The Dark Side Of Artificial Intelligence: A Systematic Literature Review And Bibliometric Analysis

Sanjay Vaid*

**Research Scholar – IIM Sambalpur and IAE Bordeaux, France, Visiting faculty, IIFT.*

Abstract
Artificial intelligence is increasingly embedded in organizational decision-making, public administration, healthcare, marketing, education, scientific communication, and digital infrastructure. This expansion has produced a growing body of research on the dark side of AI, including ethical failure, accountability gaps, strategic risk, privacy loss, surveillance, misinformation, labour disruption, and environmental burden. This study presents a systematic literature review supported by bibliometric analysis. The review is framed through PICO logic and reported using PRISMA 2020 principles. The Scopus dataset and associated evidence sources identified 86 records across the review corpus, screened 81 records after deduplication, considered 32 original Scopus papers as the main bibliometric evidence base, and included 19 studies in the core qualitative synthesis, with 15 supplementary sustainability and infrastructure records used as thematic extension evidence. The bibliometric analysis shows accelerated annual scientific production after 2022, dispersed source outlets, uneven country-level production and citation influence, fragmented collaboration structures, and a thematic map organized around artificial intelligence, risk management, risk assessment, governance, and emerging socio-technical concerns. The qualitative synthesis identifies eight domains of dark-side AI: sustainability and infrastructure burden; governance and accountability; ethics and human interests; sector-specific adoption harms; strategic and organizational risk; privacy, surveillance, and datafication; generative AI controversy and information integrity; and computational-efficiency mitigation. The review further develops a conceptual framework in which AI lifecycle activities increase infrastructure and resource demand, which mediates environmental and climate outcomes, while mitigation strategies moderate the pathway and climate feedback loops can reinforce energy and cooling pressures.

Keywords: artificial intelligence; generative AI; dark side; systematic literature review; PRISMA 2020; PICO; bibliometric analysis; governance; ethics; sustainability; AI infrastructure

1. Introduction

Artificial intelligence (AI) has shifted from a specialist computational technology to a pervasive socio-technical infrastructure, with contemporary systems shaping decision-making, communication, public service delivery, scientific production, labour processes, marketing, healthcare, finance, and platform governance (Wirtz et al., 2020; Pantano et al., 2024; Wach et al., 2023). This transformation has produced substantial benefits, but it has also intensified scholarly concern regarding the "dark side of AI", a concept used in the reviewed literature to describe harmful, unintended, inadequately governed, or insufficiently accountable effects that accompany AI adoption (Wirtz et al., 2020; Agrawal and Singh, 2025; Crocco et al., 2026). These effects include algorithmic bias, privacy invasion, surveillance, misinformation, accountability failure, strategic misjudgement, labour displacement, environmental externalities, and the escalating resource demands of AI infrastructure (Bibri and Allam, 2022; Zecca et al., 2023; Guidi et al., 2024; Li et al., 2025).

The need for a systematic literature review is clear because dark-side AI research is both recent and fragmented across governance, ethics, strategy, sectoral adoption, information integrity, and sustainability debates (Wirtz et al., 2020; Wach et al., 2023; Pantano et al., 2024; Madanchian and Taherdoost, 2025). The bibliometric evidence considered in this review indicates that publication activity increased sharply after 2022, especially in the period from 2023 to 2026, which is consistent with the recent expansion of work on generative AI risk, organizational risk analytics, ethical governance, and environmental burden (Wach et al., 2023; Keong et al., 2026; Sposato et al., 2026; Herrera et al., 2025). At the same time, the evidence base is spread across diverse journals, conferences, and disciplinary locations; consequently, AI risk is not yet organized around a single theoretical tradition but is distributed across ethics, governance, strategic management, sustainability, information systems, public administration, healthcare, marketing, engineering, and environmental studies (Morabito, 2025; Bhatt and Jhawar, 2025; Demir and Demir, 2026; Khatib et al., 2025).

This review is guided by the following research question: How does the current Scopus-indexed and supplementary evidence characterize the dark side of artificial intelligence in relation to ethics, governance, strategic risk, sectoral harm, information integrity, and environmental sustainability? The objectives are to define the review problem using PICO logic, report

the evidence-selection process using PRISMA 2020 principles, provide a standalone bibliometric analysis, synthesize the main thematic findings, develop a conceptual framework, and identify future research directions for responsible and sustainable AI governance (Cochrane, 2025). This structure follows systematic-review reporting expectations in which eligibility logic, transparent screening, corpus description, thematic synthesis, and limitations should be made explicit rather than implied (Page et al., 2021; PRISMA, 2021; Cochrane, 2025)

2. Literature Review

2.1 Conceptualizing the Dark Side of AI

The literature reviewed in this study suggests that the dark side of AI cannot be reduced to technical error; rather, it is a multi-dimensional condition in which automated systems interact with institutions, markets, users, infrastructures, and regulatory environments (Wirtz et al., 2020; Wach et al., 2023; Crocco et al., 2026). Ethical risks arise when AI systems reproduce bias, obscure accountability, or undermine human agency, while governance risks appear when organizations deploy AI faster than they build auditability, transparency, human oversight, and contestability (Morabito, 2025; Madanchian and Taherdoost, 2025). Strategic risks arise when AI changes organizational decision-making, risk appetite, and competitive behaviour without adequate institutional learning (Trzaska and Sus, 2023; Wu et al., 2026).

The reviewed and supplementary materials reinforce this broader framing by emphasizing regulatory lag, opaque model behaviour, exclusion in training datasets, social manipulation, labour displacement, and ecological burden (Agrawal and Singh, 2025; Sposato et al., 2026; Bibri and Allam, 2022). These concerns justify treating AI as a socio-technical and ecological system because harms emerge from data practices, organizational incentives, institutional decisions, cloud-service infrastructures, and lifecycle resource demands (Wirtz et al., 2020; MIT GenAI Impact Lab, 2024; Bashir et al., 2025). It is also ecological because training, inference, storage, networking, cooling, semiconductor manufacture, and hardware replacement generate material impacts that are often hidden behind apparently immaterial digital services (Guidi et al., 2024; International Energy Agency, 2025; Li et al., 2025).

2.2 Ethics, Governance, Accountability, and Human Interests

A central strand of the literature concerns the ethical and governance dimensions of AI. Studies in this area examine fairness, explainability, accountability, transparency, responsibility, human rights, leadership, and the capacity of institutions to control AI deployment (Wirtz et al., 2020; Morabito, 2025; Madanchian and Taherdoost, 2025). The core problem is not only that AI systems may produce biased or inaccurate outputs, but also that organizations may be unable to explain how systems operate, identify who is responsible for decisions, or provide meaningful remedies to affected individuals (Agrawal and Singh, 2025; Sposato et al., 2026). The dark side therefore includes institutional weakness as much as technical limitation (Crocco et al., 2026).

Governance scholarship also highlights the gap between principles and implementation. Organizations frequently endorse responsible-AI values, but implementation requires concrete mechanisms such as documentation, risk registers, model-impact assessments, independent audits, data-governance rules, environmental disclosure, escalation pathways, human oversight, and post-deployment monitoring (Wirtz et al., 2020; Madanchian and Taherdoost, 2025; Balola, 2026). Without these mechanisms, ethical AI remains aspirational and does not prevent harm at scale, especially in public administration, digital markets, and service ecosystems where affected users may have limited ability to challenge algorithmic decisions (Bibri and Allam, 2022; van der Rest and Heidary, 2024; Sposato et al., 2026).

2.3 Strategic, Organizational, and Sector-Specific Harms

The literature identifies strategic and organizational harms across several contexts. In corporate settings, AI can alter risk-taking, investment decisions, management control, supply-chain risk assessment, and competitive strategy (Kramer et al., 2022; Trzaska and Sus, 2023; Wu et al., 2026). In public administration and smart-city contexts, AI can reshape service access, eligibility decisions, surveillance practices, infrastructure planning, and citizen-state relationships (Rodriguez and Edwards, 2019; Wirtz et al., 2020; Freitas et al., 2025). In healthcare and financial services, AI raises questions about safety, clinical accountability, data privacy, explainability, patient trust, and strategic decision support (Bhatt and Jhawar, 2025; Jagadhabi, 2025; Kasim et al., 2026). In marketing and digital platforms, AI enables personalization, dynamic pricing, behavioural targeting, and content generation, but these same capabilities may intensify manipulation, exclusion, or discriminatory outcomes (Pantano et al., 2024; van der Rest and Heidary, 2024; Demir and Demir, 2026).

Sector-specific evidence is important because the dark side of AI is not uniform. The same technology may generate different harms depending on domain, user vulnerability, legal context, data sensitivity, institutional oversight, and scale of deployment (Pantano et al., 2024; Bhatt and Jhawar, 2025; Balola, 2026). A systematic review must therefore synthesize both cross-cutting mechanisms and

sector-specific manifestations, including misinformation and scientific-image falsification, strategic-risk forecasting, and algorithmic decision support (Zecca et al., 2023; Xu, 2025; Wach et al., 2023).

2.4 Environmental and Climate Impacts of AI Infrastructure

The environmental literature expands the dark-side AI agenda from social and organizational harms to infrastructure and resource impacts. Training and serving AI models require electricity for computation, storage, networking, and cooling, and these requirements connect AI governance to energy systems, water systems, and hardware supply chains (MIT GenAI Impact Lab, 2024; Bashir et al., 2025; International Energy Agency, 2025). Quantitative evidence is important because the impact is not abstract: Reported estimates suggest that the lifecycle energy consumption of large language models may range from approximately 51,773 to 62,319 MWh, while the International Energy Agency reports that global data-centre electricity demand was about 415 TWh in 2024 and could rise toward 945 TWh by 2030 under rapid AI and digital-service growth (International Energy Agency, 2025). Carbon emissions depend on the electricity mix and data-centre location, while one large-model training run can be compared with substantial lifetime automobile emissions in earlier carbon-accounting literature (Bashir et al., 2025; MIT GenAI Impact Lab, 2024). Water impacts are also material: United States data-centre cooling is commonly discussed in the range of 200–300 billion gallons per year, and water footprints depend on cooling technology, climate, regional hydrology, and the water intensity of electricity generation (Guidi et al., 2024; Ren, 2023; Li et al., 2025; Khatib et al., 2025). Hardware impacts arise from semiconductor manufacturing, rare-earth extraction, server replacement, embodied carbon, and e-waste (Panazan and Gheorghe, 2025; Guidi et al., 2024).

Prompt-level and workload-level evidence demonstrates that individual AI interactions may appear small but become material when scaled across billions of queries and continuous enterprise integration. Benchmarking studies of LLM inference, prompt-level energy profiling, OECD-oriented water-footprint analysis, MIT climate-sustainability, and data-centre sustainability research all support the same conclusion: AI sustainability cannot be assessed solely by model accuracy or per-token efficiency (Elsworth et al., 2025; Husom et al., 2024; Jegham et al., 2025; Bashir et al., 2025). It must be assessed across the lifecycle of data collection, model training, inference, deployment, data-centre operation, cooling, hardware renewal, and end-of-life disposal. The values retained in this study should therefore be read as decision-relevant estimates rather than universal constants, because electricity, water, and carbon intensity vary by model architecture, utilization rate, grid mix, cooling system, and location.

