https://doi.org/10.51408/issi2025_058

Research-Policy Alignment in AI: A Bibliometric Study of the EU AI Act

Cristian Mejia

mejia.cristian@ifi.u-tokyo.ac.jp
Institute for Future Initiatives, University of Tokyo, Tokyo (Japan)

Abstract

The rapid advancement of artificial intelligence (AI) necessitates understanding how academic research aligns with emerging regulatory frameworks. This study employs topic modeling to examine the relationship between library and information science research and AI policy priorities. We analyzed 2,795 academic publications on AI in library science and 1,005 statements from the European Union's AI Act, identifying 56 research clusters and 33 regulatory topics, respectively. Using semantic similarity measures, we mapped thematic alignments between research and policy domains. Results reveal strong concordance in areas such as governance frameworks and risk management, while highlighting gaps in regulatory implementation research and domain-specific applications. Notable mismatches include limited academic engagement with regulatory bodies and oversight mechanisms, contrasting with substantial research focus on cultural heritage and medical applications that lack direct regulatory correspondence. This study contributes a systematic methodology for evaluating research-policy alignment in emerging technologies, building on established bibliometric approaches for assessing research impact on policy. Our findings suggest the need for enhanced dialogue between researchers and policymakers while demonstrating how academic inquiry extends beyond immediate regulatory concerns.

Introduction

The unprecedented advancement of artificial intelligence (AI) technologies has prompted governments worldwide to develop comprehensive regulatory frameworks, exemplified by landmark legislation such as the European Union's AI Act (European Parliament, 2024). As researchers in information and library science, we regularly contribute to the AI knowledge base through studies on implementation, governance, ethics, and technological applications. However, there remains a critical gap in understanding whether our collective research priorities align with the aspects of AI that policymakers seek to regulate. This alignment—or potential mismatch—between academic research focus and policy concerns carries significant implications for both the effectiveness of evidence-based policymaking and the societal impact of our research. To address this knowledge gap, we propose a systematic bibliometric approach comparing research trends in library and information science with areas of interest in policy documents, providing an objective assessment of the concordance between academic interests and regulatory priorities in the rapidly evolving AI landscape.

The relationship between research and policymaking has been a subject of longstanding academic interest, traditionally examined through qualitative approaches that analyze how research findings influence policy decisions and how policy priorities shape research agendas. Ritter and Lancaster (2013) demonstrated this through a case study of drug policy, highlighting that assessing research

influence requires examining multiple channels, including direct citations in policy documents, utilization within policy processes, and dissemination through media coverage. This multi-dimensional approach acknowledges that research impact on policy extends beyond simple citation metrics and involves complex interactions between researchers, policymakers, and other stakeholders.

As the field evolved, researchers developed more systematic and quantitative methods to assess the research-policy relationship. Van Leeuwen et al. (2003) pioneered work in bibliometric approaches to evaluate research excellence and its influence on science policy, shifting from average-based impact metrics toward indicators that better reflect top-performing research. This methodological evolution was further exemplified by Debackere and Glanzel (2004), who demonstrated how bibliometric data could support major funding allocation decisions, highlighting the practical application of systematic research evaluation in policy contexts.

A significant advancement in this field has been the development of specialized databases and tools for tracking policy impact. The Overton database represents a major milestone, providing comprehensive indexing of policy documents and their academic citations (Szomszor & Adie, 2022). This development has enabled more sophisticated analyses of how research influences policy across different disciplines and jurisdictions. However, as Newson et al. (2018) revealed in their study of obesity policy documents, citation-based approaches have limitations – policy documents don't always explicitly cite their academic sources, and when they do, these citations may not accurately reflect the actual influence of research on policy development.

To address these limitations, researchers have explored innovative text-based methods to identify connections between different knowledge domains. Ittipanuvat et al. (2014) employed Literature-Based Discovery (LBD) to uncover linkages between technological developments and social issues, demonstrating how text analysis can reveal previously hidden connections between research and societal needs. Similarly, Takano and Kajikawa (2019) utilized text similarity measures to identify commercialization opportunities by comparing academic papers with patents. These approaches show how computational text analysis can uncover implicit relationships between research outputs and their practical applications, even when explicit citations are absent. Such methodologies offer promising alternatives for understanding the complex relationship between academic research and policy development, particularly in rapidly evolving fields where traditional citation metrics might lag behind the pace of innovation.

