

SANDIA REPORT

SAND2026-19577

Printed April 2026



**Sandia
National
Laboratories**

PV Operations Software Transparency: A PVMAC Industry Snapshot

Marios Theristis

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: <http://www.osti.gov/scitech>

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: <https://classic.ntis.gov/help/order-methods/>



ABSTRACT

The rapid growth of photovoltaic (PV) deployment has increased reliance on software platforms for monitoring, workflow automation, diagnostics, and performance analytics. As these tools play a central role in asset management and operations and maintenance (O&M), greater transparency in methodologies, data handling, and validation practices benefits the broader PV ecosystem. To better understand current practices and identify opportunities for improved clarity and interoperability, 24 software providers contributed detailed responses through the PV O&M Analytics Collaborative (PVMAC) initiative, the first structured questionnaire of its kind in the industry, covering onboarding, interoperability, data quality, diagnostics, AI/ML, and other operational categories. These providers represent over 1.1 TW of solar assets under management. The analysis shows broad adoption of digital twins, AI/ML, and API integrations, but also highlights challenges in onboarding processes, inconsistent definitions and methodologies, variability in key performance indicator (KPI) calculations, and limited independent validation. Greater standardization, clearer documentation, and stronger validation frameworks could improve transparency, comparability, and trust across PV operations software platforms.

ACKNOWLEDGEMENTS

The author gratefully acknowledges Dr. Clifford W. Hansen for reviewing and providing feedback on this report, as well as the following software companies and their representatives for their participation and support in this collaborative effort to advance transparency in PV operations software:

3E (Gofran Chowdhury)
Above Surveying Ltd (Adrien van den Abeele)
Alectris (Sofia Zagori)
Amperecloud (Valentina Maurer)
Denowatts (Dan Leary)
Ecosuite, Inc (Philip Court)
enSights (Shoshana Browne, Alon Mashkovich, Ed Wagner)
Enurgen, Inc. (Christopher Valdivia)
Gantner Instruments Environment Solutions GmbH (Juergen Sutterlueti)
IBM (Zach Ruland)
Mana Energy (Zoltan Milaskey)
Power Factors (Sam Costello)
PowerUQ (Chetan Chaudhari)
Proximal Energy, Inc. (Rob van Haaren)
Quadrical, Inc. (Pragya Chawla)
Raicoon GmbH (Ralf Tschanun)
Raptor Maps (Leo Kim)
Sitemark (Maxime Bossens)
Skylark Drones (Nihal Mohan)
SmartHelio PBC (Shankaransh Srivastava)
Solargrade, Inc. (Tori Clifford)
SolarSquare Energy Pvt Ltd (Abhay Tilwankar)
Stem, Inc. (Patrick Keelin)
Wattch (Ryan Babaie)

This work was supported by the U.S. Department of Energy's Office of Critical Minerals and Energy Innovation (CMEI), Office of Energy Technology (E-Tech) under the Integrated Energy Systems Office Award Number 52770. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This report describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the report do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

CONTENTS

Abstract	3
Acknowledgements.....	4
Executive Summary	8
Acronyms and Terms	9
1. Overview and Participation	10
2. Platform Types, Hardware Requirements, and Scope.....	11
3. Onboarding.....	12
3.1. Fresh install.....	12
3.2. Additional sites.....	12
3.3. Automation level.....	12
3.4. Required data inputs.....	13
4. Interoperability and Integration.....	14
4.1. Public APIs and data accessibility	14
4.2. Third-party data integration	14
4.3. Data export targets	14
4.4. Supported data formats.....	15
4.5. Temporal resolution capabilities.....	15
5. Data Quality.....	16
5.1. Incoming data validation and anomaly flagging.....	16
5.2. Missing and erroneous data handling.....	16
5.3. Transparency of data quality processes	16
6. Imaging Analytics.....	18
6.1. Supported imagery	18
6.2. Geolocation and fault overlay.....	18
7. Digital Twin Capabilities	19
7.1. Definitions of digital twin.....	19
7.2. Primary purpose of digital twin	19
7.3. Basis and capabilities of digital twins	20
8. Expected Yield and Performance Modeling.....	21
8.1. Source of expected yield	21
8.2. Transparency and adjustments.....	21
9. Key Performance Indicators	23
9.1. Technical KPIs.....	23
9.2. Energy storage KPIs.....	23
9.3. Financial KPIs	24
10. Use of AI/ML and Diagnostics.....	25
10.1. Purpose of AI/ML	25
10.2. Summary of AI/ML implementation examples.....	25
10.3. Performance issues that can be detected and classified	26
10.4. Link to workflows (e.g., work order generation or export to CMMS)	27
11. Transparency and Documentation	28
11.1. Documentation of platform's methods.....	28

11.2. Alignment with industry standards, publications or best practices	28
11.3. Reproducibility of KPIs and expected yield	29
12. Business Model and Support.....	30
12.1. Pricing models	30
12.2. Technical support	30
12.3. User training and documentation	30
13. Key Takeaways	31
References.....	32
Distribution.....	33

LIST OF FIGURES

Figure 1. Geographic distribution of assets monitored by the participating software providers (left) and their primary customer base (right).	10
--	----

This page left blank

EXECUTIVE SUMMARY

The industry lacks clear and consistent understanding of PV operations software, and many stakeholders are not fully aware of the importance of these tools for improving operational performance. This makes it difficult to assess capabilities, compare approaches, and make informed decisions. PVMAC conducted the first structured, industry-wide questionnaire of software providers to assess capabilities, integration practices, and operational functionality. The aggregated findings presented here assess industry maturity, highlight strengths and gaps, and identify areas for improvement. Rather than serving as a vendor catalogue, the report focuses on cross-cutting trends, with raw responses published separately for transparency. The table below summarizes the key findings and takeaways.

