Title

NuclearGuard AI Agent-Assisted Radiation Anomaly Detection and Operator Decision Support Using Gamma-Ray Spectral Data

Author

Fang, Yung-Sian

Publication Date

2026-06-02

UCR

AI Graduate Research Brown Bag 2026

NuclearGuard AI
Agent-Assisted Radiation Anomaly Detection and Operator Decision Support Using Gamma-Ray Spectral Data

From AI agents reshaping academic research to safety-critical decision support.

https://ucr-ai-nuclear.firebaseapp.com

Yung-Sian(Sammy) Fang

UC Riverside School of Business

M.S. in Computer Science | Applied AI | Operations & Supply Chain Analytics

yungsian.fang@email.ucr.edu

AI Agents Are Reshaping Academic Research

AI is no longer only a writing or coding assistant.
It is moving toward agentic research workflows.

Recent work defines advanced AI assistants as artificial agents that can plan and execute sequences of actions on behalf of users across domains, raising both major opportunities and governance concerns (Gabriel et al., 2024).
Surveys of agentic AI for science also describe agents as systems that coordinate tools, data, reasoning, and evaluation across scientific workflows (Gao et al., 2025).

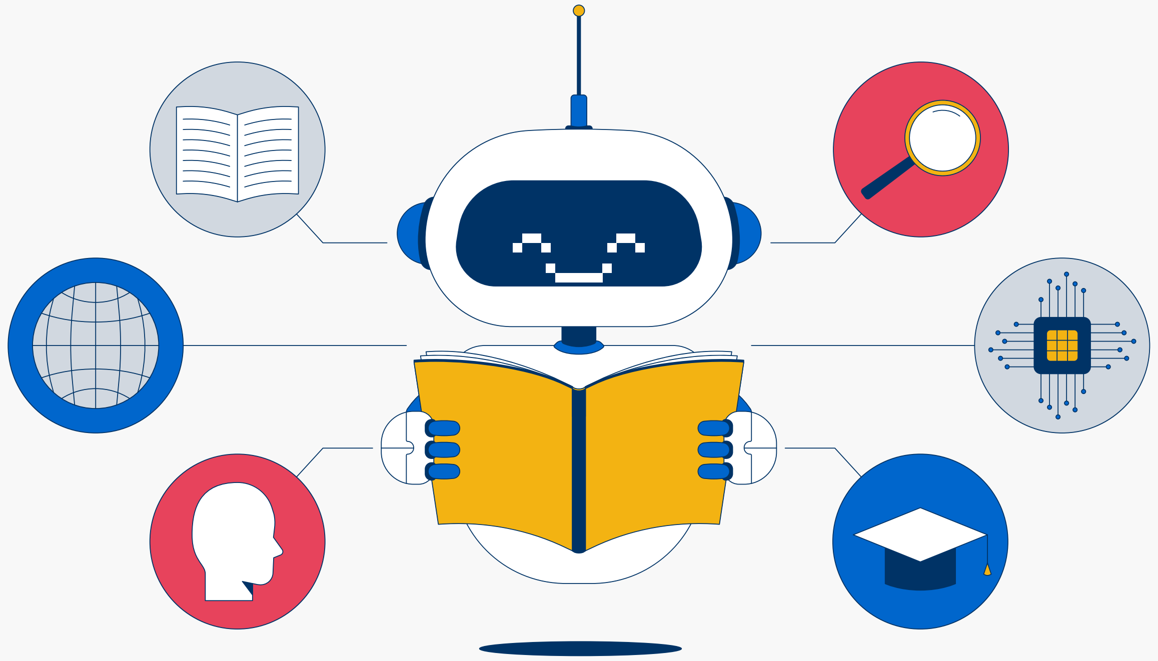
AI Agents Are Reshaping Academic Research

(Gabriel et al., 2024; Gao et al., 2025)

AI agents can **support**:

- **Data** retrieval
- **Literature** synthesis
- Experiment **planning**
- Tool calling
- Evidence interpretation
- Result summarization
- Decision documentation

Claude



The new question is not only

"Can AI generate an answer?"