These methodological approaches for analyzing research-policy relationships become particularly relevant in rapidly evolving technological domains where the need for evidence-based policymaking is crucial. Artificial intelligence represents one such domain, where the acceleration of technological capabilities has prompted unprecedented policy responses worldwide. In the past few years, we witnessed significant momentum in AI governance initiatives across different jurisdictions and international bodies. The G7 Hiroshima AI Process established the world's first international framework for AI governance (G7 Leaders, 2023), while the United Nations emphasized the need for AI regulation based on the UN Charter and Universal Declaration of Human Rights (Guterres, 2023). Organizations like

UNESCO have also contributed through their Recommendation on the Ethics of AI, offering the first global, rights-based framework for AI policy (UNESCO, 2022).

However, the European Union's AI Act, which came into force in August 2024, represents a watershed moment in AI regulation. Unlike previous policy instruments that primarily focused on ethical principles or voluntary guidelines, the EU AI Act establishes a comprehensive and legally binding regulatory framework. The Act introduces a sophisticated risk-based approach, categorizing AI applications into unacceptable, high, limited, and minimal risk levels, while also addressing the emerging challenges of general-purpose AI systems (European Parliament, 2024). Its extraterritorial scope means it affects AI providers worldwide who serve EU users, similar to the impact of GDPR on data protection practices globally (European Union, 2016).

In this context, examining how library and information science research aligns with the EU AI Act's regulatory framework becomes particularly valuable. Our field's research spans multiple dimensions of AI implementation, from technical applications in information retrieval and digital collections to broader considerations of ethics, governance, and user impact. By employing bibliometric and text analysis methods to compare research clusters with policy document clusters, we can identify areas where academic research effectively informs policy decisions and where potential gaps might exist. This systematic analysis can guide future research directions, ensure our field's relevance to policy development, and potentially reveal unique insights from our discipline that could inform future AI governance frameworks. Moreover, our methodological approach offers a replicable framework for assessing research-policy alignment in other rapidly evolving technological domains.

Data and Methods

Our analysis draws on two distinct datasets: academic publications indexed in Web of Science (WoS) and the European Union's Artificial Intelligence Act. We selected Web of Science Core Collection as our primary bibliometric data source due to its comprehensive coverage of high-quality academic literature and standardized citation tracking. WoS's detailed metadata ensure reliable bibliometric analysis, while its consistent categorization system enables precise field-specific queries.

Using the search query TS=("artificial intelligence") AND WC=("information science library science"), we extracted all document types across all available publication years. This search strategy captured articles *explicitly* acknowledging a focus on AI within the specific context of library and information science, yielding 2,795 records as of January 15, 2025. By not restricting document types or publication years, we ensured comprehensive coverage of how the field has engaged with AI-related topics over time.

For our policy analysis, we focused on the European Union's AI Act, downloaded in English from the official EU website in HTML format. The Act is structured into 12 main chapters plus a 13th chapter dedicated to amendments. While the complete legislation includes additional annexes and recitals, we rely on the main chapters to focus on the core regulatory provisions. To enable detailed content analysis, we

decomposed the Act into individual statements, treating each numerical subdivision within articles as a distinct unit of analysis. This granular approach resulted in 1,005 unique statements, providing a detailed representation of the Act's regulatory scope and requirements.

Our analytical framework employs topic modelling (Blei, 2012) to identify thematic structures within both academic publications and policy statements. For academic records, we preprocessed the data by concatenating titles and abstracts for each publication. Similarly, we prepared the policy statements by removing leading numerals while preserving the complete textual content of each regulatory provision. The topic modeling process utilized BERTopic (Grootendorst, 2022), a state-of-the-art library that leverages BERT's contextual embeddings to generate more semantically coherent topics compared to traditional approaches like LDA. To determine the optimal number of topics for each dataset, we conducted an iterative process testing configurations ranging from 10 to 200 topics, selecting the solution that maximized the coherence metric (Farea et al., 2024). This approach resulted in 56 topics for the academic dataset and 33 topics for the policy statements.

Each document was then assigned to its most probable topic, effectively creating distinct clusters within both datasets. For academic clusters, we calculated additional metrics including average publication year and mean citation count, providing temporal and impact dimensions to our analysis. We manually labeled each cluster based on careful examination of its constituent documents, considering frequent terms, representative papers, and thematic coherence.