Table 1. AI Infrastructure Pathways and Governance Implications

AI Infrastructure Pathway	Dark-Side Mechanism	Example / Alternative Use Argument	Governance Implication
Electricity demand	Training and inference increase power demand for accelerators, storage, networking, and high-availability infrastructure; reported evidence includes LLM lifecycle energy estimates of 51,773–62,319 MWh and global data-centre demand rising from approximately 415 TWh in 2024 toward 945 TWh by 2030 (International Energy Agency, 2025).	Electricity allocated to large-scale AI workloads may compete with alternative societal uses such as residential consumption, industrial production, healthcare facilities, public infrastructure, and electrification initiatives. Rising AI demand may place additional pressure on electricity grids, particularly in regions already facing energy constraints (International Energy Agency, 2025; Bashir et al., 2025).	Require energy disclosure, efficient model design, workload scheduling, renewable-energy procurement, and demand-management strategies (Jiruwala, 2024; Channi & Kumar, 2024).
Carbon emissions	Emissions vary according to grid mix, data-centre location, utilization rates, and deployment frequency (Bashir et al., 2025; MIT GenAI Impact Lab, 2024).	AI-driven emissions may offset climate gains achieved through renewable-energy deployment and decarbonization efforts if growth in computational demand exceeds efficiency improvements. Carbon-intensive AI scaling may therefore create trade-offs with national climate targets (Bashir et	Apply carbon accounting across training and inference activities and align AI expansion with low-carbon energy strategies (Ajagekar & You, 2024; Channi

		al., 2025; International Energy Agency, 2025).	& Kumar, 2024).
Water and cooling	Heat loads can increase water withdrawals and consumption where evaporative or water-dependent cooling systems are used; evidence highlights a U.S. data-centre cooling burden of approximately 200–300 billion gallons annually (Guidi et al., 2024; Ren, 2023; Li et al., 2025).	Water consumed for AI cooling may compete with agricultural production, municipal water supplies, ecosystem preservation, and drought-management priorities, particularly in water-stressed regions (Ren, 2023; Herrera et al., 2025).	Adopt water-sensitive siting decisions, cooling-efficiency standards, water-reuse technologies, and local water-stress assessments (Khatib et al., 2025; Herrera et al., 2025).
Hardware lifecycle	Specialized AI chips and servers create upstream extraction, manufacturing, replacement, and e-waste burdens (Panazan & Gheorghe, 2025; Guidi et al., 2024).	Demand for advanced semiconductors and rare-earth materials may intensify resource extraction, increase geopolitical supply-chain dependencies, and accelerate electronic waste generation. These resources could alternatively support other digital, industrial, medical, or renewable-energy technologies (Panazan & Gheorghe, 2025; Guidi et al., 2024).	Apply lifecycle assessment, circular procurement, repairability, reuse, recycling, and responsible disposal practices (Rockström et al., 2009; Ryberg et al., 2018).
Rebound effects	Efficiency gains can reduce operating costs and encourage larger models, greater adoption, and more frequent AI use, thereby increasing total resource demand (International Energy Agency, 2025; MIT GenAI Impact Lab, 2024).	Efficiency improvements do not automatically reduce environmental impacts because lower costs often stimulate additional demand. As a result, total electricity, water, and hardware consumption may continue to increase despite technological improvements (International Energy Agency, 2025; Ryberg et al., 2018).	Combine efficiency improvements with responsible scaling, proportionality assessments, and absolute-sustainability constraints (Rockström et al., 2009; Ryberg et al., 2018).

3. Methodology

3.1 Review Design

This study uses a systematic literature review design supported by bibliometric analysis. The review combines structured eligibility criteria, PRISMA 2020 reporting logic, descriptive bibliometric mapping, and qualitative thematic synthesis. The methodological purpose is to make the evidence-selection process transparent while allowing both quantitative mapping and interpretive synthesis of a rapidly developing field.

3.2 PICO Framework

The PICO framework was used to define the review problem and eligibility logic. Because this is not a clinical intervention review, PICO is adapted to a socio-technical and sustainability context. The population/problem covers AI, generative AI, AI-enabled organizational systems, public systems, and AI infrastructure. The interest/exposure is dark-side AI, including ethics, governance, strategic risk, sustainability burden, energy demand, carbon footprint, water use, and infrastructure impact. The comparator/context includes traditional governance, organizational risk management, classical computing methods, and the absence of responsible-AI controls. The outcomes include harms, risks, governance gaps, sustainability impacts, energy-efficiency implications, and mitigation requirements.

Table 2. PICO Framework

PICO element	Operational definition in this review
Population / Problem	AI, generative AI, AI-enabled organizational and public systems, and AI infrastructure such as data centres.
Interest / Exposure	Dark-side AI exposure, including ethics, governance, strategic risk, sustainability burden, energy demand, carbon footprint, and infrastructure environmental impact.
Comparator / Context	Traditional governance, organizational risk management, classical AI computing methods, or absence of responsible AI controls.
Outcome	Risks, harms, sustainability impacts, energy-efficiency implications, governance gaps, accountability issues, and mitigation requirements.

3.3 Search Strategy and Evidence Sources

The main bibliographic evidence base was a Scopus export dataset for the review. The search strategy was anchored in the query logic: ("Artificial Intelligence" OR "Generative AI") AND TITLE("sustainability" OR ethics OR governance OR risk*) AND TITLE-ABS-KEY("dark side" OR "strategic risks"). The Scopus dataset was supplemented with additional sustainability, lifecycle assessment, quantum computing, prompt-level resource-use, and bibliometric sources. These supplementary materials were incorporated to enrich the interpretation of AI infrastructure burdens, environmental impacts, and mitigation pathways, rather than to replace the core Scopus-derived evidence base.

3.4 Screening, Eligibility, and Synthesis Rules

The screening process followed PRISMA 2020 logic. Records were assessed for relevance to AI or generative AI and for explicit connection to dark-side outcomes, sustainability, governance, ethics, risk, or strategic consequences. Duplicates and records without substantive relevance were excluded. All 32 original Scopus papers were retained for bibliometric description, while 19 papers were included in the core qualitative synthesis. Fifteen supplementary sustainability and infrastructure records were used as thematic extension evidence because they provided detail on electricity, water, carbon, cooling, lifecycle, and computational-efficiency pathways.

Table 3. Screening, Eligibility, and Synthesis Rules

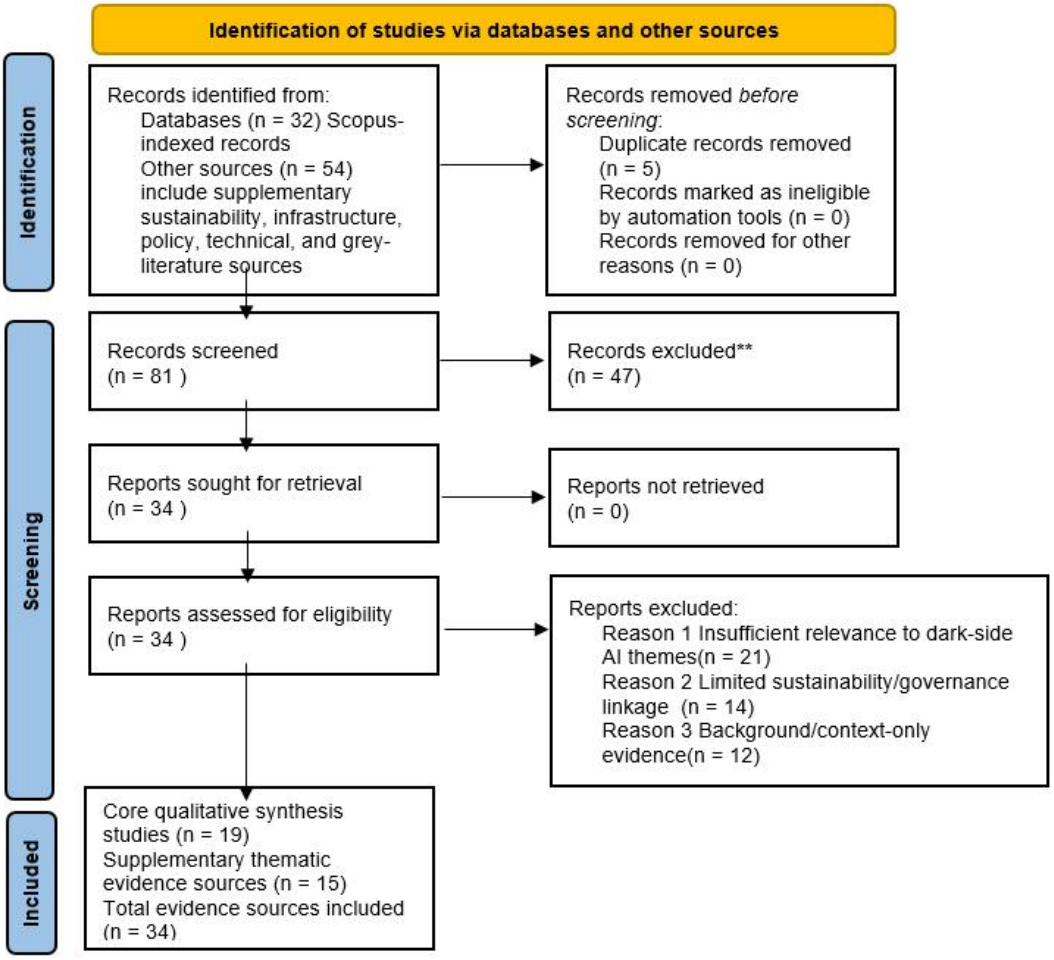
Screening component	Applied rule
Inclusion criteria	English-language Scopus records that addressed AI or generative AI and explicitly connected the technology to sustainability, ethics, governance, risk, dark-side effects, or strategic risks; records with sufficient bibliographic metadata; records relevant to at least one PICO outcome.
Exclusion criteria	Records outside the AI/generative AI domain; records lacking a substantive connection to dark-side outcomes; duplicate titles; records retained only as supplementary background when they did not meet the core qualitative synthesis threshold.
Synthesis rule	All 32 Scopus papers were considered for bibliometric description, while 19 were included in the core qualitative synthesis and 15 supplementary sustainability/infrastructure records were used as thematic extension evidence.

Table 4. Review Corpus Summary

PRISMA item	Count
Records Identified	86
Records Screened	81
Original Scopus Considered	32
Core Qualitative	19
Excluded From Core	13
Supplementary Thematic	15

Background Or Excluded	47

The PRISMA 2020 flow diagram is embedded below to make the evidence-selection pathway explicit. It reports 86 records identified, 81 records screened after deduplication, 32 original Scopus papers retained for bibliometric description, 19 studies retained in the core qualitative synthesis, and 15 supplementary sustainability and infrastructure records used for thematic extension evidence (Page et al., 2021; PRISMA, 2021).



