

Predictive AI for Socially-Aware and Legally-Compliant Research Policy in CRIS Systems

Gizem Gültekin-Várkonyi¹ and Otmane Azeroual²

¹University of Szeged, Faculty of Law and Political Sciences, Szeged, Hungary

²FernUniversität in Hagen (FeU), Hagen, Germany

gizemgv@juris.u-szeged.hu, otmane.azeroual@fernuni-hagen.de

Abstract

Current Research Information Systems (CRIS) are essential infrastructures for managing and evaluating research activities, yet they remain largely limited to retrospective analysis. This limitation becomes increasingly critical as research policy is shaped by complex socio-technical dynamics, including Open Science, Responsible Research and Innovation (RRI), diversity and inclusion goals, and regulatory frameworks such as the EU AI Act. These developments require forward-looking tools that enable decision-makers to anticipate the social, legal, and institutional impacts of policy interventions. This paper proposes a conceptual framework for integrating predictive Artificial Intelligence (AI) into CRIS systems to support socially-aware and legally-compliant research policy. The approach combines machine learning, causal modelling, agent-based simulation, and Natural Language Processing (NLP) to simulate policy scenarios and assess their impacts. Particular emphasis is placed on explainability, fairness, and regulatory compliance. The contribution lies in positioning predictive AI as a governance tool rather than an automated decision-maker, enabling evidence-based, transparent, and adaptive research policy-making within CRIS environments.

1. Introduction

Current Research Information Systems (CRIS) constitute a central infrastructure for managing, analyzing, and strategically steering research activities in universities, non-university research institutions, and research policy organizations. They aggregate heterogeneous data on projects, publications, funding, collaborations, patents, research data, as well as legal and organizational frameworks, thereby enabling reporting, evaluation, and institutional decision-making. Despite their importance, CRIS systems remain largely retrospective, focusing on the documentation and assessment of past activities rather than supporting forward-looking policy design.

At the same time, research policy is increasingly shaped by complex socio-technical dynamics. Initiatives such as Open Science, Responsible Research and Innovation (RRI), and diversity and inclusion policies require institutions to consider broader societal impacts beyond traditional performance metrics. Furthermore, regulatory developments, particularly the European Union AI Act,

introduce new requirements for risk governance in AI-supported decision-making processes, and further extends the scope of the General Data Protection Regulation (GDPR) on the main principles of data processing such as fairness, accountability and transparency (aka, FAT). These developments create a growing demand for tools that not only evaluate policies retrospectively but also enable the exploration of their potential social, legal, and institutional consequences.

In this context, predictive AI offers the potential to extend the analytical capabilities of CRIS systems (Azeroual, et al., 2025). Beyond forecasting research developments, predictive AI can be used to simulate socio-technical interactions in which policy decisions, scientific practices, legal frameworks, and societal expectations intersect. By enabling scenario-based analysis, it allows decision-makers to compare alternative policy interventions, anticipate their consequences, and support more informed strategic planning. Importantly, predictive outputs should not be interpreted as deterministic results; rather, their associated uncertainties can provide valuable insights for adaptive governance (Gstrein et al., 2024).

Methodologically, integrating predictive AI into CRIS requires combining multiple analytical approaches. Machine learning supports the identification of patterns and trends, while causal modelling enables the analysis of relationships between policy measures and their effects. Agent-based simulation allows the representation of behavioral dynamics at individual and institutional levels, and Natural Language Processing (NLP) facilitates the analysis of policy documents, regulatory texts, and qualitative data. By linking internal CRIS data with external sources—such as Open Access monitoring, funding statistics, compliance data, or social indicators—this approach enables the simulation of policy scenarios, including changes in evaluation criteria, funding mechanisms, or governance structures.

However, the use of predictive AI in research policy also raises significant social, ethical, and legal challenges (Kim et al., 2025). Models trained on historical data may reproduce or amplify existing inequalities in areas such as funding allocation, career development, or collaboration opportunities. These risks are well documented in research on algorithmic fairness (Lalor et al., 2024; Wachter et al., 2021). In addition, black-box models are problematic in policy contexts due to their lack of transparency and limited interpretability (de Bruijn et al., 2022).