Section	Summary	Takeaway
Onboarding	First-site deployment often <1 week; additional sites integrate in hours if SCADA/metadata are consistent. Automation limited (~50% semi-automated, ~8% fully automated).	Data ingestion and harmonization are key bottlenecks; scalable analytics require upstream consistency.
Interoperability and integration	APIs widely available (~70%), but documentation varies; ~30% restricted. Integration with CMMS, ERP, PV design software, forecasting, market data are common but often bilateral.	Technically connected but not standardized; inconsistent integrations limit automation and scalability.
Data quality	All platforms perform QC (range checks, cross-sensor validation, statistical filtering). Missing/erroneous data handled via filtering (~75%) and/or imputation (~67%). Transparency varies.	QC methods differ, affecting KPIs and expected yield; shared frameworks would improve consistency and trust.
Imaging analytics	~40% support RGB imagery with geolocated faults overlay on digitized plant layouts; IR, satellite, EL/PL, LiDAR less common.	Imaging is more periodic than continuous; insights depend on geospatial data rather than operational data.
Digital twin	~75% report DT capability: physics-based (~38%), field performance (~20%), AI-based (~14%), geospatial (~28%). Mainly for expected yield (~60%), less often predictive or command/control (~30-35%).	Needs clearer, standardized definitions to distinguish simulation, operational (push and receive), or predictive twins.
Expected yield	~83% calculate expected energy (physics-based, proforma, ML, hybrid). Adjustments for weather, curtailment, clipping, degradation vary. Transparency partial (~80%).	Inconsistent assumptions cause variability in reported plant performance; core source of analytics opacity.
KPIs	All platforms report KPIs. Definitions, normalization, and handling of missing data vary.	Lack of standardized KPI framework limits comparability and reproducibility.
Use of AI/ML and diagnostics	~85% use AI/ML; primarily for diagnostics, predictive maintenance, performance prediction. Additional use in data quality, imagery analysis, LLM interfaces. Strong fault detection; work order linkage ~80%.	Predictive analytics feasible only on serviceable components; methodological clarity is needed.
Transparency and documentation	Documentation access: ~45% user-accessible, ~17% public, ~25% NDA, ~10% unavailable. Reproducibility claims ~50%.	Transparency improving, but still the weakest link. Documentation and robust validation make software outcomes “auditable” and increase confidence for asset owners.
Business models and support	Pricing: MW-based subscriptions, enterprise agreements, pay-per-use. ~60% undisclosed. Support widely offered: tutorials, live sessions, certifications.	Pricing opacity limits independent evaluation; technical support generally strong.

ACRONYMS AND TERMS

Acronym/Term	Definition
AI	Artificial Intelligence
API	Application Programming Interface
BESS	Battery Energy Storage Systems
BIM	Building Information Modeling
CAD	Computer-Aided Design
C&I	Commercial and Industrial
CMMS	Computerized Maintenance Management Systems
CRM	Customer Relationship Management
DOD	Depth of Discharge
EL	Electroluminescence
EMS	Energy Management System
ERP	Enterprise Resource Planning
IEC	International Electrotechnical Commission
IR	Infrared
IRR	Internal Rate of Return
KPI	Key Performance Indicators
LiDAR	Light Detection and Ranging
ML	Machine Learning
NPV	Net Present Value
O&M	Operations and Maintenance
PL	Photoluminescence
PPA	Power Purchase Agreement
PR	Performance Ratio
PV	Photovoltaics
PVMAC	PV O&M Analytics Collaborative
REST	Representational State Transfer
RGB	Red, Green, Blue
ROI	Return of Investment
RTE	Round-Trip Efficiency
SOC	State of Charge
SOH	State of Health
SCADA	Supervisory Control and Data Acquisition
UI	User Interface

1. OVERVIEW AND PARTICIPATION

The market for photovoltaic (PV) operations software has expanded rapidly, with platforms increasingly central to monitoring, performance analytics, and operations and maintenance (O&M). Understanding the capabilities, data practices, and methodological approaches of these tools is important for performance engineers, owners, O&M providers, and other stakeholders in the industry.

To better understand current practices, the PV O&M Analytics Collaborative (PVMAC) conducted the first structured, industry-wide questionnaire of software providers covering asset management, monitoring, imaging, and analytics platforms. It was announced on August 27, 2025, via LinkedIn, email invitations, and online webform requests. Submissions were received until December 1, 2025. A total of 42 software providers were invited to respond to 56 questions around platform scope, onboarding, interoperability and integration, data quality, imaging analytics, digital twin capabilities, expected yield modeling, KPIs, AI and diagnostics, transparency and documentation, business models, and support.

In total, 24 software providers participated, collectively representing over 1.1 TW of solar assets under management across > 115,000 sites. Most respondents offer integrated monitoring and analytics, with smaller shares providing imaging analytics, asset management, or field workflows.