**Studies without a strong "dark side" AI focus

Figure 1. PRISMA 2020 flow diagram for identification, screening, eligibility, and inclusion in the dark-side AI review.

3.5 Data Extraction, Quality Appraisal, and Synthesis

Data extraction captured author, year, title, source, document type, citation information, keywords, abstract-level concepts, country or affiliation information where available, and thematic relevance. Quality appraisal focused on conceptual relevance, transparency of the argument, empirical or theoretical contribution, and connection to at least one review outcome. The synthesis proceeded in three stages. First, descriptive bibliometric patterns were examined to characterize production, sources, citations, countries, collaboration, and themes. Second, qualitative coding grouped papers into eight thematic domains. Third, the thematic domains were integrated into a conceptual framework linking socio-organizational and environmental pathways of dark-side AI.

4. Bibliometric Analysis

This chapter presents the bibliometric analysis as a standalone component of the systematic literature review. The figures are ordered according to standard bibliometric categories: annual production, citation dynamics, sources and Bradford dispersion, country production and influence, collaboration structure, keyword and thematic structure, and factorial or conceptual structure. VOSviewer-style network outputs and Biblioshiny-style descriptive outputs are interpreted as complementary evidence. The bibliometric findings should be interpreted as descriptive rather than causal, given the relatively recent nature of the research domain and the incomplete publication coverage for 2026.

4.1 Annual Scientific Production and Citation Evolution

Annual scientific production shows a clear increase after 2022. The pattern indicates that dark-side AI research accelerated alongside the public diffusion of generative AI and broader concern about AI governance. The citation curve should be interpreted cautiously because recent publications have had less time to accumulate citations, while some earlier or highly visible papers can dominate average citation values.

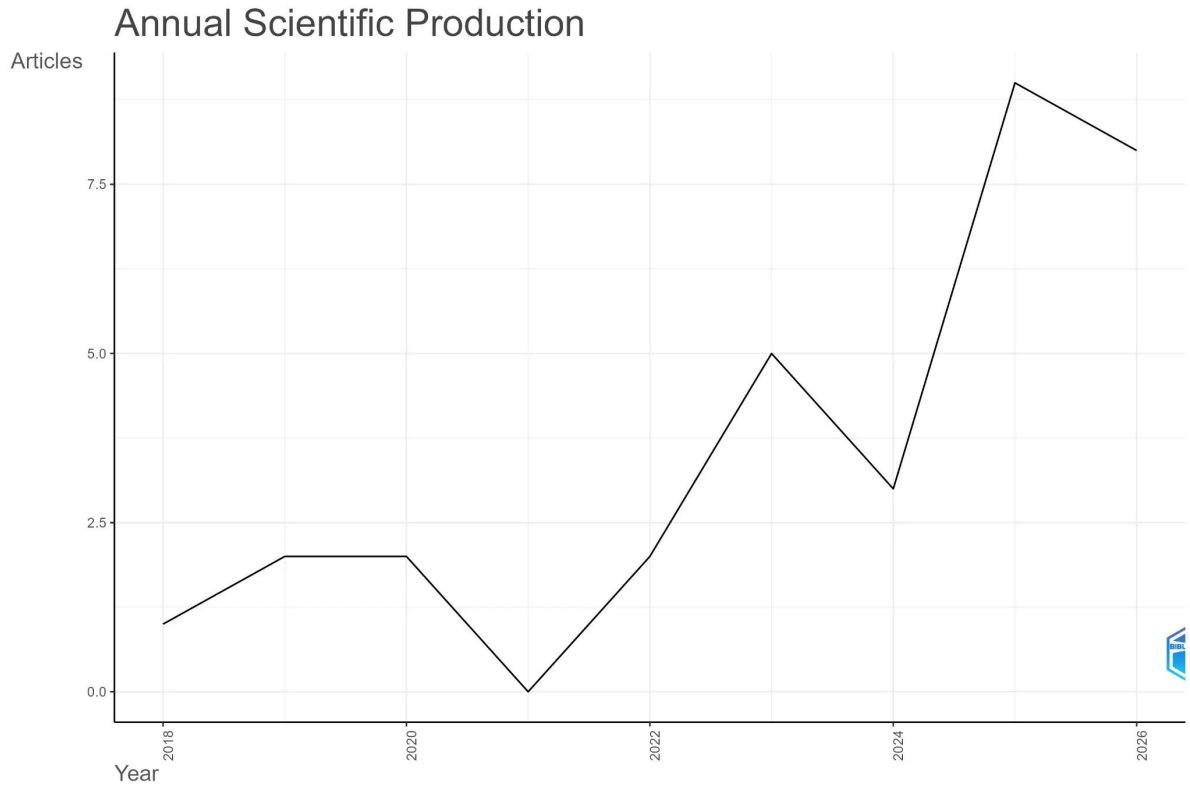


Figure 2. Annual scientific production in the Scopus-based dark-side AI corpus.

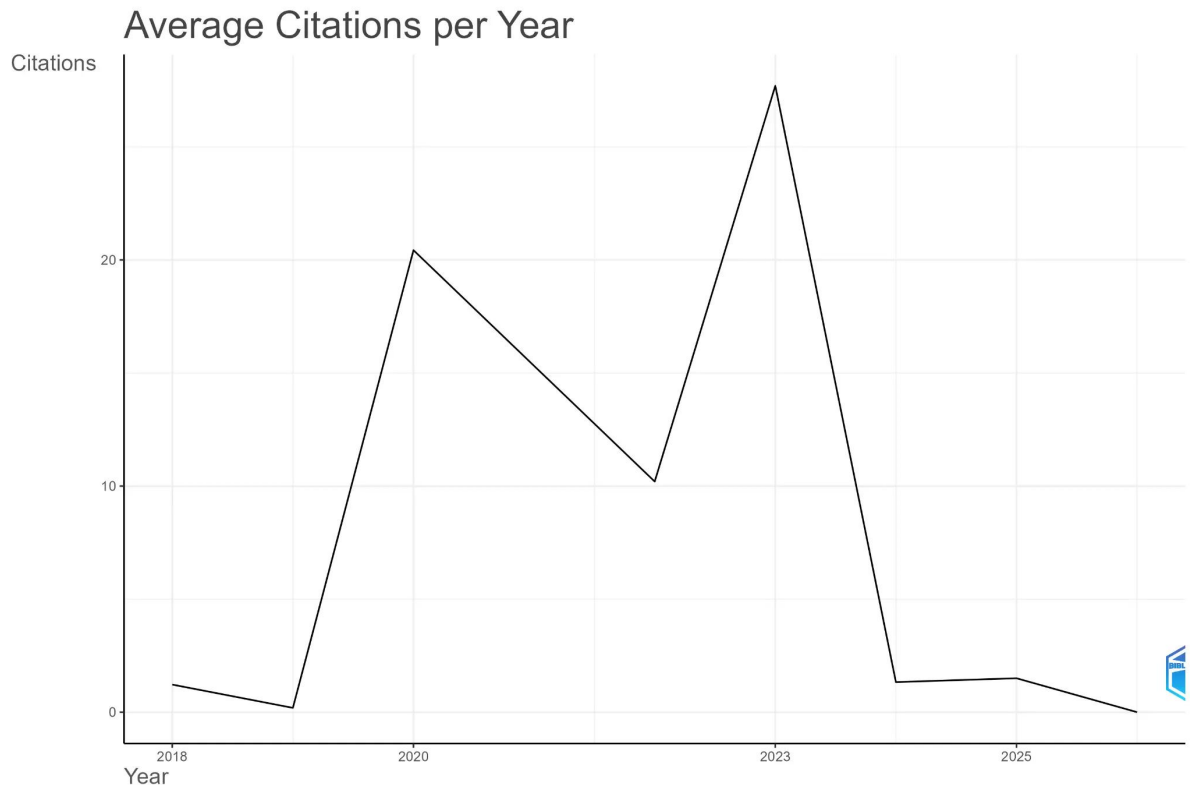


Figure 3. Average article citations per year, showing uneven citation influence across the emerging corpus.

4.2 Source Distribution and Bradford Dispersion

The source analysis shows that the literature is dispersed across many outlets rather than concentrated in one dominant journal. The most relevant sources each contribute a small number of records, and Bradford analysis indicates a broad core and peripheral distribution. This pattern is consistent with an interdisciplinary field spanning ethics, sustainability, management, computing, public administration, and sector-specific applications.

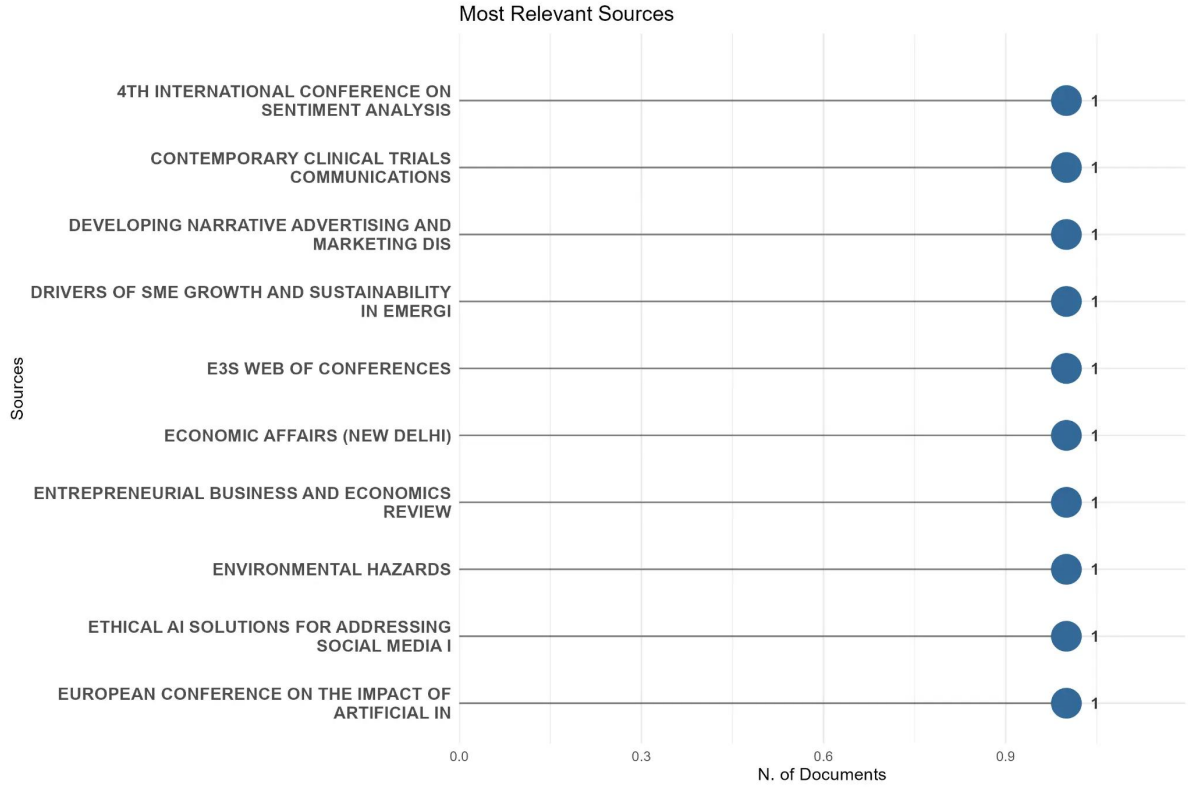


Figure 4. Most relevant sources represented in the bibliometric corpus.