The stronger question is:

Can AI **support** a **reliable**, **explainable**, and **accountable** research workflow?

From AI Tools to Human-Centered AI Systems

In safety-critical domains, **AI** should not be designed as a black-box replacement for experts.

A human-centered AI system should provide:

- High automation
- High human control
- Evidence visibility
- Clear uncertainty
- Action traceability
- Decision accountability

Theoretical foundation:

Human-Centered AI argues that **reliable**, **safe**, and **trustworthy** systems should combine automation with **meaningful human control**.

Situation Awareness = Perception + Comprehension + Projection

Shneiderman's Human-Centered AI framework argues that effective AI systems should increase automation while preserving human control, responsibility, and trustworthiness (Shneiderman, 2020). Situation awareness theory similarly emphasizes that operators in dynamic systems need perception, comprehension, and projection before making decisions (Endsley, 1995).

Why
Nuclear Radiation Monitoring Is a Strong Case

Radiation monitoring is a strong case for AI decision support because it combines:

- Continuous sensor data
- Background radiation variation
- Instrument noise
- Rare but important anomalies
- High cost of misinterpretation
- Need for human verification

The core question is not simply:
"Is radiation detected?"
The better question is:
"Is this signal normal, uncertain, or operationally significant?"

Signal = Background + Possible Source + Noise

Gamma-ray spectra can provide radionuclide "fingerprints," where energy-related spectral patterns help support identification and interpretation rather than relying only on gross count levels (IAEA, 2021). Radiation detection algorithms may operate on gross count rate, gamma-ray spectra, or both, making spectral analysis central to detection and classification tasks (Ghawaly et al., 2020).

How do we build?

```
┌───────────────┐     ┌───────────────┐     ┌────────────────┐
│ Data Sources  │ ──► │ Preprocessing │ ──► │ Feature Layer  │
└───────────────┘     └───────────────┘     └────────────────┘
 Gamma spectra         Normalization         Gross count
 Metadata              Baseline matching     Peak deviation
 Location/time         Noise handling        Fingerprint

      ┌────────────────┐     ┌────────────────┐
 ──►  │ Model Layer    │ ──► │ Decision Engine│
      └────────────────┘     └────────────────┘
       Threshold              Risk level
       Isolation Forest       Confidence score
       Autoencoder error      Recommended action

      ┌────────────────┐     ┌────────────────┐
 ──►  │ Operator UI    │ ──► │ Agent Briefing │
      └────────────────┘     └────────────────┘
       Alert queue            Situation
       Evidence charts        Evidence
       Decision buttons       Action
                              Documentation

      ┌───────────────────────────────────────┐
 ──►  │ Validation, Logging & Governance      │
      └───────────────────────────────────────┘
       Benchmark metrics | Decision log | Human review | Audit trail
```

NuclearGuard AI is structured as a modular decision-support architecture.

The system ***receives gamma-ray spectral data*** and ***monitor metadata, preprocesses the signal***, ***extracts spectral evidence***, ***applies anomaly detection models***, ***converts model outputs into risk levels***, and ***supports the operator through visualization***, ***action recommendation***, ***agent-assisted briefing, and decision logging***.

Gamma-Ray Spectral Data: The Evidence Layer

A ***gamma-ray spectrum*** is more informative than a single count.

Gross count answers:

How much radiation was detected?

Spectrum answers:

How is radiation distributed **across energy channels**?

Suspicious sources may appear as:

- Abnormal peaks
- Energy-channel deviations
- Distribution shifts
- Mismatch against background baseline

- Net spectral deviation:

$D(E_i) = x(E_i) - b(E_i)$

- Standard deviation:

$z_i = [x(E_i) - \mu_i] / \sigma_i$

$x(E_i)$ = observed count at energy channel i
μ_i = expected background mean
σ_i = background standard deviation

Gamma spectroscopy uses the relationship between energy peaks and radionuclide signatures to support identification; background subtraction and peak evaluation are therefore foundational steps in interpreting spectra (IAEA, 2021; Mirion, n.d.).

