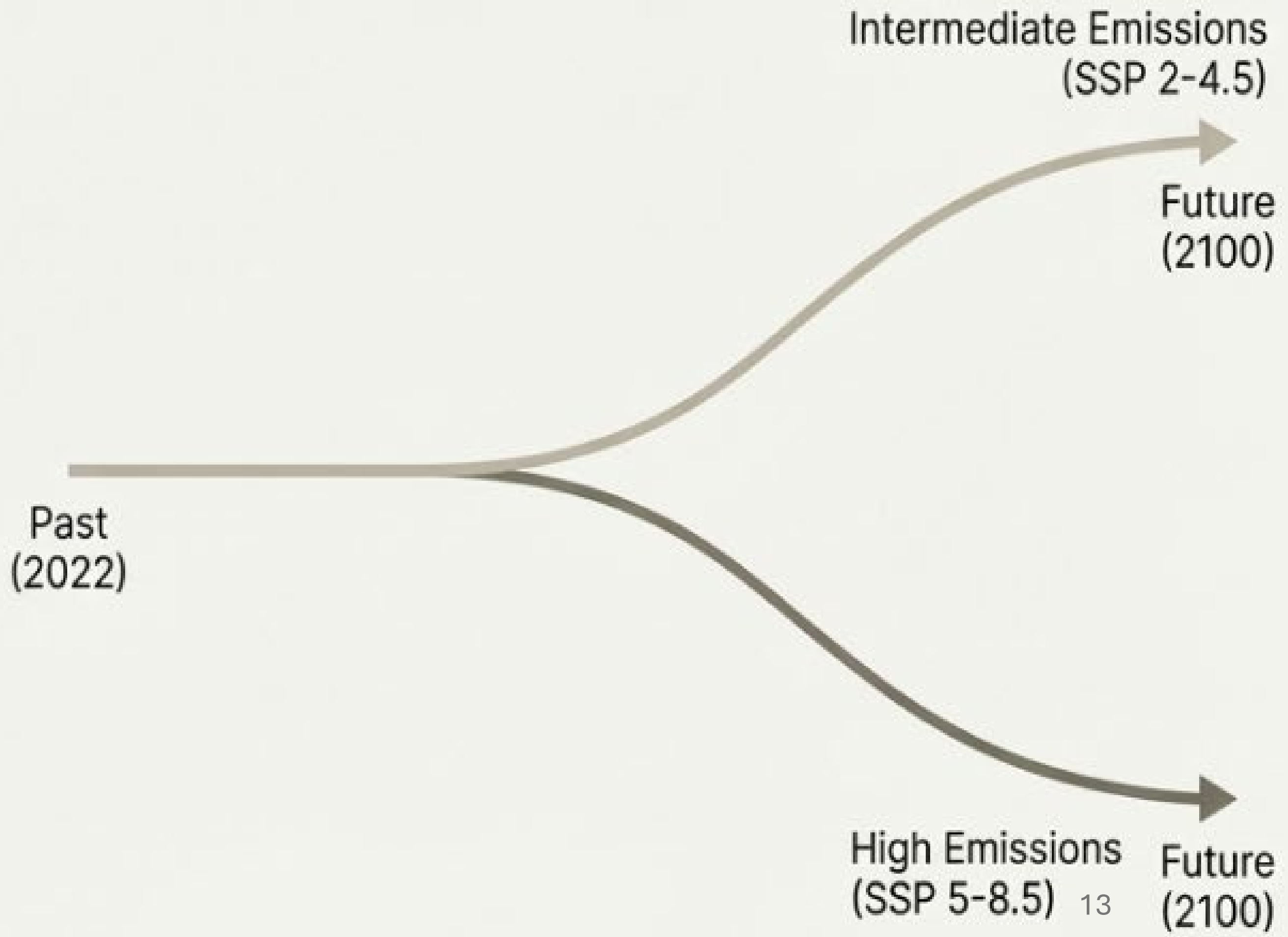
But Does This Success Hold Up in a Warmer, Drier Future?