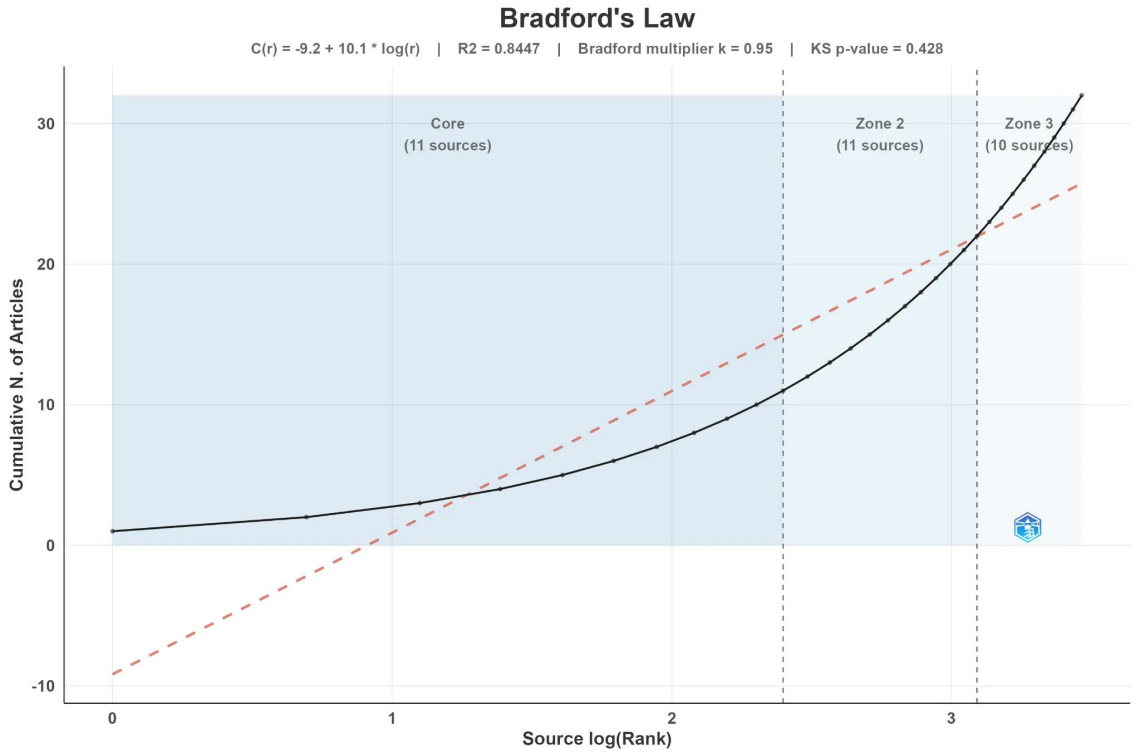


Figure 5. Bradford law visualization showing source dispersion across core and peripheral zones.

4.3 Country Production, Country Influence, and Collaboration

Country-level bibliometric analysis reveals a geographically diverse yet uneven distribution of scholarly output within the dark-side AI literature. The United Kingdom, Poland, Brazil, India, and China emerge as notable contributors to the research corpus, reflecting growing international interest in AI-related risks, governance, and sustainability concerns. Citation influence is similarly uneven, with Germany and Norway demonstrating comparatively higher citation impact, indicating a stronger influence on the intellectual development of the field. Furthermore, the country-collaboration maps reveal selective patterns of international cooperation rather than a highly integrated global research network, suggesting that scholarly collaboration remains fragmented across regions and disciplinary communities.

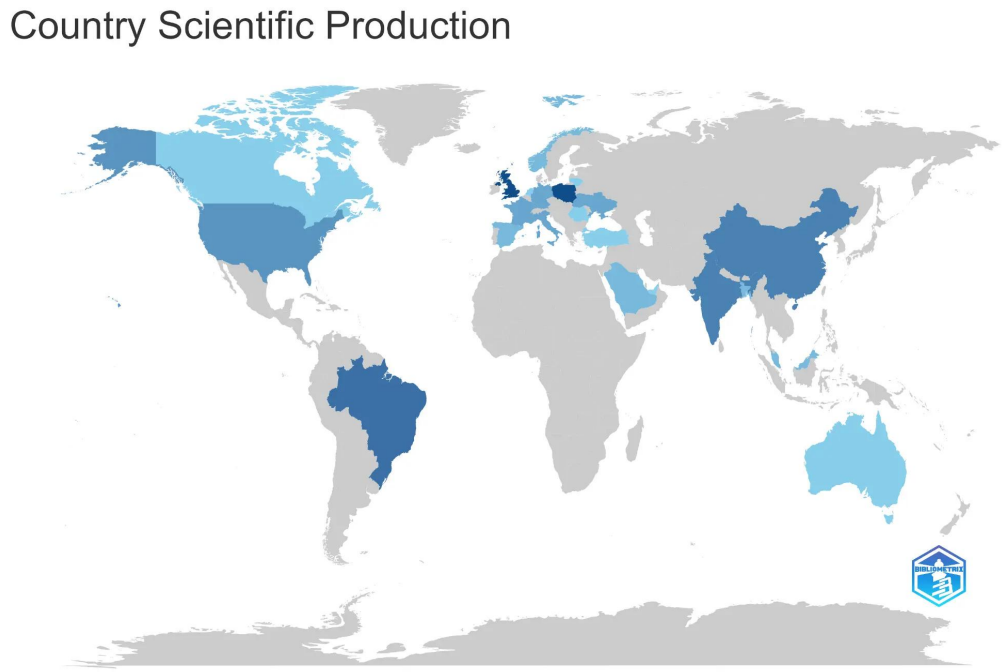


Figure 6. Country scientific production map for the dark-side AI corpus.

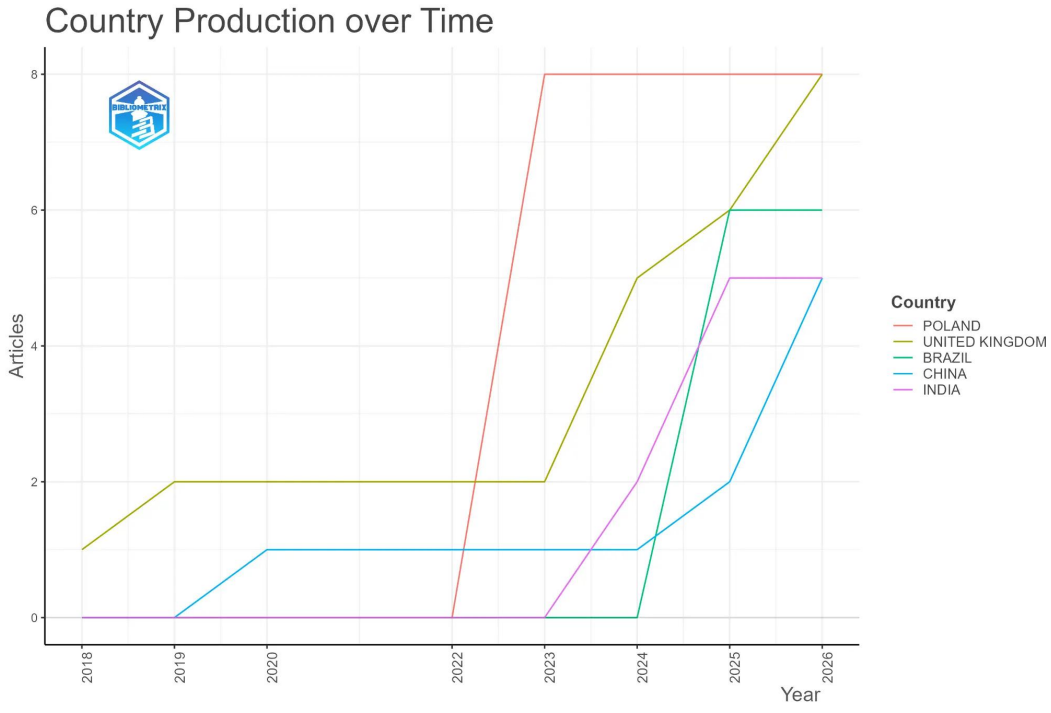


Figure 7. Country production over time for leading contributors.

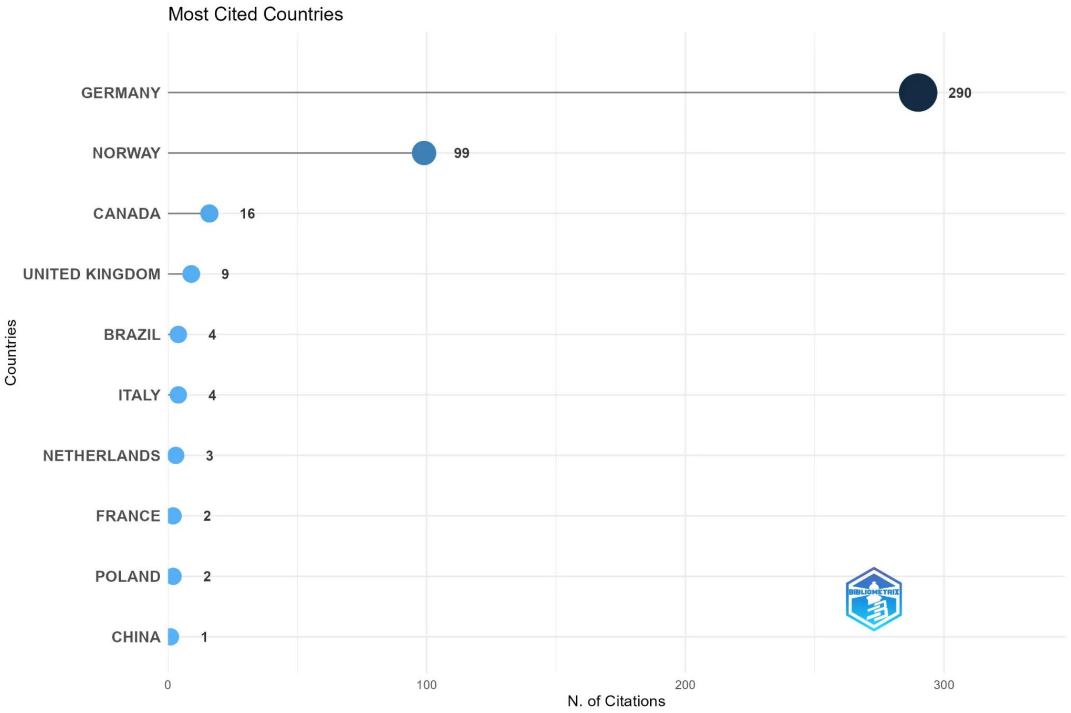


Figure 8. Most cited countries in the corpus.