To identify alignments between research priorities and policy concerns, we computed cosine similarity scores between the topic vectors of academic and policy clusters. This similarity metric captures semantic overlap between clusters, with higher scores indicating stronger thematic alignment. Cosine similarity is particularly suitable for this comparison as it normalizes for differences in document length and term frequency distributions between academic and policy texts.

This methodological framework enables systematic comparison between research focus areas and regulatory priorities, revealing both convergences and potential gaps between academic inquiry and policy development in the domain of artificial intelligence within library and information science.

Results and Discussion

Our analysis identified 56 distinct research topics from 2,795 academic publications on AI in library science, while the 1,005 statements extracted from the EU AI Act clustered into 33 regulatory topics. These two topic landscapes represent the research interests of academics and the regulatory priorities of policymakers, respectively. The results of the academic landscape analysis are presented in Figure 1 and Table 1.

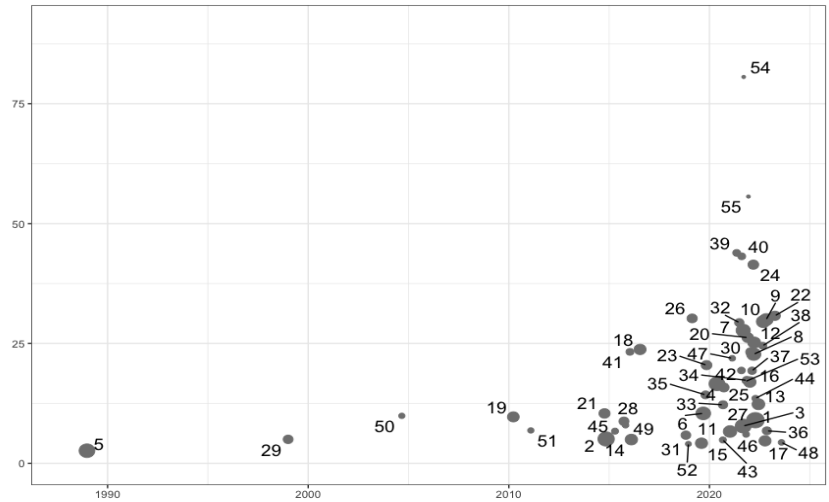

Figure 1. Temporal distribution and impact of AI research topics in library science. Each point represents a research cluster, with its position determined by average publication year (x-axis) and average citation count (y-axis). Numbers correspond to cluster IDs.

The visualization reveals the evolution of AI research within library science over the past three decades. Early research in the 1990s centered on fundamental information retrieval systems, as represented by cluster 5. The field has since undergone significant transformation, with recent research focusing on emerging technologies such as blockchain integration (cluster 55) and applications of generative AI (clusters 22, 36).

Table 1. Summary of selected research clusters on AI in library science, including the five most recent, most cited, and largest by number of documents.

Id	*Cluster name*	*Docs.*	*Ave. Year*	*Ave. Cites.*
1	Strategic Implementation of AI Technologies in Library Service Innovation	143	2,022.3	9.0
2	AI Integration and Digital Transformation in Information Management Systems	120	2,014.9	5.1
3	AI-Powered Content Analysis and Generation in Digital Media	113	2,021.7	7.8
4	Trust and Governance Frameworks for Healthcare AI Implementation	110	2,020.4	16.6
5	Applications of Artificial Intelligence in Information Retrieval Systems	110	1,989.0	2.6
9	Technology Acceptance Models and User Adoption Factors in AI-Enabled Systems	82	2,022.8	30.0
17	AI-Enhanced Peer Review Systems in Academic Publishing	61	2,022.8	4.7
22	AI-Assisted Knowledge Construction in Academic Research and Writing	49	2,023.3	30.8
24	AI Implementation Frameworks and Challenges in Organizational Systems	48	2,022.2	41.4

36	Chatbot Implementation in Libraries	36	2,022.9	6.8
39	AI-Driven Information Management Solutions During the COVID-19 Crisis	28	2,021.4	43.9
40	AI-Driven Marketing Analytics and Customer Segmentation Systems	28	2,021.6	43.2
48	Digital Literacy Evolution in the AI Era	22	2,023.6	4.4
54	Big Data Analytics Applications in Organizational Decision Support Systems	17	2,021.7	80.6
55	Blockchain Integration in Information Systems	17	2,021.9	55.6

The citation patterns reveal varying levels of scholarly impact across research topics. Big data analytics (cluster 54), blockchain applications (cluster 55), and COVID-19 related research (cluster 39) have garnered recent attention, each averaging over 40 citations per paper. Implementation frameworks (cluster 24), trust dynamics (cluster 38), and organizational impact studies (cluster 10) have also demonstrated substantial influence with moderate citation rates.

