

From Materials Discovery to Deployment: A Critical Review of Artificial Intelligence Methods Across the Sensor Development Pipeline

K Sudheendra Rao

Dayananda Sagar University

Sanjeev Kumar Shrivastava

`sanjeevkumar-ece@dsu.edu.in`

Dayananda Sagar University

Amritha Abdul Salam

Dayananda Sagar University

Sandhya N

Dayananda Sagar Academy of Technology and Management

B. S. Satyanarayana

Dayananda Sagar University

Research Article

Keywords: artificial intelligence, sensor development, physics-informed machine learning, Bayesian optimisation, digital twins, evidence-based review

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11017799/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

From Materials Discovery to Deployment: A Critical Review of Artificial Intelligence Methods Across the Sensor Development Pipeline

K. Sudheendra Rao¹, Sanjeev Kumar Shrivastava^{1*},
Amritha Abdul Salam^{2*}, Sandhya N³, B. S. Satyanarayana⁴

¹Department of Electronics and Communication Engineering, School of
Engineering, Dayananda Sagar University, Bangalore, 560 068,
Karnataka, India.

²Department of Artificial Intelligence and Machine Learning, School of
Engineering, Dayananda Sagar University, Bangalore, 560 068,
Karnataka, India.

³Department of Artificial Intelligence and Machine Learning,
Dayananda Sagar Academy of Technology and Management,
Udayapura, Bangalore, Karnataka, India.

⁴Vice Chancellor, Dayananda Sagar University, Bangalore, 560 068,
Karnataka, India.

*Corresponding author(s). E-mail(s): sanjeevkumar-ece@dsu.edu.in;
amrithaabdul-aiml@dsu.edu.in;

Contributing authors: sudheendra-ece@dsu.edu.in;
dr.sandhya1909@gmail.com; bs.satya@dsu.edu.in;

Abstract

Artificial intelligence is increasingly described not only as a means of improving sensor performance but as a force that compresses the sensor development lifecycle itself. This review examines that proposition critically across the full development pipeline, spanning materials discovery, device design, fabrication, calibration, and deployment. The evidence is organised around the specific AI techniques that operate at each stage. These include graph neural networks and generative models for materials discovery, physics-informed neural networks for design, Bayesian optimisation and self-driving laboratories for fabrication, transfer learning for calibration, and edge intelligence with digital twins for

deployment. To separate demonstrated capability from projection, a three-tier evidence framework classifies each contribution as commercially deployed, pilot-demonstrated, or conceptually projected. The analysis shows that stage-specific acceleration is well supported, most clearly in calibration and materials screening. In contrast, end-to-end pipeline compression from decade to multi-year timescales remains a projection drawn from the most data-rich domains rather than a validated general result. The review identifies where AI techniques transfer across sensor classes and where they do not, highlights the missing datasets and benchmarks that limit reproducibility, and sets out the conditions under which the acceleration thesis could be tested. The result is a maturity-calibrated map of AI in sensor engineering.

Keywords: artificial intelligence, sensor development, physics-informed machine learning, Bayesian optimisation, digital twins, evidence-based review

Corresponding authors. Sanjeev Kumar Shrivastava, sanjeevkumar-ece@dsu.edu.in, and Amritha Abdul Salam, amrithaabdul-aiml@dsu.edu.in, are the co-corresponding authors for this manuscript.

1 Introduction

Sensor engineering has entered a period in which artificial intelligence (AI) is present at every stage of development, not only in the finished device. The familiar contribution of AI to sensing is signal-level: classifiers, denoisers, and fusion models that improve what a deployed sensor can measure, an area now surveyed across flexible and intelligent sensing systems [31]. A second, more consequential claim has since emerged. It holds that AI reshapes how sensors are *built*, compressing a development cycle historically measured in years, a proposition increasingly discussed in recent surveys of AI-driven sensing [7]. This review takes that second claim as its subject.

The distinction matters because the two claims rest on very different evidence. Signal-level improvement is routinely demonstrated and widely deployed. Lifecycle compression is asserted more often than it is measured, and the strongest versions of it, in which end-to-end autonomous pipelines reduce development from fifteen years to two, extrapolate from a handful of unusually data-rich domains. Treating both claims as equally settled would misrepresent the field.

Existing reviews of AI in sensing set a useful baseline but leave a specific gap. Broad surveys catalogue where machine learning improves sensor performance and enumerate techniques across application areas [7, 31], yet they largely organise the literature by sensor type or algorithm and report demonstrations without a consistent standard for how mature each demonstration actually is. What is missing is not more coverage but calibration: a way to separate a capability that ships in products from one shown once in a laboratory, applied uniformly across the development lifecycle. The present review is positioned to fill that gap. It organises the evidence by development stage rather than by sensor type, and it applies a single explicit maturity standard to

every claim, so that the reader can see not only what AI can do in sensor development but how well each of those things is actually established.

We therefore organise this review around the sensor development pipeline viewed as a multi-stage intelligent system. Each stage, from materials discovery through deployment, is examined for the AI techniques it now employs, the strength of the evidence behind them, and whether those techniques transfer beyond the setting in which they were first shown. The pipeline framing is deliberate. It converts a scattered literature, in which individual methods are reported in isolation, into a system-level account in which the interactions and dependencies between stages become visible.

The review makes three contributions. First, it provides a technique-by-stage synthesis of AI in sensor development, rather than a catalogue of sensor types, offering a critical evaluation of each technique’s demonstrated capability and limits. Second, it introduces an explicit three-tier evidence framework that classifies every acceleration claim by maturity, so that demonstrated capability is not conflated with projection. Third, it offers a transferability analysis that identifies where methods generalise across sensor classes and where they remain domain-bound. Together these give a maturity-calibrated map intended to inform both research priorities and realistic expectations of AI in sensor engineering.

The remainder of the paper is organised as follows. Section 2 describes the review methodology, including the search strategy and the criteria used to assign evidence tiers. Section 3 sets out the development pipeline and the acceleration thesis. Section 4 examines AI techniques stage by stage. Section 5 synthesises the evidence across stages and analyses transferability. Section 6 states the limitations of the present evidence base. Section 7 identifies open questions, and Section 8 concludes.