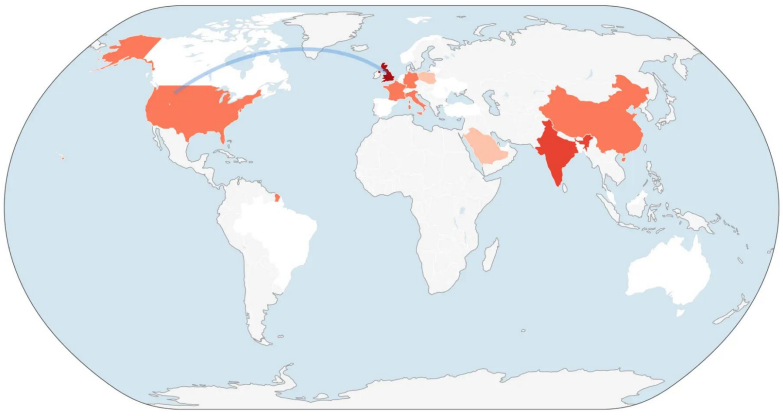


Figure 9. Country collaboration map showing international links in the corpus.

4.4 Author Collaboration and Co-Authorship Structure

The collaboration network indicates a fragmented author community. Multiple small clusters are visible, but there is limited evidence of one integrated co-authorship structure. This fragmentation is typical of an emerging interdisciplinary area in which scholars approach AI risk from different domains and methods. It also suggests that future research would benefit from cross-disciplinary teams connecting ethics, governance, sustainability, infrastructure engineering, and sectoral expertise.

Table 5. Author Collaboration Statistics

Corpus segment	Papers	Authors	Co-author links	Largest component	Authors per paper	Largest team	Isolated components
Core qualitative synthesis	19	55	84	3	2.89	8 authors	19
All 32 original Scopus papers considered	32	99	195	7	3.09	12 authors	32

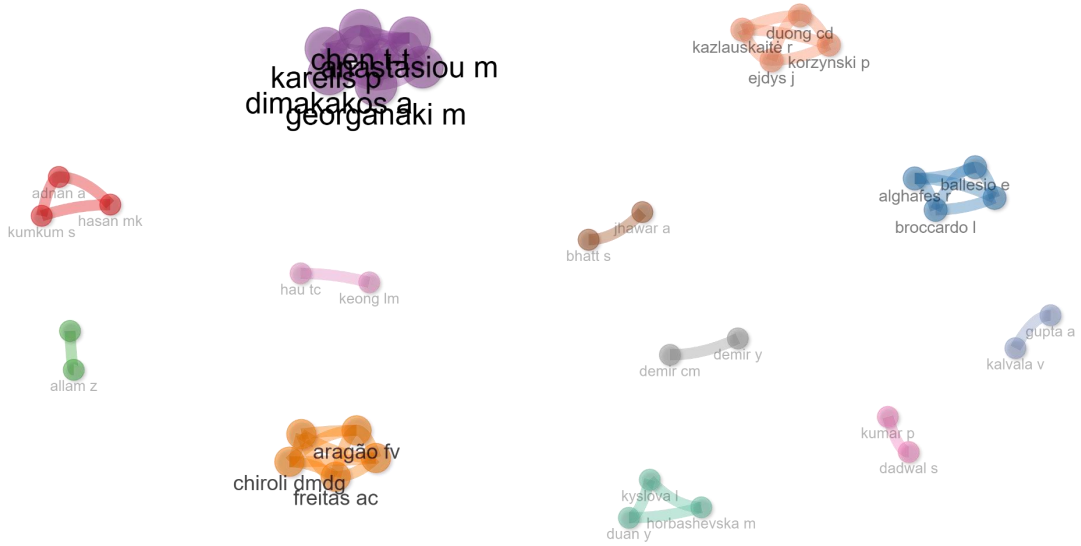


Figure 10. Collaboration network of authors in the bibliometric corpus.

4.5 Keyword Co-Occurrence, Thematic Map, and Conceptual Structure

Keyword patterns identify artificial intelligence, strategic risks, governance, risk management, ethics, and risk assessment as central terms. The thematic map positions artificial intelligence, risk management, and risks management as motor themes, while risk assessment, risk analysis, and decision support systems occupy a more basic-theme location. Governance and metaverse appear as emerging or declining themes, indicating that some subtopics remain weakly connected or underdeveloped in the corpus.

Table 6. Keyword Frequency and Co-occurrence Analysis

Keyword	Frequency
artificial intelligence	32
strategic risks	19
Governance	18
risk management	13
Ethics	10
risk assessment	10
algorithmic accountability	7
Healthcare	5
sustainability	5
Marketing	4
public administration	4

Table 7. Thematic Synthesis Results

Keyword pair	Co-occurrence strength
artificial intelligence — strategic risks	19
artificial intelligence — governance	18
artificial intelligence — risk management	13
risk management — strategic risks	11
artificial intelligence — ethics	10
artificial intelligence — risk assessment	10
governance — strategic risks	10
risk assessment — strategic risks	9

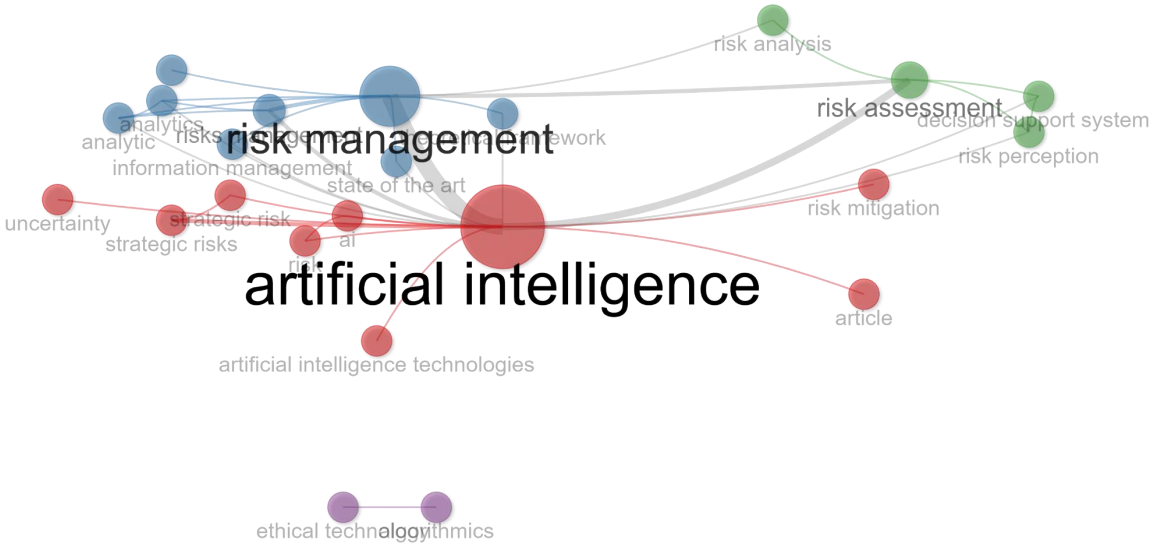


Figure 11. Keyword co-occurrence network for dark-side AI research.

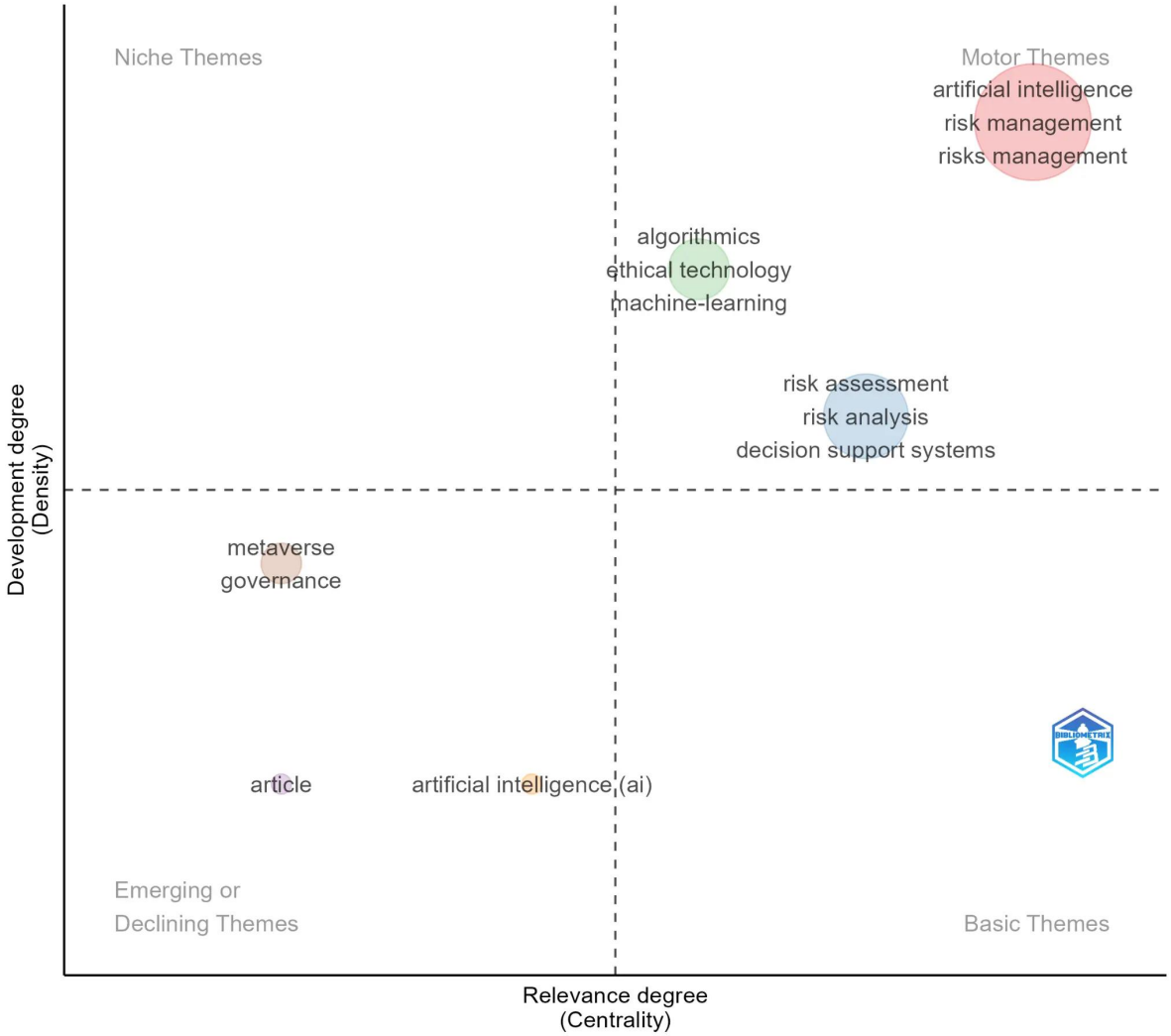


Figure 12. Thematic map showing motor, basic, niche, and emerging or declining themes.

