

Frugal AI Models Show Uneven Accuracy Loss That Threatens Reliable Environmental Decision Making

Nohman Khan

`nohman.khan@apu.edu.my`

Asia Pacific University of Technology & Innovation

Huma Sikandar

University of Technology Malaysia

Javed Ali

Sukkur IBA University

Ahmed H. Alsharif

University of Malaya

Dr. Hamad Raza

Government College University, Faisalabad

Systematic Review

Keywords: Frugal AI, Environmental sustainability, Systematic review, Green AI, resource-constrained AI, accuracy-accessibility trade-off

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11005153/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Frugal AI Models Show Uneven Accuracy Loss That Threatens Reliable Environmental Decision Making

Nohman Khan*

Strategic Research Institute (SRI), Asia Pacific University of Technology and Innovation (APU), Technology Park Malaysia, Kuala Lumpur, Malaysia
Corresponding author Email: nohman.khan@apu.edu.my

Huma Sikandar

Faculty of Management, Universiti Teknologi Malaysia (UTM), 81310, Johor Bahru, Johor, Malaysia
Email: shuma@graduate.utm.my

Javed Ali

Department of Business Administration, Sukkur IBA University, Nisar Ahmed Siddiqui Road, Sukkur, 65200, Sindh, Pakistan
javedali@iba-suk.edu.pk

Ahmed H. Alsharif

Department of Management and Marketing, Faculty of Business and Economics, Universiti Malaya, 50603 Kuala Lumpur, Malaysia.
<https://orcid.org/0000-0002-1364-3545>
ahmedalsharif07@gmail.com

Dr. Hamad Raza

Lyallpur Business School, Government College University Faisalabad, Pakistan.
Email: hamad_raza@hotmail.co.uk

Abstract

Frugal, compressed, and edge AI models produced through pruning, quantization, knowledge distillation, and TinyML design are increasingly promoted as the accessible route by which resource-constrained firms can adopt AI-driven environmental decision tools without the computational cost of full-scale models. Yet compression degrades model accuracy unevenly and unpredictably across architectures, raising concern that the same resource constraint making frugal AI attractive as an accessibility solution also makes its accuracy risk hardest to detect. This paper reports a PRISMA 2020-compliant systematic review of the literature at the intersection of frugal AI and environmental sustainability. Searches of Scopus and Web of Science returned 283 records, reducing to 224 unique records after deduplication and 31 included studies after title/abstract and full-text screening. The corpus was characterised descriptively and coded against seven a priori constructs spanning accuracy-resource trade-offs, resource-constrained context, decision-consequentiality, the AI system's own environmental footprint, uncertainty/reliability treatment, field validation, and theoretical grounding in the Green/frugal-AI literature; an exploratory topic model (Latent Dirichlet Allocation, $K = 5$) provided an independent, corroborating lens. Accuracy-resource trade-offs and resource-constrained context were reported almost universally, as required by the review's own eligibility criteria, but decision-consequentiality (54.8%), the AI system's own footprint (32.3%), field validation (48.4%), and theoretical grounding (19.4%) were each reported in fewer than three in five studies. These reporting gaps motivate the Frugal AI Environmental Efficacy (FAEE) framework, presented here as a conceptual synthesis not an empirically validated typology for locating frugal-AI-for-environment applications along axes of compression aggressiveness and decision consequentiality.

Keywords: *Frugal AI; Environmental sustainability; Systematic review; Green AI; resource-constrained AI; accuracy-accessibility trade-off*

1. Introduction

Artificial intelligence is increasingly positioned as a democratising force for environmental sustainability: lightweight, “frugal,” or compressed AI models produced through pruning, quantization, knowledge distillation, TinyML, and edge-AI design are promoted as the accessible pathway by which resource-constrained firms, including small and medium enterprises (SMEs) and organisations in developing economies, can adopt AI-driven environmental decision tools without the computational cost of full-scale models (Girdhar et al., 2025; Schwartz et al., 2020). This framing treats frugality primarily as a cost and accessibility solution, implicitly assuming that a compressed model preserves decision-relevant accuracy well enough for applications such as energy forecasting, emissions estimation, crop and waste classification, or water- and air-quality monitoring.

Recent technical evidence complicates this assumption. Compression effects are highly uneven across architectures and techniques: Paula et al. (2025) found that BERT and ELECTRA models retained accuracy above 95% after pruning and distillation, while ALBERT under quantization dropped to 65.44% accuracy on the same task family, and Hooker et al. (2020) show that compressed networks do not lose accuracy uniformly across their input distribution but concentrate that loss on a small, specific subset of harder or rarer examples a pattern invisible in an aggregate benchmark figure. Frugality, in other words, does not degrade predictably, and the firms steered toward its most aggressive settings by resource constraint are also the least likely to have the technical capacity to detect when a compressed model has quietly become unreliable. This is the frugal AI paradox motivating the present review: the same resource constraint that makes frugal AI attractive as an accessibility solution also makes its accuracy risk hardest to detect and costliest to bear when the decision it informs is environmentally consequential.

Three literatures currently run in parallel without engaging one another. The technical machine-learning literature on model compression treats accuracy-efficiency trade-offs as a domain-agnostic engineering problem, benchmarked on generic tasks rather than consequential environmental decisions (Paula et al., 2025; Hooker et al., 2020). The AI-adoption literature on SMEs and resource-constrained firms, grounded in Technology-Organization-Environment (TOE) and Resource-Based View (RBV) theorising (Tornatzky & Fleischer, 1990; Barney, 1991), examines adoption antecedents and firm-performance outcomes (Cimino et al., 2025; Muhammad et al., 2025) but treats the adopted AI tool as a black box of uniform capability, without asking whether the specific frugal variant a resource-constrained firm can actually afford behaves differently from the full-scale model its performance claims are based on. And the emerging algorithmic-environmental-justice literature documents that AI trained on sparse Global South data systematically underperforms the populations it serves (Mozaffari et al., 2026), but frames this as a data-representation problem alone, not as compounding with a second, independent inequity introduced at deployment through compression choice. No study has connected these three streams to ask a single question: when resource-constrained firms are steered toward frugal AI as their accessible route to environmental sustainability, what is the resulting distribution of decision-quality risk and does the “accessible” solution quietly compound, rather than resolve, the inequities it is meant to address?

To address this gap, the present study conducts the intersection of frugal/compressed AI and environmental sustainability, combining a theory-driven deductive coding of the resulting corpus with an exploratory topic model, and synthesising both into a novel decision-risk framework. Three objectives guided the review:

- Systematically map, using PRISMA 2020 (Page et al., 2021), the literature at the intersection of frugal, lightweight, and compressed AI and environmental-sustainability applications.
- Descriptively characterise the resulting corpus (publication trend, source distribution, environmental application domain, and compression-technique family).
- Synthesise the descriptive and thematic findings into a novel framework the Frugal AI Environmental Efficacy (FAEE) framework that extends TOE/RBV adoption theorising with an explicit decision-quality-risk dimension and specifies the boundary conditions under which frugal AI is safe versus risky for environmental decision-making.

2. Methodology

This review follows the PRISMA 2020 statement for systematic reviews (Page et al., 2021), reported against its 27-item checklist and accompanying flow diagram (Figure 1). The review question, eligibility criteria, and search strategy were finalised before screening began; the one substantive amendment made during screening retaining rather than excluding records whose environmental-sustainability linkage was ambiguous at the abstract stage, so that the full text could resolve the ambiguity is standard PRISMA 2020 practice and is documented.

Eligibility Criteria

Eligibility was defined using a population/concept/outcome structure tailored to the review's central question: whether frugal, compressed, and edge AI models preserve decision-relevant accuracy when deployed for environmental-sustainability purposes in resource-constrained settings.

Population/context: peer-reviewed studies, plus key arXiv preprints, reflecting how preprint-driven the compression literature currently is (Paula et al., 2025; Schwartz et al., 2020) reporting an AI/ML system applied to an environmental-sustainability domain (energy, water, waste, agriculture, circularity, emissions/carbon) or addressing the environmental footprint of AI itself.

Concept: the study had to engage a frugal, lightweight, compressed, edge, or resource-efficient AI approach pruning, quantization, knowledge distillation, TinyML, or edge AI (Girdhar et al., 2025).

Outcome: the study had to report at least one of an accuracy/performance/resource trade-off, a resource-constrained implementation context, or a decision-quality/reliability issue.

Exclusions: records addressing a non-environmental domain with no AI-footprint framing, containing no substantive frugal/compressed AI component, or constituting peripheral hardware/general-AI papers without sufficient sustainability relevance a criterion that, proved decisive at the full-text stage.

The search window was 2015–2026, capturing the field from the early efficient-deep-learning era through the current frugal-AI/TinyML wave.

Eligibility above is defined by topical concept and reported outcome, not by primary study design. Two of the 31 included studies are themselves narrative/technical reviews rather than primary empirical studies; both satisfy the Concept and Outcome criteria above in their own right, and are retained rather than excluded, but are flagged throughout the characteristics table (Table 1) and their construct-coding results

(Section 3.2) are interpreted as reflecting the synthesis each review reports rather than original primary evidence. A sensitivity check excluding both from the deductive coding changed every construct's explicit prevalence estimate by at most 3.4 percentage points, indicating the review's substantive findings do not depend on their inclusion.

Information Sources and Search Strategy

Records were identified from Scopus and Web of Science. The search combined four concept blocks with Boolean operators, applied to title, abstract, and keyword fields, with field-code adaptation for Web of Science:

Block 1 (Frugal/compressed AI): (“frugal AI” OR “frugal artificial intelligence” OR “lightweight AI” OR “lightweight machine learning” OR “model compression” OR “model pruning” OR quantization OR “knowledge distillation” OR TinyML OR “edge AI” OR “efficient AI” OR “green AI”)

Block 2 (Environmental sustainability): (“environmental sustainability” OR climate OR “circular economy” OR “renewable energy” OR carbon OR emissions OR “waste management” OR “resource efficiency” OR “net zero” OR “net-zero” OR “sustainable development”)

Block 3 (Resource-constrained/accessibility context): (“resource-constrained” OR SME* OR “small and medium enterprise*” OR “developing countr*” OR “low-resource” OR “digital divide”)

Block 4 (Trade-off/performance): (accuracy OR “trade-off” OR “performance degradation” OR efficacy OR reliability)

Combined string: (Block 1) AND (Block 2) AND (Block 3 OR Block 4)

The search returned 84 records from Scopus and 199 from Web of Science (283 gross records). After removing 59 duplicate records identified across both databases, 224 unique records advanced to title/abstract screening.

Identification and Title/Abstract Screening

All 224 unique records were screened against the eligibility criteria at the title/abstract level. This stage excluded 175 records, most commonly for addressing a non-environmental domain with no AI-footprint framing or for lacking a substantive frugal/compressed AI component. The remaining 49 records were classified by confidence: 36 as high-confidence matches recommended for full-text retrieval, and 13 as borderline records whose abstracts showed efficiency or trade-off evidence but where the environmental-sustainability linkage was not yet sufficiently explicit to decide without the full text, most commonly technical hardware/accelerator comparators. Consistent with PRISMA 2020 guidance to resolve genuine uncertainty at full text rather than at the abstract stage, borderline records were retained for retrieval rather than excluded outright.

Full-Text Eligibility Assessment

Of the 49 records sought for retrieval, 4 could not be retrieved and 45 were assessed in full text against the same eligibility criteria used at the abstract stage. Full-text assessment excluded 14 records.