The reported asset base is geographically distributed, with approximately 58% located in North America, 25% in Europe, and the remainder in Asia and Africa (see Fig. 1). Respondents serve a balanced mix of utility-scale and commercial & industrial (C&I) projects

Responses to each section of the questionnaire are summarized in individual sections in this document, and a compiled takeaway section appears at the end of this report. The raw responses from the software providers are available as a separate Microsoft Excel file.

Sandia National Laboratories compiled and analyzed the responses for trends and observations. However, Sandia did not independently verify or validate the accuracy of the information provided by the software providers. As such, this report reflects the responses provided by the software providers without additional verification.

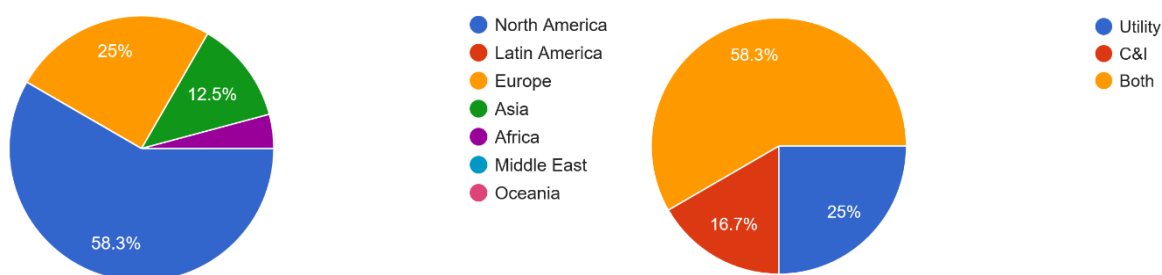


Figure 1. Geographic distribution of assets monitored by the participating software providers (left) and their primary customer base (right).

2. PLATFORM TYPES, HARDWARE REQUIREMENTS, AND SCOPE

The participating PV operations software platforms span several functional types (some overlap):

- **Integrated monitoring and analytics:** The majority (~70%) provide both real-time monitoring and performance analytics.
- **Fieldwork management platforms:** Some specialize in automated/manual field operations, including inspections, or other workflows (~10%).
- **Imaging analytics platforms:** A subset (~20%) emphasizes aerial or drone-based image collection and analysis for defect detection and plant visualization.
- **Asset management:** A few platforms (~25%) combine monitoring, analytics, field work, imaging, and portfolio-level asset management, including financial, contractual, and operational oversight.

Hardware requirements vary:

- **No proprietary hardware (~80%):** platforms run on standard browsers or mobile devices; users rely on existing Supervisory Control and Data Acquisition (SCADA), or meters.
- **Optional hardware (~15%):** certain platforms offer hardware, or third-party sensors for enhanced diagnostics or autonomous inspections.
- **Proprietary hardware (<5%):** a small number require dedicated devices for advanced control or measurement functions.

Primary use cases include:

- Daily monitoring and operational dashboards (~95%)
- Automated detection and classification of performance losses (~85%)
- Work prioritization and O&M ticketing (~75%)
- Predictive maintenance (~55%)
- Mobile field inspections (~45%)
- Financial reporting and contractual KPIs (~55%)
- Advanced use cases, e.g., autonomous operations, DER control, or AI-based compliance (~5–10%).

Typical users are PV asset owners, O&M teams, asset managers, performance engineers, and technicians, among other stakeholders.

3. ONBOARDING

Onboarding sets up the foundation for a PV system in an operations platform. In this phase, devices are defined, SCADA tags are mapped, electrical topology is constructed, and baseline models for expected energy and KPIs are established. Decisions made during onboarding affect downstream performance calculations, availability metrics, and loss attributions, making onboarding a key indicator of platform maturity and scalability.

3.1. Fresh install

Onboarding timelines for a first site vary, but several patterns emerge:

- ~54% report onboarding in < 1 week
- ~17% report < 1 month
- ~4% report 1-3 months
- ~25% respond “Depends” (hardware, portfolio size, data availability).

Most providers indicate that a first site deployment can be completed in less than one week when structured data and Application Programming Interface (API) access are available. However, approximately one quarter of respondents note that onboarding duration depends on hardware installation requirements, fleet/plant size, or completeness of site documentation and data (e.g., as-built).

The dominant bottleneck is data readiness, not analytics configuration. When site metadata and SCADA access are structured and complete, integration is fast. When metadata are inconsistent or incomplete, onboarding time increases regardless of a platform’s capability.

3.2. Additional sites

Reported onboarding times for additional sites vary considerably:

- ~21% report < 1 hour
- ~17% report 1-12 hours
- ~17% report 1 day
- ~17% report 2-7 days
- Remaining responses indicate dependency on SCADA heterogeneity, hardware, data quality or customization requirements.

Where fleets maintain consistent SCADA schemas and naming conventions, additional sites can be integrated rapidly, often within hours. However, in portfolios where naming conventions, sensor mapping, or layout documentation etc. differ from site to site, onboarding requires additional manual setup. Therefore, fleet scalability depends more on upstream data harmonization than on platform architecture alone. Even highly automated systems require site-specific adjustments.