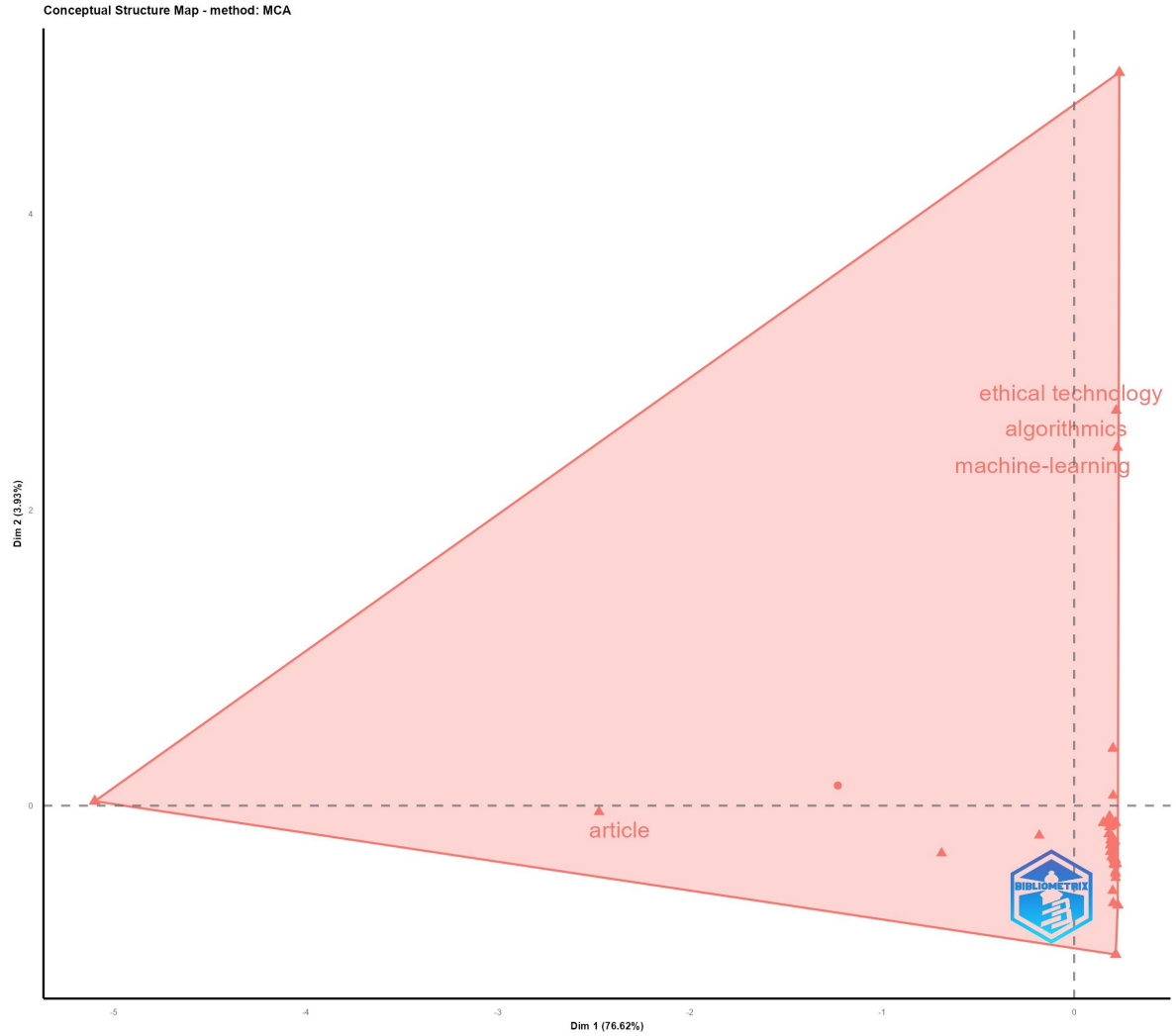


Figure 13. Factorial analysis map of conceptual proximity among bibliometric terms.

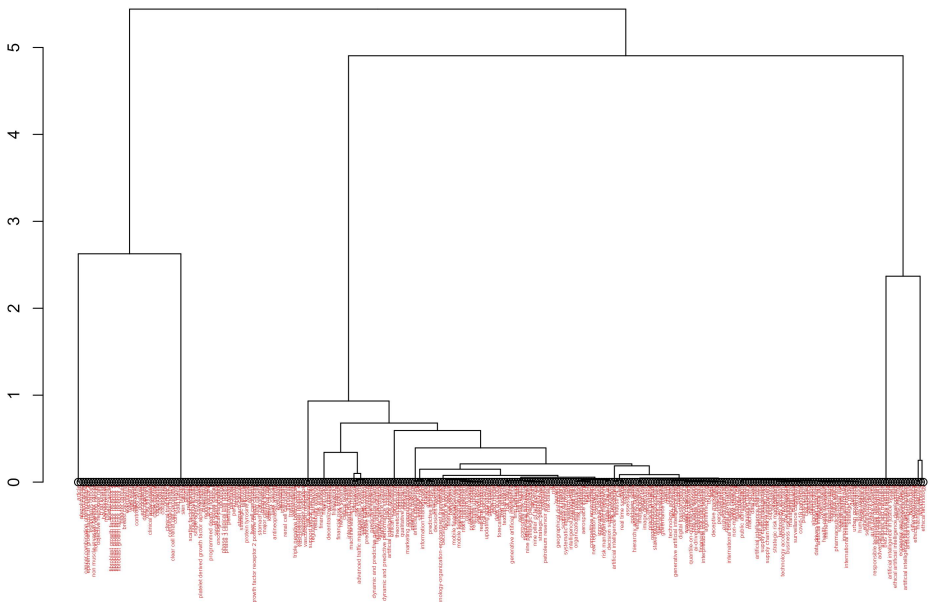


Figure 14. Factorial-analysis dendrogram showing clustered conceptual structure.

5. Findings

The thematic synthesis identifies eight domains of dark-side AI. These domains are not mutually exclusive. They interact across organizational, social, technical, and ecological levels. For example, governance weakness can increase privacy harms, strategic risk, and environmental burden when organizations deploy AI at scale without accountability or sustainability controls.

Theme	Number of supporting records	Interpretation
Sustainability, infrastructure, and energy burden	20	AI risk includes energy demand, data-centre impacts, carbon footprint, and environmental externalities.
AI governance and accountability	14	Dark-side AI is linked to governance gaps, accountability failures, algorithmic standards, and public-sector oversight.
Ethics, human interests, and leadership	11	Ethical risk concerns human welfare, responsible leadership, and moral limits on AI deployment.
Sector-specific adoption harms	8	Harms vary across public administration, healthcare, marketing, smart cities, digital markets, and service ecosystems.
Strategic and organizational risk	8	AI changes corporate risk-taking, financial decision support, strategic planning, and organizational resilience.
Privacy, surveillance, and datafication	6	AI-enabled systems intensify surveillance, profiling, data extraction, personalized pricing, and datafication.
Generative AI controversy and information integrity	4	Generative AI raises concerns about misinformation, falsified scientific images, ChatGPT controversies, and content integrity.
Quantum and computational efficiency as mitigation	2	Quantum and efficient computing appear as emerging mitigation routes for AI data-centre energy burdens.

5.1 Sustainability, Infrastructure, and Energy Burden

The first and most expansive theme is sustainability and infrastructure burden. The evidence shows that AI has material consequences through electricity demand, carbon emissions, water consumption, cooling requirements, hardware lifecycle, and e-waste. The quantitative values retained from the earlier evidence set are central to this interpretation: 51,773–62,319 MWh for large-model lifecycle energy, 415 TWh of global data-centre demand in 2024 with a possible increase toward 945 TWh by 2030, and 200–300 billion gallons per year associated with U.S. data-centre cooling. These values demonstrate why AI infrastructure must be studied as a physical system rather than as a purely digital service (International Energy Agency, 2025; Guidi et al., 2024; Ren, 2023; Li et al., 2025). The impacts are mediated by data-centre siting, grid mix, model scale, inference volume, cooling technology, and hardware turnover. The dark side of AI therefore includes the environmental cost of computation, not merely the social consequences of algorithmic decisions.

5.2 Governance, Accountability, and Ethics

The second theme combines governance, accountability, and ethics. The literature repeatedly identifies transparency, explainability, auditability, human oversight, and responsibility as necessary controls. However, the evidence also shows that these values are difficult to operationalize. AI systems may be technically complex, institutionally dispersed, or embedded in vendor platforms, making accountability difficult to assign. Effective governance must therefore operate across procurement, model design, deployment, monitoring, audit, and user remedy.

5.3 Sector-Specific and Strategic Risk

The third theme concerns sector-specific and strategic risk. AI alters decision-making in business, healthcare, public administration, education, marketing, and platform economies. In these settings, harms may arise from biased classification, unsafe recommendations, manipulative personalization, overreliance on automation, or misalignment between AI outputs and human judgment. Strategic risk also emerges when organizations treat AI adoption as inevitable and neglect proportionality, domain suitability, or long-term institutional consequences.

5.4 Privacy, Surveillance, Datafication, and Information Integrity

The fourth theme concerns privacy, surveillance, datafication, and information integrity. AI systems depend on data extraction and pattern detection. These capabilities can support useful personalization and prediction, but they can also enable monitoring, profiling, behavioural influence, discriminatory targeting, and erosion of informational autonomy. Generative AI adds a further layer by enabling synthetic text, images, and scientific or public-facing content that may be inaccurate, misleading, or difficult to verify.

5.5 Computational Efficiency, Quantum Approaches, and Mitigation

The fifth theme concerns mitigation through computational efficiency, renewable energy integration, cooling innovation, workload optimization, model compression, and quantum or hybrid quantum-classical approaches. The reviewed evidence suggests several potential pathways for reducing the environmental footprint of AI infrastructure. Quantum and hybrid quantum-classical approaches have been proposed as potential mechanisms for improving computational efficiency, with some studies indicating energy-reduction potential in the range of 12.5–30%. Similarly, advanced cooling technologies may substantially reduce data-centre energy requirements, while algorithmic optimization techniques, including model compression and pruning, can decrease unnecessary computational workloads. These estimates should be interpreted with caution, as realized benefits depend on factors such as workload characteristics, facility design, cooling systems, energy sources, baseline infrastructure efficiency, and the scale of deployment. Consequently, mitigation strategies should be evaluated within their specific operational and environmental contexts rather than as universally applicable performance benchmarks. (Ajagekar & You, 2024; Channi & Kumar, 2024; Khatib et al., 2025; Jiruwala, 2024). The literature also warns that efficiency alone is insufficient if it produces rebound effects. Sustainable AI requires absolute resource governance, lifecycle accounting, and limits on unnecessary scale as well as technical optimization.