All 14 were excluded because, on full-text reading, they showed no substantive environmental-sustainability application and no genuine AI-environmental-footprint framing. This outcome applied to 13 of the records flagged as borderline at the abstract stage, plus one record initially classified high-confidence confirming that the abstract-stage borderline flag had correctly anticipated where the eligibility criteria would not hold up under closer reading. These excluded records were predominantly hardware-accelerator or general-purpose edge-inference papers whose only connection to sustainability was a generic “power consumption” or “energy-efficient” framing used as a standard engineering metric rather than an environmental-sustainability claim; in one case the only occurrence of “sustainable” in the entire text was inside a cited reference title rather than the paper's own framing. A further record (M173) is classified as not retrieved rather than excluded, for reasons detailed. Full-text screening therefore yielded 31 included studies (Figure 1).

Screening Transparency and Data-Integrity Note

Two departures from a clean retrieval pipeline are reported here in the interest of transparency, consistent with PRISMA 2020's emphasis on documenting deviations from the review protocol (Page et al., 2021). First, 3 of the 49 records recommended for full-text retrieval could not be obtained through the review's access channels and are classified as not retrieved.

Second, the full-text PDF obtained for one further record (M173) did not match the record screened at the abstract stage: the file retrieved was a different paper, by different authors, from a different venue. Because the intended source itself was never actually obtained, this record is classified as not retrieved rather than excluded, consistent with the PRISMA 2020 distinction between reports that could not be obtained and reports that were obtained, assessed, and found ineligible (Page et al., 2021) an earlier version of this section had classified it as an exclusion, which is corrected here. Reports not retrieved therefore total 4 (3 originally unavailable + M173), reports assessed in full text total 45, and reports excluded at full text total 14. Rather than screening the substitute document as though it were the intended source which would have broken the chain of custody back to the screened 224-record corpus the record is reported in the not-retrieved category of the flow diagram (Figure 1) pending retrieval of the correct source.

PRISMA Flow Diagram

Figure 1 summarises the review's flow from identification to inclusion. Of 84 Scopus and 199 Web of Science records (283 gross, 59 duplicates), 224 unique records were screened; 175 were excluded at title/abstract; 49 were sought for full-text retrieval (36 high-confidence, 13 borderline); 4 were not retrieved (3 originally unavailable and 1 content/identity mismatch, M173 Section 2.6); 45 were assessed in full text; 14 were excluded at full text, all for insufficient environmental-sustainability/AI-footprint substance; and 31 studies were included in the review.

This review follows the PRISMA 2020 statement for systematic reviews (Page et al., 2021), reported against its 27-item checklist and accompanying flow diagram (Figure 1). The review question, eligibility criteria, and search strategy were finalised before screening began; the one substantive amendment made during screening retaining rather than excluding records whose environmental-sustainability linkage was ambiguous at the abstract stage, so that the full text could resolve the ambiguity is standard PRISMA 2020 practice and is documented.

Eligibility Criteria

Eligibility was defined using a population/concept/outcome structure tailored to the review's central question: whether frugal, compressed, and edge AI models preserve decision-relevant accuracy when deployed for environmental-sustainability purposes in resource-constrained settings.

Population/context: peer-reviewed studies, plus key arXiv preprints, reflecting how preprint-driven the compression literature currently is (Paula et al., 2025; Schwartz et al., 2020) reporting an AI/ML system applied to an environmental-sustainability domain (energy, water, waste, agriculture, circularity, emissions/carbon) or addressing the environmental footprint of AI itself.

Concept: the study had to engage a frugal, lightweight, compressed, edge, or resource-efficient AI approach pruning, quantization, knowledge distillation, TinyML, or edge AI (Girdhar et al., 2025).

Outcome: the study had to report at least one of an accuracy/performance/resource trade-off, a resource-constrained implementation context, or a decision-quality/reliability issue.

Exclusions: records addressing a non-environmental domain with no AI-footprint framing, containing no substantive frugal/compressed AI component, or constituting peripheral hardware/general-AI papers without sufficient sustainability relevance a criterion that, proved decisive at the full-text stage.

The search window was 2015–2026, capturing the field from the early efficient-deep-learning era through the current frugal-AI/TinyML wave.

Eligibility above is defined by topical concept and reported outcome, not by primary study design. Two of the 31 included studies are themselves narrative/technical reviews rather than primary empirical studies; both satisfy the Concept and Outcome criteria above in their own right, and are retained rather than excluded, but are flagged throughout the characteristics table (Table 1) and their construct-coding results (Section 3.2) are interpreted as reflecting the synthesis each review reports rather than original primary evidence. A sensitivity check excluding both from the deductive coding changed every construct's explicit prevalence estimate by at most 3.4 percentage points, indicating the review's substantive findings do not depend on their inclusion.

Information Sources and Search Strategy

Records were identified from Scopus and Web of Science. The search combined four concept blocks with Boolean operators, applied to title, abstract, and keyword fields, with field-code adaptation for Web of Science:

Block 1 (Frugal/compressed AI): (“frugal AI” OR “frugal artificial intelligence” OR “lightweight AI” OR “lightweight machine learning” OR “model compression” OR “model pruning” OR quantization OR “knowledge distillation” OR TinyML OR “edge AI” OR “efficient AI” OR “green AI”)

Block 2 (Environmental sustainability): (“environmental sustainability” OR climate OR “circular economy” OR “renewable energy” OR carbon OR emissions OR “waste management” OR “resource efficiency” OR “net zero” OR “net-zero” OR “sustainable development”)

Block 3 (Resource-constrained/accessibility context): (“resource-constrained” OR SME* OR “small and medium enterprise*” OR “developing countr*” OR “low-resource” OR “digital divide”)

Block 4 (Trade-off/performance): (accuracy OR “trade-off” OR “performance degradation” OR efficacy OR reliability)

Combined string: (Block 1) AND (Block 2) AND (Block 3 OR Block 4)

The search returned 84 records from Scopus and 199 from Web of Science (283 gross records). After removing 59 duplicate records identified across both databases, 224 unique records advanced to title/abstract screening.

Identification and Title/Abstract Screening

All 224 unique records were screened against the eligibility criteria at the title/abstract level. This stage excluded 175 records, most commonly for addressing a non-environmental domain with no AI-footprint framing or for lacking a substantive frugal/compressed AI component. The remaining 49 records were classified by confidence: 36 as high-confidence matches recommended for full-text retrieval, and 13 as borderline records whose abstracts showed efficiency or trade-off evidence but where the environmental-sustainability linkage was not yet sufficiently explicit to decide without the full text, most commonly technical hardware/accelerator comparators. Consistent with PRISMA 2020 guidance to resolve genuine uncertainty at full text rather than at the abstract stage, borderline records were retained for retrieval rather than excluded outright.

Full-Text Eligibility Assessment

Of the 49 records sought for retrieval, 4 could not be retrieved and 45 were assessed in full text against the same eligibility criteria used at the abstract stage. Full-text assessment excluded 14 records.

All 14 were excluded because, on full-text reading, they showed no substantive environmental-sustainability application and no genuine AI-environmental-footprint framing. This outcome applied to 13 of the records flagged as borderline at the abstract stage, plus one record initially classified high-confidence confirming that the abstract-stage borderline flag had correctly anticipated where the eligibility criteria would not hold up under closer reading. These excluded records were predominantly hardware-accelerator or general-purpose edge-inference papers whose only connection to sustainability was a generic “power consumption” or “energy-efficient” framing used as a standard engineering metric rather than an environmental-sustainability claim; in one case the only occurrence of “sustainable” in the entire text was inside a cited reference title rather than the paper's own framing. A further record (M173) is classified as not retrieved rather than excluded, for reasons detailed. Full-text screening therefore yielded 31 included studies (Figure 1).

Screening Transparency and Data-Integrity Note

Two departures from a clean retrieval pipeline are reported here in the interest of transparency, consistent with PRISMA 2020's emphasis on documenting deviations from the review protocol (Page et al., 2021). First, 3 of the 49 records recommended for full-text retrieval could not be obtained through the review's access channels and are classified as not retrieved.

Second, the full-text PDF obtained for one further record (M173) did not match the record screened at the abstract stage: the file retrieved was a different paper, by different authors, from a different venue. Because

the intended source itself was never actually obtained, this record is classified as not retrieved rather than excluded, consistent with the PRISMA 2020 distinction between reports that could not be obtained and reports that were obtained, assessed, and found ineligible (Page et al., 2021) an earlier version of this section had classified it as an exclusion, which is corrected here. Reports not retrieved therefore total 4 (3 originally unavailable + M173), reports assessed in full text total 45, and reports excluded at full text total 14 (Section 2.5). Rather than screening the substitute document as though it were the intended source which would have broken the chain of custody back to the screened 224-record corpus the record is reported in the not-retrieved category of the flow diagram (Figure 1) pending retrieval of the correct source.

PRISMA Flow Diagram

Figure 1 summarises the review's flow from identification to inclusion. Of 84 Scopus and 199 Web of Science records (283 gross, 59 duplicates), 224 unique records were screened; 175 were excluded at title/abstract; 49 were sought for full-text retrieval (36 high-confidence, 13 borderline); 4 were not retrieved (3 originally unavailable and 1 content/identity mismatch, M173 Section 2.6); 45 were assessed in full text; 14 were excluded at full text, all for insufficient environmental-sustainability/AI-footprint substance; and 31 studies were included in the review.

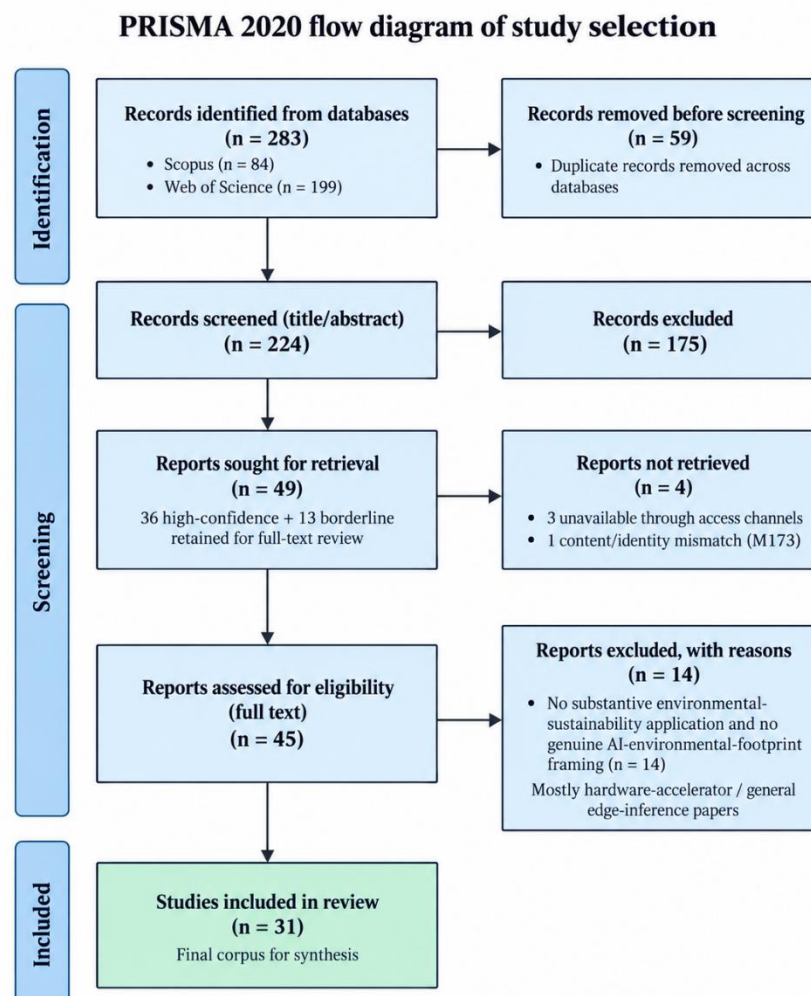


Figure 1: PRISMA inclusion and exclusion criteria

3. Results

3.1 Descriptive Results

The review's full-text screening process yielded 31 included studies. Publication activity was heavily concentrated at the recent end of the 2015–2026 search window: 2 studies were published in 2024, 8 in 2025, and 21 in 2026 (Figure 2) consistent with frugal, edge-deployed AI for environmental applications being a fast-emerging rather than an established sub-field, and with the search being run in 2026 itself.