From a legal perspective, the deployment of predictive AI must comply with emerging regulatory frameworks, particularly the EU AI Act. This regulation introduces a risk-based governance approach and requires mechanisms for transparency, accountability, and human oversight. Instruments such as Fundamental Rights Impact Assessments (FRIA) play a key role in identifying and mitigating risks prior to deployment (Malgieri & Santos, 2025; Bertaina et al., 2025). In parallel, the promotion of AI literacy (AI-L) is essential to ensure that stakeholders can understand and critically assess AI-generated outputs (Long & Magerko, 2020; Ng et al., 2021; Karaboga et al., 2023). Furthermore, whenever personal data processing takes place in any CRIS, then the FAT must also be taken into consideration in combination with the AI Act's equal relevancy. Under the GDPR, controllers must in particular conduct Data Protection Impact Assessments (DPIAs) for high-risk (personal data) processing, including large-scale profiling and automated decision-making, thereby aligning closely with the AI Act's requirement of ex-ante risk assessment for high-risk AI (Article 35 GDPR; Sartor & Lagioia, 2020). Moreover, the GDPR's accountability and data-protection-by-design obligations (Articles 24–25, Chiao, 2019) echo the AI Act's emphasis on continuous risk-management and human oversight, creating a dual-layered regime where AI-specific and data-protection rules operate in parallel.

Against this background, this paper proposes a conceptual framework for integrating predictive AI into CRIS systems that combines technological capabilities with social awareness and legal compliance. The framework positions predictive AI as a governance tool rather than an automated decision-maker, enabling institutions to simulate policy scenarios, assess their societal impacts, and support transparent and evidence-based decision-making.

The remainder of this paper is structured as follows. Section 2 reviews related work on CRIS, predictive AI, and AI governance. Section 3 presents the conceptual architecture of the proposed framework and illustrates its integration within CRIS environments. Section 4 outlines the methodological approach, followed by Sections 5 and 6, which examine social impact assessment and legal compliance. Section 7 discusses implications and limitations, and Section 8 concludes the paper.

2. Related Work

The integration of predictive AI into research information systems intersects several research domains, including CRIS development, AI-based analytics, algorithmic fairness, and AI governance. While each of these areas has been extensively studied, their combined application to research policy within CRIS environments remains limited.

Research on CRIS has primarily focused on data integration, interoperability, and their role in research evaluation and management. These systems support institutional reporting and performance assessment but are typically limited to descriptive and retrospective analyses. As a result, they provide only limited support for forward-looking policy design or scenario-based decision-making.

In parallel, advances in AI have enabled new forms of data-driven analysis. Machine learning and predictive analytics are increasingly used to identify patterns and forecast developments in complex systems. However, recent work emphasizes that predictive models should not be viewed solely as forecasting tools but also as instruments for exploring uncertainty and supporting decision-making under complex conditions (Gstrein et al., 2024; Azeroual et al., 2025).

At the same time, the use of AI in decision-making contexts has raised concerns regarding FAT and bias problems. Studies on algorithmic fairness demonstrate that predictive systems can reproduce and amplify existing inequalities when trained on biased data (Lalor et al., 2024; Wachter et al., 2021). These challenges are particularly relevant in research policy, where decisions may affect funding distribution, career trajectories, and institutional opportunities. Consequently, explainable AI (XAI) has become a key requirement to ensure transparency and accountability (de Bruijn et al., 2022; Azeroual, 2025). Research by Rebstadt et al. emphasizes that AI systems should be designed to provide explanations tailored to human levels of understanding; otherwise, complexity differentials and varying knowledge levels risk exacerbating confusion rather than alleviating it. Explainability and transparency are central pillars of responsible AI governance, integral to fairness and accountability (Papagiannidis et al., 2025). Therefore, meaningful design elements should be embedded in the AI system lifecycle in order to ensure not only legal but also ethical and operable responsible AI requirements.

Beyond technical aspects, legal and regulatory frameworks increasingly shape AI deployment. The EU AI Act institutionalises a risk-based approach that emphasises transparency, accountability, and human oversight, requiring deployers of high-risk AI systems to systematically evaluate foreseeable harms to fundamental rights, safety, and societal values prior to deployment which could be called as part of a broader human rights by design concept (Bisoyi, 2025). In practice, this obligation is operationalised through tools such as the FRIA, which structure the identification, severity-grading, and probability-assessment of impacts on rights enshrined in the EU legislation (Ajoodha&Browne, 2025). FRIAs are not isolated compliance exercises but part of an ongoing governance cycle, designed to interact with existing mechanisms such as the DPIAs under the GDPR and internal risk-management systems (Mantelero, 2024). DPIAs, in turn, help assess privacy and data-protection risks associated with AI-driven profiling and automated decision-making, particularly in contexts such as recruitment, policing, and public services governed by public law (Graux et al., 2024). Yet scholars note that DPIAs often under-address broader societal and equality-related harms, which is why the AI Act layers FRIA over DPIA to close this governance gap (Thomaidou &

Limniotis, 2025; Moniz, 2026). Complementing these procedural tools even with AI literacy (AI-L) which is emerging as a governance prerequisite, encompassing the competencies needed by developers, auditors, and end-users to understand, critically evaluate, and responsibly engage with AI systems (Hillman et al., 2025).

