# AI-driven VVUQ: Acceptability Challenges in Research and Industry



Christophe CALVIN (CEA)

*D. Besnard, C. Calvin, G. Camus, G. Daniel, C. Gouy-Pailler, C. Millet, JP Perlat*

# AI Model Reliability and Validation: Challenges and Perspectives

Forum TERATEC
20 - 21 OCTOBRE 2026
PARIS · PALAIS DES CONGRÈS

- Growing use of AI models (machine learning, deep learning, generative AI) in scientific and industrial fields: data analysis, simulation, and design.
- High performance, particularly when addressing complex, high-dimensional problems.
- Persistent limitations in reproducibility, verification, explainability, and error quantification.
- The VVUQ (Verification, Validation & Uncertainty Quantification) framework helps assess model reliability and quantify uncertainties associated with predictions.

- **Workshop objectives: review the state of the art, identify limitations and application areas, and propose research directions and use cases to improve AI model validation.**

*VVUQ encompasses scientific methods and methodologies used to assess, ensure, and enhance the credibility of numerical simulations:*

- *Verification: Checks that mathematical equations are solved and computer code is executed correctly.*
- *Validation: Compares simulation results with real-world experimental data to ensure that the model accurately represents physical reality.*
- *Uncertainty Quantification: Identifies, quantifies, and propagates uncertainties arising from input parameters, assumptions, or model limitations to assess prediction reliability.*

# Workshop Format



- Two-day workshop bringing together experts in AI, VVUQ, numerical simulation, and end users.
- Day 1: state-of-the-art overview and working sessions.
- Day 2: more technical presentations and summaries from the working groups.

| | Title | Speaker |
|---|---|---|
| | Keynote 1: Introduction to VV&UQ Approaches | Guillaume PERRIN (Université Gustave Eiffel) |
| | Keynote 2: The AI Model Development Lifecycle and Integration of VVUQ | Eiji Kawasaki (CEA/DRT) Geoffrey Daniel (CEA/DES) |
| | NUMERICAL SIMULATION AND DEEP LEARNING: Two Views of Science and Two Distinct Validation Problems | Michel BitBol (ENS Paris) |
| | Parallel Workshops (3 × 1 hour) | |
| | - Workshop 1: Verification — From Data to Systems | |
| | - Workshop 2: Validation — From Benchmarks to Real-World Use | |
| | - Workshop 3: UQ — From Uncertainty Estimation to Operational Monitoring | |

| | Title | Speaker |
|---|---|---|
| | | |
| | AI in Aeronautics: Challenges in Demonstrating Compliance | Thibaut De Cremoux (Dassault-Aviation DGT) |
| | Challenges and Research Directions for Surrogate Model VVUQ | Nicolas Bousquet (EDF R&D) |
| | How can omics biology research leverage generative artificial intelligence? (while avoiding the pitfalls of hallucinations) | Thomas Burger (CEA/DRF) |
| | Causal modeling: a gold road for more trustworthy AI | Marianne Clausel (Univ Lorraine) |
| | Formal Verification of Neural Networks with PyRAT | Augustin Lemesle (CEA/DRT) |
| | Conformal predictions: from uncertainty<br><br>quantification to AI risk control with MAPIE | Geoffray BRELURUT (Capgemini)<br><br>Nicolas BRUNEL (LaMME, ENSIIE Université Paris Saclay<br><br>Invent Lab, Capgemini) |
| | Workshop Reports (3 × 20 min) | |
| | Summary and Proposed Actions | |

# Workshop 1 Summary: Verification — From Data to Systems



- **Applying traditional verification methods to AI remains challenging:** requirements, verification targets, and evidence are still poorly defined.
- **End-to-end verification:** data, training, implementation, and operational integration.
- **Combining complementary approaches:** analytical methods, formal verification, empirical testing, benchmarks, and reference physics-based models.
- **Key challenges:** evidence traceability, representative testing in high dimensions, and re-verification as models evolve.
- **Proposed roadmap:** formalize requirements, cover the validity domain, and establish evidence-based confidence levels.

# Workshop 2 Summary: Validation — From Benchmarks to Real-World Use



- **AI models must be validated,** especially for critical and regulated applications, yet there is no methodological consensus.
- **Use-case-specific validation:** comparison against experimental or numerical references and definition of relevant acceptance criteria.
- **Challenges of AI/simulation hybridization:** validating surrogate models, individual components, and their coupling within complex systems.
- **Limitations of conventional approaches:** inadequate performance metrics, unrepresentative benchmarks, and generalization challenges (high dimensionality, extrapolation, non-i.i.d. data).
- **Complementary approaches:** incorporating physical constraints and formal methods, while addressing robustness, cybersecurity, and reproducibility.
- **Research directions:** adapt validation methods from numerical simulation to AI-specific characteristics, including how models evolve over time.

# Workshop 3 Summary: UQ — From Uncertainty Estimation to Operational



- **Defining uncertainties** is complex and depends on models, data, and use cases: distinguishing aleatoric from epistemic uncertainty.
- **A wide range of quantification methods:** Gaussian processes, Bayesian approaches, neural network ensembles, MC Dropout, conformal prediction, and latent spaces.
- **Challenges in uncertainty validation:** calibration, coverage rates, and correlation between errors and estimated uncertainties.
- **Key challenges:** scalability, out-of-distribution detection, and managing extrapolation scenarios.
- **Operational use of uncertainty:** decision support, alert triggering, expert review, or reference simulations.
- **Future directions:** integrating uncertainty into active learning, model monitoring, and operational risk management.

# Maturity level mapping

Forum TERATEC
20 - 21 OCTOBRE 2026
PARIS · PALAIS DES CONGRÈS

**Ve**

**Traçabilité, gestion de versions**
e.g.: Git, DVC, MLflow

**Vérification logicielle**
e.g.: pytest, deterministic modes

**Formal verification**
Local properties · e.g.: PyRAT (CEA)

**Va**

**Statistical validation**
i.i.d. data, excluding rare events

**Empirical interpretability**
Saliency, attributions · e.g.: Captum

**UQ**

**Gaussian-process UQ**
Low-dimensional · e.g.: OpenTURNS, URANIE

**Sensitivity via metamodels**
Sobol · e.g.: OpenTURNS, URANIE

**Calibration, conformal prediction**
Exchangeability · e.g.: MAPIE

**Out-of-distribution detection, drift**
No universal method

**Internal interpretability**
Of deep neural networks

**Rare-event guarantees**
Very low error rates

**Epistemic uncertainty**
Of deep neural networks

**Confidence score outside the validity domain**

**Identifying sources of uncertainty**
To reduce them

TRL

# Outputs and Actions

Forum TERATEC 20 - 21 OCTOBRE 2026 PARIS · PALAIS DES CONGRÈS

- **Can the VVUQ concept used in simulation be applied directly to AI?**

- The required level of verification, validation, and uncertainty quantification depends on the intended objective:
  - Validity domain of the results for decision-making
  - More detailed uncertainty analysis to understand the sources of errors and potentially correct them

- Selected Use Cases
  - Seismic sensor networks — earthquake magnitude estimation — AI model verification using PyRAT
  - AI model for developing equations of state
  - Conformal prediction approach for surrogate models
  - Using PyRAT to propagate input-data uncertainty to output results