The included studies span nine environmental application domains (Figure 2). Agriculture and precision farming was the largest single domain (9 studies, 29.0%), covering crop-health monitoring, weed identification, and pest recognition. AI's own environmental footprint studies whose primary contribution is characterising or reducing the energy/carbon cost of the AI system itself, rather than applying AI to an external environmental problem constituted a further five studies (16.1%), alongside energy/smart-grid/building applications (5 studies, 16.1%). The remaining studies addressed waste and circular-economy applications (3), and two studies each in biodiversity/wildlife monitoring, water, transport/mobility, and air quality/pollution, with one study in Earth observation/remote sensing.

Classified by dominant frugal-AI technique family (Figure 2), TinyML/edge-general approaches architectures explicitly designed or adapted for microcontroller-class deployment without being reducible to a single named compression operation were most common (15 studies, 48.4%), followed by knowledge distillation (5, 16.1%), a hybrid/multiple-technique category (4, 12.9%), quantization (4, 12.9%), pruning (2, 6.5%), and a single study (3.2%) whose efficiency claim rested on an efficient/lightweight architecture design rather than a named post-hoc compression technique.

The included studies were distributed across 25 distinct publication venues (Figure 2), indicating a fragmented rather than a consolidated publication landscape for this topic. IEEE Access was the most frequent single venue (4 studies), followed by Sensors, Technologies, and Smart Agricultural Technology (2 studies each); the remaining 21 studies were each the sole representative of their venue a pattern consistent with an emerging, cross-disciplinary research area that has not yet coalesced around a small set of specialist outlets.

Descriptive Characteristics of Included Studies

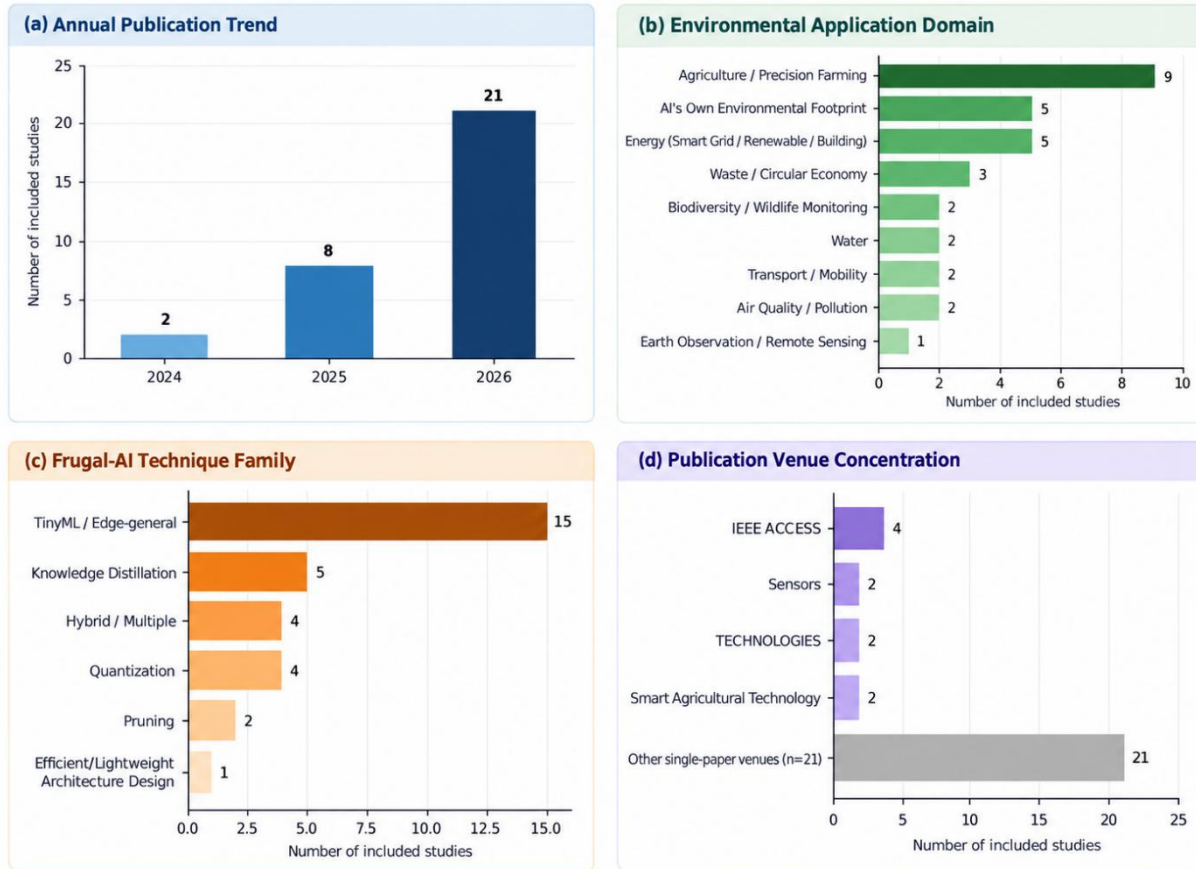


Figure 2. Descriptive characteristics of the included studies

Table 1 reports the full characteristics of all 31 included studies, publication year, environmental application domain, frugal-AI technique family, and publication source cross-referenced by the same running index. Two included studies (marked †) are themselves narrative/technical reviews rather than primary empirical studies; both satisfy the review's concept-based eligibility criteria in their own right, but their construct-coding results in Table 1 reflect the synthesis each review reports rather than original primary evidence. A sensitivity check excluding both from the deductive coding changed every construct's explicit prevalence estimate by at most 3.4 percentage points (e.g., C1 96.8%→96.6%, C4 32.3%→34.5%, C7 19.4%→20.7%), indicating the review's substantive findings do not depend on their inclusion.

Table 1: Characteristics of included studies

No.	Study (author, year)	Domain	Technique family	Source
1	Wang and Yu (2026)	Biodiversity / Wildlife Monitoring	TinyML / Edge-general	Sensors
2	Escorcia et al. (2026)	Water	TinyML / Edge-general	TECHNOLOGIES
3	Boudali and Khlil (2026)	Agriculture / Precision Farming	Knowledge Distillation	Internet of Things (The Netherlands)
4	Li et al. (2026)	Agriculture / Precision Farming	TinyML / Edge-general	COMPUTERS & ELECTRICAL ENGINEERING

5	Ali et al. (2026)	AI's Own Environmental Footprint	TinyML / Edge-general	INFORMATION
6	Rai et al. (2026)	AI's Own Environmental Footprint	Efficient/Lightweight Architecture Design	DISCOVER COMPUTING
7	di Torrepadula et al. (2026)	AI's Own Environmental Footprint	Knowledge Distillation	JOURNAL OF SYSTEMS ARCHITECTURE
8	Nabil and Leghris (2026)	Agriculture / Precision Farming	Knowledge Distillation	FUTURE INTERNET
9	Moeinnia et al. (2026)	Agriculture / Precision Farming	TinyML / Edge-general	Smart Agricultural Technology
10	Patrick et al. (2026) †	Agriculture / Precision Farming	TinyML / Edge-general	Smart Agricultural Technology
11	Sami et al. (2026)	Energy (Smart Grid / Renewable / Building)	TinyML / Edge-general	Results in Engineering
12	Attoui and Lahby (2026)	Transport / Mobility	TinyML / Edge-general	Sustainability (Switzerland)
13	Echchidmi and Bouayad (2026)	Waste / Circular Economy	Hybrid / Multiple	Technologies
14	Li et al. (2026)	Agriculture / Precision Farming	TinyML / Edge-general	AGRICULTURE-BASEL
15	Komin and Boiko (2025)	Energy (Smart Grid / Renewable / Building)	TinyML / Edge-general	Technology Audit and Production Reserves
16	Zhao et al. (2025)	Energy (Smart Grid / Renewable / Building)	Pruning	Journal of Electrical Engineering and Technology
17	Almtireen et al. (2025)	Waste / Circular Economy	TinyML / Edge-general	JOURNAL OF SUPERCOMPUTING
18	Koli et al. (2025)	Agriculture / Precision Farming	Hybrid / Multiple	SSRG International Journal of Electronics and Communication Engineering
19	Mshragi and Petri (2025)	Energy (Smart Grid / Renewable / Building)	Quantization	ARTIFICIAL INTELLIGENCE REVIEW
20	Batic et al. (2025)	Energy (Smart Grid / Renewable / Building)	Knowledge Distillation	Expert Systems with Applications
21	Zhang et al. (2025)	Biodiversity / Wildlife Monitoring	Pruning	FRONTIERS IN PLANT SCIENCE
22	Pazhanivel et al. (2024)	Air Quality / Pollution	Quantization	SENSORS
23	Sun et al. (2026)	AI's Own Environmental Footprint	TinyML / Edge-general	IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems
24	Le et al. (2026)	Earth Observation / Remote Sensing	Knowledge Distillation	IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING
25	Khan et al. (2026)	AI's Own Environmental Footprint	Hybrid / Multiple	IEEE Communications Surveys and Tutorials
26	Zhang et al. (2026)	Agriculture / Precision Farming	TinyML / Edge-general	IEEE Internet of Things Journal
27	Rojas-Hiedra et al. (2026) †	Transport / Mobility	Hybrid / Multiple	IEEE ACCESS

28	Yang et al. (2026)	Water	Quantization	IEEE ACCESS
29	Ceroli et al. (2025)	Air Quality / Pollution	Quantization	IEEE Access
30	Xan et al. (2024)	Agriculture / Precision Farming	TinyML / Edge-general	IEEE ACCESS
31	Azizi et al. (2026)	Waste / Circular Economy	TinyML / Edge-general	Journal of Engineering and Technology for Industrial Applications

3.2 Deductive Content / Thematic Synthesis

Building on the review's objectives and the proposed Frugal AI Environmental Efficacy (FAEE) framework, seven a priori constructs were defined and coded across the full text of all 31 included studies (Table 2). Each construct was coded as explicit (Y the study substantively and directly addresses the construct), partial (P the construct is present but treated only in passing, generically, or without elaboration), or absent (N). Coding combined a systematic keyword-hit digest constructed by pattern-matching each full text against a bank of construct-specific terms and extracting the surrounding context with manual verification of every hit against its source sentence, so that, for example, a paper's incidental mention of “CO2” as a sensed environmental variable (rather than as a reflection on the AI system's own carbon cost) was not miscoded as AI-footprint reflexivity.

C1 (quantified accuracy–resource trade-off) and C2 (resource-constrained deployment context specified) were explicit in 30 of 31 studies (96.8%; Table 1) unsurprising, since both properties are close to definitional of the review's Concept and Outcome eligibility criteria. The one partial case (a physics-informed biogas-monitoring system) reported strong predictive accuracy on its tailored edge model but did not report an explicit compressed-versus-baseline comparison. Because eligibility itself required a reported accuracy-resource trade-off in a resource-constrained context, C1 and C2's near-ceiling prevalence is best read as a check that the eligibility criteria were applied consistently, not as a substantive empirical finding about the literature; they are reported here for completeness and to make that circularity explicit, rather than treated as diagnostic. The constructs that carry the review's actual diagnostic content because eligibility placed no comparable floor on them are C3 through C7 below.