2 Review methodology

This review is a structured, critical synthesis rather than a quantitative meta-analysis. The aim is to map AI techniques onto sensor development stages and to assess the maturity of the evidence for each, so the method is designed for transparency and reproducibility of the selection process, not for statistical pooling of effect sizes.

2.1 Search strategy

The literature was identified through structured searches of Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library, supplemented by forward and backward citation tracing of key sources. Search strings combined a development-stage term with an AI-method term and a sensing term. Representative combinations included “materials discovery” with “graph neural network” or “generative model”; “fabrication” with “Bayesian optimisation” or “self-driving laboratory”; “calibration” with “transfer learning” or “domain adaptation”; and “deployment” with “edge AI” or “digital twin”. The primary literature search was conducted through early 2026; subsequent 2026 publications identified during reference verification were included where directly relevant. Foundational earlier work was retained where it remained the primary reference for a technique.

2.2 Inclusion and exclusion criteria

Sources were included if they reported an AI technique applied to an identifiable stage of sensor or comparable microdevice development, and if they provided either an experimental result or a clearly specified architecture. Sources were excluded if AI was used only for post-deployment signal interpretation without bearing on development, if no method detail was given, or if claims were made without supporting evidence or citation. Where several sources reported the same technique, the one with the strongest experimental grounding was cited as primary.

2.3 Stage assignment and evidence tiers

Each retained technique was assigned to the pipeline stage it primarily addresses. Techniques spanning stages, such as digital twins, are discussed at the stage where their effect is largest and cross-referenced elsewhere. Every acceleration claim was then assigned one of three evidence tiers, defined in Section 3.2. Tier assignment was based on the highest level of validation reported in the primary source, with ambiguous cases assigned to the lower tier. To illustrate the rule, transfer-learning gas-sensor calibration is placed at the pilot tier rather than commercially deployed, because its strongest evidence is a peer-reviewed experimental result rather than a product at scale, whereas convolutional-network defect inspection is placed at the deployed tier because it is in routine industrial use. This conservative rule ensures that the review does not overstate maturity.

2.4 Scope and interpretive stance

The pipeline model and the evidence-tier framework are offered as a heuristic lens for organising and appraising the literature, not as an empirical claim about a fixed or universal development process. Development practices vary across sensor classes and organisations, and the framework is intended to make that variation legible rather than to flatten it.

AI-based tools were used to assist literature identification and initial organisation during this review. Key primary sources underlying quantitative results or stage-classification claims were retrieved and checked against the original publications by the authors, who take full responsibility for the selection, interpretation, and accuracy of all material reported here.

3 The sensor development pipeline and the acceleration thesis

The preceding framing separates two claims about AI in sensing. The first, that AI provides useful tools at specific pipeline stages, is well supported experimentally. The second, that AI compresses the full pipeline from decade to multi-year timescales, is a projection drawn from the most AI-mature domains and should not be generalised without qualification.

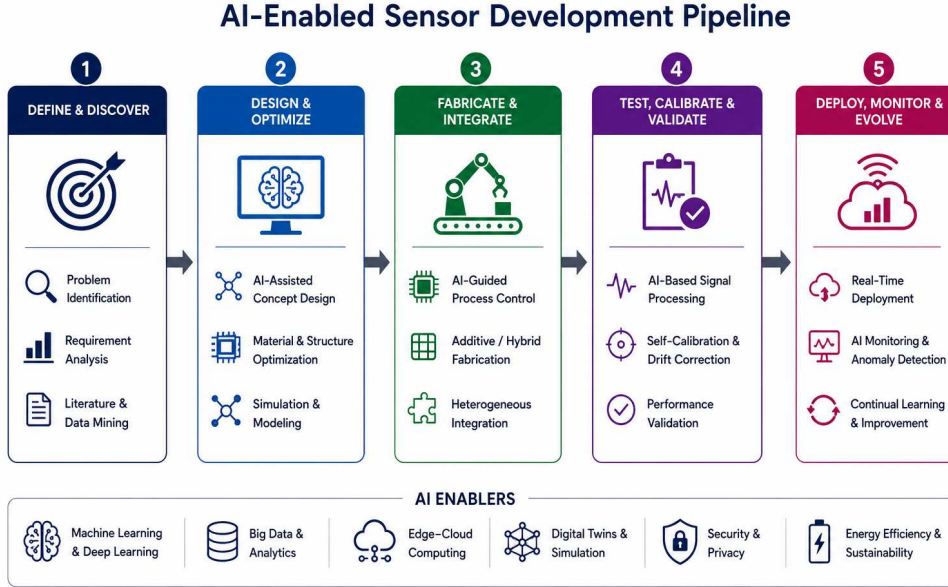


Fig. 1 The AI-enabled sensor development pipeline across five stages (define and discover, design and optimise, fabricate and integrate, test and validate, deploy and evolve), supported by cross-cutting AI enablers. The figure is conceptual.

3.1 Reframing the development pipeline

The canonical development pipeline spans five stages: define and discover, design and optimise, fabricate and integrate, test and validate, and deploy and evolve, each supported by cross-cutting AI enablers, as shown in Fig. 1. Conventional execution is lengthy: the materials development cycle from discovery to deployment has been estimated at ten to twenty years, and the Materials Genome Initiative was framed explicitly around the observation that transitioning a new material into use frequently takes twenty years or more [23, 35]. The initiative’s stated goal was to halve that time and cost, an aspirational target rather than a measured outcome [23]. Figure 2 contrasts the traditional and AI-accelerated pipelines stage by stage, illustrating where learning methods are claimed to shorten individual stages. The figure is a conceptual illustration of potential acceleration, not an empirically established timeline; no sensor development programme has yet demonstrated end-to-end compression of this kind, so specific compressed-duration figures are deliberately not asserted here.

3.2 Evidence-tier classification

Every acceleration claim in this review is classified according to a three-tier evidence framework. *Commercially deployed* denotes capabilities available in products at scale as of the time of writing; where a technique is routine in a related industry but its use specifically within sensor development is not separately documented, this is

stated rather than implied. *Pilot-demonstrated* denotes peer-reviewed experimental results or pre-production trials that have not yet reached industrial scale. *Conceptually projected* denotes architectures with proof-of-concept results whose full industrial implementation has not been independently validated. Table 1 applies this classification to each technique discussed. Several acceleration claims in the literature represent only pilot demonstrations or architectural projections; end-to-end autonomous development pipelines that would compress multi-year timelines into a fraction of their current span are not yet a validated industrial reality.

