# Bridging Explainability and Interpretability in AI-driven SCM Projects to Enhance Decision-Making

*Taoufik El Oualidi* [1*] and *Saïd Assar* [1]

[1] Université Paris-Saclay, Univ Evry, IMT-BS, LITEM
Evry, France

**Abstract.** New AI-based systems implementation in companies is steadily expanding, paving the way for novel organizational sequences. The increasing involvement of end-users has also garnered interest in AI explainability. However, AI explainability continues to be a serious concern, particularly in conventional fields of activity where end-users play an essential role in the large-scale deployment of AI-based solutions. To address this challenge, managing the close relationship between explainability and interpretability deserves particular attention to enable end-users to act and decide with confidence.

## 1 Introduction

Explainability is a sub-field of Artificial Intelligence (AI) that involves making both the operation and the results of an AI-based system humanly understandable. It is considered to be essential for building a trustful relationship with the end-user [1-2]. Moreover, if not adequately addressed, it is seen as one of the top risks for companies involved in an AI project implementation [3].

Explainable AI is defined "(...) as AI systems that can explain their rationale to a human user, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future" [4]. Within the research and data science communities, the SHAP method [5] and the LIME method [6] have attracted significant interest due to their applicability to most AI models, including deep learning. Derived from game theory, the SHAP method makes any learning model's operation and results more understandable. The SHAP method enables understanding on two levels: (1) global, by ranking each input feature of a prediction model according to its importance and influence on the result, and (2) local, by identifying the features that specifically intervene in favor or against a specific case. The LIME method uses simpler models, such as linear regressions, to obtain a simple local approximation that describes the prediction of the complex model for a specific case.

While these methods help to improve understanding of AI models, the complexity of the results and the limited user-friendliness of their output formats can be barriers to understanding for end-users. Moreover, due to the need for responsiveness specific to their field of activity, end-users may find themselves constrained by the long-run times of these explainability methods [7].

Interpretability has thus been proposed as an alternative concept for enhancing the understandability of AI-based systems. The distinction with explainability lies in the fact that “Interpretable models are ML techniques that learn more structured, interpretable, or causal models" [4]. In other words, and in a simplified way, it is then possible to say that explainability answers the question "How does the AI model work?" while interpretability focuses on “Why does the AI model suggest such and such decisions?”. This distinction has been made by many authors [8-9], although some others use the terms "explainability" and "interpretability" in an undifferentiated way [10-13], or ultimately, insist on the pre-eminence of interpretable models, especially if automated decision-making algorithms impact humans [14-15].

This lack of consensus is due, at least in part, to the heterogeneity of the areas in which explainability is addressed. However, the definition of these key concepts is essential to the structuring and advancement of research projects. This is why we propose to explore the hypothesis that managing the synergy between explainability and interpretability can positively impact the decision-making process, and we conduct a systematic literature review for this purpose.

## 2 Data selection for literature review

Due to its complexity and the need for rapid and effective decision-making, the supply chain domain represents an ideal breeding ground for using artificial intelligence (AI). Therefore, we conduct a literature review to deepen our understanding of the state of the art in AI explainability in this field. Using the Scopus database, the following query is executed (April 2, 2024) with a limitation on the articles and focusing on

* Corresponding author: taoufik.el-oualidi@universite-paris-saclay.fr

the titles, abstracts, and keywords: TITLE-ABS-KEY ( "Explainability" OR "Explicability" OR "Explainable" OR "XAI" AND ( "Decision making" OR "Decision-making" OR "Decision" ) AND ( "Supply chain management" OR "Logistic" OR "Transport") ) AND ABS ( "AI" OR "Artificial intelligence") AND ( LIMIT-TO ( DOCTYPE , "ar" ) ). After reading and analyzing, eleven articles came closest to our research theme. The majority of these articles converge on similar conclusions: the predictions generated by an AI-based solution, as well as the ability to make how the algorithm works easier to understand, are key elements in improving the decision-making process. Most of these studies rely on a dataset that feeds the AI to generate predictions. The SHAP and LIME methods are commonly used to make the workings and results of the AI algorithm understandable.