The Challenge

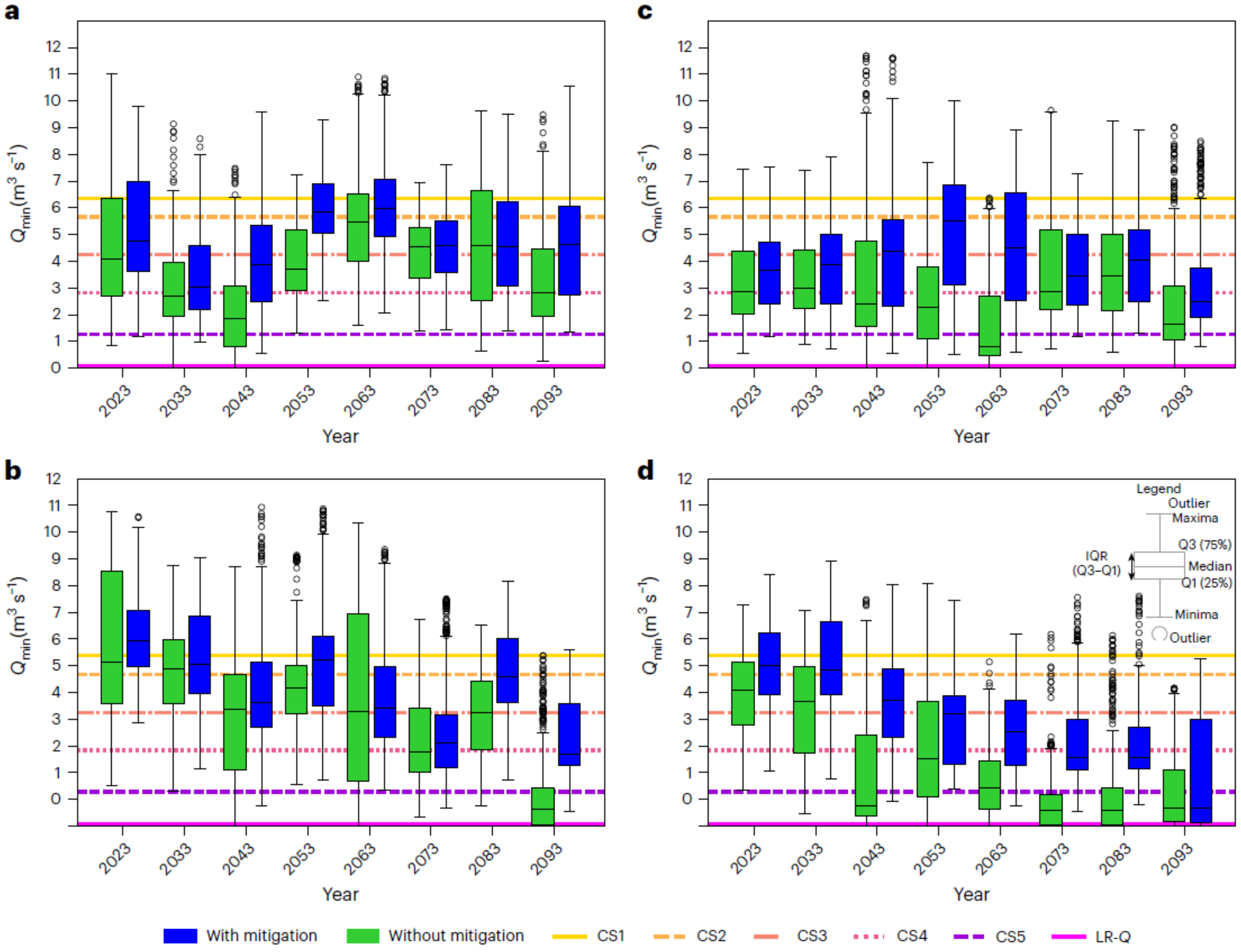
We tested the long-term effectiveness of the mitigation strategies against the stress of future climate change.

