

JULY 29, 2026

Leveraging AI for Scientific Visualization

VICTOR MATEEVITSI

Argonne National Laboratory

ATPESC 2026 · Argonne Training Program on
Extreme-Scale Computing

Q Center, St. Charles, Illinois



U.S. DEPARTMENT OF
ENERGY

Argonne National Laboratory is a
U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC.

Argonne
NATIONAL LABORATORY

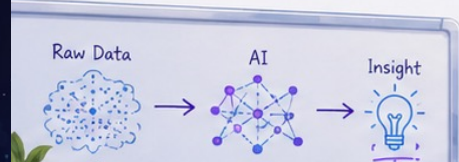


Argonne Leadership
Computing Facility

AI for ...

AI for Visualization

Turning complex data into insight.



AI sees patterns.
You see possibilities.

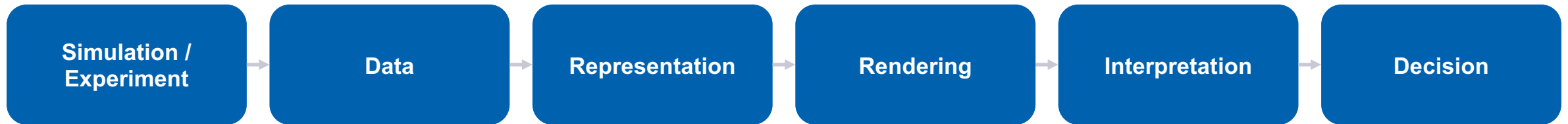
Beautiful
Insights
Ahead

VOLKER KNECHT

Kilian Hikaru Scheutwinkel • Álvaro Díaz Fernández • Bernard Jones Lozano

Every scientific result travels the same pipeline

From simulation or experiment to a decision, data passes through representation, rendering, and interpretation. Each hand-off is a chance to gain — or lose — scientific information.

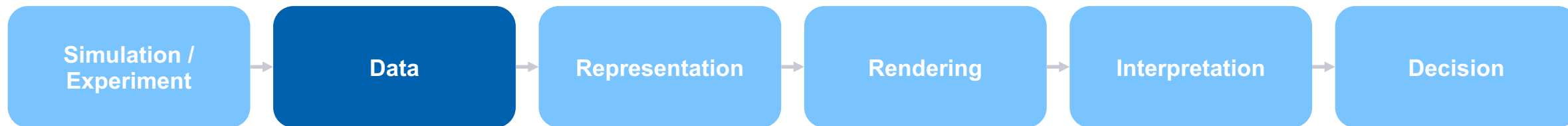


In scientific visualization the image is not the product — the traceable mapping from data to visual evidence is the product.

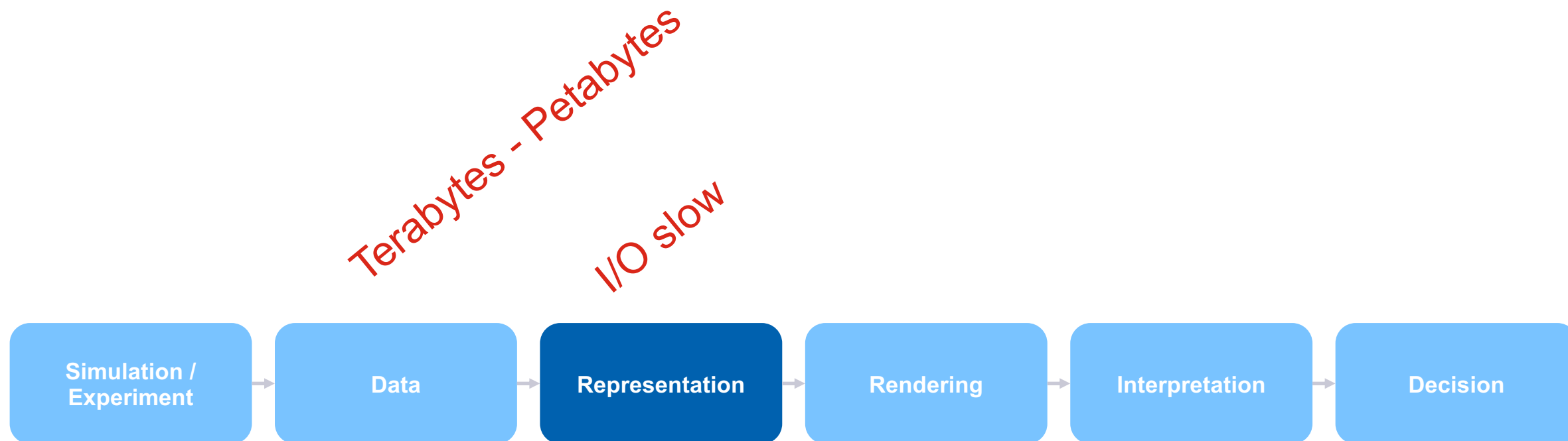
You just saw what VisIt and ParaView can do across this pipeline. My question today: what changes when AI enters it?

Where are the bottlenecks?

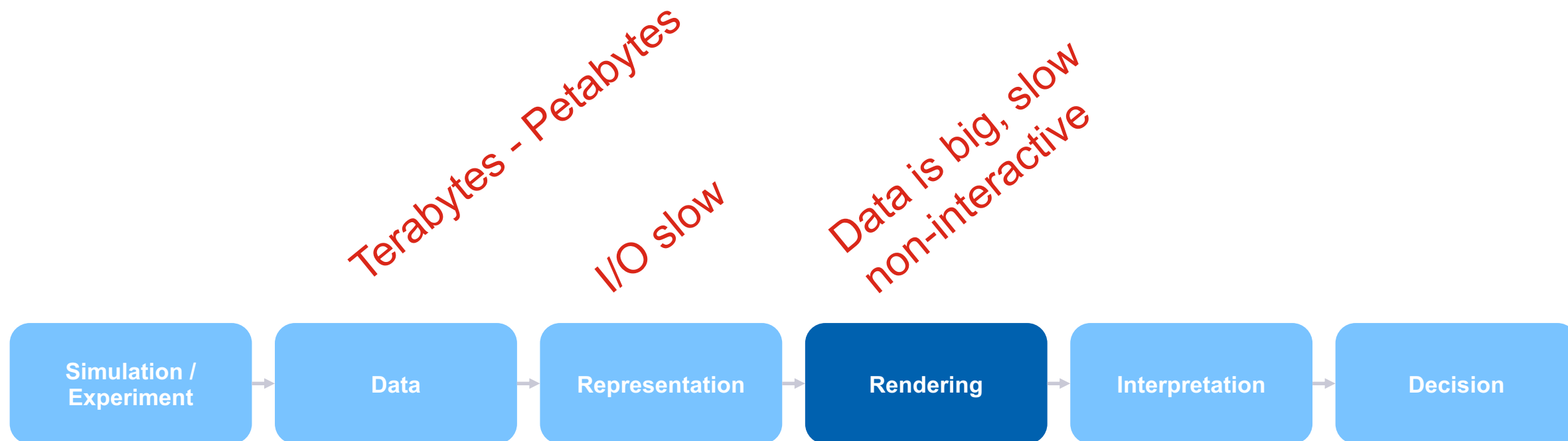
Terabytes - Petabytes



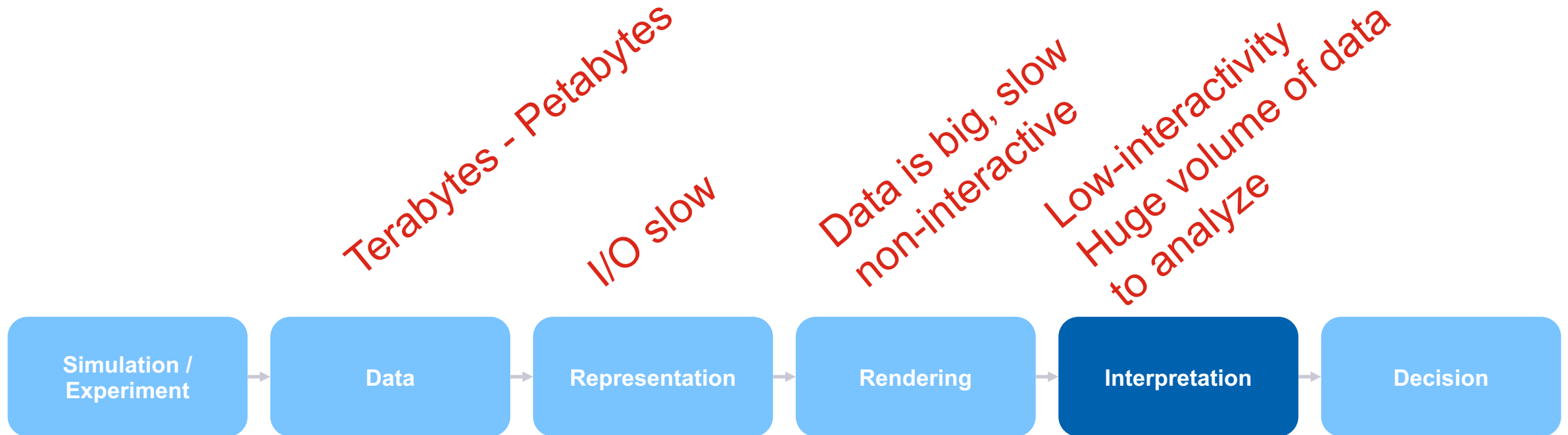
Where are the bottlenecks?