Model Layer: Anomaly Detection Methods

NuclearGuard AI compares multiple anomaly detection approaches:

1. **Gross Count Threshold**
2. **Spectral Peak Deviation**
3. **Isolation Forest**
4. **Deep Autoencoder Reconstruction Error**

Why multiple models?

- **Threshold models** are interpretable but simple.
- **Isolation Forest** detects unusual observations through isolation.
- **Autoencoders** learn normal spectral structure and flag high reconstruction error.

- **Threshold rule:**

Alert = 1 if score(x) > τ

- **Autoencoder reconstruction error:**

$r(x) = ||x - g(f(x))||^2$

- **Isolation Forest anomaly score:**

$s(x, n) = 2^{[-E(h(x)) / c(n)]}$

Where:

$h(x)$ = path length of x in isolation trees

$c(n)$ = average path length normalization term

Higher $s(x, n)$ indicates stronger anomaly likelihood

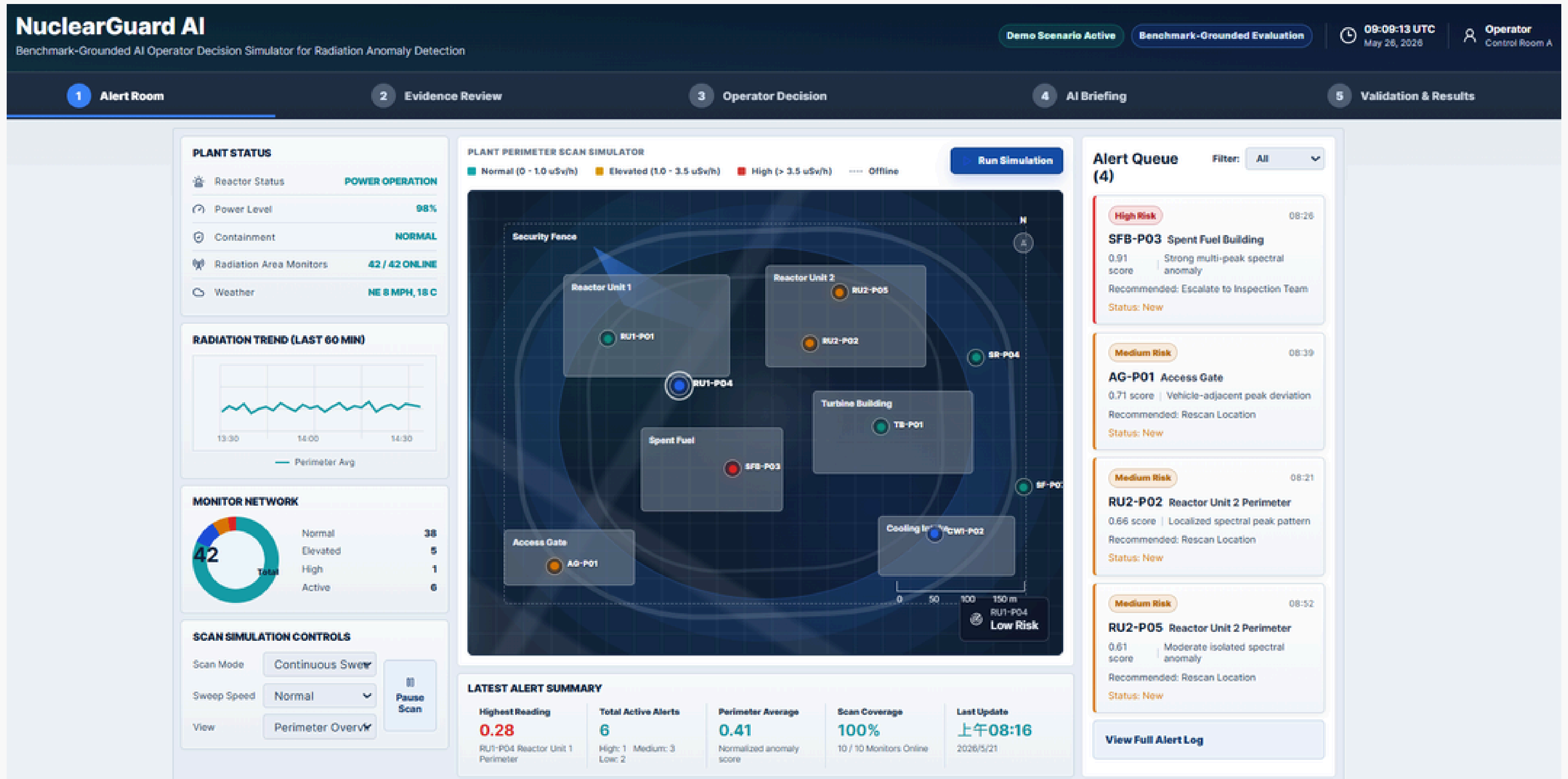
Isolation Forest detects anomalies by isolating observations rather than estimating density or distance, making it conceptually suitable for unusual-pattern detection (Liu et al., 2012). Autoencoders learn compressed representations and reconstruct inputs; reconstruction error can be used as an anomaly signal when abnormal observations are harder to reconstruct (Goodfellow et al., 2016; TensorFlow, 2024).