3.3. Automation level

With respect to the onboarding automation levels:

- ~50% report semi-automated (UI-based ingestion + manual validation): Platforms automate data import via UI or csv upload, but manual review is needed to ensure correct device mapping, hierarchy and SCADA tag alignment
- ~8% report fully automated (API-driven) where credentials and connectors allow zero-touch integration. Only two providers report fully automated onboarding without human configuration review
- ~40% report hybrid or manual processes (provider-led or customer-led).

Some platforms require completely manual onboarding, especially when site layouts or SCADA systems are nonstandard. Other platforms offer hybrid approaches adjusting the level of automation based on site complexity, existing infrastructure or customer preference. Although many platforms support API-based ingestion, most still require manual validation. Therefore, “zero-touch” onboarding remains uncommon. Manual configurations may introduce variability in plant interpretation and performance baseline assumptions. Since onboarding information affects expected yield calculations, differences introduced at this stage will propagate to KPI estimations and loss analysis.

3.4. Required data inputs

The most common data required to onboard a site include:

- SCADA or monitoring system data
- AC-level measurements
- Power meter or utility data
- Meteorological data (for platforms that run yield models)
- API credentials for third-party systems
- Plant metadata (e.g., layout, as-built documentation, spec sheets).

Less common data that improve accuracy and/or enable advanced features include:

- DC-level measurements
- Tracker rotation data
- IV curve data
- Thermography data
- Financial, legal, contractual information.

Imaging-focused platforms generally require fewer operational inputs but rely on geospatial mapping and digitized plant layouts for fault overlay. The range of required inputs reflects differences in analytical depth: platforms offering detailed loss breakdowns, weather-adjusted KPIs, or predictive diagnostics require more granular input data and metadata while simpler monitoring solutions can operate with fewer inputs but may provide less detailed insights. There is no standardized minimum onboarding dataset across providers; the depth of required data directly reflects the platform’s capabilities.

4. INTEROPERABILITY AND INTEGRATION

Interoperability determines how effectively a platform exchanges data with external systems. Integration capability affects automation, data consistency, vendor lock-in risk, and the scalability of analytics across heterogeneous portfolios.

4.1. Public APIs and data accessibility

Nearly 70% of the software products provide a public documented API. The remaining ~30% either:

- Provide API access under NDA
- Offer limited or private API access
- Do not provide a public and documented API capability.

While API availability is common, the level of documentation and openness varies. Some providers offer fully documented APIs whereas others provide access only through commercial agreements or enterprise tiers.

Software providers were also asked about data export and migration. For example, if an asset owner wants to migrate to another platform or wants to export data for their O&M provider, is this possible and is there an additional charge?

- ~60% allow free data export (e.g., CSV or API access)
- ~25% indicate additional costs for export or migration
- Remaining responses are unclear or restrictive.

Although most platforms support data export, the commercial and technical conditions vary. Differences in access structure, pricing, and implementation requirements may influence how easily data can be transferred between systems.

4.2. Third-party data integration

Platforms report integration with a broad set of external systems:

- ~90% integrate with satellite irradiance or weather data providers
- ~75% with CMMS (Computerized Maintenance Management Systems)
- ~67% with PV design platforms
- ~58% with ERP (Enterprise Resource Planning) systems
- ~50% with forecasting services
- ~45% with market pricing feeds.

Additional integrations mentioned include Energy Management System (EMS) data, thermography, CAD drawings, Customer Relationship Management (CRM) data, and equipment databases. Integrations are frequently described as API-based or custom-configured. It should be noted that integrations are often bilateral and custom configured. This increases implementation complexity and integration costs, and introduces potential risks, e.g., related to data mapping.

4.3. Data export targets

Platforms report exporting data to:

- CMMS platforms
- Asset management platforms
- ERP systems
- SCADA historians
- Robotics/autonomy platforms
- Other analytics providers
- PV design platforms.

General API-based exports are widely supported (~87%). Most platforms support outward data flow. True bidirectional synchronization (analytics ↔ maintenance ↔ financial systems) cannot be examined by the generic nature of this question.

4.4. Supported data formats

Support for standard tabular and structured data formats is widespread. Commonly supported formats are CSV, JSON, Excel, XML, whereas some platforms support CAD formats such as DWG, or PVPollada (pvc). Some platforms additionally support industry-specific design and simulation formats.

4.5. Temporal resolution capabilities

Respondents indicate variability in supported temporal resolution:

- Monitoring ingestion: sub-minute to 15-minute intervals commonly supported
- Analytics calculations: typically, 5-minute to hourly; some aggregate to daily
- Data storage: ranges from full raw data archival to aggregated.

Higher-resolution ingestion (e.g., 1-minute or sub-minute data) is reported by several platforms. However, not all platforms maintain raw-resolution storage indefinitely. Differences in temporal resolution handling can influence expected yield accuracy (e.g., due to clipping estimation accuracy) or may bias KPIs.

5. DATA QUALITY

Data quality affects the reliability of all downstream analytics. Validation rules, anomaly detection, and how missing data are handled directly influence KPI accuracy, expected yield, availability and financial calculations.

5.1. Incoming data validation and anomaly flagging

All 24 respondents indicate that they perform validation on incoming data. However, the methods applied differ in sophistication and disclosure. Commonly reported validation approaches include:

- ~80% perform range checks (e.g., physical bounds for irradiance, temperature, power)
- ~70% perform cross-sensor consistency checks
- ~63% apply statistical filters (e.g., outlier detection)
- ~63% perform model-based checks (comparison against expected performance).