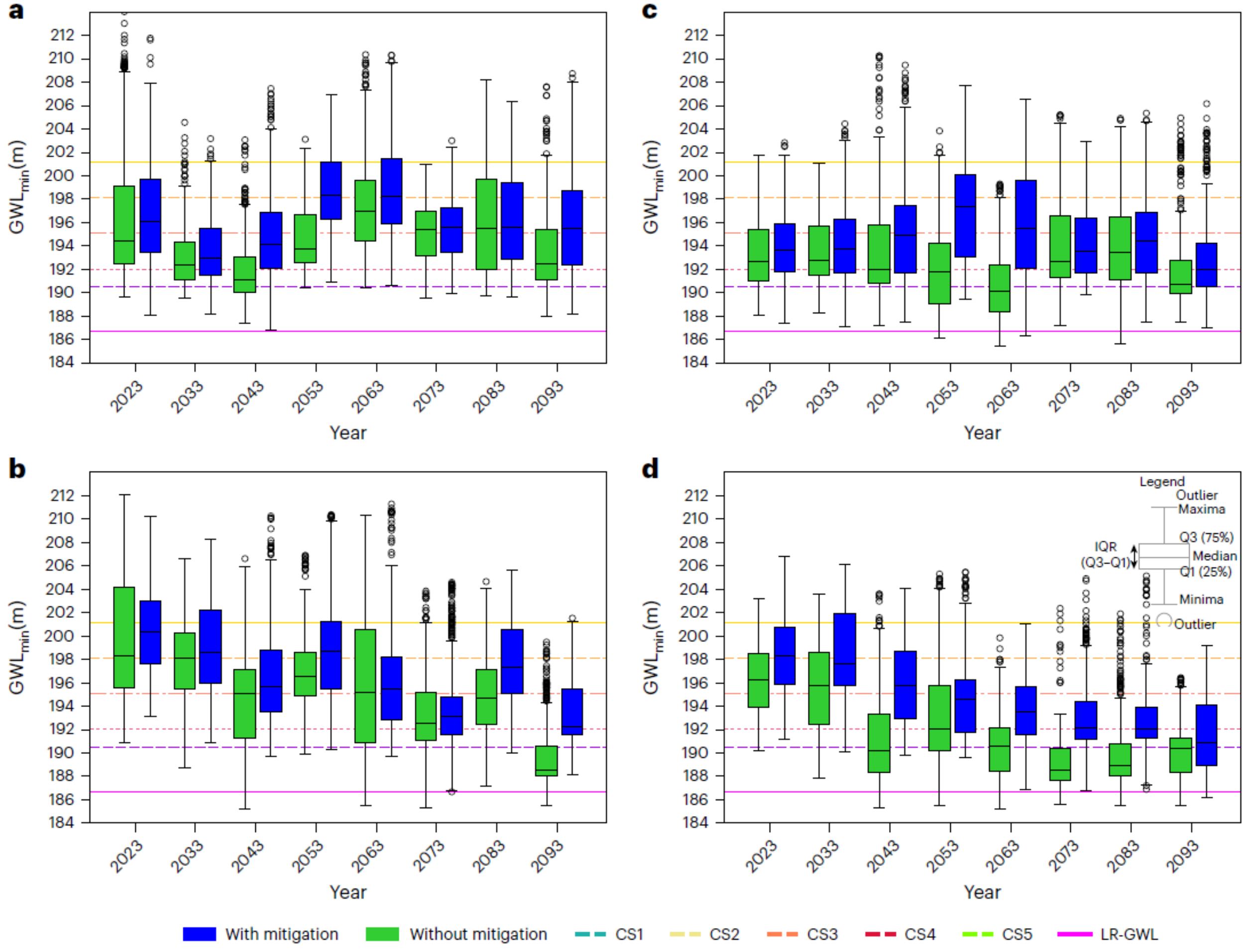
Methodology

- We used the validated AI models to project aquifer conditions out to the year 2100.
- The models were driven by customised, high-resolution climate projections (CMIP5 & CMIP6).
- We analysed two distinct futures: an intermediate-emission scenario (SSP 2-4.5) and a high-emission, “worst-case” scenario (SSP 5-8.5).