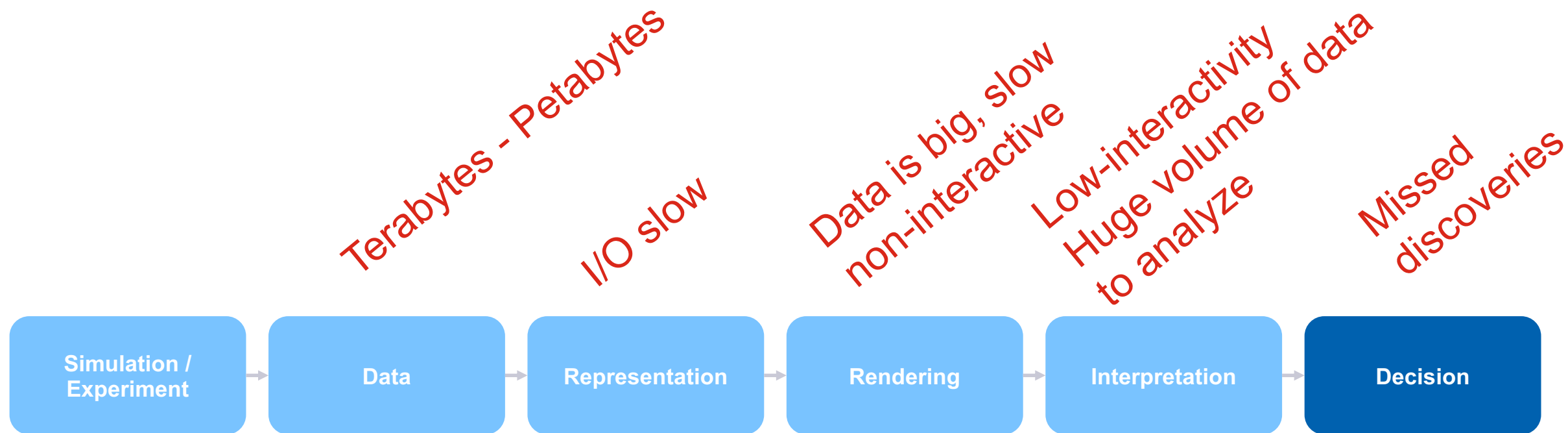
Where are the bottlenecks?



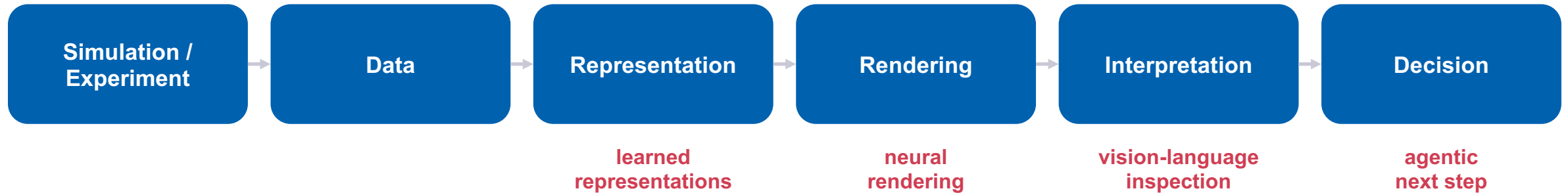
Where are the bottlenecks?



Where are the bottlenecks?



AI methods are being explored at every stage



Three sections organize the rest of this talk:

Represent — AI changes the data and rendering pipeline.

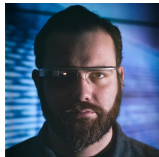
Create — AI changes how we build visualization workflows.

Reason — AI changes how we inspect and act on the result.

IDEAS

Intelligent Data Exploration Assistant for Science

ASCR – funded project



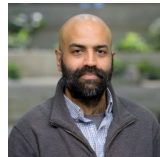
Victor Mateevitsi

ANL



Thomas Peterka

ANL



Saumil Sudhir Patel

ANL



Silvio Rizzi

ANL



David Lenz

ANL



Orcun Yildiz

ANL



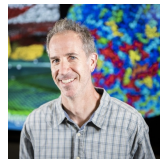
Christine Simpson

ANL



JD Emberson

ANL



Joseph A. Insley

ANL



Mengjiao Han

ANL



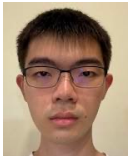
Janet Knowles

ANL



Kwan-Liu Ma

UC Davis



Kai-Chen Tung

UC Davis



Luc Renambot

UIC



Hongfeng Yu

UNL



Jianxin Sun

UNL



Cory Quammen

Kitware



Qi Wu

NVIDIA

Represent, Create, Reason — compressor, copilot, collaborator

REPRESENT

AI as compressor

Store and move less;
reconstruct more.

CREATE

AI as copilot

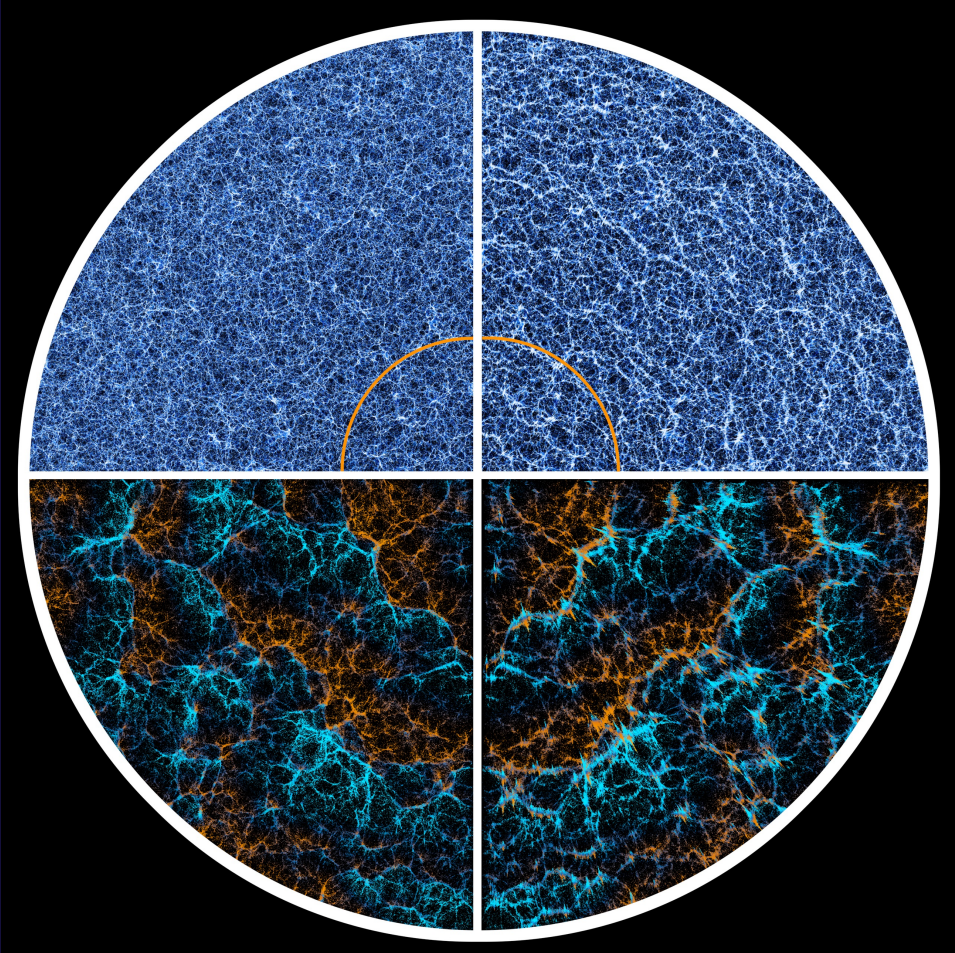
Describe intent; an agent builds
the pipeline.