Perhaps most notably, we see a marked concentration of research clusters in the 2020-2024 period, indicating an acceleration of AI-related research within library science. This temporal clustering coincides with the development and implementation of the EU AI Act, suggesting a potential alignment between academic research priorities and emerging regulatory frameworks. This synchronicity provides a valuable foundation for examining the relationship between research focus areas and regulatory priorities.

The analysis of the EU AI Act yielded 33 distinct clusters that reflect the regulatory framework's key priorities as seen in Table 2. These clusters broadly align into several core themes. The foundational elements of the Act are represented in clusters focusing on governance structures, including the establishment of the AI Office, Scientific Panel, and oversight mechanisms (clusters 4, 16, and 22). A significant portion of clusters addresses specific technical and operational requirements, such as conformity assessment procedures (cluster 5), data processing protocols (cluster 12), and logging requirements (cluster 33).

Table 2. Summary of clusters from the EU AI Act. The top largest clusters are shown.

Id	*Cluster name*	*Statements*
1	Notified Bodies	56
2	Market Surveillance and Law Enforcement Authority Framework	52
3	SME and Start-up Support Mechanisms	49
4	Governance and Advisory Bodies Structure	48
5	Conformity Assessment and Documentation Requirements	46
6	General-Purpose AI Models Classification and Risk Management	46
7	Technical Requirements for High-Risk AI Systems	45
8	Market Placement and Compliance Requirements	43
9	Stakeholder Obligations and Responsibilities	42
10	Risk Assessment and Harm Classification	40

Risk management emerges as a central theme, with dedicated clusters covering risk assessment methodologies (cluster 10), market surveillance (clusters 2 and 11), and incident reporting frameworks (cluster 32). The Act's emphasis on documentation and transparency is reflected in clusters focusing on provider obligations (cluster 30), compliance documentation (cluster 26), and certificate management (cluster 29).
Notably, several clusters specifically address emerging technologies and their regulatory implications, particularly in the context of biometric systems (cluster 18) and general-purpose AI models (cluster 6). The Act also maintains focus on practical implementation through clusters dedicated to SME support mechanisms (cluster 3), market access requirements (cluster 19), and administrative procedures (cluster 28). This clustering reveals the Act's comprehensive approach to AI regulation, balancing high-level governance principles with specific technical requirements and practical implementation considerations. The distribution of topics suggests a regulatory framework that aims to be both thorough in its coverage and pragmatic in its application.

Linkage between academic research and policy

After analyzing the individual topic landscapes, we examined the thematic alignment between academic research clusters and regulatory topics through semantic similarity analysis. When two clusters from different datasets show high similarity, this indicates that the academic research focus substantively overlaps with regulatory priorities in that area. Academic research and policy documents are written in different styles and use different vocabulary; thus, high absolute similarity is not expected. Therefore, we define as similar pairs those beyond the third quartile across all possible connections (i.e., >0.46), suggesting relative strong thematic concordance between the research focus and policy considerations. The similar pairs are shown in Figure 2.