The anomalous readings are typically flagged with:

- Automated alarms (75%)
- QA dashboards (54%).

Several providers reference proprietary validation engines or model-based checks at different hierarchy levels (e.g., string, inverter, etc.). One provider reports protocol-level validation (e.g., Modbus/TCP error detection). Imaging-focused platforms describe automated and manual QA processes for drone data whereas one provider indicates that this information is not for public release.

Overall, validation commonly combines rule-based filtering with model- and/or statistics-based consistency checks. When it comes to flagging, some platforms reference anomaly detection algorithms or ML tools under development. In a few cases, alarms are grouped and automatically prioritized (e.g., by revenue impact). Flagging mechanisms range from simple alarms to categorized event tracking systems.

5.2. Missing and erroneous data handling

Handling of missing or erroneous data typically includes:

- Filtering out (75%)
- Imputation using modeled or statistical replacement (~67%)

with a subset reporting manual backfilling, use of backup or peer data sources, categorization by duration or controlled (revise/reject) by the user.

Approaches vary in how filtered and imputed values are distinguished within downstream calculations. Treatment decisions (e.g., filtering or imputing) can bias performance KPI calculations, particularly in locations with dynamic weather conditions [1-3].

5.3. Transparency of data quality processes

The majority of software providers (~80%) indicate that their data quality processes are transparent to users via the following mechanisms:

- Per-timestamp quality flags

- Event tagging for data corrections
- Hover-level data traceability in dashboards
- Downloadable data quality reports (daily/weekly/monthly)
- Traceability of data source (raw vs. corrected)
- Undo capability for manual overrides.

However, granularity differs:

- Some platforms label every timestamp
- Others provide aggregated or generic quality reports
- A subset indicates limited or no user-facing documentation of quality handling.

Only a small number of respondents indicate that every timestamp is explicitly tagged with validation status (e.g., raw, filtered, imputed, modeled). In many cases, transparency refers to dashboard visibility or reporting features rather than reproducibility. Public or user-accessible documentation rarely includes explicit filtering thresholds or model residual criteria.

6. IMAGING ANALYTICS

Imaging analytics capabilities vary significantly across respondents. While 75% of the companies answered imaging-related questions, only a limited number operate dedicated drone or aerial inspection workflows. For many platforms, imaging functionality is either limited, integrated through third parties, or not a feature of the product.

6.1. Supported imagery

Of platforms that support imaging (~54%), reported imagery includes:

- RGB (Red, Green, Blue) imagery (~75%)
- Satellite imagery (~60%)
- Infrared (IR) imagery (~45%)

with electroluminescence/photoluminescence, Light Detection and Ranging (LiDAR) supported by less than 25%. RGB imagery is the most consistently supported type whereas infrared is sometimes supported via integration with external drone inspection providers.

Most imaging-enabled platforms focus on aerial drone-based RGB and infrared inspections due to cost and operational feasibility.

6.2. Geolocation and fault overlay

Software providers that support imaging indicated that detected faults can be geolocated and overlaid onto:

- Digitized plant layouts (e.g., inverter, string, module level) (~75%)
- Aerial orthoimagery (~55%)

Digitized plant layouts are the most common integration target. Fault mapping at string or module resolution is reported by some platforms, though the level of granularity varies depending on plant mapping accuracy and imaging resolution. Only 2 providers explicitly mention integration with 3D models.

7. DIGITAL TWIN CAPABILITIES

Digital Twin is one of the most inconsistently defined terms. 75% of participating platforms state that they offer or support a digital twin capability. However, definitions, functional scope, and technical implementation vary substantially. In the majority of cases, “digital twin” refers to an expected yield model with live data ingestion rather than a full cyber-physical representation that besides receiving, also pushes information.

7.1. Definitions of digital twin

Out of the 21 provided definitions of digital twin, one definition appears more often: ***“A real-time (sometimes dynamic) virtual replica or a digital reconstruction/representation of a physical solar asset”***.

However, what constitutes that reconstruction/representation varies significantly in architecture, granularity, fidelity and purpose. We identified four primary categories, although the percentages are approximate since some definitions overlap between categories:

1) Physics-based twins (~38%)

The most common definition describes a digital twin as a physics-based PV performance model of the plant’s electrical and/or mechanical structure. These twins simulate expected behavior using weather inputs and system configuration, often operating in parallel with the physical asset to calculate electrical parameters.

2) Field performance twins (~20%)

Some platforms define the digital twin as a real-time expected or learned power model fitted or trained on operational data. In these cases, the twin functions primarily as a performance benchmark rather than a structural reconstruction of the plant.

3) AI-based adaptive twins (~14%)

This category describes data-driven twins that continuously learn from operational behavior. These implementations use AI or machine learning to model asset response, generate granular benchmarks, and support use-case-specific diagnostics (e.g., soiling, shading, degradation).

4) Geospatial twins (~28%)

Several definitions focus on equipment-level, geospatial representations of the plant. These twins integrate electrical layouts, images, field reports, and operational data, often serving as a system of record for lifecycle management and work orchestration.