Interactive Evidence Review

AI outputs must be inspectable before they can support safety decisions.

NuclearGuard AI visualizes:

- Selected monitor location
- Risk level and anomaly score
- Observed spectrum
- Background baseline
- Difference curve
- Suspicious energy-channel peaks
- Spectral fingerprint heatmap



Operator questions:

- What changed?
- Where did it change?
- How strong is the anomaly?
- Does the evidence justify further review?

Agent Layer: Briefing & Decision Support

NuclearGuard AI includes an agent-style briefing layer.

The agent receives:

- Alert metadata
- Risk level
- Anomaly score
- Spectral evidence
- Selected operator action
- Recommended next step

The agent generates:

- Situation
- Evidence
- Recommended action
- Documentation note

Design guardrail:

The agent should not claim confirmed release.
It should frame output as decision support requiring verification.

Important decision-support logic

Recommended action = argmax_a U(a | risk, evidence, uncertainty)

- a = possible operator action
- U = expected utility of action
- risk = severity level
- evidence = spectral and model evidence
- uncertainty = model and operational uncertainty

Advanced AI assistants are defined as agents that can plan and execute sequences of actions for users, but their autonomy also creates ethical risks around alignment, misuse, overreliance, and accountability (Gabriel et al., 2024). In safety-critical systems, this means agent outputs should remain bounded, evidence-based, and reviewable.

Validation, Metrics, and Benchmark Grounding

The prototype is grounded in public radiation detection benchmark concepts.

Evaluation task:

- Background vs. source-present detection
- Anomaly prioritization
- False alarm control
- Operator review-load reduction

Relevant metrics:

- ROC-AUC
- Precision
- Recall
- F1 score
- Recall at fixed false positive rate
- Top-k precision
- Estimated review-load reduction

Precision:

Precision = TP / (TP + FP)

Recall:

Recall = TP / (TP + FN)

F1 score:

F1 = 2 × Precision × Recall / (Precision + Recall)

False positive rate:

FPR = FP / (FP + TN)

The ORNL Scientific Data benchmark provides synthetic urban gamma-ray spectra for training and testing radiation detection algorithms, with algorithms operating on gross count rates, spectra, or both (Ghawaly et al., 2020). ORNL's more recent radiation detection algorithm guidance recommends reporting performance metrics and considering background variability, shielding, detector response, and machine-learning methods (Runkle et al., 2024).

Demo metrics must be replaced with reproducible offline benchmark outputs before being presented as final research results.

Academic &Business Value

Main takeaway:

The future of AI in safety-critical systems is not full automation.

It is better evidence, clearer decisions, and more accountable ***Human-AI Collaboration***.