C5 (uncertainty/reliability treatment) was similarly near-universal (93.5% explicit), reflecting standard machine-learning reporting conventions (cross-validation, confidence intervals, robustness/ablation analysis) rather than a distinctive engagement with frugal-AI reliability specifically. The remaining four constructs, by contrast, showed substantially more variation and are the more diagnostic findings of this synthesis. Figure 3 below illustrates the deductive thematic coding for priori constructs.

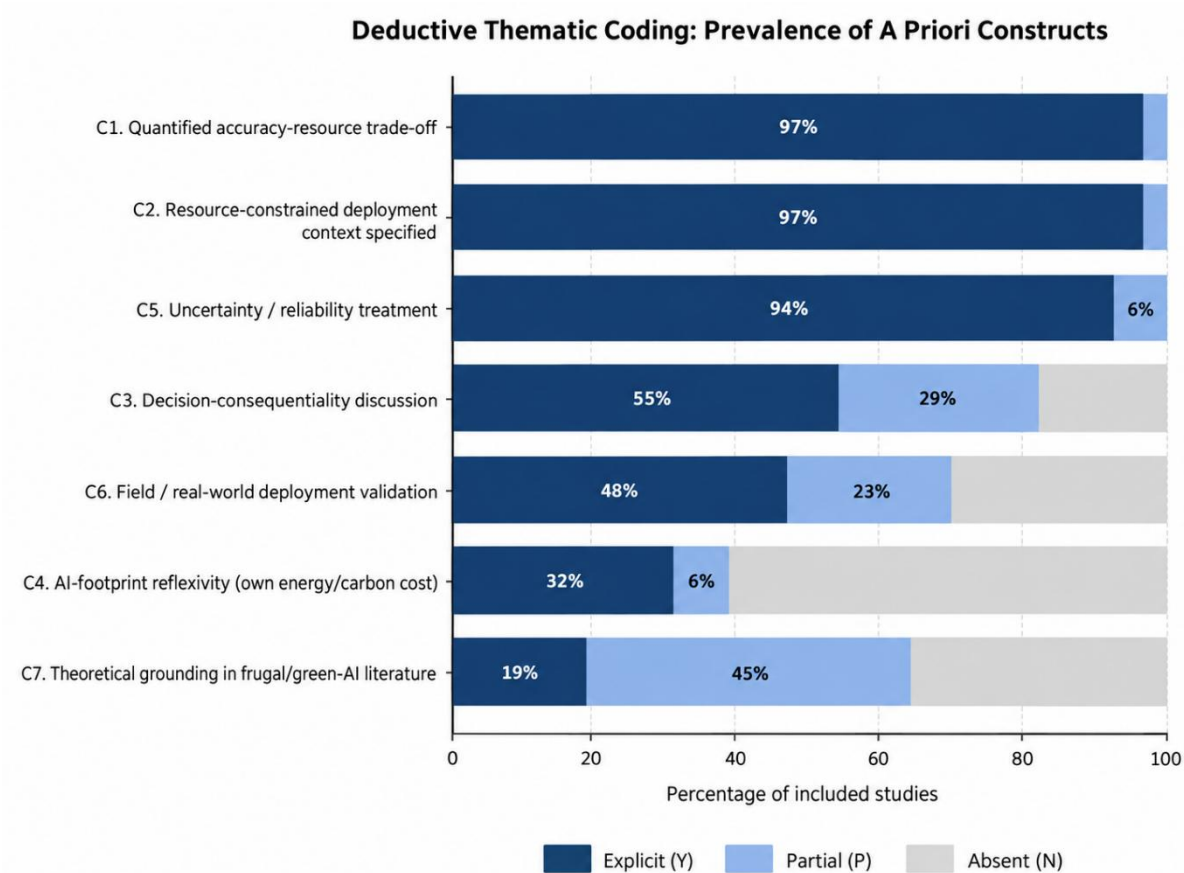


Figure 3. Deductive thematic coding

C3 (decision-consequentiality discussion) was explicit in just over half of studies (54.8%), partial in 29.0%, and entirely absent in 16.1%. Where present, this construct typically took the form of an explicit link between a specific error type (a false negative, a missed detection) and a real-world consequence for instance, Yang et al.'s (2026) characterisation of a false negative in water-infrastructure anomaly detection as a “missed attack” posing “severe safety risks” (Table 2). Its absence in nearly half of studies indicates that a substantial share of the corpus reports classification/detection performance without connecting that performance to the stakes of the decision it is meant to support.

Table 2: Deductive coding prevalence across the seven a priori constructs

Code	Construct	Explicit (Y)	Partial (P)	Absent (N)
C1	Quantified accuracy-resource trade-off	30 (96.8%)	1 (3.2%)	0 (0.0%)
C2	Resource-constrained deployment context specified	30 (96.8%)	1 (3.2%)	0 (0.0%)
C3	Decision-consequentiality discussion	17 (54.8%)	9 (29.0%)	5 (16.1%)
C4	AI-footprint reflexivity (own energy/carbon cost)	10 (32.3%)	2 (6.5%)	19 (61.3%)
C5	Uncertainty / reliability treatment	29 (93.5%)	2 (6.5%)	0 (0.0%)
C6	Field / real-world deployment validation	15 (48.4%)	7 (22.6%)	9 (29.0%)

C7	Theoretical grounding in frugal/green-AI literature	6 (19.4%)	14 (45.2%)	11 (35.5%)
----	---	-----------	------------	------------

C4 (AI-footprint reflexivity) whether a study reflects on the energy or carbon cost of the AI system itself, rather than only on the environmental application it serves was explicit in less than a third of studies (32.3%) and entirely absent in 61.3%. This is the starkest gap in the corpus: many studies that apply frugal AI to an environmental problem do not, in turn, characterise the environmental cost of the AI doing the applying. The minority that did were concentrated among the five studies whose domain classification was itself “AI’s own environmental footprint” (Figure 2), suggesting that footprint-reflexivity is currently treated as a specialised research niche rather than a standard reporting practice across frugal environmental AI generally. Stratifying C4 by environmental application domain sharpens this point: all five “AI’s Own Environmental Footprint” studies were explicit on C4 by definition, while explicit C4 reporting fell to 19.2% (5 of 26) among studies applying frugal AI to an external environmental problem agriculture (2 of 9), energy (0 of 5), waste (0 of 3), water (1 of 2), transport (1 of 2), biodiversity (0 of 2), air quality (0 of 2, one partial), and Earth observation (1 of 1) indicating that outside the studies whose stated topic is the AI system’s own footprint, footprint-reflexivity is closer to one study in five than one in three.

C6 (field/real-world deployment validation) was explicit in 48.4% of studies, partial in 22.6%, and absent in 29.0% meaning roughly three in ten included studies validate their frugal model entirely offline or in simulation, despite the review’s Population/context criterion requiring a resource-constrained deployment framing. Explicit field validation ranged from single hardware-latency benchmarks to genuine multi-site trials, such as Moeinnia et al.’s (2026) laser-weeding robot field-tested in “Vancouver, Canada, and Arusha, Tanzania” (Table 3).

C7 (theoretical grounding in frugal/green-AI literature) was the least prevalent construct: only 19.4% of studies explicitly engaged the conceptual Green AI/frugal-AI literature (e.g., citing Schwartz et al., 2020, or explicitly framing the work as “Green AI” or “sustainable AI”), 45.2% showed partial engagement (typically using “TinyML” only as a technique label, without situating the work within a broader sustainability-of-AI research program), and 35.5% showed no such grounding at all. Combined with the C4 finding, this indicates that most of the reviewed literature treats compression as an engineering means to a deployment end rather than as part of a deliberate environmental-AI research agenda precisely the conceptual gap the FAEE framework is intended to address.

Table 3: Representative textual evidence illustrating the deductive coding

Construct	Study	Representative excerpt
C1	Li et al. (2026) (WAGENet)	“a 31× compression ratio, while maintaining 99.4% model-fidelity retention”
	Moeinnia et al. (2026) (LiteWeed)	YOLOX-nano was selected for deployment despite a lower mAP@0.5:0.95 of 0.32 versus 0.44 for YOLOX-s, trading accuracy for inference latency
C3	Yang et al. (2026) (water-CPS anomaly detection)	“a false negative (e.g., missed attack) poses severe safety risks”
	Rai et al. (2026) (traffic-sign DF-CNN)	minimizing false negatives is described as “an essential requirement for safety-critical” recognition
C4	Khan et al. (2026) (xG wireless compression survey)	explicitly framed as “Model Compression for Sustainable AI”, reducing “the carbon footprint of AI operations”

	Sun et al. (2026) (carbon-aware accelerator model)	an analytical “carbon-aware” performance model quantifying the carbon footprint of AI accelerator design choices
C6	Moeinnia et al. (2026) (LiteWeed)	“Field trials in Vancouver, Canada, and Arusha, Tanzania” validating the robotic platform under real agronomic conditions
	Komin and Boiko (2025) (HEMS edge aggregation)	“Tests on Raspberry Pi 5 (16 GB) showed 97.2% availability over 72 hours”
C7	di Torrepadula et al. (2026) (KD for low-energy AIoT)	“According to Schwartz et al. [4], the amount of computational...” — explicit engagement with the Green AI literature
	Yang et al. (2026) (Green-WSADE)	framework “inspired by the Green AI idea”

3.3 Exploratory Topic Model

Preprocessing and Model Specification

Latent Dirichlet Allocation (LDA) was employed as the primary computational method for thematic synthesis, selected over conventional narrative coding because it provides a probabilistic, replicable, and data-driven decomposition of latent thematic structures embedded within the corpus. This analysis uses standard LDA rather than Structural Topic Modelling (STM); STM extends LDA by incorporating document-level covariates, such as publication year, directly into the topic prevalence model, but the present implementation did not include such covariates. Temporal patterns in topic prevalence were instead examined post hoc by aggregating fitted topic probabilities by publication year.

Search-K Diagnostic

Following the STM searchK convention, four diagnostics were computed for each candidate K in the range 2–8 (Figure 4): (a) semantic coherence, using the UMass intrinsic coherence measure (Mimno et al., 2011) computed from the model's own top-10-word document co-occurrence statistics; (b) exclusivity, a FREX-style statistic giving the mean share of each top-word's total probability mass that is attributable to its assigned topic rather than shared with other topics; (c) held-out log-likelihood, estimated by document completion withholding a random 20% of tokens from every document before training, then scoring the withheld tokens under each document's inferred topic distribution and the trained topic-word distributions; and (d) residual dispersion, a Pearson-residual multinomial-dispersion statistic comparing observed word counts to those implied by the fitted model (values above 1 indicate residual overdispersion relative to a well-specified multinomial mixture).

Held-out log-likelihood and residual dispersion both improved monotonically as K increased from 2 to 8 (Figure 4) an expected pattern, since additional topics give the model more flexibility to fit both held-out and in-sample word counts, and provide no basis on their own for selecting K. Coherence and exclusivity, which penalise over-fragmentation, told a more informative story: exclusivity declined steadily from K = 2 onward, while coherence reached its worst (most negative) value at K = 4 before partially recovering. Marginal gains in held-out likelihood and dispersion beyond K = 5 were small relative to the gains from K = 2 to K = 5, consistent with a diminishing-returns elbow around K = 5.