Key implication. The value of the pipeline account lies less in any single timeline figure than in making explicit, stage by stage, which claims rest on deployed evidence and which on projection.

4 AI techniques across the development pipeline

4.1 Materials discovery and generative design

Materials discovery is where the acceleration thesis makes its boldest claim: that data-driven screening can compress candidate evaluation from years of synthesis to days of computation. The evidence here is strongest at the level of property prediction. Graph neural networks (GNNs) represent a crystal as a graph of atoms and bonds and learn to predict functional properties directly from structure. The crystal graph convolutional network of Xie and Grossman established that this representation predicts density-functional-theory properties across diverse compositions while remaining interpretable [36]. Later architectures extended the approach into a broad family now reviewed as a mature subfield [27]. The scale now achievable is substantial. Merchant et al. trained graph networks on large stability datasets and reported an order-of-magnitude improvement in discovery efficiency, identifying about 2.2 million structures stable with respect to previous work, including 381,000 entries on the updated convex hull, of which 736 had been independently experimentally realised [21]. The novelty of some of these structures has since been questioned, a reminder that large-scale generation still requires expert validation before a candidate counts as a genuine discovery. For sensor work, this matters because transducer performance often depends on precisely the properties, band gaps, formation energies, and piezoelectric or dielectric responses, that these models predict.

Generative models push from prediction toward design. Variational autoencoders and related generators propose targeted candidate structures rather than screening fixed libraries, and inverse-design formulations select for a desired property directly [3]. Physics-informed neural networks (PINNs) constrain such predictions to physically admissible solutions by embedding governing physical equations into the learning problem [26]. In principle this approach applies to the coupled physical fields that sensors exploit, from electromagnetic transduction to microfluidic transport and chemical reaction-diffusion, though sensor-specific PINN demonstrations remain comparatively sparse. Physics-informed surrogates have been used to reduce the number of high-fidelity simulation calls needed for MEMS geometry optimisation relative to design-of-experiments [24]. Inverse design has also been applied directly at the

device level: Liu et al. demonstrated a property-to-structure method for flexible pressure sensors that maps a target response onto physical geometric parameters using a reduced-order model and a surrogate-assisted iterative search, predicting many candidate structures that were then experimentally validated across several material systems [17]. This is a sensor-specific case where AI outputs a physical design that is built and tested, rather than a material property alone.

The gap between this activity and validated sensor outcomes remains the central caveat. Most demonstrations report property prediction or candidate generation, not fabricated sensors with measured performance that matches the prediction. It is worth distinguishing several senses of discovery that are easily conflated: identifying a new material or composition, identifying a previously unreported material-property relationship, and optimising an existing material or device. Most generative work addresses the first; the second is rarer and more directly relevant to sensing. A pertinent example is the work of Gao et al., who combined density-functional theory with machine learning to predict the gas response of a material for which no such response had been reported, then confirmed that predicted response experimentally [10]. This is best described not as the discovery of a new material but as the AI-assisted prediction of a previously unreported functional response, subsequently validated, which is a stronger and more specific claim than property prediction alone. The evidence tier for generative inverse design is therefore pilot, and it varies by domain, with the strongest results in general materials informatics rather than in delivered devices.

Key implication. Generative design is real and, for property prediction, well validated, but its translation into sensor-specific materials with confirmed device performance is still emerging rather than established.

4.2 Fabrication optimisation

At the fabrication stage the question shifts from what to make to how to make it reliably, and here the evidence is more concrete because the optimisation targets are measurable yields and process parameters. Bayesian optimisation (BO) treats fabrication as a search over a high-dimensional process space too costly to characterise exhaustively, using a probabilistic surrogate to choose each next experiment. Applied to MEMS manufacturing, BO and related AI methods have delivered domain-specific yield and quality gains [24], although the magnitudes vary and do not transfer across platforms.

The most striking fabrication-stage results come from self-driving laboratories that close the loop between synthesis, characterisation, and model-guided decision-making. Because autonomous experimentation has matured faster in general materials chemistry than in sensor fabrication, several of the foundational demonstrations below are drawn from adjacent materials domains rather than from sensors directly; they establish the method, while sensor-specific closed-loop evidence remains comparatively scarce. Burger et al. demonstrated a mobile robot that ran autonomously for eight days, performing 688 experiments in a ten-variable space and identifying photocatalyst mixtures six times more active than the starting formulations [6]. MacLeod et al. reported a self-driving platform that optimised the optical and electronic properties of thin films, a material class directly relevant to optical and flexible sensors

[19]. Langner et al. pushed throughput further, fabricating thousands of films per day and mapping a multi-component parameter space with milligram-scale material use [15]. Szymanski et al. extended the paradigm to solid-state synthesis, originally reporting 41 compounds from 58 targets over 17 days of continuous operation [32]. Following publication, the novelty and diffraction-based identification of the products were contested, and a peer-reviewed author correction subsequently clarified that the materials were new to the prediction platform rather than necessarily new to science, and that re-analysis confirmed 36 of 40 reported successes, with four inconclusive [33]. The episode illustrates that autonomy at the bench still depends on careful human verification of what was actually made.

These are genuine, peer-reviewed pilot demonstrations, not projections. Their limitation is scope. Each operates within a defined chemistry and a well-instrumented process, and none yet constitutes a universal multi-stage pipeline spanning all sensor technologies. That universal pipeline, distinct from these bounded successes, remains conceptual and unvalidated at industrial scale.

Closer to the sensor itself, the same closed-loop paradigm has been applied directly to sensor formulation. Ai et al. demonstrated a knowledge-aware, Bayesian-optimisation-driven robotic workflow for iterative colorimetric gas-sensor development, integrating algorithmic candidate selection with automated formulation and testing across a formulation space of up to nineteen variables [1]. The reported platform reached an ammonia detection limit of about 50 ppb over a dynamic range of 0.5 to 500 ppm within roughly one week of automated closed-loop optimisation. Unlike the materials-chemistry systems above, this is a sensor-specific automated design-build-test-learn demonstration, although initial reagent preparation and experimental setup still required human intervention, and it indicates that robotic closed-loop optimisation is beginning to move from adjacent domains into sensor fabrication itself, though still at single-laboratory scale.

