

Bridging Governance and Sustainability

Evidence-Based Standards for Deploying Agentic AI Systems in Real-World Contexts

PRISMA 2020 Systematic Literature Review

Chinyere Okechukwu

PhD Student, Systems Engineering & Informatics
University of Hradec Kralove, Czech Republic

Supervisory Committee: Prof. Edison Pignaton De Freitas (Halmstad) · Prof. Dante A. C. Barone (UFRGS)
Prof. Wilk Oliveira (Tampere) · Prof. Pavel Bachman (UHK)

Research Problem & the Gap

Study Aim:

To systematically synthesize the empirical literature on agentic AI architecture, deployment, and governance (2015–2026) to develop a validated, sustainability-weighted taxonomy, benchmarking suite, and deployment playbook — bridging agentic AI's technical capability with evidence-based, responsible governance.

Governance guidance for agentic AI is presence BUT it is fragmented, fast-multiplying, empirically unvalidated and no cross-domain synthesis

What changed

- Since late 2025, multiple governance instruments have shipped: Singapore IMDA's Model AI Governance Framework for Agentic AI (Jan 2026), NIST's AI Agent Standards Initiative (Feb 2026), and EU AI Act Art. 14/15 obligations extended to autonomous agents.
- Papagiannidis, Mikalef & Conboy (2025, J. Strategic Information Systems) published a peer-reviewed responsible-AI-governance review and framework.
- Enterprise playbooks (IBM, WEF, Databricks 'State of AI Agents 2026') now exist — but remain vendor- or sector-specific, not empirically benchmarked.

What remains genuinely open

- No harmonized, cross-domain definition of "agentic AI" — surveys still conflate symbolic and neural lineages (Abou Ali & Dornaika, 2025).
- No validated, multi-dimensional benchmarking suite that weighs sustainability alongside technical performance.
- No systematic, evidence-based synthesis connecting deployment outcomes (success/failure) to the governance instruments above.
- This is the corrected gap the SLR now targets: not absence of governance discourse, but absence of empirical, cross-domain synthesis of it.

Gap → Contribution → Bridge

Existing work answers pieces of each research question in isolation. The contribution is the synthesis and empirical validation across the pieces.

	The Gap	Our Contribution	How It Bridges the Gap
RQ1 Architecture	Three competing taxonomies (Abou Ali & Dornaika; Khapre et al.; Sapkota et al.) with no shared vocabulary, and none map onto the language regulators use.	A taxonomy translating technical architecture into the data-flow / decision-authority language regulators and enterprises already use.	Turns RQ1 findings into terms RQ3 governance tools can actually reference.
RQ2 Applications	Sector-specific case evidence (finance, supply chain, government) with no cross-domain comparison of success/failure factors.	A cross-domain synthesis identifying which deployment factors generalize versus stay sector-specific.	Supplies the empirical evidence base the Deployment Playbook stands on.
RQ3 Governance	Parallel, non-interoperable instruments (IMDA, EU AI Act, NIST, one academic framework); prescriptive, not empirically tested; sustainability absent as a criterion.	Governance & Definition Taxonomy + 5-dimension Benchmarking Suite (sustainability-weighted) + Evidence-Based Deployment Playbook.	First framework to test governance guidance against real outcomes and formally weight sustainability.

Research Questions

1

Architecture

What are the primary definitional characteristics and architectural paradigms distinguishing Agentic AI from traditional and generative systems?

2

Applications

What real-world deployment contexts and sector-specific implementation patterns are observed in science and industry, and which factors separate successful from failed deployments?

3

Governance

What frameworks, standards, and benchmarking approaches currently exist to audit safety, alignment, robustness, and sustainability — and how empirically validated are they?

Theoretical Framework

No single theory captures Agentic AI's technical, governance, diffusion, and sustainability dimensions at once — the review is framed by this integrative six-layer model instead.