This literature review highlights three main themes related to the explainability of AI in the field of supply chain management. First, the decision-support capability of AI-based systems is recognized as a significant asset applicable to various activities in this field, ranging from predicting ship detentions [15] to optimizing electric vehicle infrastructure [16]. Second, the predominant approach involves developing sophisticated explainability algorithms, such as SHAP and LIME, emphasizing visual representation. Third, human expertise [17-19], and tacit knowledge [20] are important factors to be considered with AI explainability. This underlines that AI explainability provides important but incomplete answers without human input.

The literature dealing with AI explainability and decision-making for SCM concerns is emerging as an essential field of study for this domain, which often operates in an uncertain and just-in-time context. Recent research shows the use of various methods to make AI-based solutions more comprehensible. This literature review highlights the approaches, methods, and limits of AI explainability in the field of SCM. These articles explore learning models (e.g., XGBoost, neural network, Random Forest) for which AI explainability methods such as SHAP and LIME are mainly used. In SCM studies, AI explainability is characterized by the desire to identify the variables that influence the predictions or, more generally, the results of an AI-based system. For example, [17] uses SHAP to evaluate the impact of anomaly codes on ship-detention decisions, while [21] proposes an explainable model with SHAP for evaluating food safety and quality.

While Figure 1. illustrates the methods currently used in SCM, other approaches, such as counterfactual [22] and example-based explanations [23], could overcome these barriers and deliver benefits to end-users. Counterfactual explanations provide users with actionable insights by pointing out the minimal changes that allow a model's decision to be modified. The operation of the AI model can be made more understandable by explicitly showing how different inputs can lead to different results. This makes it easier for users to relate the internal workings of the AI model to their working context. This method leverages familiarity and relativity, reducing the cognitive load on

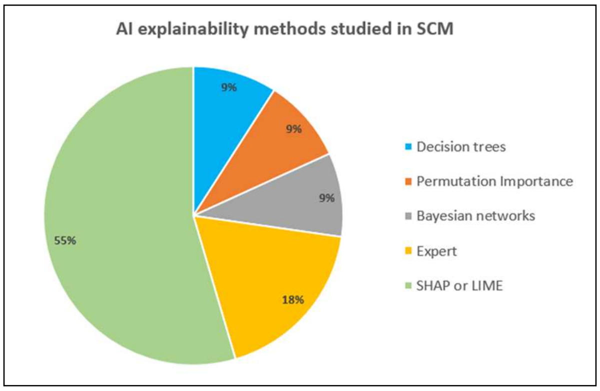


**Fig. 1.** Main AI explainability methods studied in SCM.

users trying to interpret the rationale behind AI predictions. Using familiar and relevant concepts, this method makes it easier for users to interact with AI, reducing the cognitive effort required to understand the principles of AI prediction. Similarly, by presenting specific instances that have influenced the model's decision, example-based explanations can make the abstract decision-making process more tangible. Thus, users can relate to real-world examples, which helps them understand complex model behaviors more intuitively.

Although the explainability of AI shows significant potential for improving the decision-making process in SCM, its full integration seems to require greater involvement of end users and adaptation to future hybrid classical-quantum models [24]. Therefore, our study seeks to strike a balance by exploring the end-user perspective, which implies a more global and contextual approach to enhance the integration of their requirements into the decision-making process. By leveraging the strengths of counterfactual and example-based explanations, AI-based systems can become more understandable, interactive, and adaptable to the evolving needs of users in an often uncertain SCM environment. This approach can align the development of AI-based systems with the practical requirements of end-users, ensuring that the explainability of AI becomes a practical tool that improves decision-making along the SCM value chain [25].

# 3 A complementary and aligned approach

## 3.1 Relationship between Explainability and Interpretability

The success of an IS project must be approached holistically. Its various aspects (technological, organizational, and human) must be considered, and their interactions must be managed throughout the project lifecycle [26-27]. Explainability and interpretability have become essential features of new AI-based systems, contributing to their democratization, especially in decision-support contexts of supply chain management. Whereas these two concepts are often used interchangeably, they serve different purposes.

They should interact in synergy to ease the use of AI systems and enhance end-user confidence.

Both concepts are indispensable in AI-based projects, particularly in supply chain management, to ensure trusted AI systems and effective use. Explainability allows AI developers and data scientists to fine-tune algorithms by discovering possible biases or errors. At the same time, interpretability means the assurance that decision-makers can leverage AI insights without needing deep technical knowledge of how it operates. These two elements have to be aligned in such a way that they form a system that can perform its functions accurately and with actionable, understandable outputs that would become easily applicable in real life.