This quantitative pattern was cross-checked against qualitative inspection of the resulting topic-word lists at K = 4 through K = 7. At K = 4 and K = 5, all topics were substantively coherent and mutually

distinguishable. At $K = 6$, one topic began to fragment into a generic mixture of technical terms (“devices”, “anomaly”, “memory”, “distribution”, “cost”) without a clear unifying theme; at $K = 7$, two of the seven topics were dominated by such generic, cross-cutting vocabulary one a mixed methodology/evaluation topic (“test”, “evaluation”, “pipeline”, spanning traffic-sign and acoustic-monitoring studies with no substantive link between them) and one a near-duplicate of the hardware/energy topic already present at lower K

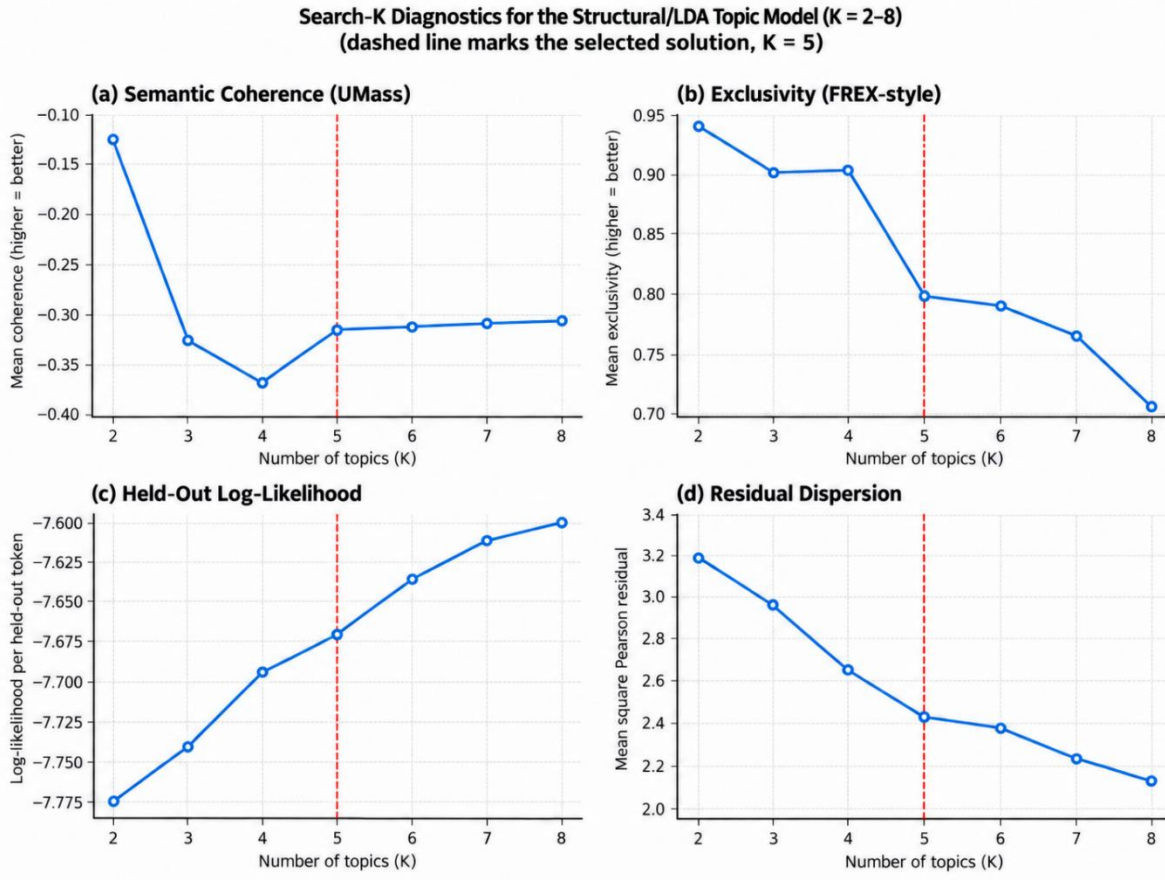


Figure 4. Search-K diagnostics for $K = 2-8$

Topic Proportions and FREX Terms

The five-topic solution resolved into a mixture of domain-specific and method-specific topics, summarised by their FREX terms the words that are simultaneously frequent within a topic and disproportionately exclusive to it (frequency-exclusivity weighted; Table 4) and by their mean corpus-level proportion (Figure 5).

T1: Bioacoustics / Animal Monitoring (12.1%). Built on acoustic-sensing and animal-activity vocabulary (“animal”, “animals”, “forest”, “behavior”, “sampling”), this topic aligns closely with the review's biodiversity/wildlife-monitoring and livestock-monitoring domain studies.

T2: Water-Air Quality Forecasting (12.4%). Built on time-series forecasting and sensor-fusion vocabulary (“forecasting”, “fog”, “gru”, “quality”, “cloud”, “attention”), this topic groups the review's water- and air-quality monitoring studies, which share a common architectural vocabulary

(recurrent/attention-based sequence models, fog/cloud-edge hybrid deployment) distinct from the vision-based classification studies in T3.

T3: Model Compression Methods (16.8%). The clearest method-defined (rather than domain-defined) topic, dominated by compression-technique vocabulary (“student”, “teacher”, “compression”, “pruning”, “distillation”, “quantization”). Studies loading heavily on T2 include several of the review’s AI-footprint-domain and wireless/communications studies, for which the compression method itself rather than a specific downstream environmental application is the paper’s primary contribution.

T4: Vision-Based Crop, Weed & Waste Classification (20.4%). Built on computer-vision and classification vocabulary (“disease”, “weed”, “images”, “waste”, “plant”, “agriculture”), this topic closely tracks the review’s largest single domain category, agriculture/precision farming, together with the waste/circular-economy studies that share the same image-classification architecture family (lightweight CNN/YOLO variants).

Table 4: FREX (frequency-exclusivity weighted) top 10 terms per topic

Topic	Label	Top 10 FREX terms	Mean exclusivity
T1	Bioacoustic / Animal Monitoring	articles, predictive, literature, forest, total, animals, behavior, sampling, animal, random	0.926
T2	Water-Air Quality Forecasting	forecasting, air, seq, fog, gru, quality, production, cloud, acoustic, monitoring	0.941
T3	Model Compression Methods (KD/Pruning/Quantization)	student, teacher, compression, pruning, distillation, knowledge, proc, wireless, int, nilm	0.909
T4	Vision-Based Crop, Weed & Waste Classification	disease, weed, images, image, waste, plant, classes, agriculture, module, class	0.916
T5	Energy, Power & Predictive Hardware Management	process, management, information, anomaly, power, computing, control, resource, devices, engineering	0.948

T5: Energy, Power & Predictive Hardware Management (38.3%). The largest topic by corpus share, built on comparatively abstract, general-purpose engineering vocabulary (“process”, “management”, “power”, “computing”, “control”, “resource”, “devices”, “engineering”) rather than the sharply domain-specific nouns that characterise T1, T2 and T4. Notably, despite this more generic-sounding vocabulary, T4 returned the highest mean exclusivity of the five topics (0.948; Table 2) indicating that, empirically, this particular combination of abstract engineering terms is used in a concentrated way by a fairly coherent subset of studies (chiefly energy/smart-grid/building-management and general hardware-benchmarking papers, Section 3.1) rather than being scattered evenly across the whole corpus. T4’s outsized share nonetheless suggests it also functions partly as a cross-cutting topic capturing shared resource-efficiency and hardware-deployment framing used across many otherwise domain-distinct studies a known feature of LDA solutions fit to small, lexically homogeneous technical corpora, reported here rather than over-interpreted as a single narrow substantive theme. Figure 5 Mean Corpus level topic proportions.

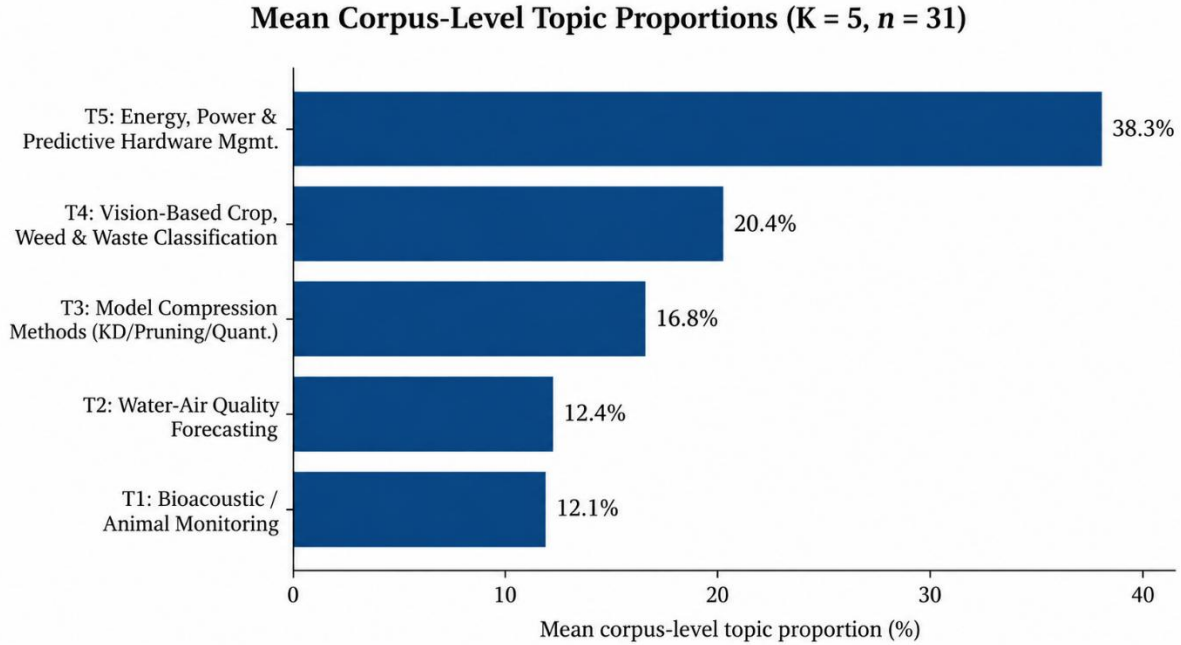


Figure 5. Mean corpus-level topic proportions, K = 5.

Corpus-to-Topic Reference Mapping

Table 5 reports the full document-topic mapping for all 31 included studies: each study's dominant topic and its complete estimated topic-proportion vector (θ), rounded to three decimal places. Dominant-topic proportions ranged from 0.369 to 0.761, indicating that most studies load predominantly on a single topic rather than mixing evenly across all five consistent with a corpus of relatively focused, single-contribution technical papers rather than broad multi-topic reviews. Cross-referencing this table against the deductive domain coding shows that the topic model's method/lexical axis of variation is only partially aligned with the review's a priori environmental-domain coding: for example, the two air-quality/pollution studies (Pazhanivel et al., 2024, a fog-enabled forecasting system, and Cerioli et al., 2025, a microplastics-detection compiler) share the same domain classification but load on different dominant topics (T2 and T5 respectively), and the agriculture/precision-farming domain the review's largest splits its dominant-topic assignment across T4 (the majority, vision-based classification studies), T5 (hardware/energy-framed studies), and T0 (a livestock-sensing review). This divergence is an informative finding in its own right: it indicates that the corpus's inductive lexical structure is organised as much around how efficiency is engineered and reported (vision-classification framing versus compression-method framing versus energy/hardware framing) as around which environmental problem is being addressed a distinction the deductive domain coding alone could not have revealed.