Despite these developments, existing research often addresses technical, social, or legal aspects separately. There is a lack of integrated frameworks that combine predictive analytics, simulation, social impact assessment, and legal compliance within CRIS environments. In particular, the use of predictive AI as a governance tool for research policy remains insufficiently explored.

This paper addresses this gap by proposing a conceptual framework that integrates predictive AI into CRIS systems as a socio-technical and governance-oriented approach, advancing CRIS from retrospective data infrastructures toward forward-looking systems for responsible research policy.

3. Conceptual Architecture

To address the limitations of retrospective CRIS systems and enable forward-looking, socially-aware, and legally-compliant research policy, this paper proposes a multi-layer conceptual architecture that integrates predictive AI, simulation capabilities, and governance mechanisms within existing CRIS environments.

3.1 Architecture Overview

The proposed architecture is designed as an integrative extension of CRIS systems rather than a standalone analytical tool. This ensures compatibility with existing data infrastructures, institutional workflows, and governance frameworks. The architecture follows a structured pipeline that transforms heterogeneous research data into actionable decision-support outputs.

As illustrated in Figure 1, the architecture consists of four main layers:

- Input Layer (Data Sources)
- Predictive AI Layer (Models & Analytics)
- Simulation & Scenario Layer (Scenario Testing)
- Output & Decision Support Layer (Decision Support)

In addition, a cross-cutting Governance & Compliance Layer ensures that all analytical processes are aligned with principles of fairness, transparency, legal compliance, and human oversight. A feedback loop connects outputs back to the input layer, enabling continuous learning and system improvement.

The Figure 1 illustrates a multi-layer architecture integrating CRIS data sources, predictive AI models, and simulation mechanisms to support socially-aware and legally-compliant research policy. The pipeline connects heterogeneous data inputs with machine learning, causal modelling, agent-based simulation, and NLP techniques, enabling scenario testing and decision support. A cross-cutting governance layer ensures fairness, transparency, human oversight, and compliance with regulatory frameworks such as the EU AI Act.

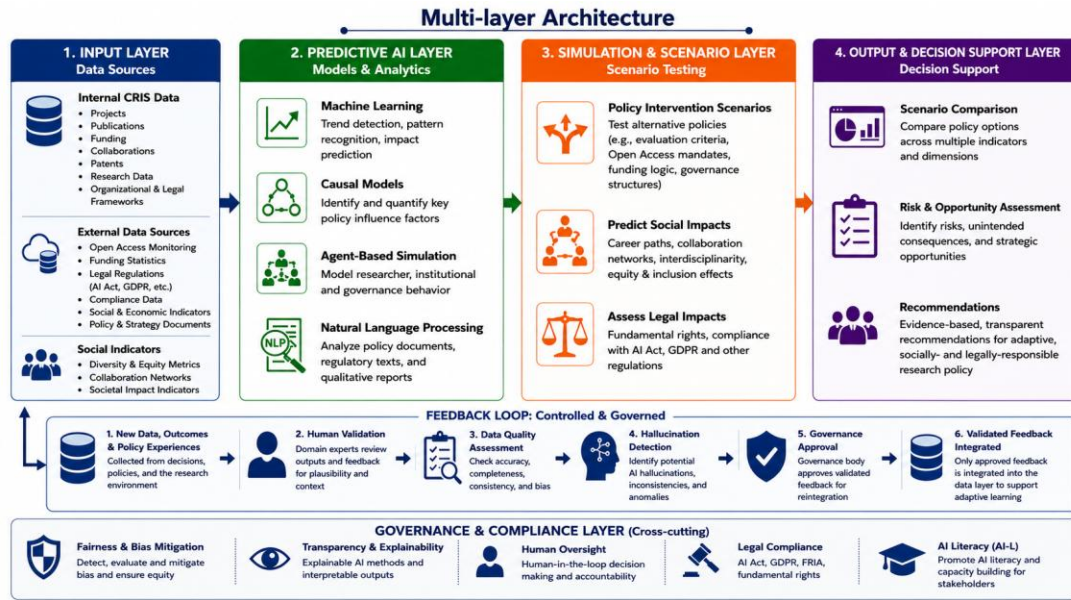


Figure 1: Predictive AI for Socially-Aware Research Policy in CRIS Systems.