REASON

AI as collaborator

The agent sees the render and
helps decide what's next.

Our running task: find and validate a rare collision in the cosmic web



Task: locate a bullet-cluster-like collision in HACC-scale cosmology data — and prove it is real.

The same task drives all three sections:

Represent — does a compact model preserve the density ridges and collision morphology we are hunting for?

Create — can an agent build the slice / threshold / candidate-inspection workflow without silently using the wrong field?

Reason — can a vision agent flag a candidate and request the measurement that confirms or rejects it?

Petabytes of multivariate particle data; more views than any person can inspect. That is the pressure behind every choice.

Image: HACC Farpoint lightcone. ALCF Visualization Team science-image library. The bullet-cluster search is our illustrative running example.

Part I — Represent

AI as compressor: change the data and rendering pipeline.



U.S. DEPARTMENT OF
ENERGY

Argonne National Laboratory is a
U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC.

Argonne
NATIONAL LABORATORY

Argonne Leadership
Computing Facility

At exascale, data movement often dominates exploration

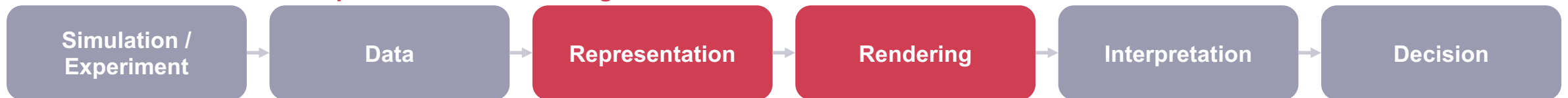
At exascale, we can generate far more data than we can store, move, or explore interactively.

The classic response: write less to disk.

The AI response: keep a compact learned model you can query and re-render on demand.

Render or store less — reconstruct more.

This section lives here → Representation + Rendering



A representation is a model you query — not a grid you store.

STORED — a grid

Explicit array: one value per grid point

N samples become N values on disk

Access = read from disk; move everything to use anything

Resolution fixed at write time

QUERIED — a model

Compact model: neural weights, spline control points, or learned primitives

Evaluate anywhere, on demand — no precomputed grid required

Move the model (small); reconstruct the data (big)

Why it matters

Query any location, any resolution — without loading the full field

Shift work from storage and bandwidth to computation

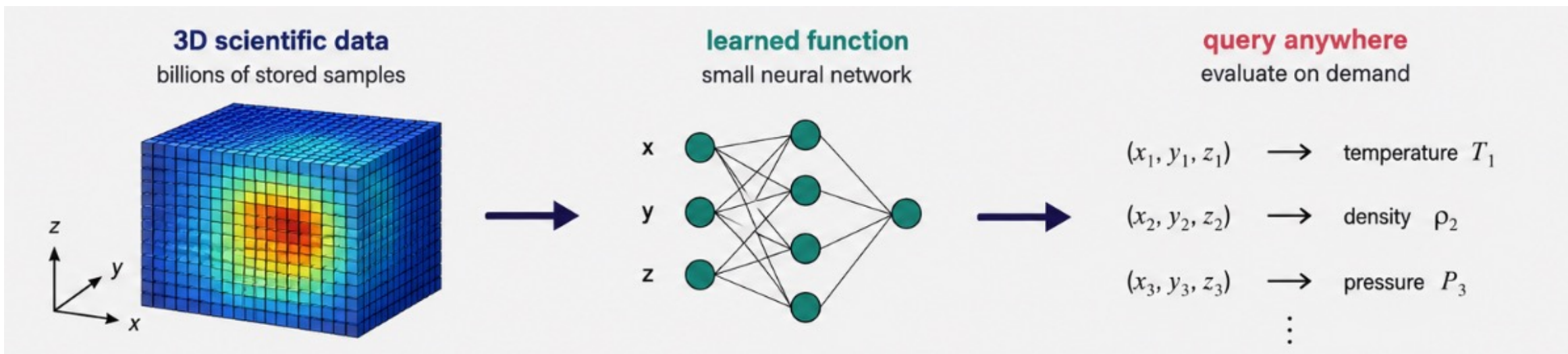
Enable interactive exploration at exascale scale

Three instances follow — Implicit neural representations (INR), Gaussian splats, and Multivariate Functional Approximation (MFA): each a different way to be compact, and a different way to be wrong.

An INR replaces billions of samples with a learned function

AI intervention: may smooth or interpolate away rare localized structure

Implicit neural representation (INR): a small neural network that maps a coordinate to a field value. Instead of a stored grid, you keep the weights and evaluate the field anywhere, on demand.



The trade

Preprocessing + training now, for a compact, continuous, queryable model later.

Scientific example

A 3D simulation can store temperature, density, or pressure as a learned function instead of a giant volume.

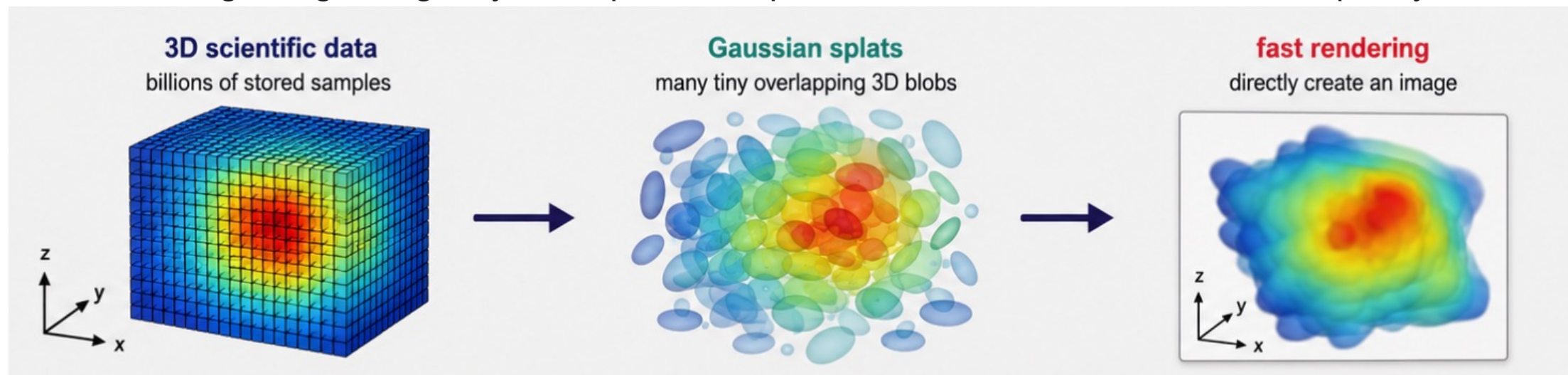
The risk

It is an approximation. Smooth by construction — it can quietly round off sharp, rare structure.

Gaussian splats replace the grid with many tiny 3D blobs

AI intervention: learns many fuzzy 3D blobs – can smooth over thin or rare structure

Gaussian splatting represents a 3D scientific field with many small overlapping Gaussian blobs. Instead of storing a huge 3D grid, you keep the blob parameters and render the structure quickly from them.



What it is

A Gaussian splat is a small fuzzy 3D blob with a position, size, shape, and value. Many splats overlap to represent the whole field.

Scientific example

A temperature or density field can be approximated by many splats that preserve the main structures and render quickly for interactive viewing.