Metric	Value	Meaning
ROC-AUC	**0.9**	Strong model discrimination
Recall @ 5% FPR	**0.71**	Better anomaly capture under controlled false alarms
Top-10 Precision	**0.8**	Higher-quality alert prioritization
Review-Load Reduction	**36%**	Less low-value manual review

Academic value

- AI agents for safety-critical research workflows
- Data interpretation, evidence synthesis, and decision documentation
- Radiation anomaly detection as a concrete AI research case
- Explainability, situation awareness, and human-centered AI
- Reproducible benchmark evaluation pathway

Business and product value

- Lower signal overload
- Better alert prioritization
- Stronger decision documentation
- Faster escalation and review
- More accountable human-AI collaboratio

Machine-learning alarm analysis has practical value in radiation monitoring because nuisance alarms can create significant operational workload, and performance evaluation should consider field review burden rather than only abstract classification accuracy (Labov et al., 2025). This supports the product framing of NuclearGuard AI as workflow improvement, not only model development.

Furture

Product Builder

&

Cross Researcher

References

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32–64. https://doi.org/10.1518/001872095779049543

Gabriel, I., Manzini, A., Keeling, G., Hendricks, L. A., Rieser, V., Iqbal, H., Tomašev, N., Ktena, I., Kenton, Z., Rodriguez, M., El-Sayed, S., Brown, S., Akbulut, C., Trask, A., Hughes, E., Bergman, A. S., Shelby, R., Marchal, N., Griffin, C., … Isaac, W. (2024). The ethics of advanced AI assistants. arXiv. https://doi.org/10.48550/arXiv.2404.16244

Gao, L., Wang, Z., Chen, J., Zhao, Y., Zhang, H., & Liu, Y. (2025). Agentic AI for scientific discovery: A survey of progress, challenges, and future directions. arXiv. https://arxiv.org/abs/2503.08979

Ghawaly, J. M., Jr., Nicholson, A. D., Peplow, D. E., Willis, M. J., & Archer, D. E. (2020). Data for training and testing radiation detection algorithms in an urban environment. Scientific Data, 7, Article 328. https://doi.org/10.1038/s41597-020-00672-2

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. https://www.deeplearningbook.org/

International Atomic Energy Agency. (2021). Gamma and alpha spectrometry for workplace monitoring. https://www.iaea.org/

Labov, S. E., Keillor, M. E., Davis, K. L., Nattress, J. A., & Jarman, K. D. (2025). Improving operational performance using machine learning alarm analysis for radiation portal monitors. Nuclear Instruments and Methods in Physics Research Section A, 1068, 169740. https://doi.org/10.1016/j.nima.2024.169740

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2012). Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data, 6(1), 1–39. https://doi.org/10.1145/2133360.2133363

Mirion. (n.d.). Nuclear measurement fundamental principles: Spectrum analysis. Retrieved May 26, 2026, from https://www.mirion.com/discover/knowledge-hub/articles/education/nuclear-measurement-fundamental-principle-spectrum-analysis

Runkle, R. C., Archer, D. E., Ghawaly, J. M., Nicholson, A. D., Peplow, D. E., & Willis, M. J. (2024). Metrics and methods for radiation detection algorithm development. Oak Ridge National Laboratory. https://info.ornl.gov/sites/publications/Files/Pub193589.pdf

Sahoh, B., & Choksuriwong, A. (2023). The role of explainable artificial intelligence in high-stakes decision-making systems: A systematic review. IEEE Access, 11, 28704–28719. https://doi.org/10.1109/ACCESS.2023.3259041

TensorFlow. (2024). Intro to autoencoders. https://www.tensorflow.org/tutorials/generative/autoencoder

AI Graduate Research Brown Bag 2026

Thank You

NuclearGuard AI
Agent-Assisted Radiation Anomaly Detection and Operator Decision Support Using Gamma-Ray Spectral Data

Yung-Sian(Sammy) Fang

UC Riverside School of Business
M.S. in Computer Science | Applied AI | Operations & Supply Chain Analytics
yungsian.fang@email.ucr.edu