3.2 Input Layer (Data Sources)

The input layer integrates heterogeneous data from multiple sources. This includes internal CRIS data such as projects, publications, funding, collaborations, patents, and research data, as well as organizational and legal information. In addition, external data sources are incorporated, including Open Access monitoring, funding statistics, regulatory documents (e.g., AI Act, GDPR), compliance data, and policy or strategy documents.

A distinctive element of this layer is the inclusion of social indicators, such as diversity and equity metrics, collaboration networks, and broader societal impact indicators. These data are essential for enabling socially-aware policy analysis. The integration of these diverse data sources requires preprocessing, harmonization, and feature engineering to ensure analytical usability.

3.3 Predictive AI Layer (Models & Analytics)

The predictive AI layer applies multiple analytical methods to extract patterns and generate insights. As shown in Figure 1, this layer combines:

- Machine Learning, for trend detection, pattern recognition, and impact prediction
- Causal Models, to identify and quantify key policy influence factors
- Agent-Based Simulation, to model researcher, institutional, and governance behavior
- NLP, to analyze policy documents, legal texts, and qualitative reports

These methods complement each other by capturing different dimensions of the research system. Predictive outputs are interpreted as probabilistic insights rather than deterministic forecasts, allowing decision-makers to explore uncertainty.

3.4 Simulation & Scenario Layer (Scenario Testing)

Building on the analytical results, the simulation layer enables the evaluation of policy interventions through scenario testing. This includes the modelling of alternative policy options, such as changes in evaluation criteria, Open Access mandates, funding logic, or governance structures.

The framework supports three main analytical perspectives:

- Policy Intervention Scenarios, enabling the comparison of alternative policy designs
- Social Impact Prediction, including effects on career paths, collaboration networks, interdisciplinarity, and equity
- Legal Impact Assessment, including compliance with the AI Act, GDPR, and fundamental rights considerations

This layer is particularly important for identifying not only intended outcomes but also potential unintended consequences of policy interventions.

3.5 Output & Decision Support Layer

The final layer translates analytical and simulation results into actionable decision-support outputs. These include:

- Scenario comparison, allowing the evaluation of different policy options
- Risk and opportunity assessment, identifying potential unintended effects and strategic opportunities
- Policy recommendations, providing evidence-based and transparent guidance for decision-makers

These outputs are designed to support different stakeholders, including research managers, institutional leadership, and policy-makers.

3.6 Governance & Compliance Layer (Cross-cutting)

A key feature of the architecture is the cross-cutting governance and compliance layer, which spans all stages of the pipeline. This layer ensures that predictive AI is applied in a responsible and legally compliant manner.

It includes the following components:

- Fairness and bias mitigation, to detect and reduce structural inequalities
- Transparency and explainability, through interpretable AI models
- Human oversight, ensuring accountability in decision-making
- Legal compliance, including alignment with the AI Act, GDPR, and FRIA requirements
- AI-L, supporting stakeholders in understanding and interpreting AI outputs

By embedding these mechanisms directly into the architecture, governance is treated as an integral component rather than an external constraint.

3.7 Feedback Loop and Adaptive Learning

The proposed architecture incorporates a controlled feedback loop that connects decision-support outputs back to the data layer, enabling continuous learning and iterative refinement. However, AI-generated outputs are not automatically reintegrated into subsequent learning cycles. Instead, newly generated data, policy outcomes, and user feedback undergo human validation, data quality assessment, and governance-based review before being incorporated into the system.

This controlled feedback mechanism is designed to prevent error propagation, AI hallucinations, model drift, and the reinforcement of historical biases, thereby ensuring that adaptive learning remains trustworthy and aligned with legal and ethical requirements. By combining human oversight, explainability, and continuous monitoring, the framework supports responsible model evolution while maintaining transparency, accountability, and regulatory compliance.

Consequently, predictive AI is conceptualized not as a static analytical system, but as a continuously governed decision-support capability that evolves through validated evidence, institutional learning, and responsible AI governance.