Ai et al. represents the most automated end of a spectrum of sensor-specific closed-loop work, and it is useful to distinguish the degrees of physical coupling the other examples show. Yang et al. applied active learning to strain-sensor design, training a classifier on 351 compositions and then fabricating 125 physical sensors across twelve active-learning loops, with experimental results selecting each successive round and data augmentation extending a scarce dataset to more than ten thousand virtual points [38]. Here the loop is real but human-executed rather than robotic. Sugiura et al. closed the loop physically in a different way, using Bayesian optimisation with visual feedback to 3D-print a conductive-hydrogel soft sensor, evaluate it, and iteratively refine its morphology, demonstrated on a sensorised glove [30]; this is physical-in-the-loop morphology optimisation rather than fully autonomous fabrication, since motion and testing remain partly manual. Huang et al. addressed formulation rather than geometry, combining machine learning, molecular fingerprints, and Bayesian optimisation over 1745 literature data points to prospectively design ion-selective-electrode membranes, then fabricating Na^+ , Mg^{2+} and Al^{3+} sensors whose measured responses fell within 8.2% of the ideal Nernstian slope [12]. Read together, these establish that sensor-specific closed-loop development now spans human-executed active learning,

physical-in-the-loop morphology tuning, formulation design, and robotic design-build-test-learn, but that the degree of automation, and therefore of demonstrated maturity, differs sharply between them.

Alongside these research systems, one fabrication-stage use of AI is already commercially routine: convolutional-network visual inspection for defect detection on wafers and printed substrates, which is the most mature vision-AI role in semiconductor and sensor manufacturing [24].

Key implication. Fabrication AI is closest to deployment where the process space is narrow and well instrumented, and the self-driving laboratory is its most convincing embodiment, but generality across processes is demonstrated nowhere yet.

4.3 Calibration and deployment intelligence

Transfer learning for cross-device calibration is among the more experimentally mature AI contributions to sensor deployment. Neural calibration models trained on a reference device and adapted to a target device via small-calibration-set fine-tuning have demonstrated up to 93% reduction in per-device calibration data and time in gas-sensor applications, at the cost of increased prediction error relative to full calibration, with cross-device error within twice the RMSE of full-calibration performance [28]. A complementary line of work reframes calibration as a quasi-parallelised process, using tree-based regression to identify each sensor’s working point dynamically and reporting a 23.8% reduction in calibration duration on MEMS inertial sensors [25].

The harder version of this problem is drift, where a sensor’s response changes over time or differs between nominally identical units. Domain adaptation addresses it by aligning the feature distributions of source and target devices. A recurring practical obstacle is that classical domain adaptation needs data from the drifted target distribution, which is often unavailable before deployment. Recent work removes that requirement: a target-domain-free convolutional network learns domain-invariant features from the source alone and compensates drift in gas sensors without target data during training [41]. Related electronic-nose methods use a small set of shared transfer samples to realign drifted responses at both feature and decision levels [39], and optimal-transport formulations adapt calibration for mass-produced soft sensors and for long-term signal drift in a single device [14]. Deployment-stage anomaly detection using autoencoders and statistical process control complements these methods, monitoring drift and flagging calibration degradation without manual reference checks [22]. Digital-twin models, which are physics-augmented virtual representations of sensors, extend the same idea toward predictive maintenance and adaptive recalibration and are examined in Section 4.6.

The common limitation across these methods is that each is validated on a specific sensor class and drift regime. Reported accuracies are strong, but a technique proven for electronic-nose drift does not transfer unchanged to inertial or optical sensors, so the maturity is real but bounded.

Key implication. Calibration is where AI acceleration is most firmly established, because the task is well posed and reference data are available, yet even here the strongest results remain specific to the sensor class on which they were shown.

4.4 LLM-assisted design workflows

Large language models (LLMs) applied to design represent the newest and least settled part of the pipeline. The strongest evidence comes from adjacent fields rather than sensor design itself. In chemistry, Coscientist, a GPT-4-driven agent, autonomously planned and optimised reactions using tools for literature search, code execution, and experimental control [4], and ChemCrow augmented an LLM with expert-designed tools to plan syntheses of target compounds [5]. These show that LLM agents can convert natural-language goals into executable experimental plans, a capability directly relevant to sensor design workflows.

The honest counterweight is that rigorous benchmarking finds current agents far from reliable. A central failure mode is hallucination: large language models can generate fluent but factually wrong outputs, proposing synthesis steps, material parameters, or design rationales that are plausible in form yet incorrect in substance. This is particularly consequential in a design setting, where an unverified hallucinated parameter can propagate into a fabricated device before it is caught, and where the models are poor judges of their own errors: in the ChemCrow evaluation, a GPT-4 assessor could not reliably distinguish correct chemical reasoning from confident mistakes [5]. On ScienceAgentBench, a suite of 102 real tasks drawn from peer-reviewed studies, the best baseline agent solved 32.4% of tasks independently, while a reasoning model reached 42.2% under substantially higher inference cost [8]. A similar pattern appears by analogy in other domains where LLM agents have been studied more extensively, such as software engineering, namely real capability on scoped subtasks alongside limited reliability on open-ended problems [18]. For sensor design specifically, direct evidence remains sparse and proof-of-concept, and no cross-device quantitative benchmark yet exists, nor any reported hallucination-rate characterisation for sensor-design tasks. The evidence tier is therefore conceptual for device-design generalisation, and at most pilot for narrow, well-bounded tasks.

Key implication. LLM design assistance is a credible direction with demonstrated agent capability in neighbouring fields, but for sensor design it is unmeasured; hallucination and the absence of reliable self-evaluation mean benchmark results elsewhere caution against reading demonstrations as reliability.