### 3.2 Complementarity of AI and human expertise

Using AI-based SCM systems opens huge opportunities by automating challenging tasks like demand forecasting, inventory management, and route optimization. However, how such systems can succeed depends fundamentally on whether they work with human expertise most of the time, in cases where knowledge is silent and contextual insights are vital. Managers and logisticians often emphasize that no AI system, however advanced, can replace human intuition altogether or consider all variables in real-time decision-making [28]. Human intuition is best suited to integrate the tacit knowledge [29] that lies beyond the analysis of even finely tuned AI models in sensitive or novel situations.

This complementary relationship between AI and human expertise is nowhere more manifest than in the aspect of interpretability: AI systems can process large volumes of data and provide optimized results, but those results must be interpretable by humans if they are to take appropriate action. This synergy is particularly important in dynamic fields, such as supply chain management, where decision-makers need to rapidly assess and act on shifting conditions, such as supply and demand.

### 3.3 Visualization and Interpretability

Among the main findings of recent research into AI-based decision support systems [30] is a preference for visual formats for presenting results. Already outlined in management studies [31], this preference shows that visual representations (graphs and maps) have a positive effect on the interpretability of complex and voluminous data. This shows the trend towards more visual explanations to make AI system results more accessible and usable.

Visual representation is, therefore, a powerful means for bridging the gap between explainability and interpretability for AI. These representations make it very easy to communicate information in its simple form so that users will focus on implications from a contextualized results point of view rather than algorithms producing those results. The ability to understand this becomes even more important in SCM, where the decision maker needs to evaluate one option of AI recommendations against others as quickly as possible.

Practical Challenges and Need for Synergy: While AI explainability and interpretability are quite critical in building confidence and making the insights produced yield adequate decision-making, many problems beset their practical implementation. One of the main limitations is the periodicity at which the data feeding into an AI system has been updated. Supply chain data hardly comes in real-time [32] since most of the sources feeding an AI system do so naturally. This is because delayed data updates impede informed decisions, particularly in an environment whose conditions might change in a few minutes. This represents a significant challenge when implementing a new AI system within the SCM field.

Another challenge is that most AI systems are developed in a one-size-fits-all paradigm that fully neglects diverse user type needs [33]. While technical specialists would like detailed explainability, for end-users, high-level interpretability is often sufficient. Such different needs can be balanced only through flexible design of AI that shall include user feedback and personalization of explanations according to the specific context and user requirements.

## 4 Explainability, Interpretability, and bounded rationality

To support this need to reconcile explainability and interpretability within a coherent framework of actions, the concept of bounded rationality reminds us that individuals operate under cognitive and informational constraints [34]. Thus, while new AI-based systems enable the processing of large volumes of data through an automated learning process, they may paradoxically introduce new forms of complexity and opacity for end-users. For example, an automated learning system makes recommendations by incorporating a carrier's new pricing that was updated an hour ago. This situation could either reduce or increase the cognitive and informational constraints on the end-user. In fact, the decision patterns used by logisticians to plan the least-cost transportation of goods could be less efficient due to the lack of real-time interpretability of AI recommendations. Thus, although the explainability of AI remains unchanged (because its operation is always the same), its relationship to interpretability is evolving. Therefore, the challenge is to present this evolution to the end-user so that they can act with confidence in the best interests of the enterprise.

This example highlights the need to understand how AI works (an aspect of explainability) and why results or recommendations are suggested (an aspect of interpretability). In a context as dynamic as SCM, where real-time adjustments are often required, the ability to make the AI system's results humanly understandable becomes a source of competitive advantage for the company. Therefore, if explainability and interpretability are to be considered in a complementary way, the contextualization of results [35] becomes an essential element for end-users. An operational contextualization of the results then enables end-users to make a link between the reality on the ground and the company's goals. This gives the AI system an additional advantage: anchoring trust in everyday practice.

## 5 Proposal for a theoretical model

Through a theoretical model (Fig. 2), we emphasize the close relationship between explainability and interpretability in order to understand how these elements influence each other when it comes to developing end-user confidence.