7.2. Primary purpose of digital twin

Among platforms that report digital twin capability, the twins serve four primary purposes:

- Simulation of expected performance (60%)
- Real-time status mirroring and deviation detection (30%)
- Predictive analytics (30%)
- Work planning and dispatch (30%)

Most platforms’ digital twin capabilities are for simulation and modeling. In these cases, the twin is mainly used to estimate expected yield under given conditions. Overall, while digital twin

implementations vary structurally, the dominant purpose across platforms is performance modeling, with operational monitoring and predictive functions layered on top.

7.3. Basis and capabilities of digital twins

Digital twins are primarily built on simulation models (~50%), often using physics-based models. A significant portion (~35%) combines empirical models fitted to historical operational data, while ~15% rely on CAD/BIM-based 3D structures or site-specific as-built information. About 20% additionally incorporate real-time sensor inputs or ML-driven behavior models, creating dynamic, continuously updating representations.

Regarding capabilities:

Historical and geospatial analytics: ~95% support visualization of historical performance and faults geographically or temporally.

Real-time data integration: ~85% of digital twins ingest live operational data from the plant, enabling immediate visibility of performance and device status.

Command and control features: Only ~35% can push commands to equipment (e.g., request IV curves, reset inverters).

While simulation or physics-based modeling remains the foundation for most digital twins, the most capable platforms combine simulation, empirical data, and real-time sensor inputs to enable monitoring, predictive analysis, and limited operational control. Capabilities vary widely, and more advanced twins require higher implementation effort but provide richer insights.

8. EXPECTED YIELD AND PERFORMANCE MODELING

Expected yield modeling is a foundational component of PV operations platforms. Approximately 83% of participating providers report calculating expected energy. The remaining respondents are primarily imaging-only platforms or tools focused on asset management without energy modeling functionality. While expected yield calculation is widely implemented, modeling methodologies, assumptions, and transparency levels vary significantly.

8.1. Source of expected yield

Respondents reported the following modeling approaches:

- 1) Physics-based models trained on operating data (~55%)
- 2) Physics-based models not trained on operating data (~10%)
- 3) Proforma model from commissioning/design (~35%)
- 4) Machine learning or regression models trained on past data (~20%)
- 5) Machine learning or regression models trained on proforma model (0%)

Approximately ~20% of platforms use hybrid approaches combining physics-based, proforma, and machine learning or regression models. These methodological differences influence performance interpretation. For example:

- A model calibrated on early operational data that already include degradation may normalize that degradation and become unable in detecting it.
- A static proforma baseline will not reflect real-world weather conditions, operational changes, or evolving loss factors.
- Data-driven models depend heavily on training windows and feature selection.

As a result, identical data may yield different loss breakdowns depending on the chosen modeling framework and calibration strategy. Expected yield is therefore not standardized across platforms, complicating benchmarking and performance loss attribution.

8.2. Transparency and adjustments

Most platforms (~80%) allow users to compare multiple expected yield estimates, such as from comparing a proforma design model to an adjusted yield. Expected yield is commonly adjusted for:

- Weather effects (nearly all platforms)
- Inverter clipping (~74%)
- External curtailment (~68%)
- Proforma losses/derates (~74%)
- Grid or external losses (~58%)
- Data resolution effects, e.g., hourly vs subhourly bias (~60%)

Additional adjustments may include temperature effects, spectral corrections, degradation, availability, shading, snow, soiling, tracking, and terrain. The sequence and logic of these adjustments is not standardized. For example:

- Curtailment may be removed before or after KPI calculations
- Clipping may be explicitly modeled via DC/AC limits or assumed as energy loss
- Degradation may be applied as a fixed annual rate or estimated dynamically

Even with identical data, expected yield (and therefore KPIs) can vary substantially across platforms depending on the applied methodology and the adjustments considered. This reinforces that expected yield estimates are highly model- and adjustment-dependent and should be interpreted carefully when assessing plant performance, state of health, and relevant KPIs. Without standardized correction sequences or transparent documentation of model assumptions, cross-platform comparisons are unreliable.

9. KEY PERFORMANCE INDICATORS

All platforms report calculating key performance indicators (KPIs), though coverage and flexibility vary. While IEC 61724 is frequently referenced, implementation consistency is limited.

9.1. Technical KPIs

Commonly reported performance KPIs include:

- ~80% Performance Ratio (PR)
- ~80% specific yield (kWh/kWp)
- ~75% time-based availability
- ~60% energy performance index
- ~50% weather adjusted PR
- ~50% energy-based availability
- ~40% capacity factor

Additional or customized metrics were also reported such as:

- IEC 61724 based KPIs
- Reference yield and inverter-level indices
- DC capacity loss indicators
- Irradiance- or temperature-weighted parameters
- Custom KPI engines allowing user-defined metrics

9.2. Energy storage KPIs

Nearly half of respondents report calculating KPIs for energy storage systems. Commonly reported storage metrics include:

- State of Charge (SOC)
- State of Health (SOH)
- Round-Trip Efficiency (RTE)
- Depth of Discharge (DOD)
- Cycle count
- Capacity fade
- Availability
- Thermal parameters (cell/container temperature)
- DC voltage and current

Some platforms also report:

- Imbalance metrics
- Predictive fault indicators

- Digital twin—simulated SOC or voltage
- Forecasted savings or EMS-derived metrics

9.3. Financial KPIs

Over 70% of the participating software calculate financial KPIs. These typically include:

- Actual vs expected revenue
- Revenue loss
- Lost energy value calculated using the power purchase agreement (PPA) rate
- Capture price and capture ratio
- Contract breach indicators
- Return of investment (ROI), Internal Rate of Return (IRR), Net Present Value (NPV)

Some platforms support ingestion of market pricing or dynamic tariffs, while others assume flat PPA rates.