4.5 Edge AI and continual adaptation in deployment

Deployment is where AI must run on the sensor itself, under power, memory, and latency limits that cloud models never face. The enabling advance is the compression of deep models to microcontroller scale. Lin et al. showed with MCUNet that co-designing the network and its inference engine allows ImageNet-scale accuracy on a commercial microcontroller using several times less memory than quantised mobile networks [16]. This tiny-machine-learning approach, now surveyed as a fast-moving field spanning quantisation, pruning, and architecture search [29], is what makes on-sensor inference a commercially deployed capability rather than a laboratory curiosity.

A more radical line moves computation into the sensor itself. In-sensor and near-sensor computing place processing at or beside the transducer, cutting the redundant data movement that dominates power and latency in conventional pipelines [43]. A

landmark demonstration is the reconfigurable two-dimensional-material image sensor of Mennel et al., which senses and classifies simultaneously at nanosecond timescales, processing on the order of tens of millions of samples per second without a separate processor [20]. Such results remain small-array laboratory devices, but they define the direction of travel, and the emerging notion of the sensor as an integrated sense-and-compute unit is now framed as a design paradigm in its own right [34].

Keeping that on-device intelligence accurate over time is harder. Industrial environments impose demands unforeseeable at design time: shifting temperatures, varying chemical backgrounds, changing vibration spectra, and evolving electromagnetic interference. A model trained on a static distribution degrades under these conditions in ways that are hard to detect without monitoring. Continual-learning architectures that absorb new data without catastrophic forgetting are under active investigation, and transfer learning, domain adaptation, and federated learning across sensor fleets [37] each partially address distributional shift. None is universally effective, and the choice remains application-specific. The harder constraint is systemic: on-device adaptation needs update and monitoring infrastructure that most fielded sensor systems do not yet have.

Key implication. Edge inference is now a deployed capability, and in-sensor computing points further ahead, but robust deployment intelligence over time depends less on any single algorithm than on the monitoring and update infrastructure that most fielded systems still lack.

4.6 Digital twins as a potential connective architecture

The architecture most often proposed to give coherence to AI acceleration across all pipeline stages is the physics-based digital twin. Digital twins vary substantially in scope and fidelity; the implementations most relevant here are continuously updated computational models that ingest measurement data, maintain a state estimate of the physical device, and generate predictions of future behaviour under varied operating conditions [9, 11]. In sensor development, digital twins of this kind can enable virtual commissioning before deployment, provide physics-grounded constraints that prevent unphysical predictions, support predictive maintenance by flagging incipient failure, and accelerate calibration through physics-derived parameter estimates.

The most mature evidence sits in industrial predictive maintenance, from which the sensor-domain case borrows. Physics-based twins have been used to estimate remaining useful life by tuning a simulation model to field data and propagating degradation [2], and systematic reviews now catalogue predictive-maintenance twins across manufacturing, power, aerospace, and related sectors [42]. Within sensing, the piezoelectric-vibration-sensor twin of Fu et al. applies the same degradation-modelling logic directly to a sensor’s remaining useful life, reducing the remaining-useful-life prediction error to about eight hours [9], and gas-turbine pilots that fuse a mechanism model with measurement data report low modelling error after iterative calibration [11], though the reported error metrics are not directly comparable across these studies.

Two honest qualifications follow. First, this account synthesises three literatures, sensor-specific twins, calibration modelling, and industrial asset twins, rather than

Table 1 AI techniques across sensor development stages with evidence classification. Tiers are assigned conservatively to the highest validation level reported in the primary source.

Stage	AI technique	Representative method	Demonstrated outcome	Evidence level	Key limitation
Materials discovery	Generative inverse design	GNN + VAE on property databases	Property-targeted generation; no sensor-material evidence yet	Pilot (lab)	Training-data bias; feasibility unvalidated
Materials discovery	Self-driving laboratory	Robotics + Bayesian loop	Autonomous synthesis–test loops	Pilot	Sub-industrial; wet-chemistry only
Device design	Physics-informed NNs	Embedded PDE in DNN	Fewer FEA calls for MEMS geometry	Pilot (lab)	Training instability: physics vs. data
Device design	LLM design agents	LLM + simulation API	Shorter early-stage iteration	Conceptual / PoC	No cross-class benchmark
Fabrication	Bayesian optimisation	GP surrogate on fab parameters	MEMS yield gains (domain-dependent)	Pilot	Degrades at boundaries; known spaces only
Fabrication	CNN quality inspection	Defect classification on wafers	Full-throughput inspection at volume	Deployed (semiconductor mfg.)	Rare defects need costly labels; sensor-specific use not separately documented
Calibration	Transfer learning	Domain adaptation	Up to 93% time cut; < 2× RMSE (gas)	Pilot	Degrades at high variance; needs fleet
Calibration	Federated learning	On-device update; no raw data	Continuous fleet calibration	Pilot / Conceptual	Comms overhead; adversarial clients
Deployment	Edge AI inference	Quantised NNs on MCU/NPU	Low-latency, low-power inference	Deployed (general edge)	On-chip memory; thermal limits
Deployment	Digital twin	Physics–AI state estimation	Low reported error post-calibration; downtime cuts	Pilot (industrial)	Needs continuous data; high setup

resting on a single sensor-domain result. Sectoral reviews indicate AI predictive maintenance can cut downtime, but the magnitude varies and lacks a sensor-domain benchmark, so it is treated qualitatively. Second, the main technical constraint is fidelity degradation when a sensor drifts outside its build conditions, which requires continuous field-data updating that is compute-intensive at scale.

Key implication. The digital twin is a potential connective architecture for the pipeline rather than a demonstrated one; its sensor-domain evidence is largely borrowed from industrial asset management and still awaits a dedicated benchmark.