Table 5: Corpus-to-topic reference mapping: dominant topic and full topic-proportion vector

No.	Study (author, year)	Dominant	θ_0	θ_1	θ_2	θ_3	θ_4
1	Wang and Yu (2026) A Sensor-Based TinyML Acoustic Monitoring S...	T2	0.117	0.598	0.037	0.138	0.111
2	Escorcía et al. (2026) AIoT Ecosystem for Intelligent Water Qualit...	T5	0.137	0.318	0.038	0.058	0.449
3	Boudali and Khilil (2026) An intelligent data-driven framework for re...	T4	0.072	0.072	0.159	0.401	0.296

4	Li et al. (2026) Artificial intelligence of things for smart...	T5	0.297	0.130	0.033	0.089	0.452
5	Ali et al. (2026) Benchmarking Energy Efficiency of Supervise...	T1	0.534	0.026	0.036	0.031	0.373
6	Rai et al. (2026) Deep fusion convolutional neural network fo...	T4	0.084	0.029	0.044	0.436	0.407
7	di Torrepadula et al. (2026) Distilling knowledge for low-energy AIoT	T3	0.073	0.027	0.425	0.081	0.393
8	Nabil and Leghris (2026) EcoTomHybridNet: Policy-Guided Adaptive CNN...	T4	0.112	0.050	0.134	0.490	0.214
9	Moeinnia et al. (2026) LiteWeed: An affordable edge-AI robotic pla...	T4	0.090	0.030	0.040	0.485	0.356
10	Patrick et al. (2026) Review of movement sensor applications in l...	T1	0.655	0.045	0.072	0.022	0.205
11	Sami et al. (2026) SmartBioGas-PIoT: Physics-informed edge AI ...	T1	0.127	0.470	0.017	0.034	0.350
12	Attioui and Lahby (2026) Sustainable Traffic Congestion Forecasting ...	T5	0.104	0.332	0.062	0.055	0.447
13	Echchidmi and Bouayad (2026) TinyML for Sustainable Edge Intelligence: P...	T3	0.079	0.064	0.466	0.273	0.119
14	Li et al. (2026) WAGENet: A Hardware-Aware Lightweight Netwo...	T4	0.052	0.051	0.116	0.579	0.203
15	Komin and Boiko (2025) — A PRIVACY-PRESERVING EDGE DATA AGGREGATION ...	T5	0.099	0.180	0.045	0.020	0.656
16	Zhao et al. (2025) — Data Model Compression Method Based on Iter...	T5	0.055	0.055	0.388	0.038	0.464
17	Almtireen et al. (2025) — Deep transfer learning for sustainable wast...	T4	0.084	0.031	0.045	0.443	0.396
18	Koli et al. (2025) — Edge AI-Powered System Architecture for Alo...	T4	0.083	0.157	0.027	0.388	0.346
19	Mshrugi and Petri (2025) — Fast machine learning for building manageme...	T5	0.088	0.077	0.052	0.022	0.761
20	Batic et al. (2025) — Interpretability and reliability-driven kno...	T5	0.111	0.063	0.338	0.041	0.446
21	Zhang et al. (2025) — YOLOv8 forestry pest recognition based on i...	T4	0.053	0.022	0.082	0.476	0.368
22	Pazhanivel et al. (2024) — Design and Enhancement of a Fog-Enabled Air...	T1	0.037	0.586	0.026	0.016	0.335
23	Sun et al. (2026) An Analytical Model for Performance-Carbon ...	T3	0.062	0.039	0.500	0.014	0.384
24	Le et al. (2026) GLUSE: Enhanced Channelwise Adaptive Gated ...	T3	0.055	0.030	0.414	0.209	0.291
25	Khan et al. (2026) Model Compression for Sustainable AI in xG ...	T3	0.061	0.017	0.618	0.044	0.260
26	Zhang et al. (2026) TinyML-Enabled IoT Edge Framework With Know...	T4	0.051	0.021	0.287	0.369	0.272
27	Rojas-Hiedra et al. (2026) Toward Sustainable Mobility: A Comprehensiv...	T5	0.085	0.048	0.138	0.029	0.700
28	Yang et al. (2026) Transformer-Based Anomaly Behavior Analysis...	T5	0.093	0.090	0.164	0.038	0.615
29	Cerioli et al. (2025) Efficient Detection of Microplastics on Edg...	T5	0.083	0.133	0.245	0.061	0.477
30	Xan et al. (2024) Effective Edge Solution for Early Detection...	T5	0.070	0.046	0.113	0.353	0.418
31	Azizi et al. (2026) EDGE-AI EMPOWERED WASTE DETECTION: OPTIMIZI...	T4	0.035	0.021	0.049	0.597	0.298

Taken together, the descriptive, deductive, and exploratory topic-model results converge on a consistent picture. Descriptively, the corpus is a young, fragmented literature concentrated in agriculture and dominated by TinyML/edge-general approaches. Deductively (Section 3.2), the literature is near-uniform in reporting an accuracy-resource trade-off and a resource-constrained deployment context — both close to definitional of the review's eligibility criteria — but shows substantial, and theoretically consequential, variation in whether it discusses the consequentiality of the decisions its models support (55% explicit),

reflects on the AI system's own environmental footprint (32% explicit), validates performance under real field conditions (48% explicit), or situates itself within the conceptual Green AI/frugal-AI literature (19% explicit). The exploratory topic model (Section 3.3) corroborates this picture from an independent, inductive angle: the corpus organises lexically around how efficiency is engineered and reported (vision-classification, compression-method, and energy/hardware framings) at least as much as around which environmental domain is targeted, and the largest topic (T4, 38.3%) is characterised by comparatively generic efficiency/management language rather than deep domain or theoretical specificity. Collectively, these findings supply the empirical basis, developed further in the Discussion, for the review's central Frugal AI Paradox argument and for the proposed Frugal AI Environmental Efficacy (FAEE) framework's contention that current practice foregrounds whether compression preserves accuracy while under-examining whether that accuracy is sufficient for the decision-consequentiality of its application context.

4. Discussion

This review three objectives: to systematically map the literature at the intersection of frugal/compressed AI and environmental sustainability using PRISMA 2020. Following the PRISMA 2020 statement (Page et al., 2021), the search of Scopus and Web of Science returned 283 gross records (84 and 199 respectively), which reduced to 224 unique records after de-duplication, 49 recommended for full-text retrieval after title/abstract screening, and 31 included studies after full-text eligibility assessment. This gives the review a systematically documented, replicable evidence base at the frugal-AI/environmental-sustainability intersection, rather than a hand-picked or convenience sample.

In addition, objective 2 was likewise fully achieved, extended by a theory-driven deductive coding of seven a priori constructs across the full corpus and corroborated by an exploratory topic model. The corpus is young and fast-emerging (21 of 31 studies published in 2026 alone, the year the search itself was run), concentrated in agriculture and precision farming (29.0% of studies) and dominated methodologically by TinyML/edge-general approaches (48.4%) over any single named post-hoc compression operation, and scattered across 25 largely non-overlapping venues the profile of a field still forming (Girdhar et al., 2025) rather than one with an established reporting convention. The deductive coding produced the paper's central empirical finding: accuracy-degradation risk is not discussed uniformly across the corpus, and the specific properties that would let a reader assess it decision-consequentiality (C3, 54.8% explicit), the AI system's own footprint (C4, 32.3%), field/real-world validation (C6, 48.4%), and explicit theoretical grounding in the Green AI/frugal-AI literature (C7, 19.4%) are each reported in fewer than three in five studies, and in the case of C4 and C7, in fewer than one in three. The topic model's descriptive results are directionally consistent with this: the corpus's single largest, most lexically self-contained topic (T5, Energy/Power/Predictive Hardware Management, 38.3% of the corpus) is defined by generic efficiency-management vocabulary rather than deep domain or decision-context specificity, though this observation is not put forward as statistical corroboration of the coding result.

Finally, objective 3 was achieved as a conceptual synthesis, not as an empirical validation. The FAEE framework, introduced at the proposal stage, is retained and developed using the deductive-coding findings above as motivating evidence for why its Risky Frugality zone is a live concern but, as makes explicit, FAEE's two axes (compression aggressiveness and decision consequentiality) were not directly, quantitatively measured for each included study or for the corpus as a whole. Section 4 maps the C3/C4/C6/C7 reporting gaps onto what information each FAEE axis would require to assess a study, and shows that this information is often missing, which is a finding about the corpus's reporting practices, not a demonstration that any given study, or the corpus collectively, has been empirically placed in a specific FAEE zone.

Towards the Frugal AI Environmental Efficacy (FAEE) Framework: A Conceptual Synthesis

The study proposal introduced the Frugal AI Environmental Efficacy (FAEE) framework, which organises the accuracy-risk space along two axes compression aggressiveness and decision consequentiality and identifies Risky Frugality (high compression aggressiveness applied to a consequential environmental decision) as a critical, currently invisible zone: the zone a resource-constrained adopter is pushed into precisely because frugal AI is marketed to them as an affordable accessibility solution, without a reliable

way for a non-specialist to know whether their particular compressed model has quietly become unreliable. Figure 7 depicts the framework's four zones. FAEE is presented here as a conceptual framework motivated by, and elaborated using, the deductive-coding findings below not as an empirically validated typology. Neither compression aggressiveness nor decision consequentiality was directly, quantitatively measured for the 31 included studies in this review; no claim is made that any individual study, or the corpus as a whole, has been empirically located within a specific FAEE zone. What follows instead is an argument that the information needed to make that placement is often missing from the published record a gap-in-reporting finding, not a location finding.

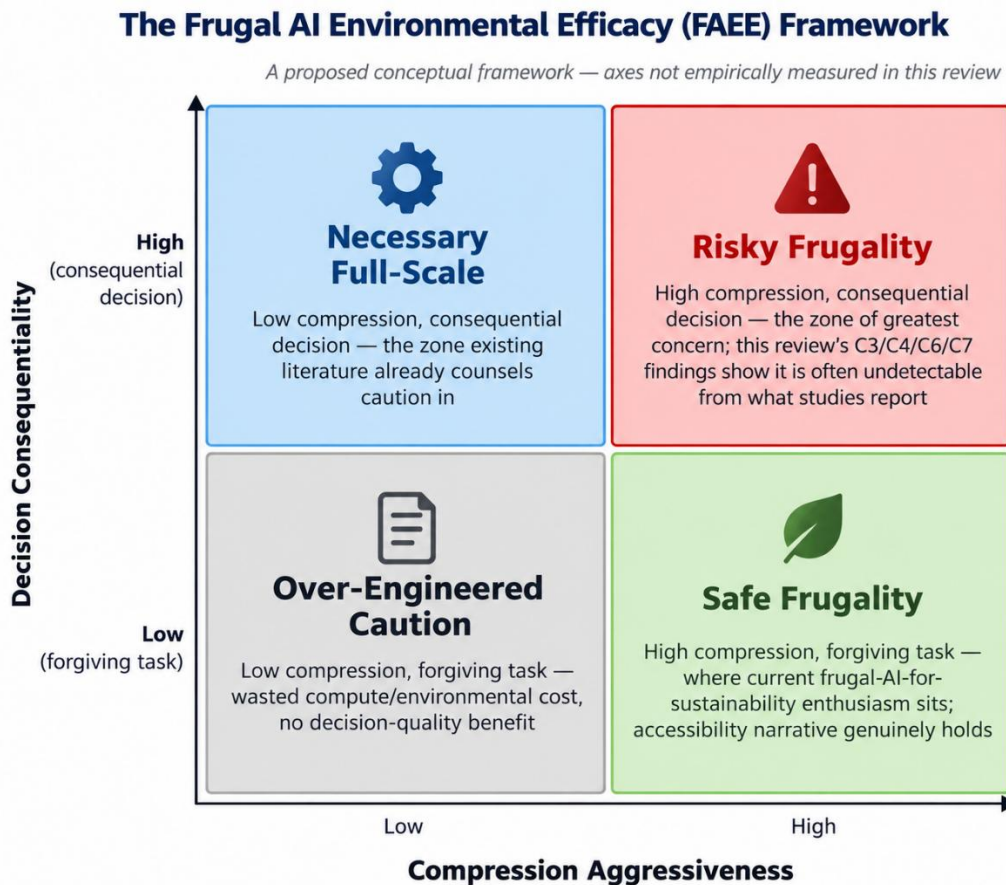


Figure 7. The Frugal AI Environmental Efficacy (FAEE) framework

Read against the FAEE axes, the four low-prevalence constructs are not four unrelated gaps; they are the specific pieces of information a study would need to report before anyone the authors, a reader, or a resource-constrained adopter could locate it on the FAEE grid at all. C1 and C2 place a study on the map's starting square (some accuracy-resource trade-off, in some resource-constrained context), but placement on the two axes that actually separate Safe Frugality from Risky Frugality requires more: C4 (own-footprint reflexivity, 32.3% explicit) is offered here as an indirect indicator of whether a study frames its own compression choice reflexively enough to situate it on the compression-aggressiveness axis, rather than a direct measurement of aggressiveness itself; C3 (decision-consequentiality, 54.8% explicit, 16.1% entirely absent) is treated as an indicator, not a measurement, of where a study sits on the consequentiality axis; and C6 (field/real-world validation, 48.4% explicit, 29.0% entirely absent) indicates whether the accuracy

figure being reported would plausibly hold up outside benchmark conditions in the deployment context that gives the consequentiality axis its meaning. None of C3, C4, or C6 was designed or validated as a measurement instrument for the FAEE axes; each is, at best, a proxy visible in what a paper chooses to report. A study silent on all three can still report an impressive compression ratio and a clean benchmark accuracy number and, on the evidence of C1/C2, nearly every study in this corpus does exactly that while remaining, on the published record, unlocatable: it is not possible to tell, from the paper alone, whether it describes Safe Frugality or Risky Frugality. C7 (explicit theoretical grounding in the Green AI/frugal-AI literature, 19.4%) compounds this: most of the corpus does not even use the conceptual vocabulary that would let a study recognise itself as needing to answer these questions in the first place.