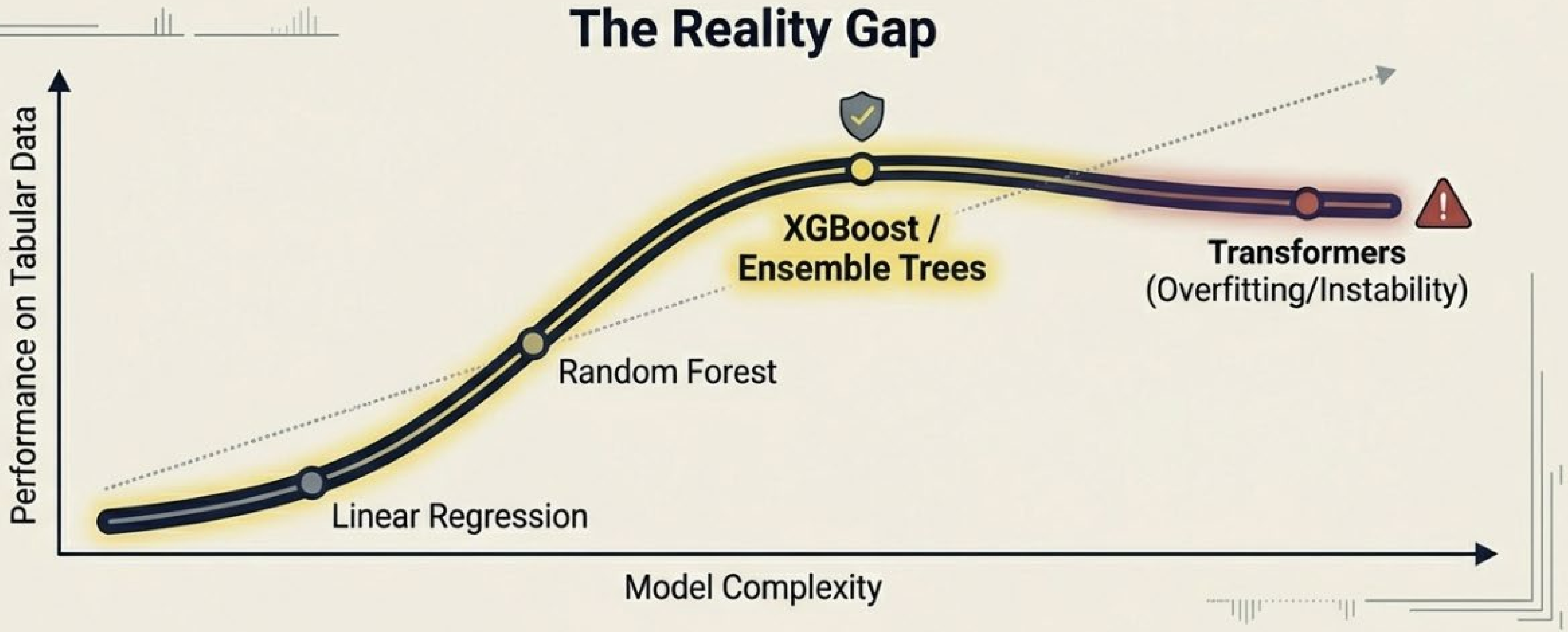


Critical Spring Flows and Aquifer Levels Projected To Remain Sustained and Higher Under Future Climate Scenarios





Beyond the Hype: A Lesson for Hydro-Informatics

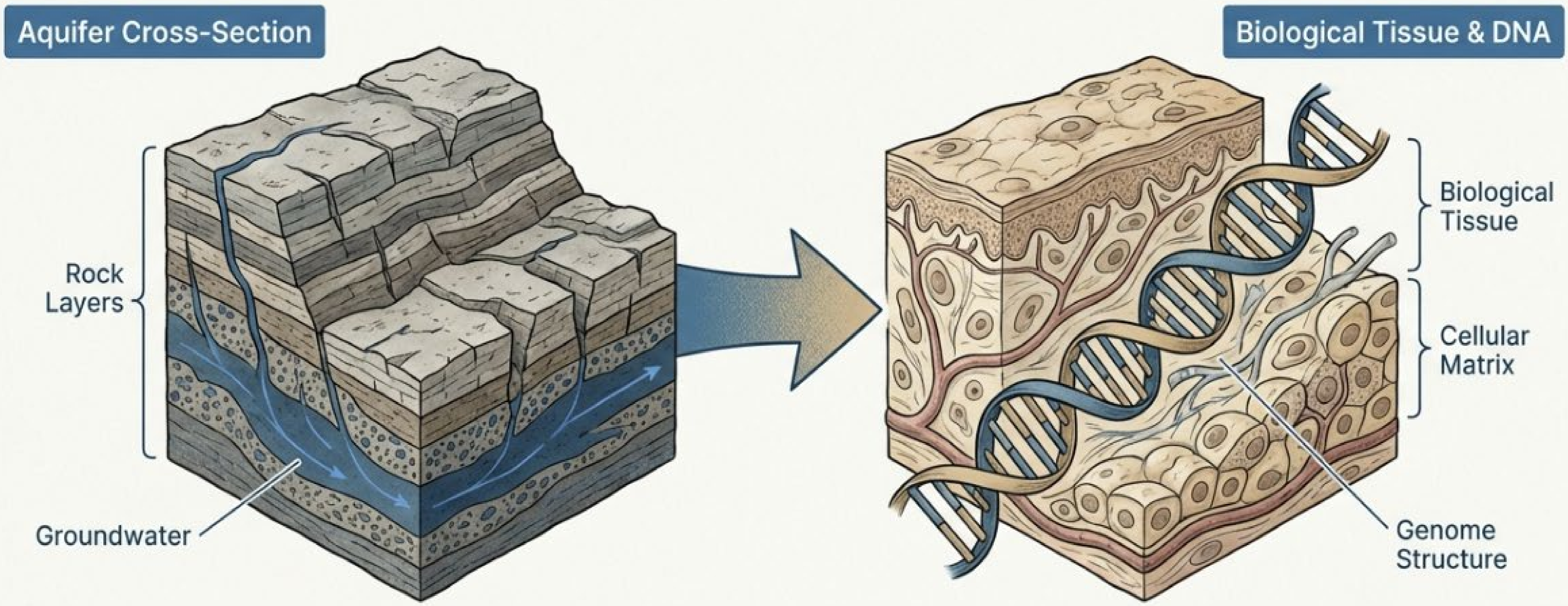


The Myth: Newer architectures automatically yield better results.

The Reality: In site-specific, tabular hydrological forecasting, traditional Machine Learning (Ensemble Trees) often outperforms Deep Learning.

Future Focus: Research should prioritize data quality and feature engineering over architectural novelty.

Universality of the Framework: From Aquifers to Oncology



Dr. Chakraborty applies these same principles—Causal and Explainable AI—to cancer bioinformatics. Just as hydrology models must respect mass balance, biomedical models must respect biological consistency. Whether the system is the Edwards Aquifer or the human genome, the requirement for interpretability remains constant.