10. USE OF AI/ML AND DIAGNOSTICS

A significant majority of respondents (~85%) report using AI and/or ML within their platforms. However, the depth, scope, and technical rigor of these implementations vary widely.

10.1. Purpose of AI/ML

Respondents report AI/ML applications in the following domains:

- Anomaly detection (~60%)
- Fault classification (~60)
- Predictive maintenance (~55%)
- Performance prediction (~55%)
- Data quality and/or ingestion (~50%)
- Imagery analysis (~27%)

A smaller subset (~15–20%) report using:

- LLM-based interfaces for alert interpretation
- Contract analytics
- Chat-based queries
- Snow effects

10.2. Summary of AI/ML implementation examples

The software providers were asked to provide an example of AI/ML usage. The examples provided fall into five main categories:

1) Predictive (time-series based)

Regression models, LSTM, ANN/CNN architectures, or other unspecified models to predict:

- Plant yield
- Inverter and connector failures
- BESS faults based on SOC/SOH deviations
- Wind turbine faults based on thermal anomalies

2) Analysis of imagery

Software providers use image-based ML models (e.g., YOLO, custom pipelines) to identify:

- Module or other defects in image or video footage
- Anomalies in radiometric thermal maps
- Construction deviations (e.g., piling alignment)

3) Detection and classification

Some platforms apply detection and classification models:

- On device logs

- For data quality
- To detect deviations or underperformance from baseline (e.g., inverter shutdown, transformer overheating, etc.)
- To detect shading patterns.

These models attempt to identify fault and underperformance rather than only flag deviations.

4) LLM-based interfaces

A subset of platforms uses large language models to:

- Summarize O&M logs and field reports
- Compose site performance summaries
- Analyze contracts and extract key information
- Interact with analytics capabilities using voice or chat.

These applications are workflow-oriented rather than performance-model-driven.

5) Model infrastructure and deployment

One respondent describes a “model zoo” infrastructure allowing:

- Iterative ML model development
- Selective deployment by asset managers
- Modular integration of analytics tools

This reflects a platform-level ML architecture rather than a single-use case.

Technically, implementations range from traditional regression and gradient boosting to deep learning (LSTM, CNN) and LLM-based systems. The sophistication varies, but most use cases center on detecting deviations and predicting asset failures rather than autonomous optimization. Furthermore, prediction of failures is reported on wind and BESS with the only PV-related fault predictions on connectors and inverters.

10.3. Performance issues that can be detected and classified

The majority report detection of core performance losses. The following issues are detected by most respondents and are generally quantified in terms of energy loss:

- Soiling (~90%)
- Snow (~75%)
- Inverter issues (~80%)
- String issues (~80%)
- Tracker issues (~75%)
- Curtailment (~70%)
- Ground faults (~65%)
- Grid issues (~67%)

Degradation is detectable and quantified by ~70%, for the following:

- PID identification
- “Accelerated degradation” detection
- Module degradation

Not all platforms estimate degradation dynamically; some assume a constant rate. “Shading” is another broader category that is reported (~80%) and includes:

- Vegetation
- Row-to-row shading
- Fixed-object shading

Module-level defects are mainly reported by thermography providers and can include:

- Hotspots
- Cracks
- Delamination
- Bypass diode failures
- Junction box / lightning damage

However, quantification of power loss from module-level defects is less consistently demonstrated unless it is linked to string-level modeling.

Other issues that were mentioned are related to connectors (~40%). Notably, ~10-15% of respondents either do not classify root causes or provide alerts without automated attribution, indicating that detection does not necessarily imply performance loss allocation.

10.4. Link to workflows (e.g., work order generation or export to CMMS)

Approximately 80% of platforms link diagnostics to work order generation, either:

- Directly within the platform, or
- Via export/API to external CMMS systems.

11. TRANSPARENCY AND DOCUMENTATION

Transparency and documentation practices vary significantly across respondents. While many platforms describe advanced analytics capabilities, fewer provide detailed methodological documentation, model disclosures, or validation frameworks.

11.1. Documentation of platform's methods

Access to documentation of KPI calculations, yield modeling, analytics varies:

- Available to users (directly in platform or provided materials): ~45%
- Publicly available: ~17%
- Shared upon request (typically under NDA): ~25%
- Not shared (~10%)

Most platforms provide some level of documentation, but full public transparency is less common.

11.2. Alignment with industry standards, publications or best practices

The following standards, methodologies, and best practices are cited:

- IEC 61724 (-1, -2, -3) for monitoring and energy evaluation
- IEC 62446-3 for thermographic inspection
- IEEE 2030.5 for secure communication
- ASTM E2848 for capacity reporting
- Solar Power Europe reports on best practices
- NABCEP certifications
- NREL/Sandia methodologies for metrics and O&M best practices
- pvlib open-source library for PV performance modeling
- Guide on Uncertainty Measurements
- NASA uncertainty guidance
- ISO 17025 accreditation

Several also claim:

- Independent validation
- Peer review
- In-house R&D validation on test PV plants

Approximately 55% of software providers report using pvlib, which may reflect both its accessibility as a free, open-source tool and its transparent, peer-reviewed modeling framework. While adoption suggests some preference for validated and transparent models, it may also be influenced by practical considerations such as cost and ease of use.