6. Discussion

6.1 Integrating Socio-Organizational and Environmental Risk

The central contribution of this review is to integrate socio-organizational and environmental risk into one dark-side AI framework. The reviewed evidence shows that ethical, governance, strategic, informational, and ecological harms are connected. A system that is opaque and weakly governed may simultaneously produce biased outputs, increase surveillance, encourage over-automation, and expand computation without environmental accountability. Conversely, responsible AI governance must include both social safeguards and infrastructure safeguards.

6.2 Conceptual Framework: Environmental and Climate Impacts of AI

The proposed conceptual framework treats AI lifecycle activity as the independent variable. This includes data collection, model training, inference, deployment, and data-centre operations. These lifecycle activities increase infrastructure and resource demand through four mediating pathways: energy demand, water consumption and cooling, hardware lifecycle burden, and data-centre operations. The dependent outcomes are environmental and climate consequences, including carbon emissions, water stress, e-waste, resource depletion, and climate feedback effects. Mitigation strategies moderate the pathway by reducing energy intensity, cooling demand, hardware turnover, and inefficient computation. Such strategies include renewable energy integration, algorithm optimization, model compression, advanced cooling technologies, neuromorphic chips, and quantum or hybrid quantum-classical computing.

The framework also includes a reinforcing climate feedback loop. Warming can increase cooling demand, which can increase energy use when data centres operate in heat-stressed locations or rely on carbon-intensive electricity. The framework therefore positions AI sustainability as a dynamic system rather than a static footprint. This means that future AI governance should assess not only current resource use but also how scaling, climate conditions, grid constraints, cooling choices, and rebound effects may change the risk pathway over time.

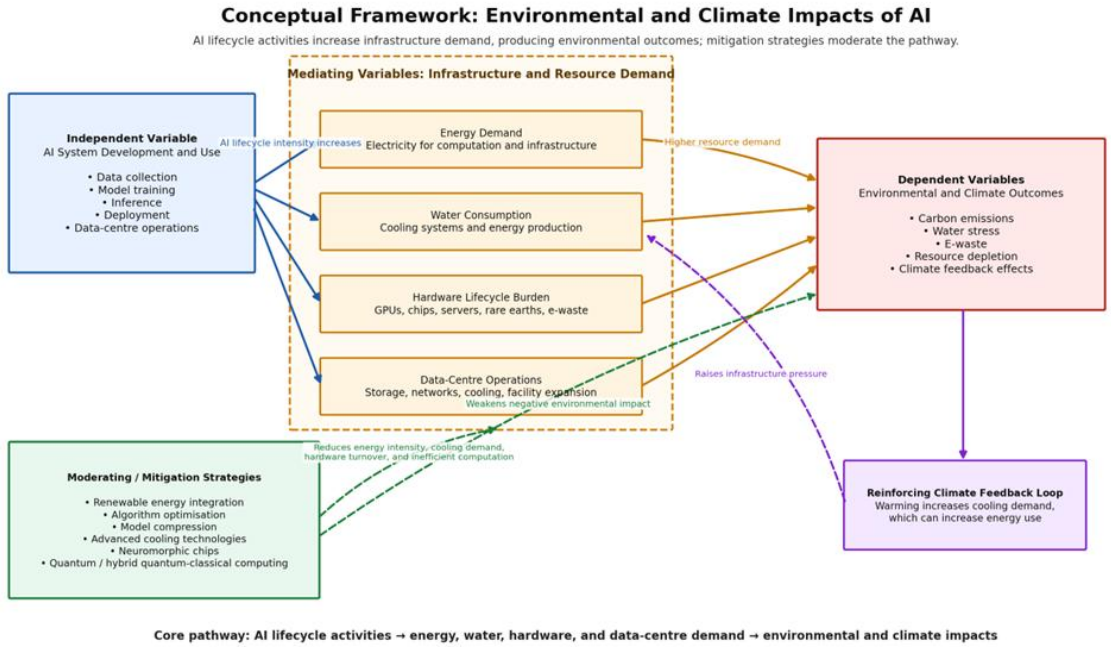


Figure 15. Conceptual framework for environmental and climate impacts of AI, linking AI lifecycle activities, mediating infrastructure demand, environmental outcomes, mitigation strategies, and reinforcing climate feedback loops.

6.3 Implications for Responsible AI Governance

The findings imply that responsible AI governance should be broadened. Standard concerns about fairness, transparency, privacy, and accountability remain essential, but they should be joined with sustainability disclosure, energy and water measurement, lifecycle assessment, hardware accountability, data-centre siting review, carbon-aware workload scheduling, and proportionality tests for high-intensity AI use. The numerical evidence strengthens that implication: when data-centre electricity demand is measured in hundreds of terawatt-hours, model energy in tens of thousands of megawatt-hours, and cooling-related water demand in hundreds of billions of gallons, governance cannot treat sustainability as a peripheral reporting issue. Governance should also distinguish between necessary and discretionary AI deployment. Not all AI use has equal social value, and sustainability governance requires evaluating whether computational costs are justified by public, organizational, or scientific benefit.

6.4 Limitations

This review has several limitations. First, the bibliometric evidence base was derived from a Scopus dataset comprising 32 original records, which may not fully capture all relevant publications available across other databases such as Web of Science, IEEE Xplore, ACM Digital Library, or PubMed. Second, bibliometric indicators should be interpreted descriptively because the field is relatively recent and publication data for 2026 remain incomplete. Third, keyword co-occurrence and thematic mapping rely on the quality and consistency of bibliographic metadata and may not fully reflect concepts discussed only within the full text of publications. Fourth, supplementary literature from sustainability, environmental studies, and quantum-computing research was consulted solely to support discussion, interpretation, and conceptual development. These sources were not part of the systematic review corpus and were excluded from all bibliometric and quantitative analyses. Finally, the findings reflect the literature available up to the date of data extraction. Publications indexed after the search cutoff date were not captured and therefore fall outside the scope of this review. Future studies may benefit from larger, multi-database datasets and longitudinal analyses as the field continues to evolve.

7. Conclusion and Future Research

This systematic literature review demonstrates that the dark side of AI is best understood as a multi-dimensional socio-technical and ecological risk field. The evidence shows rapid publication growth after 2022, dispersed publication sources, uneven country-level influence, fragmented collaboration networks, and a conceptual structure organized around artificial intelligence, risk management, governance, ethics, and risk assessment. The thematic synthesis identifies sustainability burden, governance and accountability, ethics and human interests, sector-specific harms, strategic risk, privacy and surveillance, generative AI information

integrity, and computational-efficiency mitigation as central domains.

The review concludes that responsible AI cannot be limited to model performance or technical fairness. It must also address institutional accountability, strategic risk governance, human interests, data power, information integrity, electricity demand, carbon emissions, water consumption, cooling, hardware lifecycle, rebound effects, and the environmental cost of computation. The proposed conceptual framework conceptualizes AI lifecycle activities as antecedents of infrastructure and resource demand, which act as mediating mechanisms linking AI development and deployment to environmental outcomes. Governance interventions and mitigation strategies are incorporated as moderating factors that influence the strength and direction of these relationships.

Future research should advance the field in six directions. First, researchers should build larger cross-database corpora using Scopus, Web of Science, IEEE Xplore, ACM Digital Library, PubMed, policy repositories, and preprint databases. Second, empirical studies should measure prompt-level and workload-level electricity, carbon, water, and hardware-lifecycle impacts across model sizes, data-centre locations, cooling systems, and grid mixes. Third, scholars should develop absolute-sustainability indicators that connect AI efficiency to planetary boundaries, water stress, mineral extraction, e-waste, and rebound effects. Fourth, quantum computing and hybrid quantum-classical AI should be evaluated as mitigation pathways under transparent lifecycle assumptions. Fifth, governance studies should test integrated mechanisms that combine ethical AI, environmental disclosure, lifecycle assessment, auditability, and accountability. Sixth, comparative international research should examine why dark-side AI scholarship and collaboration are concentrated in particular countries and how governance capacity differs across jurisdictions.

8. Bibliography

This bibliography separates methodological and supplementary sources from the original Scopus corpus references used for bibliometric traceability. The first subsection records PRISMA, PICO, sustainability, prompt-level resource, data-centre, quantum, and lifecycle sources. The second subsection preserves the original Scopus records considered in the review.

8.1 Methodological, Sustainability, Quantum, and Prompt-Level Sources

1. 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., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71
2. PRISMA. (2021). PRISMA 2020 statement. https://www.prisma-statement.org/prisma-2020-statement
3. Cochrane. (2025). Cochrane Handbook for Systematic Reviews of Interventions, Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-03
4. Cochrane Library. (n.d.). About PICO. https://www.cochranelibrary.com/about-pico
5. Ajagekar, A., & You, F. (2024). Variational Quantum Circuit Learning-Enabled Robust Optimization for AI Data Center Energy Control and Decarbonization. Advances in Applied Energy. https://doi.org/10.1016/j.adapen.2024.100179
6. Channi, H. K., & Kumar, P. (2024). Green Data Centers and Renewable Energy. Practice, Progress, and Proficiency in Sustainability, 161-186. https://doi.org/10.4018/979-8-3693-6567-0.ch009
7. Guidi, G. M., Dominici, F., Gilmour, J., Butler, K. A., Bell, E. F., Delaney, S., & Bargagli-Stoffi, F. J. (2024). Environmental Burden of United States Data Centers in the Artificial Intelligence Era. https://doi.org/10.48550/arxiv.2411.09786
8. Herrera, M., Xie, X., Menapace, A., Zanfei, A., & Brentan, B. (2025). Sustainable AI infrastructure: A scenario-based forecast of water footprint under uncertainty. https://doi.org/10.31223/x55m86
9. Jiruwala, A. (2024). The Power Paradox: A Review of the Challenges and Solutions to the Energy Efficiency of AI and Cloud Computing. International Journal of Engineering and Advanced Technology, 14(2), 11-18. https://doi.org/10.35940/ijeat.b4554.14021224
10. Khatib, L., Pham, A., Ahmed, K., & Frenkel, V. S. (2025). Data Centers and Water: Challenges and Solutions for Sustainable Cooling. Journal American Water Works Association, 117(8), 48-53. https://doi.org/10.1002/awwa.2504
11. Elsworth, D., et al. (2025). How much energy does Google's AI use? We did the math. Google Cloud Blog. https://cloud.google.com/blog/products/infrastructure/measuring-the-environmental-impact-of-ai-inference/
12. Ren, S. (2023). How much water does AI consume? OECD.AI Wonk.