Layer	Theory (key citations)	What it explains
Architectural frame	Sociotechnical Systems Theory (Trist & Bamforth, 1951; Bostrom & Heinen, 1977)	Architecture and human oversight must be coded together — grounds RQ1
Governance motivation	Agency Theory (Jensen & Meckling, 1976; Eisenhardt, 1989)	Why oversight is needed: principal-agent problem under information asymmetry
Governance design	Responsive Regulation + IRGC Risk Governance + RRI (Ayres & Braithwaite, 1992; Renn, 2008; Stilgoe et al., 2013)	What structurally adequate oversight looks like — feeds risk tiers in RQ3
Sector-variation explanation	TOE Framework + Institutional Theory (Tornatzky & Fleischer, 1990; DiMaggio & Powell, 1983)	Why governance maturity differs systematically by sector — grounds RQ2
Connective thread	Knowledge-Based View / SECI Model (Grant, 1996; Nonaka & Takeuchi, 1995)	Agents as knowledge-conversion systems — links to the review's own IM/KM lens
Sustainability value layer	Green IS/BAO + Triple Bottom Line + NRBV, tempered by Jevons Paradox (Melville, 2010; Elkington, 1997; Hart, 1995)	Environmental cost as a governance criterion, not an afterthought — grounds RQ3

Positioning Against Closely Related Reviews

Dimension	Abou Ali & Dornaika (2025) — PRISMA-guided SLR, arXiv:2510.25445	Khapre et al. (2026) multi-agent survey	This dissertation (PRISMA-guided SLR → Framework Development-sustainability-inclusive)
Core lens	Dual-paradigm taxonomy (Symbolic vs. Neural)	Six-dimension technical taxonomy (architecture, cognition, security, evaluation)	Information & knowledge-management lens: data flows, decision structures, ERP/enterprise integration
Governance treatment	Notes governance deficit as a finding	Covers security/safety alignment technically	Governance as a primary deliverable - a validated taxonomy and playbook, not a side finding
Sustainability	Not addressed	Not addressed	A scorecard grading agentic AI on 5 metrics: performance, safety, governance, sustainability, explainability
Output	Descriptive taxonomy + roadmap	Technical survey	Actionable: Governance & Definition Taxonomy, Benchmarking Suite, Evidence-Based Deployment Playbook

Search Strategy & PRISMA 2020 Process

Search string architecture

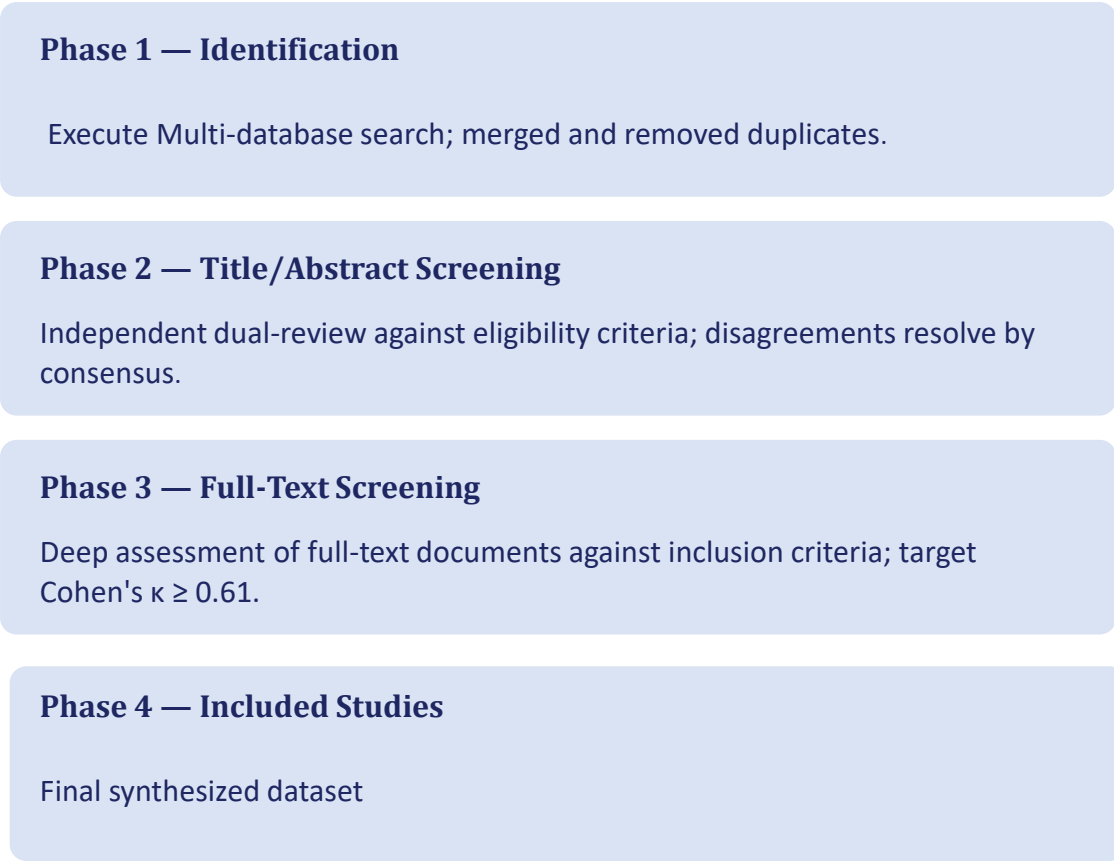
```
("agentic AI" OR "agentic artificial intelligence" OR "agent-based AI" OR "autonomous agents" OR "LLM-based agents")
AND
(architecture* OR framework* OR evaluation* OR application*
OR deploy* OR challenge* OR governance*)
```