A small, exploratory secondary analysis sharpens this picture further, with an important caveat: technique family (whether a study applies a named, post-hoc compression operation versus an architecture-level or general approach) is used below as a coarse, indirect proxy for compression aggressiveness, not a validated measure of it – a named compression operation is not necessarily more aggressive, in the sense FAEE intends, than an architecture designed for efficiency from the outset, and this review did not independently quantify aggressiveness (e.g., compression ratio or parameter reduction) for each study. With that caveat, grouping the 31 studies by whether their technique family is a named, post-hoc compression operation applied to an existing model (knowledge distillation, quantization, or pruning; $n = 11$) versus an architecture-level or general approach (TinyML/edge-general design, efficient/lightweight architecture, or a hybrid of multiple techniques; $n = 20$), decision-consequentiality reporting is essentially identical between the two groups (54.5% vs. 55.0% explicit) but field/real-world validation is not: only 27.3% of the named-compression subgroup report field validation, versus 60.0% of the architecture-level/general subgroup, and own-footprint reflexivity is somewhat lower in the named-compression subgroup as well (27.3% vs. 35.0%). This pattern offered descriptively, not as a statistically tested effect, given the modest and uneven subgroup sizes involved (individual technique cells range down to $n = 1$ –2; Table 4) and the coarseness of the technique-family proxy itself is consistent with, though it does not confirm, the possibility that studies using named post-hoc compression are, on this corpus's own evidence, less often subjected to the real-world check that would let a reader assess whether a given study is closer to Safe or Risky Frugality.

Table 4: Exploratory post hoc comparison

Technique-aggressiveness group	n	C3 explicit (consequentiality)	C4 explicit (footprint)	C6 explicit (field validation)	C7 explicit (theory)
Named post-hoc compression (knowledge distillation, quantization, pruning)	11	54.5%	27.3%	27.3%	18.2%
Architecture-level / general (TinyML/edge-general, efficient/lightweight design, hybrid)	20	55.0%	35.0%	60.0%	20.0%

Situating the Findings in Existing Literature

The corpus's near silence on the AI system's own environmental footprint (C4, 32.3%) is a gap of translation rather than a gap of available knowledge. Concern about the energy and carbon cost of machine learning models in their own right is well established elsewhere in the field Strubell, Ganesh and McCallum (2019) first drew sustained attention to it for large NLP models, and Schwartz et al.'s (2020) Green AI agenda explicitly called for reporting efficiency alongside accuracy as a standard, not optional, practice. That this

norm has not propagated into the frugal-AI-for-environment literature, despite that literature's explicit efficiency orientation, echoes Girdhar et al.'s (2025) observation that frugal AI remains conceptually fragmented from the broader Green AI research programme a fragmentation this review's C7 finding (19.4% explicit theoretical grounding) makes directly visible at the level of individual studies rather than only at the level of the field's self-description.

The review's field-validation finding (C6, 48.4% explicit; and the sharper named-compression-subgroup gap in Table 4) is consistent with, and helps explain the practical stakes of, Hooker et al.'s (2020) technical finding that compression does not degrade a model's accuracy uniformly across its input distribution: pruned and quantized networks disproportionately lose accuracy on a small, specific subset of harder or rarer examples, a pattern invisible in an aggregate benchmark accuracy figure and detectable only through targeted, often field-realistic, stress-testing. A compressed model can therefore report benchmark accuracy statistically indistinguishable from its full-scale baseline while performing markedly worse on precisely the atypical cases a rare disease presentation in a crop-health image, an unusual noise signature in a bioacoustic recording, an edge-case emissions reading most likely to carry real decision consequences. Read together with Paula et al.'s (2025) finding, already motivating this review's proposal, that some compressed architectures lose over 30 percentage points of accuracy under aggressive quantization while others lose almost none, this technical literature gives a concrete mechanism for why C6's field-validation gap is not a minor reporting omission: without it, exactly the failure mode Hooker et al. (2020) and Paula et al. (2025) describe is structurally undetectable from the paper alone.

The finding also speaks back to the AI-adoption literature this review's proposal positioned itself against. Technology-Organization-Environment and Resource-Based View theorising (Tornatzky & Fleischer, 1990; Barney, 1991), as applied to SME and resource-constrained-firm AI adoption (Cimino et al., 2025; Muhammad et al., 2025), examines adoption antecedents and firm-performance outcomes but, as the proposal argued, treats the adopted AI tool as a black box of uniform capability. This review's technical-literature-side findings now substantiate that critique empirically rather than only conceptually: the corpus of technical studies that such an adopting firm would draw on to select a frugal AI tool is itself largely silent on the very information own-footprint, decision-consequentiality, field-validated reliability an adoption decision would need. The black box is not simply an oversight of the adoption literature; it is, on this review's evidence, substantially a black box in the underlying technical literature the adoption literature assumes is legible.

Finally, the corpus's field-validation and footprint gaps carry a distributive dimension already flagged conceptually in the review's proposal via the algorithmic-environmental-justice literature (Mozaffari et al., 2026), which documents that AI trained on sparse Global South data systematically underperforms for the populations it is meant to serve. If as this review's evidence on named-compression techniques suggests the studies least likely to be field-validated are also disproportionately the ones offering the most aggressive, lowest-cost compression, then the adopters for whom frugal AI's affordability case is strongest are structurally the ones receiving the least assurance that the resulting model still works for their context. This is the same compounding-inequity mechanism Mozaffari et al. (2026) describe for training-data representation, now identified at the point of deployment-time compression choice rather than only at the point of data collection.

Theoretical Implications

- Extending adoption theory with a decision-quality-risk dimension. TOE and RBV accounts of AI adoption in resource-constrained firms (Tornatzky & Fleischer, 1990; Barney, 1991; Cimino et al., 2025; Muhammad et al., 2025) can be extended, as the FAEE framework proposes, with an explicit decision-quality-risk dimension not merely whether a firm adopts frugal AI, but whether the specific compression aggressiveness of the tool it can afford is matched to the consequentiality of the decision it is used for. This review's findings give that extension an empirical anchor: the technical literature such adoption decisions draw on does not, at present, reliably supply the information the extended framework requires.
- Repositioning Green AI reporting around consequentiality, not only aggregate cost. The Green AI agenda (Schwartz et al., 2020) has focused chiefly on reporting the computational and environmental cost of AI research itself. This review's findings suggest a complementary reporting dimension is needed specifically for AI applied to environmental decisions: not only how much a model costs to run, but how consequential the decisions it informs are, and whether its accuracy has been validated at a level of realism matched to that consequentiality. FAEE operationalises this as a second axis alongside compression aggressiveness rather than as a replacement for existing Green AI cost-reporting norms.
- A bridge between two literatures that, per the review's proposal, run in parallel. By coding the same corpus against both a technical property (compression aggressiveness, proxied here by technique family) and a decision-theoretic property (consequentiality, footprint-reflexivity, field validation), this review offers one of the first corpus-level empirical bridges between the technical model-compression literature and the environmental-AI-adoption literature, operationalising a link the proposal argued was, until now, asserted conceptually rather than demonstrated.

Practical and Policy Implications

For researchers publishing frugal-AI-for-environment work, the four low-prevalence constructs identified here (C3, C4, C6, C7) amount to a minimal reporting checklist that would let a reader locate a study on the FAEE grid: state explicitly how consequential an erroneous decision would be in the deployment context targeted; report the AI system's own energy/carbon cost, or explain why this is not applicable; validate the compressed model under field or field-realistic conditions rather than benchmark conditions alone, particularly when a named post-hoc compression technique (distillation, quantization, or pruning) has been applied; and situate the contribution explicitly within the Green AI/frugal-AI research programme rather than treating efficiency as a purely local engineering detail.

For practitioners and resource-constrained organisations evaluating a frugal AI tool for an environmental decision, this review's findings suggest a small set of due-diligence questions worth asking before adoption: has this model's accuracy been validated under conditions resembling my actual deployment context, not only a public benchmark; is the compression technique used a named, aggressive post-hoc operation (distillation, quantization, pruning) rather than an architecture designed for efficiency from the outset and if so, has real-world validation specifically been reported, since this review finds that reporting is markedly less common for that group; and how consequential would an error be for the decision this tool is meant to support. Absence of clear answers to these questions is not proof that a tool sits in the Risky Frugality zone, but it is, on this review's evidence, the modal condition of the published literature such tools are drawn from.

For journals, reviewers, and funders the findings support treating explicit decision-consequentiality statements and field-validation reporting as review criteria for frugal-AI-for-sustainability submissions, analogous to how reporting checklists have been adopted elsewhere in the AI literature for other under-reported properties. For standards bodies and procurement policymakers, the FAEE grid suggests a possible certification or labelling approach comparable in spirit to model cards but organised around environmental-decision consequentiality specifically that would let a resource-constrained adopter identify a tool's FAEE zone without needing to independently audit the underlying technical literature.

Limitations

- Small-sample subgroup analysis. The technique-aggressiveness cross-tabulation in Section 4.2 (Table 4) is exploratory and descriptive; the named-compression subgroup ($n = 11$) and its constituent technique cells (as low as $n = 1-2$) are too small to support a formal statistical test, and the pattern should be treated as suggestive.
- Corpus coverage. The search was restricted to Scopus- and Web of Science-indexed, predominantly English-language venues as of the 2026 search date, grey literature, non-English regional practice, and practitioner (non-peer-reviewed) implementations most directly relevant to SME adoption contexts are likely underrepresented. Search coverage itself is a further, unresolved limitation: the review has not yet run a diagnostic or expanded search to test whether the current Scopus/Web of Science strategy missed relevant venues or terms, and the corpus should be read with that open question in mind pending such a check.
- Cross-sectional snapshot. Given the corpus's sharp recency skew, this review is a snapshot of a field still forming; both the descriptive profile and the reporting gaps identified here are likely to shift as frugal-AI-for-environment research matures and reporting conventions develop.