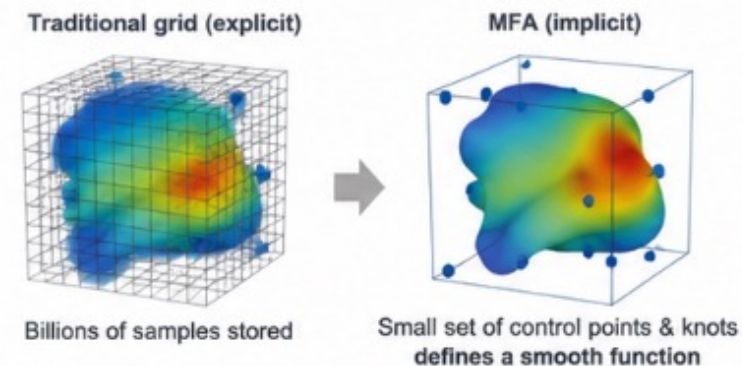
The risk

It is an approximation. Soft blobs can blur thin boundaries, sharp peaks, or rare localized features.

MFA replaces the grid with a smooth, compact function

AI intervention: approximates scientific 3D fields with smooth splines — can round off sharp, rare structure

MFA represents a 3D scientific field (e.g., temperature, pressure, velocity) as a smooth function built from a small set of control points and basis functions. You can evaluate the field anywhere, on demand, without storing a huge grid.



1. What it is

MFA uses multivariate B-splines (or NURBS) to approximate a 3D field.



Field value = weighted sum of smooth basis functions centered on control points.

Result: continuous, compact, query-anywhere.

2. How it works (simple view)

A few control points + knots



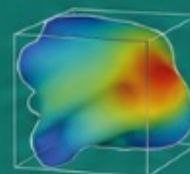
+

Smooth basis functions



=

Sum (blend) them together



Evaluate the field at any location (x, y, z)
→ get value(s) (e.g., temperature, velocity).

3. Why it matters



Massive compression

From billions of grid samples to thousands of control points.



Fast, on-demand evaluation

Query the field anywhere without loading a giant dataset.

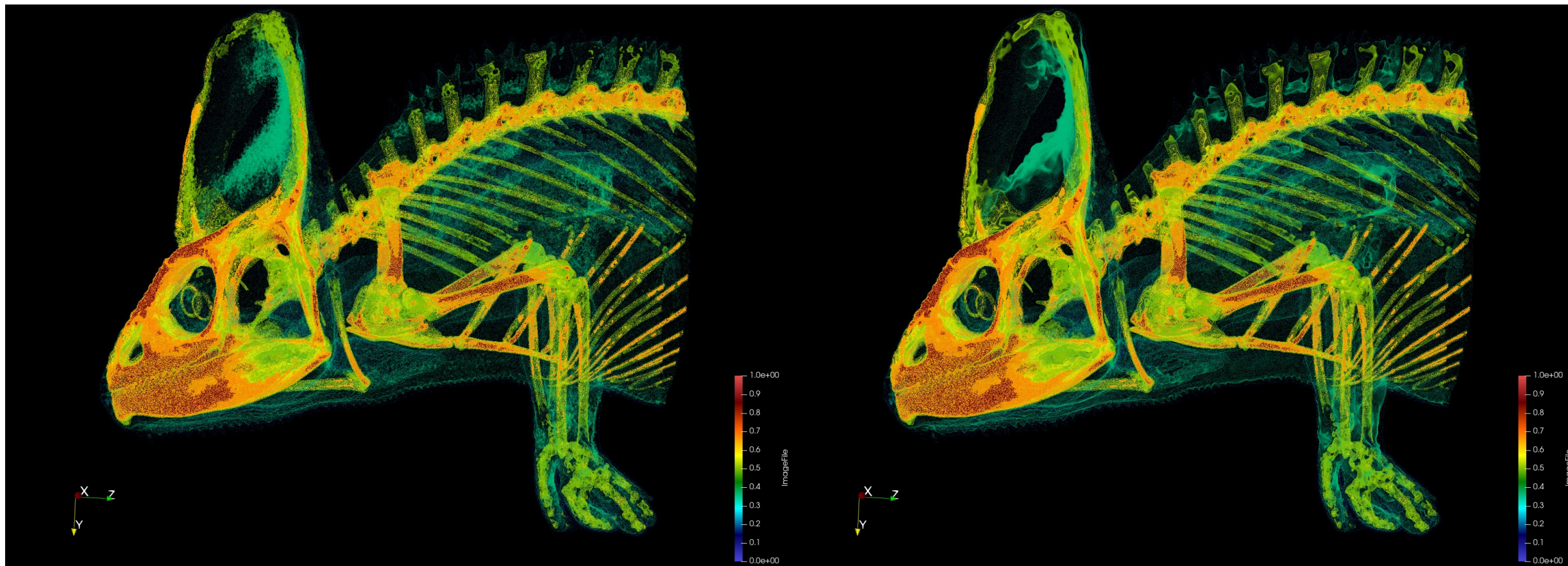


Smooth by design

Great for capturing overall trends and structures with fewer degrees of freedom.

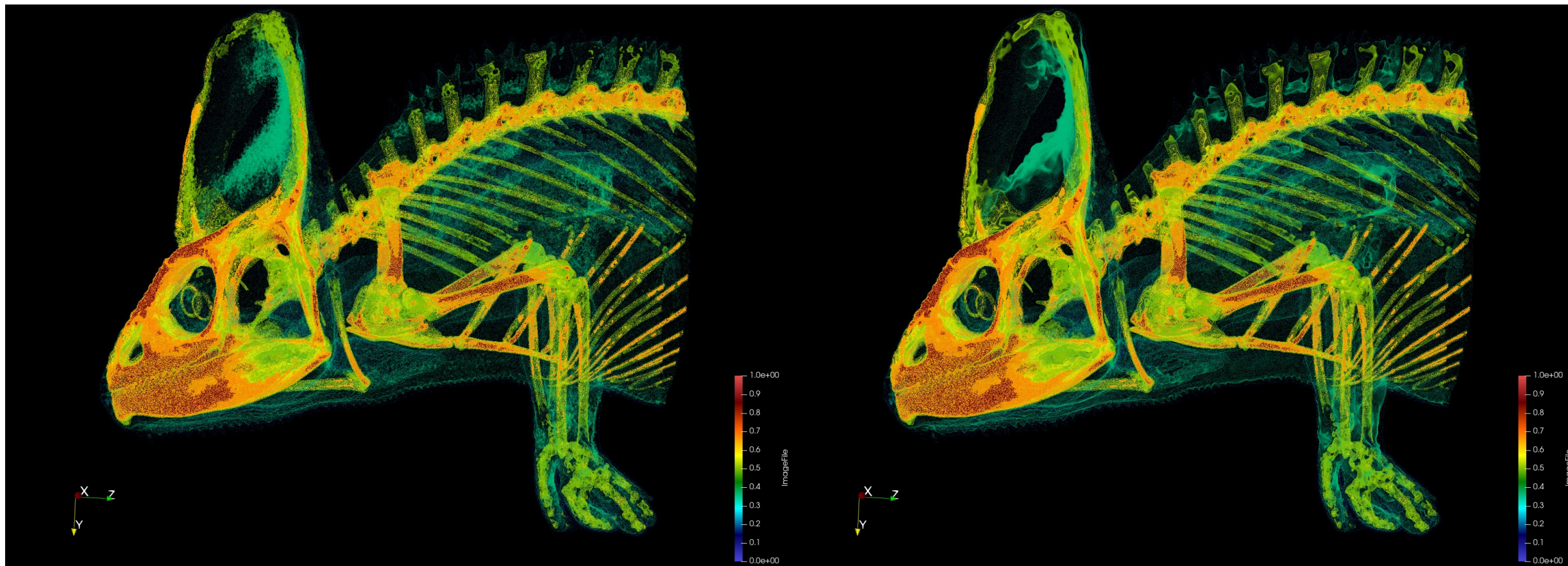
Peterka et al., Foundations of Multivariate Functional Approximation for Scientific Data, IEEE LDAV 2018.

A real example: The Chameleon Dataset with Gaussian Splats



Which one is raw data rendering (Ground Truth) and which one a Gaussian Splat?