Databases searched

Cluster	Sources
Technology & Engineering	IEEE Xplore
Multidisciplinary	Scopus, Web of Science
Applied & Open Access	MDPI, Taylor & Francis
Biomedical & Social	PubMed Central, SSRN

Quality proxy for preprints: ≥10 citations required for inclusion.

PRISMA 2020 flow



Search String Architecture — Blocks by Database

Base Capture (AND with every block below): “agentic AI” OR “AI agent(s)” OR “autonomous agent(s)” OR “LLM agent(s)” OR “large language model agent(s)”

Block(s)	Theme	Representative terms
1	Definitional & Architectural Core	architecture, framework, taxonomy, planning
2	Multi-Agent Systems & Orchestration	multi-agent, orchestration, coordination
3a–3i (9 blocks)	Sector Deployment	healthcare, finance, government, supply chain, cybersecurity, agriculture, pharmaceutical, software/IT, FMCG
4	Governance, Regulation & Standards	governance, AI Act, compliance, oversight, accountability
5	Benchmarking, Evaluation & Safety	benchmark, evaluation, red teaming, robustness
6	Sustainability & Environmental Cost	sustainability, energy, carbon, water footprint

14 blocks × 7 databases (IEEE Xplore, Scopus, WoS, MDPI, Taylor & Francis, PubMed/PMC, SSRN) — each block run and logged separately per database, then merged and deduplicated before screening. Field tags per platform (Scopus TITLE-ABS-KEY, WoS TS=, PubMed [tiab]).

Sample Search Strings — Same Logic, Platform-Specific Syntax

Block 6 — Sustainability & Environmental Cost (RQ3), shown across three platforms as a worked example. All 14 blocks × 7 databases are logged verbatim in the Search Log workbook (Exact Search Strings tab) for full reproducibility.