Future Research Directions

- Primary, prospective empirical studies that directly test the FAEE grid: field trials comparing compressed and full-scale model decision outcomes across a deliberately varied range of decision-consequentiality levels, rather than relying on retrospective literature-reporting gaps as a proxy.
- Development and validation of a standardised frugal-AI-for-environment reporting instrument, building on the seven constructs coded here, potentially adopted as a submission checklist by journals and special issues in this area.
- Qualitative, interview-based research with SME and resource-constrained adopters to test whether Risky Frugality is experienced in practice that is, whether adopters actually encounter undetected accuracy degradation in consequential decisions rather than only inferred from publication-level reporting gaps.
- Extension of the algorithmic-environmental-justice literature (Mozaffari et al., 2026) to compression-induced inequity specifically, ideally with primary data on which resource-context adopters are being deployed into which compression-aggressiveness levels, for which decisions.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Data availability

Not applicable

Funding Declaration:

No funding used

Clinical Trial Number

Not Applicable

References

- 1) Ali, A., Qamar, R., Asif, R., & Hina, S. (2026). Benchmarking Energy Efficiency of Supervised Machine Learning Models on Multi-Domain Classification Datasets. Information. <https://doi.org/10.3390/info17070652>
- 2) Almtireen, N., Abuhejleh, A. A., Ryalat, M., Elmoaqet, H., & Al-refai, G. (2025). Deep transfer learning for sustainable waste management: Real-time waste segregation apparatus using a two-phase CNN framework. Journal of Supercomputing. <https://doi.org/10.1007/s11227-025-07931-2>
- 3) Attioui, M., & Lahby, M. (2026). Sustainable Traffic Congestion Forecasting Through Lightweight Explainable AI and TinyML Edge Deployment: A Casablanca Case Study. Sustainability (Switzerland). <https://doi.org/10.3390/su18094439>
- 4) Azizi, S., Bendahmane, A., & Tlemsani, R. (2026). Edge-AI empowered waste detection: Optimizing YOLOv11 for resource-limited devices. Journal of Engineering and Technology for Industrial Applications. <https://doi.org/10.5935/jetia.v12i59.3594>
- 5) Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120.
- 6) Batic, D., Tanoni, G., Principi, E., Stankovic, L., Stankovic, V., & Squartini, S. (2025). Interpretability and reliability-driven knowledge distillation for non-intrusive load monitoring on the edge. Expert Systems with Applications. <https://doi.org/10.1016/j.eswa.2025.128837>
- 7) Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022.

- 8) Boudali, I., & Khilil, E. H. (2026). An intelligent data-driven framework for real-time crop health monitoring using knowledge distillation. *Internet of Things (The Netherlands)*. <https://doi.org/10.1016/j.iot.2026.102041>
- 9) Cerioli, A., Petrosino, L., Sasso, D., Laroche, C., Piechowiak, T., Pezzarossa, L., Merone, M., Vollero, L., & Sabatini, A. (2025). Efficient Detection of Microplastics on Edge Devices With Tailored Compiler for TinyML Applications. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3567816>
- 10) Cimino, A., Corvello, V., Troise, C., Thomas, A., & Tani, M. (2025). Artificial intelligence adoption for sustainable growth in SMEs: An extended dynamic capability framework. *Corporate Social Responsibility and Environmental Management*, 32(5), 6120–6138.
- 11) di Torrepadula, F. R., Maisto, V., Cilaro, A., & Mazzocca, N. (2026). Distilling knowledge for low-energy AIoT. *Journal of Systems Architecture*. <https://doi.org/10.1016/j.sysarc.2026.103692>
- 12) Echchidmi, M., & Bouayad, A. (2026). TinyML for Sustainable Edge Intelligence: Practical Optimization Under Extreme Resource Constraints. *Technologies*. <https://doi.org/10.3390/technologies14040215>
- 13) Escorcia, G. R. C., Vera-Londoño, L., & Perez-Taborda, J. A. (2026). AIoT Ecosystem for Intelligent Water Quality Monitoring Through Edge Processing and Generative Artificial Intelligence. *Technologies*. <https://doi.org/10.3390/technologies14050296>
- 14) Girdhar, N., Raj, A., Sharma, D., Singh, V., Doucet, A., & Renz, M. (2025). A comprehensive review of frugal artificial intelligence: Challenges, applications, and the road to sustainable AI. *Soft Computing*, 29(13), 4823–4856.
- 15) Hooker, S., Courville, A., Clark, G., Dauphin, Y., & Frome, A. (2020). What do compressed deep neural networks forget? *arXiv:1911.05248*.
- 16) Khan, F. M. A., Hallaq, M., Abou-Zeid, H., Erak, O., Waqar, O., Hassan, S. A., Alhussein, O., & Hossain, E. (2026). Model Compression for Sustainable AI in xG Wireless Networks: Recent Advances, Challenges, and Future Directions. *IEEE Communications Surveys and Tutorials*. <https://doi.org/10.1109/COMST.2026.3682638>
- 17) Koli, S., Baloni, D., Shukla, S., Gehlot, A., Singh, R., Joshi, L. M., & Kumar, S. (2025). Edge AI-Powered System Architecture for Aloe Vera Plant Disease Detection. *SSRG International Journal of Electronics and Communication Engineering*. <https://doi.org/10.14445/23488549/IJECE-V12I5P128>
- 18) Komin, A., & Boiko, O. (2025). A privacy-preserving edge data aggregation for TinyML energy forecasting in households. *Technology Audit and Production Reserves*. <https://doi.org/10.15587/2706-5448.2025.339277>
- 19) Le, T. D., Ha, V. N., Nguyen, T. T., Tran, D. D., Hung, N. K., Garces-Socarras, L. M., Merlano-Duncan, J. C., & Chatzinotas, S. (2026). GLUSE: Enhanced Channelwise Adaptive Gated Linear Units SE for Onboard Satellite Earth Observation Image Classification. *IEEE Journal of Selected*

Topics in Applied Earth Observations and Remote Sensing.
<https://doi.org/10.1109/JSTARS.2026.3716862>

- 20) Li, L. L., Liu, C. Z., Cao, K. J., Zhou, Z. H., & Alrowaily, M. (2026). Artificial intelligence of things for smart agriculture with optimal embedded edge intelligence. *Computers & Electrical Engineering*. <https://doi.org/10.1016/j.compeleceng.2026.111039>
- 21) Li, Y. J., Huang, Y. Q., Lu, Y. C., Kuang, M. Q., Wu, Y. H., Guo, D. F., Fan, Z. Q., Yang, L., & Zhang, Y. X. (2026). WAGENet: A Hardware-Aware Lightweight Network for Real-Time Weed Identification on Low-Power Resource-Constrained MCUs. *Agriculture-Basel*. <https://doi.org/10.3390/agriculture16101086>
- 22) Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262–272). Association for Computational Linguistics.
- 23) Moeinnia, H., Armstrong, D., Angelozzi, J., Theys, G., Rezaeian, K., Wen, Y., Scott, D., Sulle, E., & Kim, W. S. (2026). LiteWeed: An affordable edge-AI robotic platform for laser weed management in resource-constrained environments. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2026.102340>
- 24) Mozaffari, A., Duarte, A., Teckentrup, L., Materia, S., Charnley, G. E. C., Palma, L., Baulenas Serra, E., Bojovic, D., Checchia, P., Carreric, A., & Doblas-Reyes, F. (2026). The rise of AI in weather and climate information and its impact on global inequality. *arXiv*. <https://arxiv.org/abs/2603.05710>
- 25) Mshragi, M., & Petri, I. (2025). Fast machine learning for building management systems. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-025-11226-6>
- 26) Muhammad, S. S., Dey, B. L., Kamal, M. M., Samuel, L., & Alzeiby, E. A. (2025). Digital transformation or digital divide? SMEs' use of AI during global crisis. *Technological Forecasting and Social Change*, 217, Article 124184.
- 27) Nabil, O., & Leghris, C. (2026). EcoTomHybridNet: Policy-Guided Adaptive CNN-Transformer Inference for Resource-Aware Edge-Based Tomato Leaf Disease Classification. *Future Internet*. <https://doi.org/10.3390/fi18050271>
- 28) Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71.
- 29) Patrick, B., Kanjo, E., & Kaiwartya, O. (2026). Review of movement sensor applications in livestock animal activity recognition: Communications, data collection practices, and edge-AI solutions. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.atech.2026.101986>

- 30) Paula, E., Soni, J., Upadhyay, H., & Lagos, L. (2025). Comparative analysis of model compression techniques for achieving carbon efficient AI. *Scientific Reports*, 15, Article 23461.
- 31) Pazhanivel, D. B., Velu, A. N., & Palaniappan, B. S. (2024). Design and Enhancement of a Fog-Enabled Air Quality Monitoring and Prediction System: An Optimized Lightweight Deep Learning Model for a Smart Fog Environmental Gateway. *Sensors*. <https://doi.org/10.3390/s24155069>
- 32) Rai, M., Kesarwani, A., & Gangwar, M. (2026). Deep fusion convolutional neural network for accurate and carbon efficient traffic sign recognition. *Discover Computing*. <https://doi.org/10.1007/s10791-026-10246-2>
- 33) Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, 91(2), 1–40.
- 34) Rojas-Hiedra, D., Espinoza-Andaluz, M., Barzola-Monteses, J., & Guarochico-Moreira, V. (2026). Toward Sustainable Mobility: A Comprehensive Review of EV Battery Optimization via AI and Hardware Management. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2026.3710122>
- 35) Sami, F. K., Nisa, Z. S., Bijoy, M. H. I., Ahmed, F., Arefin, M. S., Dhar, P. K., & Shimamura, T. (2026). SmartBioGas-PIoT: Physics-informed edge AI and TinyML for real-time monitoring and methane rate prediction in anaerobic digesters. *Results in Engineering*. <https://doi.org/10.1016/j.rineng.2026.110657>
- 36) Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63.
- 37) Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645–3650). Association for Computational Linguistics.
- 38) Sun, J., Yi, X., Symons, A., Gielen, G., Eeckhout, L., & Verhelst, M. (2026). An Analytical Model for Performance-Carbon Co-Optimization of Edge AI Accelerators. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. <https://doi.org/10.1109/TCAD.2025.3602746>
- 39) Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington Books.
- 40) Wang, Z., & Yu, G. (2026). A Sensor-Based TinyML Acoustic Monitoring System for Edge-Side Animal Sound Recognition on Resource-Constrained Microcontrollers. *Sensors*. <https://doi.org/10.3390/s26133972>
- 41) Xan, C. J., Nugroho, H., Eswaran, S., & Siang, T. F. (2024). Effective Edge Solution for Early Detection of Rice Disease on ARM- M Microcontroller. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3470970>

- 42) Yang, Z. L., Shi, Q. X., Shao, S. C., & Ranganathan, P. (2026). Transformer-Based Anomaly Behavior Analysis for CPS in Water Infrastructure. IEEE Access. <https://doi.org/10.1109/ACCESS.2026.3698479>
- 43) Zhang, L. N., Yu, S. P., Yang, B., Zhao, S., Huang, Z. Y., Yang, Z. Y., & Yu, H. L. (2025). YOLOv8 forestry pest recognition based on improved re-parametric convolution. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2025.1552853>
- 44) Zhang, Y., Lu, Y., Martinez-Rau, L. S., Fan, Z., Qiu, Q., O'Flynn, B., & Bader, S. (2026). TinyML-Enabled IoT Edge Framework With Knowledge Distillation for Weed Classification. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/JIOT.2026.3679508>
- 45) Zhao, S., Li, S., Gao, P., Ge, S., Zhang, Z., & Shu, L. (2025). Data Model Compression Method Based on Iterative Pruning and Distillation for Edge Non-Intrusive Load Monitoring. *Journal of Electrical Engineering and Technology*. <https://doi.org/10.1007/s42835-025-02355-7>