4. Methodological Approach

The proposed framework follows a socio-technical and design-oriented methodological approach, integrating multiple AI techniques to support predictive and simulation-based research policy analysis within CRIS environments. The objective is not to develop a single predictive model but to combine complementary methods that capture different dimensions of complex research systems. Such methodological approach has already been assessed in the current literature with a specific emphasize on AI design (Riesen, 2025; Hedfeld, 2026; Herrman, 2026). In this article, we propose a practical approach in integrating different data sources to create a model which could operationalize the workflows, harmonize the content and finally offer a predictive outcome for a research policy in line with the legal and ethical principles.

At the core of the methodology lies the integration of heterogeneous data sources. Internal CRIS data, such as publications, projects, funding records, and collaboration networks, are combined with external datasets, including Open Access monitoring, legal regulations, compliance data, and socio-economic indicators. This integrated data foundation enables a more comprehensive representation of research ecosystems and policy contexts.

The analytical component of the framework consists of four complementary methodological approaches. First, machine learning techniques are used to identify patterns, trends, and correlations in historical data. These models support the detection of emerging research areas, shifts in collaboration structures, and potential policy effects. However, rather than providing deterministic predictions, these models generate probabilistic insights that reflect uncertainty and variability in research systems.

Second, causal modelling is employed to identify and quantify relationships between policy interventions and their outcomes. Unlike purely predictive models, causal approaches aim to uncover underlying mechanisms and dependencies, enabling a more robust interpretation of policy impacts. This is particularly relevant for distinguishing correlation from causation in complex research environments.

Third, agent-based simulation is used to model the behavior of individual actors and institutions within the research system. These simulations capture dynamic interactions between researchers, organizations, and policy frameworks, allowing the exploration of emergent effects that cannot be derived from static models alone. By simulating behavioral responses to policy changes, this approach provides insights into potential long-term system dynamics.

Fourth, NLP is applied to analyze unstructured textual data, including policy documents, legal regulations, and qualitative reports. NLP techniques enable the extraction of relevant information, identification of regulatory requirements, and alignment of predictive models with legal and policy frameworks. This is particularly important in the context of the EU AI Act, where legal texts must be systematically interpreted and integrated into decision-making processes.

These methodological components are combined within a unified workflow that supports scenario-based analysis. Policy interventions, such as changes in evaluation criteria, funding allocation, or Open Access mandates, can be modelled as input parameters. The system then generates simulated outcomes across multiple dimensions, including social impact, institutional behavior, and legal compliance.

A key feature of the methodology is the integration of explainability and transparency mechanisms. Explainable AI techniques are used to provide insights into model behavior, feature importance, and decision pathways. This is essential for ensuring that outputs are interpretable and can be critically assessed by decision-makers. In addition, fairness-aware modelling approaches are incorporated to detect and mitigate biases in data and predictions, addressing concerns related to equity and inclusion.

Finally, the methodology includes iterative feedback loops that enable continuous refinement of models and policy scenarios. New data, outcomes, and user feedback are incorporated into the system, allowing adaptive learning and ongoing improvement. This iterative approach aligns with the understanding of predictive AI as a governance tool that supports learning and decision-making under uncertainty, rather than providing fixed or definitive answers.

5. Social Impact Assessment

An essential contribution of the proposed framework lies in extending CRIS-based analysis toward the assessment of social impacts of research policy. Conventional CRIS implementations primarily focus on measurable outputs such as publications or funding volumes, offering limited insight into how policy decisions shape research practices and institutional dynamics.

By integrating predictive and simulation-based approaches, the framework enables the exploration of how policy interventions may affect the research system over time. For example, changes in evaluation criteria or funding mechanisms can influence career trajectories, collaboration structures, and the distribution of opportunities across disciplines. Rather than relying solely on historical indicators, the approach supports the analysis of potential future developments under varying assumptions.

Particular attention is given to equity-related dimensions. Predictive analyses can reveal structural patterns in career progression and identify whether specific policy measures systematically advantage or disadvantage certain groups (Ajunwa, 2021; Trigo et al., 2024; Dubey&Vachher, 2025; Ahmad et al., 2026;). Similarly, network-based approaches allow the examination of collaboration dynamics and interdisciplinarity, providing a more differentiated understanding of knowledge exchange within research systems.

At the same time, the use of predictive models introduces risks related to bias and fairness. Since CRIS data reflect historical developments, they may encode structural imbalances that influence model outputs. Without appropriate mitigation strategies, predictive systems may reinforce existing inequalities (Lalor et al., 2024; Wachter et al., 2021). To address this, the framework incorporates fairness-aware modelling and transparent evaluation mechanisms.