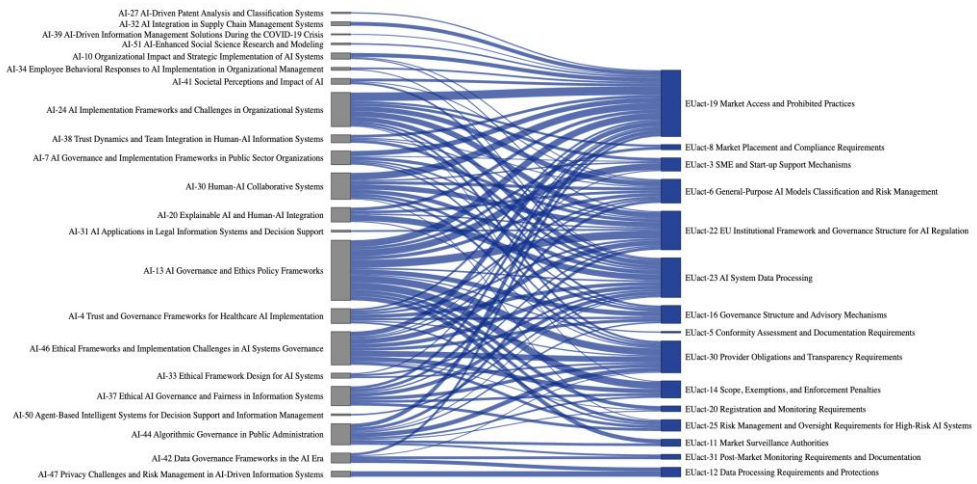

Figure 2. Semantic linkages between AI research topics in library science (left) and EU AI Act regulatory clusters (right). The width of connecting lines represents the strength of thematic similarity between clusters. Only connections with similarity scores above 0.46 are shown.

The Sankey diagram reveals several notable alignments between research and policy domains. A particularly strong connection exists between academic research on "AI Governance and Ethics Policy Frameworks" (cluster 13) and the regulatory focus on "Market Surveillance and Law Enforcement Authority Framework" (cluster 2). This alignment suggests that academic research has been actively engaging with governance challenges that policymakers consider crucial.

Another significant match appears between research on "Trust Dynamics and Team Integration in Human-AI Information Systems" (cluster 38) and the regulatory cluster on "Risk Management and Oversight Requirements for High-Risk AI Systems" (cluster 25). This pairing indicates that academic investigations into human-AI interaction and trust align well with regulatory concerns about risk management in high-stakes AI applications.

However, the visualization also reveals areas where academic research and regulatory focus may not fully align, as evidenced by clusters with few or no strong connections. This pattern suggests opportunities for future research to address emerging regulatory priorities.

Research gaps in relation to regulatory clusters

The analysis reveals notable gaps between academic research priorities and certain regulatory focuses. Particularly striking is the limited academic engagement with regulatory bodies and administrative frameworks, which are central to clusters 1 ("Notified Bodies") and 4 ("Governance and Advisory Bodies Structure") of the EU AI Act. While these clusters detail the operational mechanics of AI oversight - including the roles of notified bodies in conformity assessment and the structure of advisory forums - our bibliometric analysis shows minimal research addressing these institutional aspects within library and information science.

This mismatch likely stems from the traditionally technical and user-focused nature of library science research, which has emphasized practical implementations and user interactions with AI systems rather than regulatory mechanisms. However, this gap presents valuable research opportunities. Future studies could examine how information institutions interact with regulatory bodies, how conformity assessments impact information services, and how library and information science expertise could inform the development of AI governance structures. Additionally, research investigating the role of libraries and information centers as potential intermediaries in the regulatory framework could provide valuable insights for both policymakers and practitioners.

Regulatory gaps in relation to research clusters

The analysis also reveals areas where academic research has developed substantial focus that is not directly reflected in the regulatory framework. For instance, clusters AI-15 ("AI-Enabled Digital Collection Management in Cultural Institutions") and AI-16 ("Clinical Applications of AI in Medical Diagnosis and Prognosis") represent significant research streams with limited corresponding regulatory attention in the EU AI Act.

In the case of digital collection management (AI-15), this mismatch likely reflects the specialized nature of cultural heritage applications, which may not warrant specific regulatory attention despite their importance to the library and information science community. The research in this area focuses on practical implementations and professional practices that fall under broader regulatory categories rather than requiring dedicated regulatory frameworks.

Similarly, while medical AI applications (AI-16) represent a crucial research area within our field, their regulation is primarily addressed through specialized healthcare frameworks and medical device regulations rather than the general-purpose AI Act. This suggests that some domain-specific AI applications, though important in academic research, may be better governed through sector-specific regulatory instruments rather than general AI legislation.

Conclusion

In this study, we employed bibliometric analysis and topic modeling to examine the alignment between academic research in library science and AI policy priorities as reflected in the EU AI Act. By analyzing 2,795 academic publications and 1,005 policy statements, we identified 56 research clusters and 33 regulatory topics, enabling a systematic comparison of thematic focus areas through semantic similarity measures.

Our findings resonate with previous research on evidence-based policymaking. As Van Leeuwen et al. (2003) emphasized the need for sophisticated metrics to evaluate research excellence, our analysis provides a quantitative framework for assessing research-policy alignment. Our proposal also highlights the role of bibliometrics in providing new angles that may facilitate the work of policymakers (Kajikawa, 2022). The identification of both matches and mismatches in our results supports Ritter and Lancaster's (2013) assertion that research influence on policy operates through multiple channels and complex interactions.