Scopus (TITLE-ABS-KEY field tag)

```
TITLE-ABS-KEY(("agentic AI" OR "AI agent" OR "AI agents" OR "autonomous agent" OR "autonomous agents" OR "LLM agent" OR "LLM agents" OR "large language model agent" OR "large language model agents") AND (sustainable OR sustainability OR "energy consumption" OR "carbon footprint" OR "environmental impact" OR "resource efficiency" OR "green AI"))
```

Web of Science (TS= field tag)

```
TS=(("agentic AI" OR "AI agent" OR "AI agents" OR "autonomous agent" OR "autonomous agents" OR "LLM agent" OR "LLM agents" OR "large language model agent" OR "large language model agents") AND (sustainable OR sustainability OR "energy consumption" OR "carbon footprint" OR "environmental impact" OR "resource efficiency" OR "green AI"))
```

PubMed / PMC ([tiab] field tag)

```
("agentic AI"[tiab] OR "AI agent"[tiab] OR "AI agents"[tiab] OR "autonomous agent"[tiab] OR "autonomous agents"[tiab] OR "LLM agent"[tiab] OR "LLM agents"[tiab] OR "large language model agent"[tiab] OR "large language model agents"[tiab]) AND (sustainable[tiab] OR sustainability[tiab] OR "energy consumption"[tiab] OR "carbon footprint"[tiab] OR "environmental impact"[tiab] OR "resource efficiency"[tiab] OR "green AI"[tiab])
```

Inclusion & Exclusion Criteria

Inclusion Criteria

- Empirical studies (Observational, case studies with reported outcomes)
- Simulation-based architecture blueprints
- Goal-driven, multi-step planning architectures
- Peer-reviewed articles, or credible preprints with ≥ 10 citations

Exclusion Criteria

- Agentic AI mentioned only as a passing example, not the study focus
- Studies unrelated to agentic or autonomous-agent systems
- Gray literature, blogs, news articles, unpublished manuscripts
- Non-English studies; secondary/tertiary reviews (tracked separately for discussion)
- Opinion pieces without empirical grounding
- Pure generative AI with no agentic component; rule-based chatbots

Note on terminology: earlier drafts of this criteria set used an undefined abbreviation ("EERs"). This has been corrected here to plain-language criteria for tonight's presentation.

Verified Reference Anchors (Preliminary Corpus)

13 real, DOI/arXiv-verifiable sources identified to date, mapped to each research question. Full corpus screening is ongoing.

RQ1 — Architecture

- Abou Ali & Dornaika (2025). Agentic AI: A Comprehensive Survey. [\[link\]](#)
- Khapre et al. (2026). Multi-agent architectures survey. [\[link\]](#)
- Sapkota, Roumeliotis & Karkee (2025). AI Agents vs. Agentic AI. [\[link\]](#)

RQ2 — Applications

- Zhao et al. (2026). ESG Intelligence Agent & Benchmark. [\[link\]](#)
- Gosmar et al. (2025). Agentic AI Sustainability Assessment, Supply Chain. [\[link\]](#)
- WEF (2026). Agentic AI Government Readiness Framework. [\[link\]](#)

RQ3 — Governance

- Papagiannidis, Mikalef & Conboy (2025). Responsible AI governance review. [\[link\]](#)
- Singapore IMDA (2026). Model AI Governance Framework for Agentic AI. [\[link\]](#)
- EU AI Act, Regulation (EU) 2024/1689, Art. 14 & 15. [\[link\]](#)
- AGILE Index 2025. [\[link\]](#)
- Paduraru et al. (2026). Trace-Based Assurance Framework. [\[link\]](#)
- IBM (2026). Agentic AI Governance Playbook. [\[link\]](#)
- Brookings/CMU/Berkeley (Oct 2025). How can we best evaluate agentic AI? [\[link\]](#)

Expected Research Outcomes

Governance & Definition Taxonomy

A peer-validated taxonomy reconciling fragmented governance instruments (IMDA, EU AI Act, NIST) with academic definitions of agentic AI — answers RQ1 & RQ2.

Robust Benchmarking Suite

A simple scorecard grading agentic AI systems on 5 metrics — (1) Technical Performance, (2) Safety & Reliability, (3) Governance & Compliance, (4) Sustainability, (5) Explainability & Auditability — testing whether existing benchmarks (AgentBench, WebArena, GAIA-style) under-weight the last three. Answers RQ3.