Another critical aspect concerns unintended consequences. Policy interventions designed to improve efficiency or performance may trigger behavioral adaptations that are not captured by

traditional indicators. Simulation-based analysis enables the identification of such effects in advance, thereby supporting more robust and reflective policy design.

Overall, the integration of social impact assessment expands the analytical scope of CRIS systems, allowing decision-makers to consider not only performance outcomes but also the broader societal implications of research policy.

6. Legal Compliance and Governance

The integration of predictive AI into CRIS systems requires careful consideration of legal and governance requirements, particularly in the context of European regulatory frameworks. Research policy operates within a highly regulated environment, and AI-supported decision-making must comply with evolving legal standards.

A central reference point is the European Union AI Act, which introduces a risk-based approach to AI governance. Depending on their application, predictive models used in research policy may be classified as high-risk systems, requiring enhanced transparency, documentation, and human oversight (Ebers et al., 2021; Gstrein et al., 2024). This is especially relevant when such systems influence funding allocation, evaluation processes, or strategic decision-making.

Within this framework, FRIA plays a key role as a structured mechanism for assessing potential risks prior to deployment. It requires organizations to evaluate not only technical performance but also the broader societal and legal implications of AI systems, including their impact on affected groups (Malgieri & Santos, 2025; Bertaina et al., 2025). In CRIS contexts, this implies that predictive models must be assessed in terms of both accuracy and potential harm.

In addition, AI-L is essential to ensure that stakeholders can interpret and critically evaluate AI outputs. Decision-makers must understand the limitations, uncertainties, and assumptions underlying predictive models in order to use them responsibly (Long & Magerko, 2020; Ng et al., 2021). This reinforces the need to treat AI outputs as supportive inputs rather than authoritative decisions. Human oversight remains a fundamental requirement and if combined with the AI-L, human oversight could be meaningful. For example, as the emerging researches highlight challenges such as automation bias, where users may over-rely on algorithmic outputs (Laux & Ruschemeier, 2025), AI-L could help overcoming the overreliance and could contribute to meaningful human oversight. The importance of combining technical safeguards with organizational measures and user awareness could together contribute to broader principles of responsible AI.

Finally, the AI Act emphasizes that decisions with significant consequences should not be fully automated. In research policy, where decisions often involve normative considerations, predictive AI should augment rather than replace human judgement.

Taken together, legal compliance should be understood as an integral design principle. Embedding transparency, accountability, and risk assessment directly into CRIS-based AI systems contributes to their legitimacy and supports responsible research governance.

7. Discussion

The framework presented in this paper demonstrates how predictive AI can expand CRIS systems from retrospective reporting tools toward forward-looking infrastructures for research policy analysis. By combining predictive modelling with simulation-based approaches, the system enables the exploration of alternative policy scenarios and their potential consequences.

However, predictive AI should not be interpreted as providing definitive answers. Model outputs are inherently shaped by data quality, modelling assumptions, and uncertainty. Their value lies in

supporting structured reflection and making potential developments visible, rather than prescribing decisions.

The utility of predictive AI is determined not solely by technical performance, but also by its usability, interpretability, and the degree of institutional trust it commands. Lack of design elements away from these principles point to a challenge relevant to practical adoption rather than technical and legal feasibility. For such systems to be effectively implemented, decision-makers must be able to comprehend model outputs and incorporate them into established workflows. Accordingly, the successful deployment of predictive AI requires the alignment of technical solutions with existing organizational practices. This framework also emphasizes that technical, social, and legal dimensions are closely interconnected. Issues such as bias, fairness, and regulatory compliance cannot be treated separately but must be addressed within a unified design approach. This reinforces the understanding of predictive AI as a socio-technical system embedded in governance structures.

An important consideration concerns the governance of adaptive feedback mechanisms. Although continuous learning enables predictive models to evolve over time, uncontrolled feedback loops may reinforce erroneous predictions, historical biases, or AI-generated hallucinations. Consequently, feedback data should only be incorporated following appropriate validation procedures, including human oversight, data quality assessment, and governance-based approval. Such safeguards are essential to maintain the reliability, transparency, and legal compliance of predictive AI within CRIS environments. Finally, the conceptual nature of the framework represents a limitation. While it provides a structured foundation, empirical validation is required to assess its practical feasibility and impact. Future work should therefore focus on pilot implementations and user-centered evaluations.