The study reveals both encouraging alignments and notable gaps between academic research and regulatory priorities. While we found strong concordance in areas such as governance frameworks and risk management, significant disparities emerged in others. These findings underscore the need for enhanced dialogue between researchers and policymakers in shaping AI governance within information environments. Similarly, the presence of research clusters with limited regulatory correspondence demonstrates how academic inquiry naturally extends beyond immediate regulatory concerns to address domain-specific challenges.

This research makes a novel contribution by providing a systematic, quantitative methodology for evaluating the relationship between research priorities and regulatory frameworks in rapidly evolving technological domains. Our approach offers a replicable framework for assessing research-policy alignment that could be applied to other emerging technologies and regulatory contexts.

Several limitations and opportunities for future research exist. First, our analysis focuses solely on the EU AI Act; incorporating other regulatory frameworks such as the Council of Europe AI Treaty and White House Executive Orders would provide a more comprehensive view. Second, expanding the analysis beyond library science

or broadening the AI-related search terms could offer wider perspectives on research-policy alignment. Finally, analyzing research clusters by country of origin could reveal geographical variations in research priorities and their relationship to national policy approaches.

Acknowledgments

This research is supported by a Grant in Aid for Early Career Scholars from the Japan Society for the Promotion of Science [24K16438].

References

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Debackere, K., & Glänzel, W. (2004). Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key. *Scientometrics, 59*(2), Article 2. https://doi.org/10.1023/B:SCIE.0000018532.70146.02

European Parliament. (2024). *Regulation 2024/1689. Artificial Intelligence Act* (OJ L, 2024/1689, 12.7.2024). https://eur-lex.europa.eu/eli/reg/2024/1689/oj#document1

European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)* (32016R0679; Official Journal of the European Union). https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

Farea, A., Tripathi, S., Glazko, G., & Emmert-Streib, F. (2024). Investigating the optimal number of topics by advanced text-mining techniques: Sustainable energy research. *Engineering Applications of Artificial Intelligence, 136*, 108877. https://doi.org/10.1016/j.engappai.2024.108877

G7 Leaders. (2023). *Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI system*. G7 Hiroshima Summit 2023.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794; Version 1). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Guterres, A. (2023, July 18). *UN Secretary-General's remarks to the Security Council on Artificial Intelligence*. https://www.un.org/sg/en/content/sg/speeches/2023-07-18/secretary-generals-remarks-the-security-council-artificial-intelligence

Ittipanuvat, V., Fujita, K., Sakata, I., & Kajikawa, Y. (2014). Finding linkage between technology and social issue: A Literature Based Discovery approach. *Journal of Engineering and Technology Management, 32*, 160–184. https://doi.org/10.1016/j.jengtecman.2013.05.006

Kajikawa, Y. (2022). Reframing evidence in evidence-based policy making and role of bibliometrics: Toward transdisciplinary scientometric research. *Scientometrics, 127*(9), Article 9. https://doi.org/10.1007/s11192-022-04325-6

Newson, R., Rychetnik, L., King, L., Milat, A., & Bauman, A. (2018). Does citation matter? Research citation in policy documents as an indicator of research impact – an Australian obesity policy case-study. *Health Research Policy and Systems, 16*(1), 55. https://doi.org/10.1186/s12961-018-0326-9

Ritter, A., & Lancaster, K. (2013). Measuring research influence on drug policy: A case example of two epidemiological monitoring systems. *International Journal of Drug Policy*, *24*(1), 30–37. https://doi.org/10.1016/j.drugpo.2012.02.005

Szomszor, M., & Adie, E. (2022). Overton: A bibliometric database of policy document citations. *Quantitative Science Studies*, *3*(3), 624–650. https://doi.org/10.1162/qss_a_00204

Takano, Y., & Kajikawa, Y. (2019). Extracting commercialization opportunities of the Internet of Things: Measuring text similarity between papers and patents. *Technological Forecasting and Social Change*, *138*, 45–68. https://doi.org/10.1016/j.techfore.2018.08.008

UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence* (Programme and Meeting Document SHS/BIO/PI/2021/1). https://unesdoc.unesco.org/ark:/48223/pf0000381137

Van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & Van Raan, A. F. J. (2003). The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics*, *57*(2), Article 2. https://doi.org/10.1023/A:1024141819302