5 Cross-stage synthesis and transferability

Read across stages, the evidence supports a nuanced assessment. Operationally mature AI capabilities are evident in industrial visual inspection and embedded edge inference, whereas sensor-specific transfer-learning calibration remains experimentally demonstrated at pilot scale. Pilot-demonstrated contributions include Bayesian process optimisation, self-driving-laboratory workflows in electrochemistry and photocatalysis, and digital-twin predictive maintenance. Conceptually projected claims include

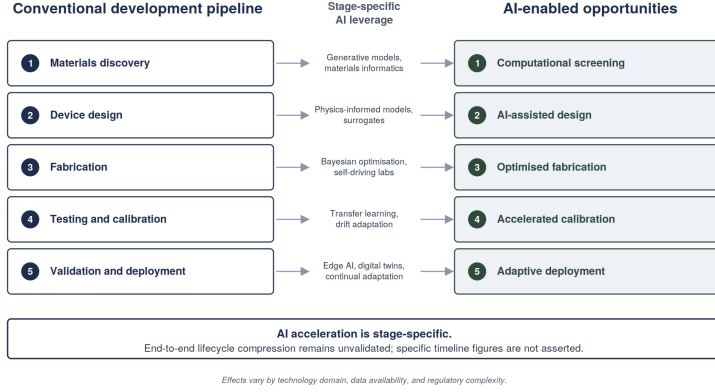


Fig. 2 Stage-specific opportunities for machine learning across the sensor development pipeline. Each conventional stage (left) is paired with the AI methods most associated with it (centre) and the resulting AI-enabled opportunity (right). The figure is conceptual: it locates where acceleration is claimed, not how much. AI acceleration is stage-specific, and end-to-end lifecycle compression remains unvalidated, so specific timeline figures are deliberately not asserted. Effects vary by technology domain, data availability, and regulatory complexity.

full-pipeline autonomous development from materials to regulatory submission, LLM-guided device design at sensor-class generality, and the full end-to-end compression projections.

Viewed as a system, the pipeline shows a clear pattern: the stages with the best-posed tasks and richest reference data, calibration and parts of fabrication, are where AI is most mature, while the stages that depend on open-ended generation or cross-domain transfer remain projections. The dependencies between stages, not any single stage, are what a systems-level reading makes visible.

Transferability is uneven and is the central practical finding. Techniques succeed where the task is well posed and reference data are available, as in calibration, and where the search space is narrow and instrumented, as in specific fabrication processes. They transfer poorly where the physics differs sharply between sensor classes, where labelled data are scarce, or where deployment conditions diverge from training distributions. A method demonstrated for gas-sensor calibration does not carry over unchanged to a quantum magnetometer, and a Bayesian-optimisation gain on one MEMS process does not predict the gain on another. The maturity map in Table 1 should therefore be read stage by stage and domain by domain, not as a single verdict on “AI in sensing”.

Key implication. The useful question is not whether AI accelerates sensor development, but which technique, at which stage, for which sensor class, and with what evidence.

A further distinction cuts across every stage: predicted performance is not the same as manufacturable performance. Jiang et al. made this explicit for gas-sensing materials, showing that candidates ranked highly for sensitivity alone need not be reproducible across nominally identical devices, and building a risk-constrained model that screens for reproducibility alongside sensitivity, learned over roughly

Table 2 Synthesised quantitative evidence for AI across the sensor development pipeline. Figures are as reported in the cited primary sources; tiers follow the classification in Section 3.2. Entries marked with a dagger (†) report results whose novelty or analysis has been publicly contested, and are included with that caveat.

Stage	Method (source)	Reported result	quantitative	Tier	Principal caveat
Materials discovery	Crystal graph CNN [36]	Formation-energy MAE ≈ 0.039 eV/atom, near-DFT		Pilot	Property prediction, not device
Materials discovery	Large-scale GNN [21]†	Order-of-magnitude gain; $\sim 2.2\text{M}$ structures identified, 381,000 on updated convex hull		Pilot	Novelty of structures disputed
Materials discovery	Generative model [40]	Generated structures $> 2\times$ more likely new and stable		Pilot	General materials, not sensors
Device design	Physics-informed NN [26]	PDE-constrained solution from sparse data		Pilot	Training stability
Fabrication	Self-driving laboratory [6]	688 experiments in 8 days; $6\times$ activity gain		Pilot	Single chemistry
Fabrication	Autonomous synthesis [32, 33]†	36 of 40 successes confirmed post-correction		Pilot	Novelty reframed; identification contested
Calibration	Transfer learning [28]	Up to 93% calibration-time cut; $< 2\times$ RMSE		Pilot	Gas sensors; degrades at high variance
Calibration	Target-domain-free adaptation [41]	Drift compensation without target data		Pilot	Cross-batch accuracy moderate
Deployment	TinyML on MCU [16]	$> 70\%$ ImageNet top-1 on a commercial MCU; $3.5\times$ less SRAM, $5.7\times$ less Flash		Deployed	On-chip memory limits
Deployment	In-sensor vision [20]	Classification/encoding at ~ 20 million bins/s		Pilot	Small array; lab demonstration
Deployment	Digital-twin RUL [9]	RUL prediction error reduced to 8 h		Pilot	Borrowed from asset management; metrics not cross-comparable
Design agents	LLM benchmark [8]	Best agent solves 32–42% of tasks		Conceptual	Far from reliable

nine thousand measurements on paired metal-loaded tin-oxide MEMS sensors [13]. The reproducibility-risk model reached an area under the curve of 0.911, and the reproducibility-aware ranking kept device-to-device variation typically below twenty per cent. This reframes a large part of the acceleration question: a sensor candidate is valuable in practice only if AI-predicted sensitivity survives fabrication variability, so performance-reproducibility co-optimisation, not sensitivity alone, is the more demanding and more useful target.

Table 2 consolidates representative quantitative results across the pipeline, with each entry’s maturity tier and principal caveat, so that the headline numbers can be read together with the conditions under which they hold.

6 Limitations of the evidence base

Several limitations qualify the conclusions above. First, much of the strongest quantitative evidence rests on single-source demonstrations that have not been independently replicated, so reported magnitudes should be treated as indicative. Second, the field lacks shared datasets, benchmarks, and reporting standards for AI-in-sensing, which

prevents like-for-like comparison across studies and inflates the apparent generality of individual results. This scarcity is not incidental but structural, and it is arguably the single most important constraint on maturity across the pipeline. Modern AI methods are data-intensive, whereas sensor fabrication and manufacturing are characteristically data-poor: process data are proprietary, batches are small and costly to produce, and failed runs are rarely published, so the large, labelled, openly shared corpora that drive progress in image or language tasks have no equivalent in most sensor-development settings. Much of the field’s methodological effort, transfer learning to reuse data across devices, active learning to spend experiments sparingly, and data augmentation to synthesise training points [28, 38], is in effect a response to this shortage rather than a demonstration of abundance. The asymmetry helps explain why data-rich stages such as calibration reach deployment while data-poor stages such as end-to-end fabrication remain at pilot scale, and it implies that closing the data gap may accelerate the field more than any new algorithm. Third, several cross-stage claims, particularly full-pipeline compression, synthesise evidence from adjacent fields such as general materials informatics and industrial asset management, and their transfer to sensor development is assumed rather than demonstrated. Fourth, this review is a structured critical synthesis, not a quantitative meta-analysis, so it does not pool effect sizes and cannot rule out selection effects in the primary literature. Finally, the pipeline model is a heuristic lens; real development processes vary and do not always follow its stages in order.