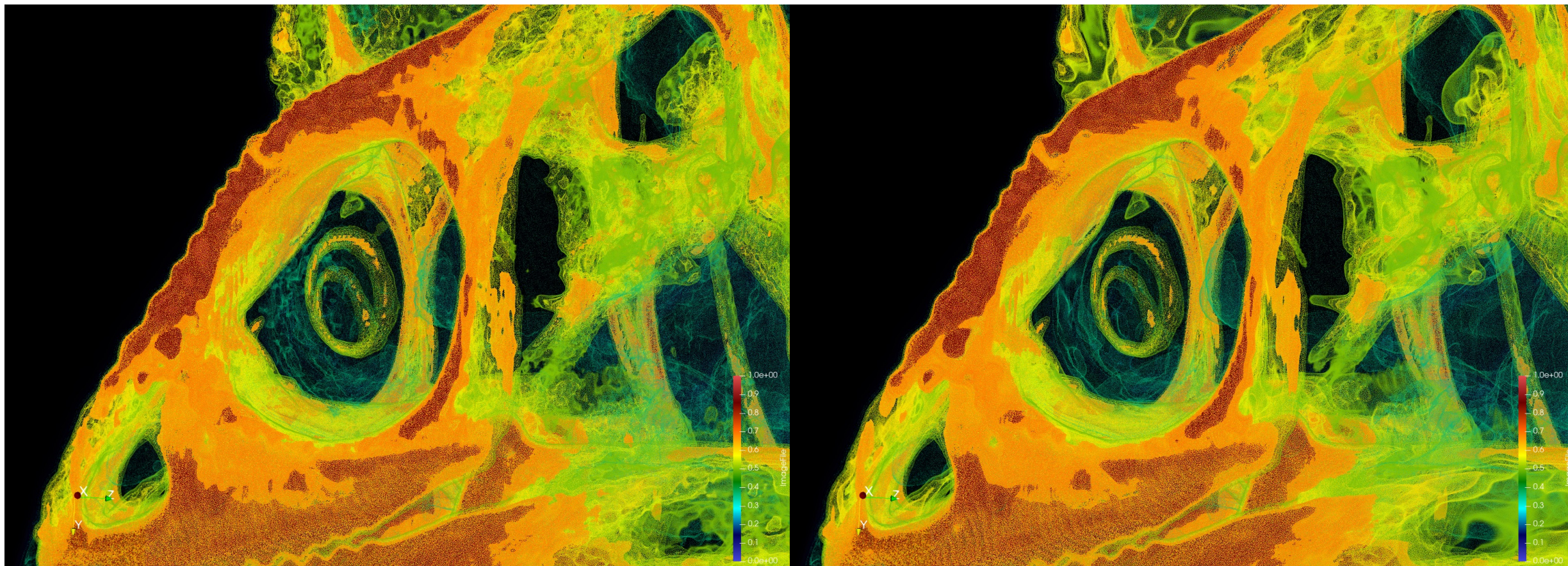
A real example: The Chameleon Dataset with Gaussian Splats



Ground Truth

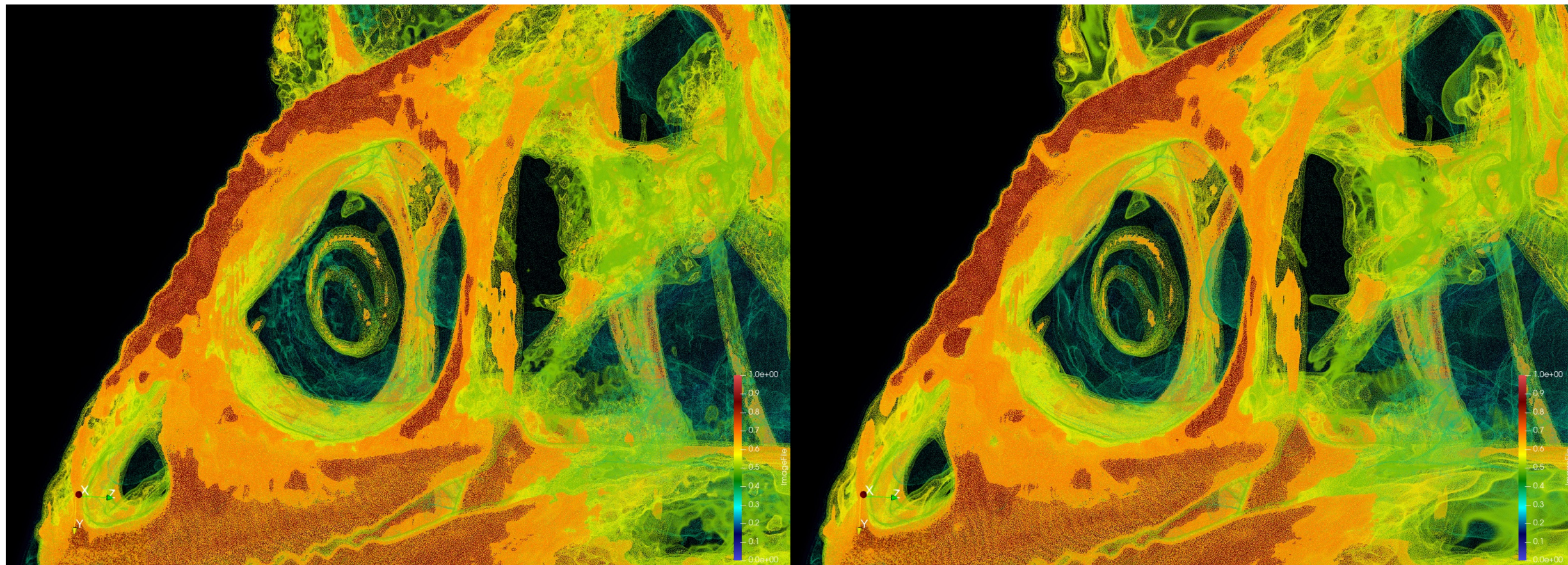
Gaussian Splat

A real example: The Chameleon Dataset with Gaussian Splats



Which one is raw data rendering (Ground Truth) and which one a Gaussian Splat?

A real example: The Chameleon Dataset with Gaussian Splats



Ground Truth

Gaussian Splat

Part II — Create

AI as copilot: change how we build the visualization workflow.



U.S. DEPARTMENT OF
ENERGY

Argonne National Laboratory is a
U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC.



Argonne Leadership
Computing Facility

The scientist knows the question, not every ParaView call

“Slice the density field at $z = 0$, threshold the top 1%, and color by temperature.”

That is a clear scientific intent. Turning it into correct code means knowing the right filters, point-vs-cell data, array names, and parameters.

An LLM coding agent can bridge intent and API — if it has the right context.

This section lives here → building and interacting with the pipeline



End goal: Even harder

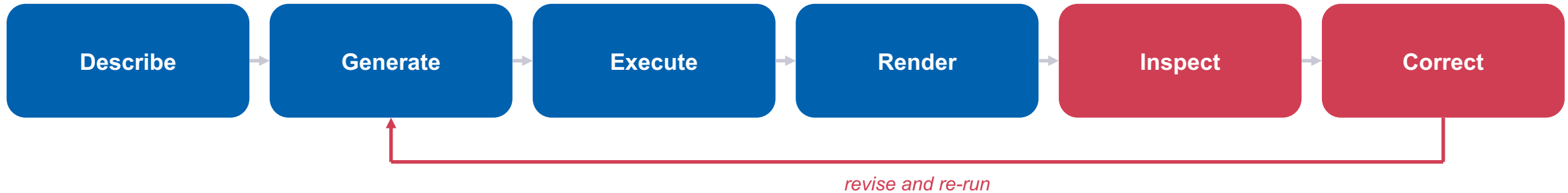
“Find me the bullet clusters.”

- Agent needs to figure out
 - What is a bullet cluster?
 - How will it find it?
 - How will it visualize it?
 - How will it validate it?
 - Etc.

This section lives here → building and interacting with the pipeline



A visualization agent needs more than an LLM



The loop:

Describe → Generate → Execute → Render → Inspect → Correct

Inspect + Correct are the hard part — and the scientific part.

Context the agent must reach

- Docs + examples · dataset metadata + arrays
- A real execution environment · compiler / runtime errors
- The rendered output · performance signals
- Scientific constraints · human feedback

Code that runs is not a visualization that is correct

92.2% of recorded failures fail at code execution.

The loud 92%: the code crashes, ParaView throws, you know immediately.

The dangerous ~6%: runnable code that renders a convincing picture and answers the question wrong — no crash, no error.

Grounded in NL2SciVis: 15 atomic ParaView tasks $\times$ 4 models $\times$ 14,000+ trials; two gates — did it run, and did it produce the right method, field, and parameters?