8. Conclusion

This paper has introduced a conceptual framework for integrating predictive AI into CRIS systems to support socially-aware and legally-compliant research policy. By combining machine learning, causal modelling, agent-based simulation, and NLP, the framework enables the analysis of policy scenarios across social, institutional, and legal dimensions.

The main contribution of this paper is in positioning predictive AI as a governance-oriented tool that enhances, rather than replaces, existing CRIS functionalities. The approach supports more transparent and evidence-based policy processes by enabling the exploration of potential developments and associated uncertainties. At the same time, the paper highlights that effective implementation requires the integration of technical capabilities with social and legal considerations. Data quality, fairness, explainability, and regulatory compliance are not peripheral issues but core elements of system design.

Future research should focus on implementing and evaluating the proposed framework in real-world settings. This includes developing prototypes, conducting pilot studies, and examining how predictive AI can be meaningfully integrated into institutional decision-making. Furthermore, future work should investigate robust feedback governance mechanisms, including human-in-the-loop validation, hallucination detection, data quality assessment, and continuous monitoring strategies, to ensure that adaptive AI systems remain reliable, transparent, and legally compliant throughout their lifecycle. Such efforts are essential to assess the practical value of predictive AI for adaptive, trustworthy, and responsible research policy.

References

- Ahmad, A., Vallès, Y., & Idaghdour, Y. (2026) Bias in AI systems: integrating formal and socio-technical approaches. *Front. Big Data* 8:1686452. <https://doi.org/10.3389/fdata.2025.1686452>
- Ajoodha, T. & Browne, J. (2025). Fundamental Rights Impact Assessments: What are they? How do they work? CEDPO AI and Data Working Group Micro-Insights Series January 2025. <https://cedpo.eu/wp-content/uploads/CEDPO-micro-insight-paper-fundamental-rights-impact-assessments.pdf>
- Ajunwa, I. (2021). An auditing imperative for automated hiring. *Harvard Journal of Law & Technology*, 34(2), 621–685.
- Azeroual, O., Nacheva, R., Nikiforova, A., & Störl, U. (2025). A CRISP-DM and Predictive Analytics Framework for Enhanced Decision-Making in Research Information Management Systems. *Informatica*, 49(18). <https://doi.org/10.31449/inf.v49i18.5613>
- Azeroual, O. (2025b). Beyond Black Boxes: Adaptive XAI for Dynamic Data Pipelines. In *Proceedings of the 17th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (Vol. 2, pp. 428-437). <https://doi.org/10.5220/0013736100004000>
- Bertaina, S., Biganzoli, I., Desiante, R., Fontanella, D., Inverardi, N., Penco, I.G., & Cosentini, A.C. (2025). Fundamental rights and artificial intelligence impact assessment: A new quantitative methodology in the upcoming era of AI Act. *Computer Law & Security Review*, Volume 56, 106101, ISSN 2212-473X. <https://doi.org/10.1016/j.clsr.2024.106101>
- Bisoyi, A. (2025). The EUai Act and Human Rights by Design Obligations. In *European Yearbook on Human Rights 2025*. Leiden, The Netherlands: Brill | Nijhoff. https://doi.org/10.1163/9789004751026_017
- Chiao, V. (2019). Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice. *International Journal of Law in Context*, 15(2), 126–139. <https://doi.org/10.1017/S1744552319000077>
- de Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, 39(2). <https://doi.org/10.1016/j.giq.2021.101666>
- Dubey, S., & Vachher, L. (2025). Stride Towards Fair AI: Addressing Algorithmic Prejudices in Human Resource Analytics. *Abhigyan*, <https://doi.org/10.1177/0970238525138>
- Ebers, M., Hoch, V. R., Rosenkranz, F., Ruschemeier, H., & Steinrötter, B. (2021). The European Commission's proposal for an artificial intelligence act—a critical assessment by members of the robotics and ai law society (rails). *J*, 4(4), 589-603. <https://doi.org/10.3390/j4040043>
- Gstrein, O.J., Haleem, N., & Zwitter, A. (2024). General-purpose AI regulation and the European Union AI Act. *Internet Policy Review*, 13(3). <https://doi.org/10.14763/2024.3.1790>
- Graux, H., Gryffroy, P., Gad-Nowak, M., & Boghaert, L. (2024). The role of artificial intelligence in processing and generating new data An exploration of legal and policy challenges in open data ecosystems. European Commission, Directorate-General for Communications Networks, Content and Technology, Unit G.1 Data Policy and Innovation. <https://data.europa.eu/sites/default/files/report/The%20Role%20of%20Artificial%20Intelligence%20in%20Processing%20and%20Generating%20New%20Data-1-2.pdf>
- Hedfeld, P. (2026). AI as a socio-technical actor: rethinking definitions for ethics and governance. *AI Ethics* 6, 254. <https://doi.org/10.1007/s43681-026-01123-1>
- Herrmann, T. (2026). Organizational Practices and Socio-Technical Design of Human-Centered AI. arXiv. https://doi.org/10.1007/978-981-97-8440-0_66-1
- Hillman, N., Holmes, W., & Duarte, T. (2025). A rapid review of AI literacy frameworks, with policy recommendations. Royal Society report. <https://royalsociety.org/-/media/policy/projects/ai-in-education/hillman-et-al-a-rapid-review-of-ai-literacy-frameworks.pdf>