7 Open questions and research priorities

The analysis points to a research agenda defined less by new algorithms than by the infrastructure needed to test existing ones. Priorities include shared sensor-development datasets and public benchmarks that would allow AI acceleration claims to be compared and replicated; sensor-domain digital-twin evaluations with agreed error and downtime metrics; systematic study of when calibration and fabrication models transfer across sensor classes and when they fail; and regulatory frameworks capable of evaluating AI-generated designs, without which pipeline compression cannot reach regulated sensor markets. Progress on these fronts would convert projected acceleration into measurable acceleration.

8 Conclusion

Artificial intelligence now operates at every stage of the sensor development pipeline, but the strength of the evidence varies sharply across stages. Industrial visual inspection and embedded edge inference are operationally mature, while sensor-specific calibration, materials discovery, self-driving laboratories, and digital-twin maintenance rest on credible pilot evidence; and full-pipeline compression remains a projection from the most data-rich domains. The contribution of this review is not a single headline figure but a maturity-calibrated map that keeps demonstrated capability distinct from projection and identifies where techniques transfer and where they do not. If the field is to convert a persuasive thesis into an established result, its central task is

not another algorithm but the shared datasets, benchmarks, and regulatory pathways that would let the acceleration claim finally be measured.

Declarations

Competing interests. The authors declare no competing interests.

Funding. No external funding was received for this work.

Author contributions. Author contribution details are provided on the separate title page and are omitted here for double-blind review.

Data availability. No new experimental dataset was generated. The review used published literature as its evidence base, and all sources analysed are cited in the reference list.

Use of AI tools. The authors used AI-based tools to assist with literature identification and manuscript preparation. Key primary sources underlying quantitative or stage-classification claims were retrieved and checked against the original publications by the authors, who take full responsibility for the content and accuracy of the manuscript. Readers should consult primary sources for critical details.

Acknowledgements. [Anonymised for double-blind review.]

References

- [1] Ai Z, Zhang L, Chen Y, Long Y, Li B, Dong Q, Wang Y, Jiang J (2024) On-demand optimization of colorimetric gas sensors using a knowledge-aware algorithm-driven robotic experimental platform. *ACS Sens* 9(2):745–752. <https://doi.org/10.1021/acssensors.3c02043>
- [2] Aivaliotis P, Georgoulas K, Chrysosouris G (2019) The use of digital twin for predictive maintenance in manufacturing. *Int J Comput Integr Manuf* 32(11):1067–1080. <https://doi.org/10.1080/0951192X.2019.1686173>
- [3] Ballard Z, Brown C, Madni AM, Ozcan A (2021) Machine learning and computation-enabled intelligent sensor design. *Nat Mach Intell* 3:556–565. <https://doi.org/10.1038/s42256-021-00360-9>
- [4] Boiko DA, MacKnight R, Kline B, Gomes G (2023) Autonomous chemical research with large language models. *Nature* 624:570–578. <https://doi.org/10.1038/s41586-023-06792-0>
- [5] Bran AM, Cox S, Schilter O, Baldassari C, White AD, Schwaller P (2024) Augmenting large language models with chemistry tools. *Nat Mach Intell* 6:525–535. <https://doi.org/10.1038/s42256-024-00832-8>
- [6] Burger B, Maffettone PM, Gusev VV et al (2020) A mobile robotic chemist. *Nature* 583:237–241. <https://doi.org/10.1038/s41586-020-2442-2>

- [7] Chen L, Xia C, Zhao Z, Fu H, Chen Y (2024) AI-driven sensing technology: review. *Sensors* 24(10):2958. <https://doi.org/10.3390/s24102958>
- [8] Chen Z, Chen S, Ning Y et al (2024) ScienceAgentBench: toward rigorous assessment of language agents for data-driven scientific discovery. *arXiv:2410.05080*. <https://doi.org/10.48550/arXiv.2410.05080>
- [9] Fu C, Gao C, Zhang W (2023) A digital-twin framework for predicting the remaining useful life of piezoelectric vibration sensors with sensitivity degradation modeling. *Sensors* 23(19):8173. <https://doi.org/10.3390/s23198173>
- [10] Gao S, Cheng Y, Chen L, Huang S (2024) Rapid discovery of gas response in materials via density functional theory and machine learning. *Energy Environ Mater* 8(1):e12816. <https://doi.org/10.1002/eem2.12816>
- [11] Hu M, He Y, Lin X, Lu Z, Jiang Z, Ma B (2023) Digital twin model of gas turbine and its application in warning of performance fault. *Chin J Aeronaut* 36(3):449–470. <https://doi.org/10.1016/j.cja.2022.07.021>
- [12] Huang Y, Zhong S, Gan L, Chen Y (2024) Development of machine learning models for ion-selective electrode cation sensor design. *ACS ES&T Eng* 4(7):1702–1711. <https://doi.org/10.1021/acsestengg.4c00087>
- [13] Jiang Y, Lin W, Yao L et al (2026) A risk-constrained machine learning method for reproducible and sensitive gas-sensing material screening. *J Hazard Mater* 515:142900. <https://doi.org/10.1016/j.jhazmat.2026.142900>
- [14] Kim D, Kwon J, Jeon B, Park YL (2020) Adaptive calibration of soft sensors using optimal transportation transfer learning for mass production and long-term usage. *Adv Intell Syst* 2(6):1900178. <https://doi.org/10.1002/aisy.201900178>
- [15] Langner S, Hase F, Perea JD et al (2020) Beyond ternary OPV: high-throughput experimentation and self-driving laboratories optimize multicomponent systems. *Adv Mater* 32(14):1907801. <https://doi.org/10.1002/adma.201907801>
- [16] Lin J, Chen WM, Lin Y, Cohn J, Gan C, Han S (2020) MCUNet: tiny deep learning on IoT devices. In: *Advances in neural information processing systems*, vol 33, pp 11711–11722. <https://doi.org/10.48550/arXiv.2007.10319>
- [17] Liu Z, Cai M, Hong S et al (2024) Data-driven inverse design of flexible pressure sensors. *Proc Natl Acad Sci USA* 121(28):e2320222121. <https://doi.org/10.1073/pnas.2320222121>
- [18] Liu J, Wang K, Chen Y et al (2026) Large language model-based agents for software engineering: a survey. *ACM Trans Softw Eng Methodol*. <https://doi.org/10.1145/3796507>