Model choice alone swung pass rates by 48 points.

Mathai, Han, Knowles, Mateevitsi, Rizzi, Childs. NL2SciVis, EuroVis 2026.

ChatVis: the correction loop as a working agent

An LLM writes a ParaView Python script, runs it, reads the error, and rewrites — until it executes correctly.

ChatVis is an iterative assistant (Argonne): describe an operation in natural language → generate a script → execute → detect & extract the error → re-prompt the LLM to correct → repeat.

What it is

An LLM-driven ParaView Python generator with an automated error-detection-and-correction loop. A working system, not a benchmark. Code on GitHub.

The result

On 5 canonical visualization scenarios, ChatVis produced the correct script every time (vs ground truth). The same LLMs, unassisted, failed on every one.

The lesson

What made it work was the execute-observe-correct loop — iteration and real error feedback — not a bigger model.

Mallick, Yildiz, Lenz, Peterka. ChatVis: Automating Scientific Visualization with an LLM. SC'24 Workshops (arXiv:2410.11863); extended IEEE LDAV 2025.

GENIUS turns natural speech into visualization and HPC jobs

An AI assistant that lets a scientist speak to the workflow — ask questions, manipulate 3D models, and launch jobs on ALCF supercomputers.

GENIUS (General Experimentation and Natural Interface Utility System): an Argonne AI assistant combining large language models with immersive mixed reality (XR) and natural speech — converse with your data, your models, and the supercomputer.

The problem

The HPC workflow is fragmented: submit sims remotely, retrieve via command line, then switch to specialized desktop viz tools — skills many researchers lack.

What GENIUS does

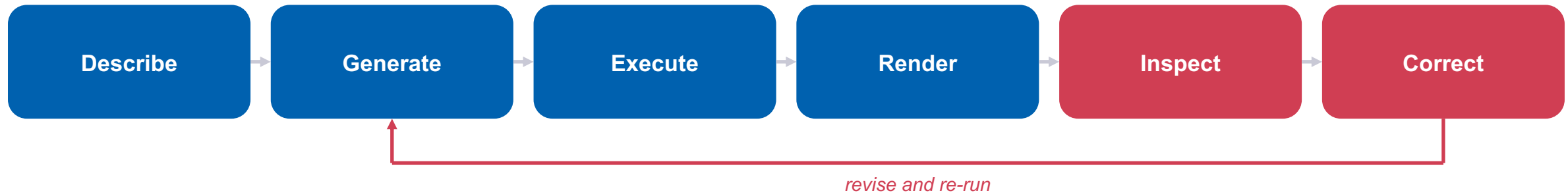
Through natural speech: ask questions, visualize and manipulate 3D models in mixed reality, and execute jobs on ALCF supercomputers — one conversational interface.

Why it matters here

It closes the arc: compact representations and agent-built pipelines become things a scientist can simply talk to and steer.

Massa, Shaik, Ta, Han, Insley, Knowles, Mateevitsi, Papka, Rizzi, Shilpika. GENIUS: AI Powered Assistant for Scientific Research. IEEE eScience 2025.

The workflow is the lesson, not the model of the month



What made the corrections work was context and iteration, not a bigger model: real execution, real error messages, and a human who knew the science.

Real Argonne systems — ChatVis and GENIUS — show this in practice.

Where we are: we can now generate the candidate-inspection workflow — but runnable code is not yet a correct measurement.

Mallick et al., ChatVis, SC'24 Workshops; Peterka et al., IEEE LDAV 2025.

Part III — Reason

AI as collaborator: change how we inspect and act on the result.



U.S. DEPARTMENT OF
ENERGY

Argonne National Laboratory is a
U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC.



Argonne Leadership
Computing Facility

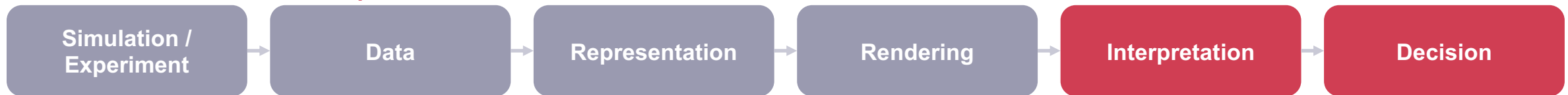
What changes when the agent can see what it rendered?

Until now the render was the end of the line — a picture for a human. A vision-capable agent can treat the render as input: describe it, compare it, question it, and propose the next step.

In agentic systems the image can become feedback, not just the deliverable.

For 2D plots this feedback is demonstrated. For 3D HACC-scale scientific reasoning, it is active research.

This section lives here → Interpretation + Decision



Four levels: describe, compare, investigate, act

1 Describe

Name visible structures; summarize a view.

Demonstrated (2D charts)

2 Compare

Contrast timesteps, parameters, ensembles, viewpoints.

Demonstrated / emerging

3 Investigate

Request another slice, camera, threshold, or measurement.

Emerging

4 Act

Recommend or trigger the next analysis or run.

Research frontier

Levels 1–2: Bendeck & Stasko, IEEE TVCG 2024; MatPlotAgent, ACL Findings 2024. Levels 3–4: VizGenie, IEEE TVCG 2026 (emerging).

Production

Use AI today with your own work.



U.S. DEPARTMENT OF
ENERGY

Argonne National Laboratory is a
U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC.

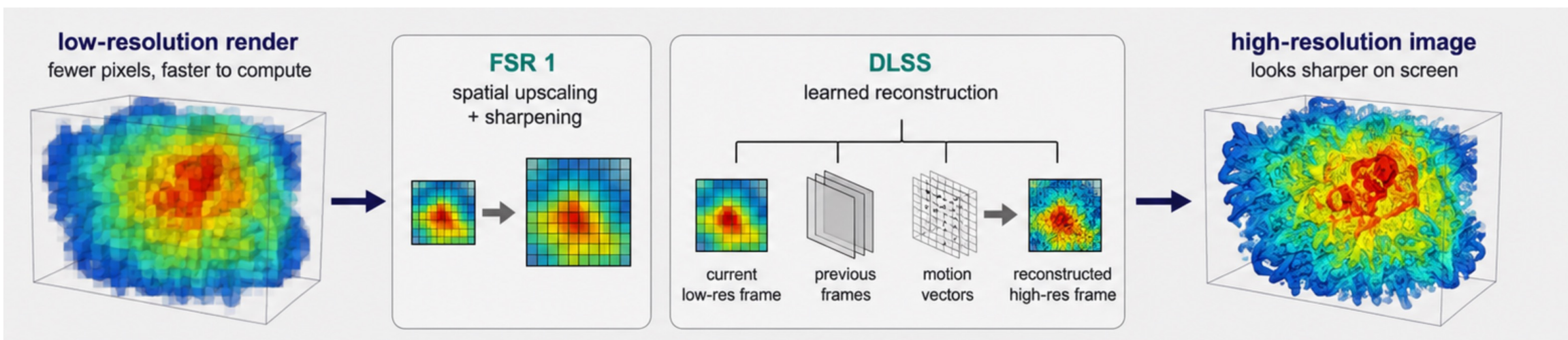
Argonne 
NATIONAL LABORATORY

Argonne Leadership
Computing Facility

Super-resolution renders fewer pixels—and reconstructs the rest

AI intervention: reconstructs pixels from fewer pixels — can sharpen , hallucinate or miss thin structure

Super-resolution starts from a lower-resolution render and creates a higher-resolution image. The goal is to save rendering time while preserving a sharp image for analysis and interaction.



What it is

Render fewer pixels now, then reconstruct a large image later. This increases frame rate and reduces rendering cost.

FSR1 vs DLSS

- FSR1: classic spatial upscaling and sharpening; fast and simple; not AI
- DLSS: neural-network reconstruction g prior training d temporal information

The risk

It is an approximation. Fine edges, tiny features, or rare structures may be softened, shimmer, or be hallucinated.