https://oecd.ai/en/wonk/how-much-water-does-ai-consume
13. Li, P., Yang, J., Islam, M. A., & Ren, S. (2025). Making AI Less Thirsty: Uncovering and Addressing the Secret Water Footprint of AI Models. https://arxiv.org/abs/2304.03271
14. Jegham, N., et al. (2025). How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference. https://arxiv.org/abs/2505.09598
15. Husom, E. J., et al. (2024). The Price of a Prompt: Profiling Energy Use in Large Language Models Inference. https://arxiv.org/abs/2407.16893
16. Bashir, N., et al. (2025). Explained: Generative AI's environmental impact. MIT News. https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117
17. MIT GenAI Impact Lab. (2024). The Climate and Sustainability Implications of Generative AI. https://mit-genai.pubpub.org/pub/8ulgrckc/release/2
18. International Energy Agency. (2025). Energy and AI: Energy demand from AI. https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai
19. Rockström, J., et al. (2009). A safe operating space for humanity. Nature, 461, 472-475. https://doi.org/10.1038/461472a
20. Ryberg, M. W., Owsianiak, M., Richardson, K., & Hauschild, M. Z. (2018). Development of a life-cycle impact assessment methodology linked to the Planetary Boundaries framework. Ecological Indicators, 88, 250-262. https://doi.org/10.1016/j.ecolind.2017.12.065
21. Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., De Wit, C. A., Hughes, T., Van Der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., ... Foley, J. A. (2009). A safe operating space for humanity. Nature, 461(7263), 472–475. https://doi.org/10.1038/461472a
22. Santarius, T. (2012). Green growth unravelled: How rebound effects baffle sustainability targets when the economy keeps growing. Wuppertal Institute for Climate, Environment and Energy.
23. Richardson, K., Steffen, W., Lucht, W., Bendtsen, J., Cornell, S. E., Donges, J. F., Drüke, M., Fetzer, I., Bala, G., Von Bloh, W., Feulner, G., Fiedler, S., Gerten, D., Gleeson, T., Hofmann, M., Huiskamp, W., Kummu, M., Mohan, C., Nogués-Bravo, D., ... Rockström, J. (2023). Earth beyond six of nine planetary boundaries. Science Advances, 9(37), eadh2458. https://doi.org/10.1126/sciadv.adh2458
24. Freitag, C., Berners-Lee, M., Widdicks, K., Knowles, B., Blair, G. S., & Friday, A. (2021). The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. Patterns, 2(9), 100340. https://doi.org/10.1016/j.patter.2021.100340

8.2 Bibliography of Original Scopus Papers Considered

1. Agrawal A.; Singh J.. (2025). The dark side of AI: Risks, ethics, and safeguarding human interests. Ethical AI Solutions for Addressing Social Media Influence and Hate Speech. https://doi.org/10.4018/979-8-3693-9904-0.ch007
2. Balola Y.H.A.. (2026). Government as a Strategic Actor: A Framework for Assessing Regulatory Risk in Digital Markets. Thunderbird International Business Review. https://doi.org/10.1002/tie.70137
3. Bhatt S.; Jhawar A.. (2025). Navigating the dark side of metaverse adoption in healthcare: ethical challenges, risks, and strategic interventions. International Journal of Pharmaceutical and Healthcare Marketing. https://doi.org/10.1108/IJPHM-12-2024-0138
4. Bibri S.E.; Allam Z.. (2022). The Metaverse as a Virtual Form of Data-Driven Smart Urbanism: On Post-Pandemic Governance through the Prism of the Logic of Surveillance Capitalism. Smart Cities. https://doi.org/10.3390/smartcities5020037
5. Burdyak P.. (2023). Model of risk management in railway transport. E3S Web of Conferences. https://doi.org/10.1051/e3sconf/202340206017
6. Chen D.. (2020). The possibility of evolution from non-strategic risk to strategic risk: The role of strategic assets. IOP Conference Series: Materials Science and Engineering. https://doi.org/10.1088/1757-899X/806/1/012045
7. Crocco E.; Ballesio E.; Broccardo L.; Alghafes R.. (2026). Welcome to the dark side: unveiling the potential risks related to the implementation of Artificial Intelligence. Journal of Risk Research. https://doi.org/10.1080/13669877.2025.2611954
8. Dadwal S.; Kumar P.; Verma R.; Singh G.. (2024). Drivers of SME growth and sustainability in emerging markets. Drivers of SME Growth and Sustainability in Emerging Markets. https://doi.org/10.4018/979-8-3693-0111-1
9. Demir C.M.; Demir Y.. (2026). Limits and Risks: The Dark Side of AI and Ethical Quandaries in Developing Marketing Stories. Developing Narrative Advertising and Marketing Discourse in the AI Era. https://doi.org/10.4018/979-8-3373-5666-2.ch009
10. Duan Y.; Shuplat O.; Matsuka V.; Lukash S.; Horbashevska M.; Kyslova L.. (2023). Risk Management Strategy for International Investment Projects of an Innovative Enterprise in the Context of Industry 4.0. Economic Affairs

(New Delhi). https://doi.org/10.46852/0424-2513.4.2023.16
11. Freitas A.C.; Oliveira C.R.S.D.; Gomes P.F.D.O.; Chiroli D.M.D.G.; Gerônimo B.M.; Zola F.C.; Aragão F.V.. (2025). Risk management in smart cities: influence analysis using DEMATEL. Environmental Hazards. https://doi.org/10.1080/17477891.2025.2467637
12. Jagadhabi N.. (2025). Explainable AI-Driven Decision Support for Strategic Risk Management in Financial Services. International Academic Journal of Science and Engineering. https://doi.org/10.71086/IAJSE/V12I3/IAJSE1254
13. Kalvala V.; Gupta A.. (2025). Integrating Machine Learning and Statistical Models in Enterprise Risk Analysis. 4th International Conference on Sentiment Analysis and Deep Learning, ICSADL 2025 - Proceedings. https://doi.org/10.1109/ICSADL65848.2025.10933366
14. Kasim A.; Anastasiou M.; Loizou E.; Dimakakos A.; Georganaki M.; Mourelatos I.; Karelis P.; Esposito M.; Zhou H.; Chen T.-T.; Skaltsas D.; Stockman P.. (2026). Strategic risk assessment in oncology: Utilizing single-agent activity to boost combination therapy approvals. Contemporary Clinical Trials Communications. https://doi.org/10.1016/j.conctc.2025.101590
15. Keong L.M.; Ling C.P.; Seong C.M.; Hau T.C.. (2026). AI-Driven Predictive Analytics in Business Risk Management. Studies in Systems, Decision and Control. https://doi.org/10.1007/978-3-032-11612-3_53
16. Kramer K.J.; Mousavi D.; Schmidt M.. (2022). Strategic Supply Chain Risk Management Artificial Intelligence and Big Data to Support Strategic Supply Chain Risk Management; [Strategisches Supply-Chain-Risikomanagement Einsatz von Künstlicher Intelligenz und Big Data zur Unterstützung des strategischen Supply-Chain-Risikomanagements]. ZWF Zeitschrift fuer Wirtschaftlichen Fabrikbetrieb. https://doi.org/10.1515/zwf-2022-1055
17. Kumkum S.; Tapan Mahmud M.; Adnan A.; Hasan M.K.. (2026). Impact of artificial intelligence on decision-making quality in mobile financial services in Bangladesh: the mediating role of risk mitigation. Journal of Decision Systems. https://doi.org/10.1080/12460125.2026.2620371
18. Madanchian M.; Taherdoost H.. (2025). Ethical theories, governance models, and strategic frameworks for responsible AI adoption and organizational success. Frontiers in Artificial Intelligence. https://doi.org/10.3389/frai.2025.1619029
19. Morabito, G. (2025). The intricate balance between artificial intelligence, ethics, and algorithmic standards. In D. Marino & M. A. Monaca (Eds.), *Generative artificial intelligence and fifth industrial revolution* (pp. 33–46). Springer. https://doi.org/10.1007/978-3-031-73880-7_4
20. Panazan O.; GheorgheC.. (2025). Strategic Risk Spillovers from Rare Earth Markets to Critical Industrial Sectors. International Journal of Financial Studies. https://doi.org/10.3390/ijfs13030156
21. Pantano E.; Marikyan D.; Papagiannidis S.. (2024). The dark side of artificial intelligence for industrial marketing management: Threats and risks of AI adoption. Industrial Marketing Management. https://doi.org/10.1016/j.indmarman.2023.11.008
22. Pokrovskaya M.. (2019). Risk mitigation based on innovative solutions. Proceedings of the International Astronautical Congress, IAC.
23. Rodriguez, E., & Edwards, J. S. (2020). AI in smart cities development: A perspective of strategic risk management. In M. M. Cruz-Cunha, E. Portela, & R. Rijo (Eds.), *Data analytics and AI* (pp. 133–150). Auerbach Publications. https://doi.org/10.1201/9781003019855-8
24. Skvortsova M.; Grout V.. (2018). Basic approaches to assessing risks and threats in decision support systems. Proceedings of the 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2018. https://doi.org/10.1109/EIConRus.2018.8317397
25. Sposato M.; Dittmar E.C.; Portillo J.P.V.. (2026). Navigating the dark side of AI in service ecosystems: an ethical leadership framework for risk mitigation; [穿越服务生态系统中人工智能的阴暗面：基于伦理领导力的风险缓解框架]. Service Industries Journal. https://doi.org/10.1080/02642069.2026.2643384
26. Trzaska R.; Sus A.. (2023). Industry 4.0 business strategic risks based on the scalability 4.0 concept. Artificial Intelligence area. Procedia Computer Science. https://doi.org/10.1016/j.procs.2023.10.319
27. Wach K.; Duong C.D.; Ejdys J.; Kazlauskaitė R.; Korzynski P.; Mazurek G.; Paliszkiewicz J.; Ziemba E.. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. Entrepreneurial Business and Economics Review. https://doi.org/10.15678/EBER.2023.110201
28. Wirtz B.W.; Weyerer J.C.; Sturm B.J.. (2020). The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public

Administration. International Journal of Public Administration. https://doi.org/10.1080/01900692.2020.1749851

29. Wu Y.; Zheng H.; Li D.. (2026). Generative Artificial Intelligence and Corporate Risk-Taking: A Stylized Theoretical Framework with Evidence. Proceedings of 2026 2nd International Conference on Generative Artificial Intelligence and Digital Media, GADM 2026. https://doi.org/10.1145/3803730.3803756
30. Xu B.. (2025). Artificial Intelligence-Driven Strategic Risk Forecasting: Integration of Random Forest and GBDT Models with Real-Time Updates. ISMSIT 2025 - 9th International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings. https://doi.org/10.1109/ISMSIT67332.2025.11267847
31. Zecca P.A.; Marcella R.; Andrea B.; Marina P.; Marina B.; Mario R.. (2023). The Dark Side of Artificial Intelligence: The Possible Risk of Falsifying Images for Scientific Articles. Microscopy and Microanalysis. https://doi.org/10.1093/micmic/ozad093
32. van der Rest J.-P.; Heidary K.. (2024). Personalised pricing under the current European legal framework: a call for ethics, sustainability and responsibility in the age of artificial intelligence. Research in Hospitality Management. https://doi.org/10.1080/22243534.2024.2357037