- Karaboga, M., Runge, G., & Metzger, F. (2023). Meeting report: „Data Sharing–Datenkapitalismus by Default?“. Conference, 2023, Berlin, DE. *TATuP–Journal for Technology Assessment in Theory and Practice*, 32(3), 96-97. <https://doi.org/10.14512/tatup.32.3.96>
- Kim, B. J., Jeong, S., Cho, B. K., & Chung, J. B. (2025). AI Governance in the Context of the EU AI Act: A Bibliometric and Literature Review Approach. arXiv preprint arXiv:2502.03468. <https://doi.org/10.48550/arXiv.2502.03>
- Lalor, J. P., Abbasi, A., Oketch, K., Yang, Y., & Forsgren, N. (2024). Should fairness be a metric or a model? A model-based framework for assessing bias in machine learning pipelines. *ACM Transactions on Information Systems*, 42(4), 1-41. <https://doi.org/10.1145/3641276>
- Laux, J., & Ruschemeier H. (2025). Automation Bias in the AI Act: On the Legal Implications of Attempting to De-Bias Human Oversight of AI. *European Journal of Risk Regulation*. Published online 2025:1-16. <https://doi.org/10.1017/err.2025.10033>
- Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3313831.3376727>
- Mantelero, A. (2024) The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template, *Computer Law & Security Review*, Volume 54, 106020, ISSN 2212-473X, <https://doi.org/10.1016/j.clsr.2024.106020>
- Malgieri, G., & Santos, C. (2025). Assessing the (severity of) impacts on fundamental rights. *Computer Law & Security Review*, 56, 106113. <https://doi.org/10.1016/j.clsr.2025.106113>
- Moniz, G. C. (2026). DPIAs’ role in fundamental rights governance. *Information & Communications Technology Law*, 35(1), 1–20. <https://doi.org/10.1080/13600834.2025.2495441>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- OECD (2025). Empowering learners for the age of AI: An AI literacy framework for primary and secondary education (Review draft). OECD. Paris. <https://ailiteracyframework.org>
- Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, Volume 34, Issue 2, 101885, <https://doi.org/10.1016/j.jsis.2024.101885>
- Rebstadt, J., Remark, F., Fukas, P., Meier, P., & Thomas, O. (2022). Towards Personalized Explanations for AI Systems: Designing a Role Model for Explainable AI in Auditing. *Wirtschaftsinformatik 2022 Proceedings*. <https://aisel.aisnet.org/wi2022/ai/ai/2>
- Riesen, E. (2025). A sociotechnological-system approach to AI ethics. *AI & Soc* 40, 6231–6243. <https://doi.org/10.1007/s00146-025-02335-5>
- Sartor, G. & Lagioia, F. (2020). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, European Parliamentary Research Service (EPRS). <https://cris.unibo.it/handle/11585/763225>
- Thomaidou, A., & Limniotis, K. (2025). Navigating Through Human Rights in AI: Exploring the Interplay Between GDPR and Fundamental Rights Impact Assessment. *Journal of Cybersecurity and Privacy*, 5(1), 7. <https://doi.org/10.3390/jcp5010007>
- Trigo, A., Stein, N., & Paulo Belfo, F. (2024). Strategies to improve fairness in artificial intelligence: A systematic literature review. *Education for Information*, 40(3), 323-346. <https://doi.org/10.3233/EFI-240045>
- Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, 41, 2021, 105567. <https://doi.org/10.1016/j.clsr.2021.105567>