Super-resolution renders fewer pixels—and reconstructs the rest



Super-resolution renders fewer pixels—and reconstructs the rest

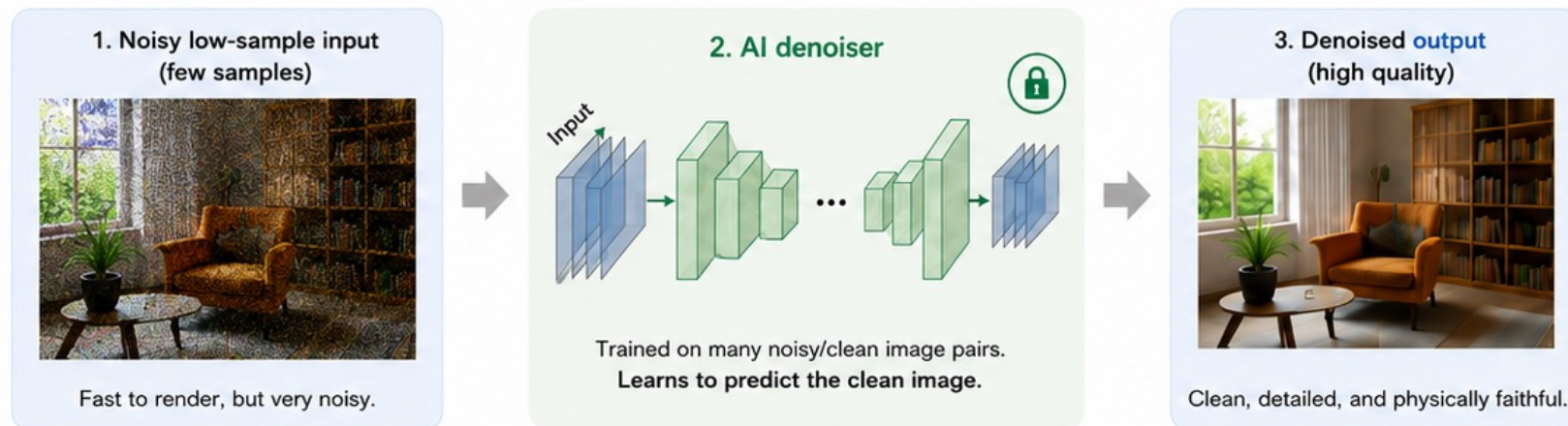


Path-traced denoising

AI intervention: removes noise from few-sample renders while preserving detail and physical correctness

Path-tracing denoising uses an AI model to remove Monte Carlo noise from low-sample renders, producing a cleaner image much faster than brute-force sampling alone.

How it works



What it is

A path-tracing denoiser is an AI model that takes a noisy low-sample render and predicts a cleaner image.

Example (same scene)



Why it matters



Big speedups

Render with far fewer samples (10–100x less) and get results much faster.



Preserves detail

Recovers fine details, edges, textures, and lighting while removing noise.



Physically aware

Trained on path-traced data to respect lighting, materials, and energy.