Evidence-Based Deployment Playbook

Risk categories, failure modes, and operational boundaries for safe, loop-based execution, grounded in documented deployment outcomes rather than vendor claims — answers RQ3.

Status & Next Steps



Bridging to the GFDT Poster: A Shared Robust AI Agenda

This SLR and the companion GFDT poster (Okechukwu & Bureš, ISSS 2026 — Margaret Mead Memorial Award) are separate studies with one shared ask: robust, accountable AI. Below is where they meet.

	The GFDT Poster (this symposium)	This SLR (in progress)	How They Connect
1 Robustness	Names a Robustness Limit: lab-trained yield-forecasting and anomaly-detection models can fail in chaotic smallholder fields. Lists AI robustness benchmarking as a collaboration ask.	A 5-metric Robust Benchmarking Suite (Technical Performance, Safety & Reliability, Governance & Compliance, Sustainability, Explainability & Auditability) built to test agentic-AI systems, RQ3.	The SLR’s scorecard is the concrete instrument to operationalise the poster’s robustness ask — applying it to GFDT’s forecasting and anomaly-detection models once field data exists.
2 Data governance	L4 (Data & Knowledge) layer calls for FAIR+CARE data governance and federated, cross-regional data sharing — not yet formally classified.	Governance & Definition Taxonomy reconciling fragmented instruments (IMDA, EU AI Act, NIST) with academic definitions of agentic AI, RQ1–RQ2.	The same taxonomy can classify GFDT’s own autonomous components (irrigation/dosing decisions) against emerging regulatory categories before field deployment.
3 Near-term ELLIIT ask	Immediate priorities: evaluate the hybrid DES–ABM framework, benchmark model generalisation under climate/data uncertainty, prototype federated FAIR+CARE governance.	Immediate priorities: PROSPERO registration, full-text screening, and initial validation of the Robust Benchmarking Suite (Q3 2026–early 2027).	One shared ELLIIT collaboration: the SLR supplies the evaluation method, the GFDT poster supplies the real-world agri-food testbed to apply it to.

01 ABOUT ME

Chinyere Chidera Okechukwu

Information & Knowledge Management Researcher

- ❖ PhD Student, Systems Engineering & Informatics | University of Hradec Králové, Czech Republic
- ❖ Principal Investigator, SPEV Research Project 2026
- ❖ Visiting Scholar, ELLIIT Focus Period

RESEARCH PERSPECTIVE

Knowledge systems that support trustworthy AI, responsible data use and measurable sustainability outcomes.

AWARD

MARGARET MEAD MEMORIAL AWARD

ISSS 2026 · 70th Annual Meeting

Outstanding pre doctoral research in systems science

MY WORK AT THE ELLIIT FOCUS PERIOD

Robust AI, information governance and sustainability

Research and publications connecting technical reliability with responsible implementation in organisations and society.

02 CURRENT RESEARCH

Knowledge and information management in sustainability reporting

Robust AI and digital twins for Agriculture 5.0

Governance, benchmarking and responsible deployment of agentic AI

*Focus: transparent evidence, reproducible
evaluation and sustainability aware systems*

03 PUBLISHED WORKS

2026

Intelligent Database Systems for Sustainable Development: Real Time ESG Monitoring in African FMCG Operations

2026

From Black Box to Glass Box: Algorithmic Impression Management and Cultural Moderation in Enterprise AI

2026

AI as an Information and Knowledge Management Capability

2026

Energy Data Ecosystems for Sustainable Energy Mix Optimization

2026

Socio Technical Evolution toward Sustainable Agriculture 5.0

2025

AI in Multinational Corporations: Strategic Integration and Applications

2025

Digital Marketing Risks in Aging Populations

chinyere.okechukwu@uhk.cz

[linkedin.com/in/chinyere-chidera-okechukwu-mba-ca-04670a8a](https://www.linkedin.com/in/chinyere-chidera-okechukwu-mba-ca-04670a8a)



ORCID
0009 0008 8088 6193

ELLIIT · HALMSTAD 2026

Thank you