11.3. Reproducibility of KPIs and expected yield

When asked whether documentation is sufficient to reproduce KPIs and expected yield (assuming access to raw and processed data):

- **Yes:** ~50%
- **Partial:** ~25%
- **No:** ~25%

Some limitations were mentioned:

- Proprietary data cleaning pipelines
- Incomplete documentation for some KPIs
- Depending on user's skill because of complex yield modeling

While most platforms reference industry standards (especially IEC 61724), documentation transparency is heterogeneous. Roughly half claim reproducibility of KPIs and expected yield, but a meaningful share either provide documentation only under NDA, acknowledge partial coverage, or rely on proprietary preprocessing steps that limit full methodological transparency. Alignment with standards is commonly asserted; explicit methodological detail sufficient for independent replication is less demonstrated.

12. BUSINESS MODEL AND SUPPORT

Platforms offer a range of pricing models, including subscription-based, per-MW, and custom enterprise agreements, often with optional onboarding or pay-per-use fees. Technical support, training, and documentation are typically included, with live sessions, tutorials, and knowledge bases provided to facilitate adoption and effective platform use.

12.1. Pricing models

Platforms use a variety of pricing approaches:

- Onboarding/setup fees: Commonly per MW or per site, with amounts ranging from a few hundred USD/EUR to \$30k+ for large-scale plants. Some platforms offer optional premium onboarding.
- Subscription fees:
 - Per MW/year: Most common (~75%); examples range from ~\$150–\$250/MW/yr for smaller plants, sometimes scaling with total fleet size.
 - Per month: Some platforms charge monthly subscription based on MW under management or per-user access.
 - Custom enterprise agreements: Many platforms provide bespoke pricing for large clients or fleets >1 GW.
 - Pay-per-use: Some platforms offer fees per report, inspection, or analysis.
- Example scenario (100 MW plant in California): Onboarding ~\$0k–\$30k; annual subscription ~\$24k–\$30k, depending on platform and contract type.

Most providers (~60%) do not publicly disclose pricing; a few platforms base costs on hardware included or on modular feature selection.

12.2. Technical support

Nearly all platforms include technical support in subscription pricing. Level of support varies:

- Basic support: included in standard subscription
- Premium support: optional, for additional cost (e.g., hands-on implementation, extended hours)

12.3. User training and documentation

Most platforms provide training and guidance, typically including:

- Built-in tutorials or guided onboarding
- Live training sessions (virtual, in-person, or hybrid)
- Knowledge base or documentation

Some platforms offer certification courses or deeper hands-on training for premium subscribers. Pricing is flexible and often modular, scaling with plant size, fleet size, and selected features. Support and training are widely offered, but most platforms did not disclose detailed costs, making pricing largely opaque and difficult to evaluate independently.

13. KEY TAKEAWAYS

Onboarding and data readiness drive scalability: Consistent SCADA mapping, metadata, and structured data ingestion are more critical than platform automation alone. Standardized data structures and harmonized naming conventions accelerate deployment.

Integration is feasible but fragmented: APIs, exports, and third-party connections exist, yet inconsistent or bilateral integrations limit automation and scalability. Adoption of open APIs, standard taxonomies, and clear integration protocols would enable seamless interoperability.

Data quality and expected yield modeling are core sources of variability: Differences and opacity in QC processes and yield modeling propagate through KPIs, performance loss attribution, and financial outputs. Harmonizing expected yield and energy accounting and generating best practices for KPIs would improve reproducibility and trust.

Digital twins provide richer insights but vary widely: Most are physics-based simulations; only a few combine real-time data, predictive or command/control elements. Clear definitions are needed to distinguish simulation, operational (push/receive) and predictive twins.

AI/ML is primarily for anomaly detection: Most platforms use AI/ML to detect deviations or predict failures on select components (e.g., connectors, inverters, BESS, wind), with limited predictive capability for broader PV operations. Linking AI/ML outputs to workflows is common. Autonomous optimization is rare. Standardized reporting of model scope, performance, and limitations could guide more effective use of AI/ML.

Transparency and documentation remain limited: Pricing, methodologies, and reproducibility are inconsistent. Improved disclosure, open documentation and independent validation would enable confidence and comparability.

REFERENCES

- [1] M. Theristis, K. Anderson, J. Ascencio-Vasquez, and J. S. Stein, "How Climate and Data Quality Impact Photovoltaic Performance Loss Rate Estimations," *Solar RRL*, vol. 8, no. 2, p. 2300815, 2024, doi: <https://doi.org/10.1002/solr.202300815>.
- [2] S. Lindig *et al.*, "Review of Technical Photovoltaic Key Performance Indicators and the Importance of Data Quality Routines," *Solar RRL*, vol. 8, no. 24, p. 2400634, 2024, doi: <https://doi.org/10.1002/solr.202400634>.
- [3] S. Lindig *et al.*, "Technical Key Performance Indicators for Photovoltaic Systems: Challenges and Best Practices," IEA-PVPS T13-28:2024, 2024.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	1911	sanddocs@sandia.gov

This page left blank



Sandia
National
Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.