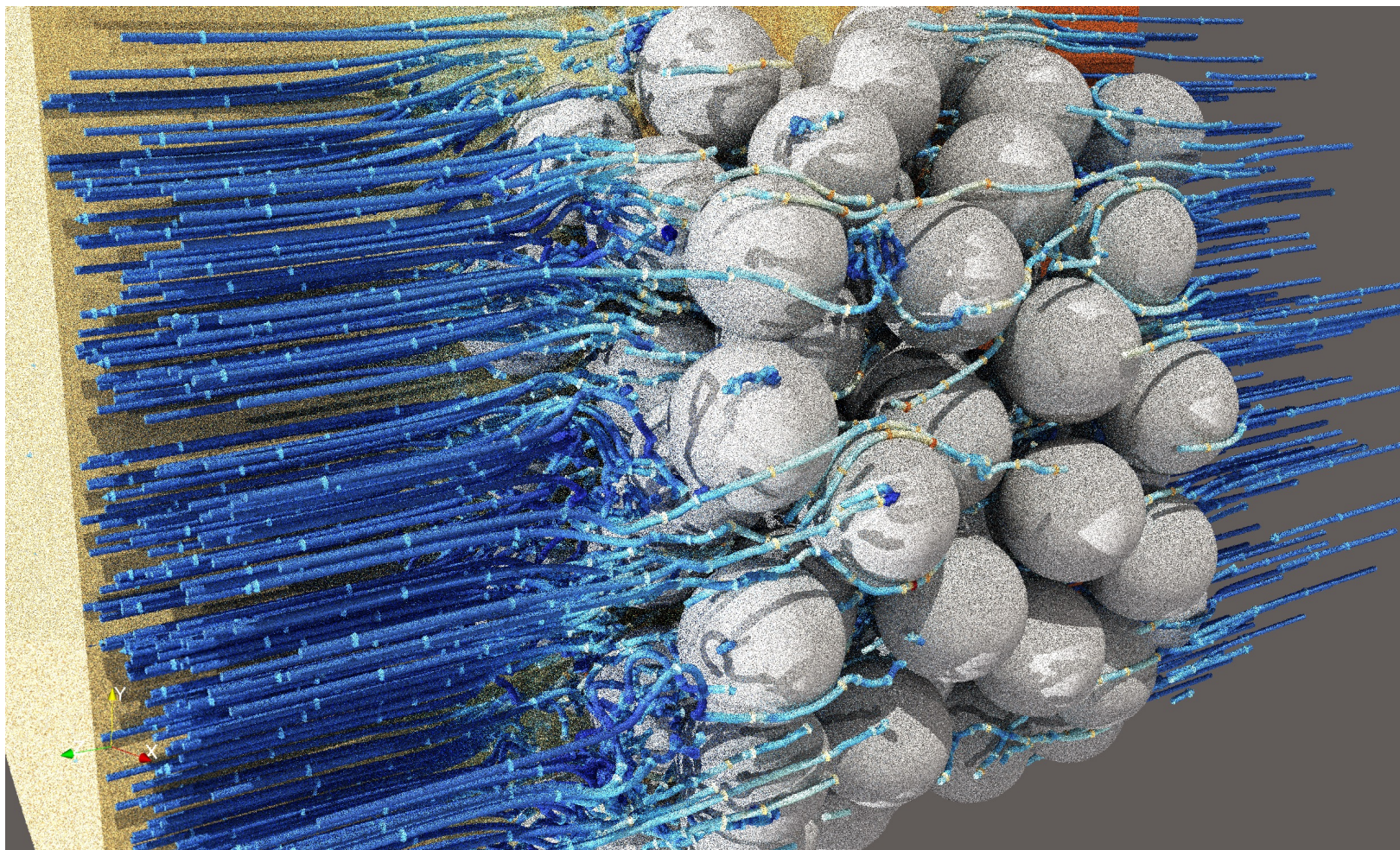


Why people use it

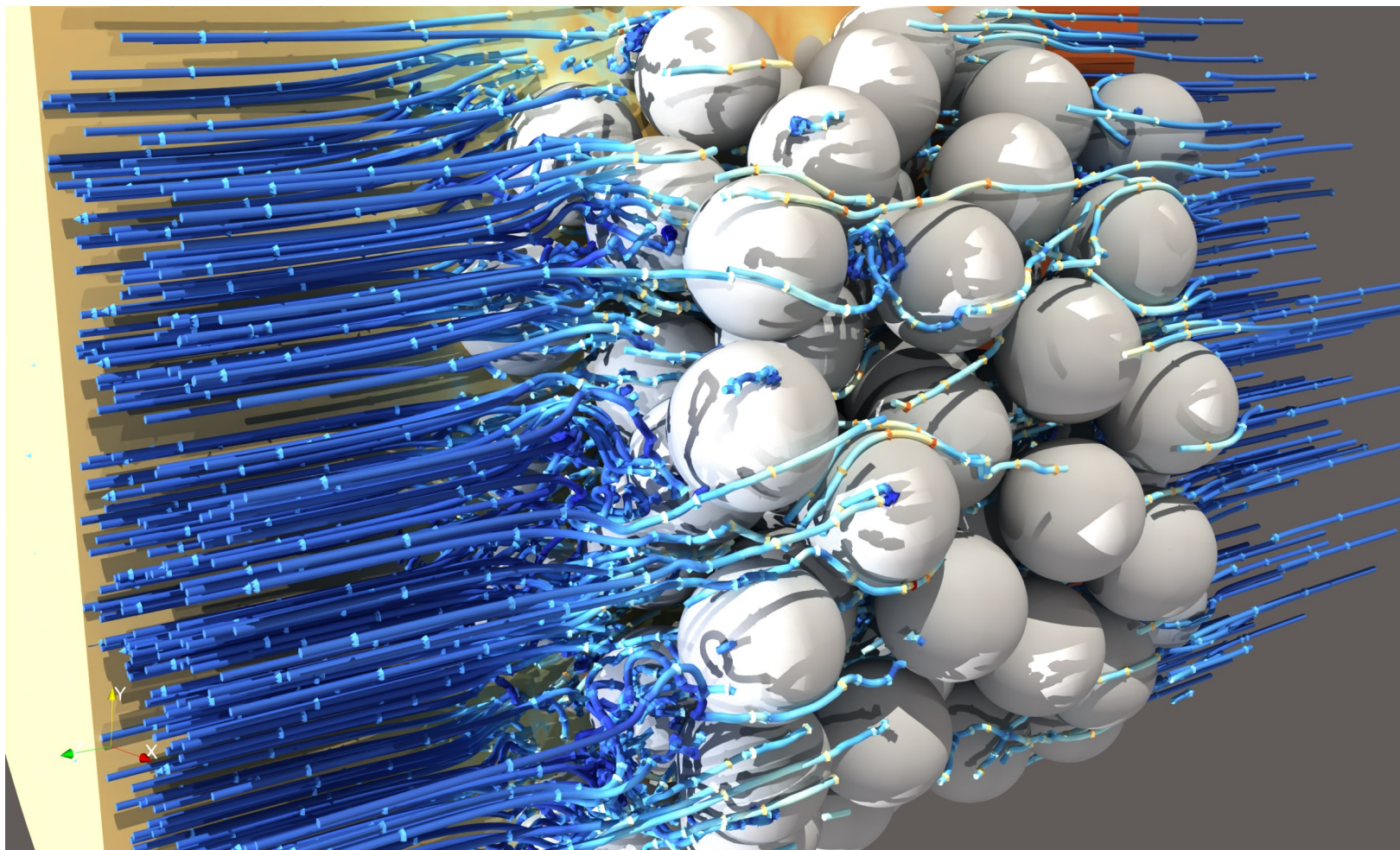
- Real-time interactive visualization
- Faster rendering with fewer samples
- Cleaner images for exploration
- Improves visual quality while saving time

Path-tracing denoising uses machine learning to predict a clean image from a noisy render.

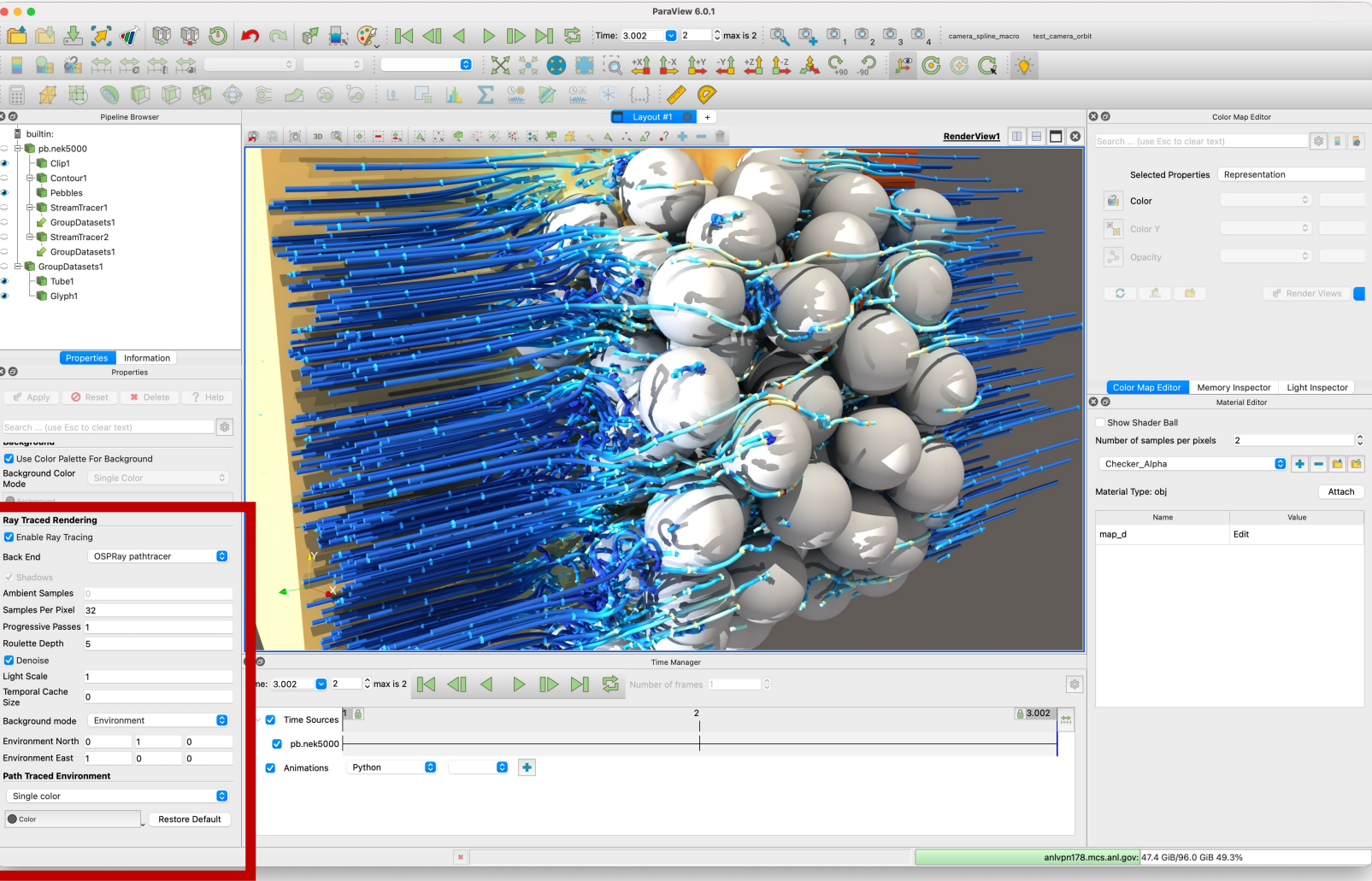
Path-traced denoising



Path-traced denoising



Path-traced denoising



Ray Traced Rendering

☒ Enable Ray Tracing

Back End OSPRay pathtracer

☒ Shadows

Ambient Samples 0

Samples Per Pixel 32

Progressive Passes 1

Roulette Depth 5

☒ Denoise

Light Scale 1

Temporal Cache Size 0

Background mode Environment

Environment North 0 1 0

Environment East 1 0 0

Path Traced Environment

Single color

☐ Color

Restore Default

Generative imagery: fine for communication, not for evidence

Appropriate

- Brainstorming visual encodings
- Storyboarding a figure
- Clearly labeled explanatory illustrations
- Exploring a visual metaphor

Dangerous

- Filling in missing scientific data
- Adding unmarked detail
- Using an image with no provenance
- Confusing plausibility with fidelity

Generative AI can suggest what an explanation might look like. That is not the same as having visualized the data.

LLMs and Agents



Framework & Conclusion

Four questions you can take back to your own work.



U.S. DEPARTMENT OF
ENERGY

Argonne National Laboratory is a
U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC.

Argonne 
NATIONAL LABORATORY

Argonne Leadership
Computing Facility

Four questions before you add AI to a visualization

1. What bottleneck are we solving?

Data movement, storage, render cost, software complexity, attention, or interpretation?

2. What is the AI allowed to change?

The representation, resolution, code, encoding, interpretation, or the decision?

3. How could it be wrong?

Lost feature, invented detail, bad code, misleading encoding, misread pattern, failure at scale?

4. How will we validate it?

Error maps, feature metrics, compare to original, quantitative follow-up, multiple views, expert review, provenance.

Thank you

VICTOR MATEEVITSI
Argonne National Laboratory
vmateevitsi@anl.gov

This research used resources of the Argonne Leadership Computing Facility, which is a U.S. Department of Energy Office of Science User Facility operated under contract DE-AC02-06CH11357. This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR), for the project Argonne Base: Scientific Data Management and Visualization to Advance AI for Science, under Contract DE-AC02-06CH11357.



U.S. DEPARTMENT OF
ENERGY

Argonne National Laboratory is a
U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC.

Argonne
NATIONAL LABORATORY

Argonne Leadership
Computing Facility