- [19] MacLeod BP, Parlane FGL, Morrissey TD et al (2020) Self-driving laboratory for accelerated discovery of thin-film materials. *Sci Adv* 6(20):eaaz8867. <https://doi.org/10.1126/sciadv.aaz8867>
- [20] Mennel L, Symonowicz J, Wachter S, Polyushkin DK, Molina-Mendoza AJ, Mueller T (2020) Ultrafast machine vision with 2D material neural network image sensors. *Nature* 579:62–66. <https://doi.org/10.1038/s41586-020-2038-x>
- [21] Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED (2023) Scaling deep learning for materials discovery. *Nature* 624:80–85. <https://doi.org/10.1038/s41586-023-06735-9>
- [22] Nassif AB, Talib MA, Nasir Q, Dakalbab FM (2021) Machine learning for anomaly detection: a systematic review. *IEEE Access* 9:78658–78700. <https://doi.org/10.1109/ACCESS.2021.3083060>
- [23] National Science and Technology Council (2014) Materials genome initiative strategic plan. Executive Office of the President, Washington, DC
- [24] Podder I, Fischl T, Bub U (2023) Artificial intelligence applications for MEMS-based sensors and manufacturing process optimization. *Telecom* 4(1):165–197. <https://doi.org/10.3390/telecom4010011>
- [25] Podder I, Fischl T, Bub U (2024) Smart calibration and monitoring: leveraging artificial intelligence to improve MEMS-based inertial sensor calibration. *Complex Intell Syst* 10:7451–7474. <https://doi.org/10.1007/s40747-024-01531-y>
- [26] Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 378:686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- [27] Reiser P, Neubert M, Eberhard A et al (2022) Graph neural networks for materials science and chemistry. *Commun Mater* 3:93. <https://doi.org/10.1038/s43246-022-00315-6>
- [28] Robin Y, Amann J, Goodarzi P, Schneider T, Schütze A, Bur C (2022) Deep learning based calibration time reduction for MOS gas sensors with transfer learning. *Atmosphere* 13(10):1614. <https://doi.org/10.3390/atmos13101614>
- [29] Somvanshi S, Islam MM, Chhetri G et al (2025) From tiny machine learning to tiny deep learning: a survey. *ACM Comput Surv* 58(7):168. <https://doi.org/10.1145/3776588>
- [30] Sugiura S, Hardman D, Thuruthel TG, Hasegawa Y, Iida F (2023) Closed-loop optimization of soft sensor morphology using 3D printing of electrically conductive hydrogel. *Adv Intell Syst* 5(11):2300152. <https://doi.org/10.1002/aisy.202300152>

- [31] Sun T, Feng B, Huo J et al (2023) Artificial intelligence meets flexible sensors: emerging smart flexible sensing systems driven by machine learning and artificial synapses. *Nano-Micro Lett* 16:14. <https://doi.org/10.1007/s40820-023-01235-x>
- [32] Szymanski NJ, Rendy B, Fei Y et al (2023) An autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature* 624:86–91. <https://doi.org/10.1038/s41586-023-06734-w>
- [33] Szymanski NJ, Rendy B, Fei Y et al (2026) Author correction: an autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature* 650:E1. <https://doi.org/10.1038/s41586-025-09992-y>
- [34] Warden P, Stewart M, Plancher B, Katti S, Reddi VJ (2023) Machine learning sensors. *Commun ACM* 66(11):25–28. <https://doi.org/10.1145/3586991>
- [35] White A (2012) The materials genome initiative: one year on. *MRS Bull* 37(8):715–716. <https://doi.org/10.1557/mrs.2012.194>
- [36] Xie T, Grossman JC (2018) Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys Rev Lett* 120:145301. <https://doi.org/10.1103/PhysRevLett.120.145301>
- [37] Xu D, Li T, Li Y, Su X, Tarkoma S, Jiang T (2021) Edge intelligence: empowering intelligence to the edge of network. *Proc IEEE* 109(11):1778–1837. <https://doi.org/10.1109/JPROC.2021.3119950>
- [38] Yang H, Li J, Lim KZ et al (2022) Automatic strain sensor design via active learning and data augmentation for soft machines. *Nat Mach Intell* 4:84–94. <https://doi.org/10.1038/s42256-021-00434-8>
- [39] Yi R, Yan J, Shi D et al (2021) Improving the performance of drifted/shifted electronic nose systems by cross-domain transfer using common transfer samples. *Sens Actuators B Chem* 329:129162. <https://doi.org/10.1016/j.snb.2020.129162>
- [40] Zeni C, Pinsler R, Zugner D et al (2025) A generative model for inorganic materials design. *Nature* 639:624–632. <https://doi.org/10.1038/s41586-025-08628-5>
- [41] Zhang Y, Xiang S, Wang Z, Peng X, Tian Y, Duan S, Yan J (2022) TDACNN: target-domain-free domain adaptation convolutional neural network for drift compensation in gas sensors. *Sens Actuators B Chem* 361:131739. <https://doi.org/10.1016/j.snb.2022.131739>
- [42] Zhong D, Xia Z, Zhu Y, Duan J (2023) Overview of predictive maintenance based on digital twin technology. *Heliyon* 9(4):e14534. <https://doi.org/10.1016/j.heliyon.2023.e14534>

- [43] Zhou F, Chai Y (2020) Near-sensor and in-sensor computing. *Nat Electron* 3:664–671. <https://doi.org/10.1038/s41928-020-00501-9>