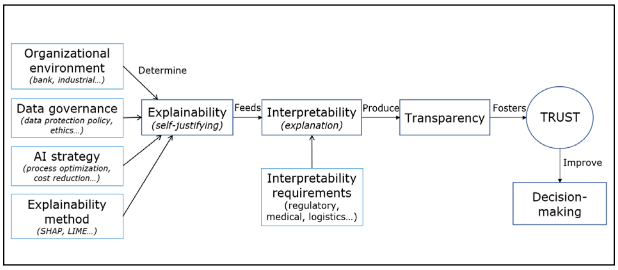


**Fig. 2.** Impact of explainability on decision-making: a theoretical model.

This model could be useful in regulated or complex sectors such as banking, healthcare, and logistics, where decisions have to be justified and understandable by end-users with no in-depth technical expertise. Therefore, this model provides a framework for guiding the design, development, and implementation of AI systems by considering requirements expressed by end-users and providing visualization of such concepts.

To ensure the success of AI-based projects, a well-calibrated combination of explainability and interpretability is required:

• Explainability (self-justifying): "In the field of artificial intelligence, explainability is the ability to relate and make understandable the elements that an AI system considers to produce a result" [36]. Key variables such as the organizational environment, data governance, AI strategy, and explainability methods (e.g., SHAP, LIME) contribute to determining explainability and making the operation of an AI model comprehensible. These variables clarify the underlying explainable AI process that led to a decision.

• Interpretability (Explanation): Distinct from explainability, interpretability is closely related to explanation [37]. The interpretability represents the way in which users perceive and understand the results produced by the AI model. Interpretability requirements, such as regulatory or healthcare constraints, influence this process. Interpretability focuses more on the "why" behind the decisions made by AI.

• Transparency: It results from the combination of explainability and interpretability. More transparency [38] enables users to understand how and why decisions are made, creating an environment where they can trust the AI-based system.

• Trust and decision-making: The level of trust stakeholders develop during their algorithmic interactions is a qualitative outcome [39]. Transparency enhances users' trust in the AI-based system, which improves decision-making by enabling them to rely on AI results more confidently.

The proposed theoretical model requires to be tested to confirm its validity. Therefore, several approaches can be considered to test the model's validity:

• Empirical validation: Case studies in different areas such as banking, logistics, and healthcare could be carried out to observe how the combination of explainability-interpretability is shaped and influences trust and decision-making. These studies could analyze AI-based systems deployed in real environments to verify the validity of the model's hypotheses.

• User-centered approach: This model can be enriched with information regarding users' specific requirements [40]. For instance, some users require visual explanations, while others want more detailed and technical justifications. A user-oriented model would allow AI systems to adapt better to different user profiles.

• Updating explainability methods: Explainability methods such as SHAP and LIME are fast-developing. New tools shall be put into the model to ensure they meet the interpretability requirements for various contexts.

• Regulatory compliance: Add increasing regulatory requirements of AI system transparency, such as the new European AI Act [41], and test whether the model can meet the latest constraints in transparency and trust.

## 6 Conclusion

A well-calibrated combination of explainability and interpretability is required to implement AI-based projects that satisfy all stakeholders successfully. The explainability of AI guarantees better technical knowledge without systematically ensuring a perfect and complete mastery of the algorithmic operation. Combined with interpretability, this ensures that AI results are actionable, relevant, and meaningful for end-user decision-making. Therefore, the proposed conceptual model provides a foundation for a global response to the challenges of AI explainability. This approach provides the elements needed to improve decision-making in complex environments such as supply chain management. As AI continually evolves and improves, maintaining a dynamic balance between explainability and interpretability will be one of the essential elements an organization must consider to fully appreciate its investments in AI.

## References

[1] E. Glikson et A. W. Woolley, « Human trust in artificial intelligence: Review of empirical research », *Academy of Management Annals*, vol. 14, nº 2, p. 627-660, 2020.

[2] K. A. Hoff et M. Bashir, « Trust in automation: Integrating empirical evidence on factors that influence trust », *Human factors*, vol. 57, nº 3, p. 407-434, 2015.

[3] « AI Index Report 2023 – Artificial Intelligence Index ». [En ligne]. Disponible sur: https://aiindex.stanford.edu/report/

[4] D. Gunning et D. Aha, « DARPA's explainable artificial intelligence (XAI) program », *AI magazine*, vol. 40, nº 2, p. 44-58, 2019.

[5] S. M. Lundberg et S.-I. Lee, « A unified approach to interpreting model predictions », in *Proceedings of the 31st international conference on neural information processing systems*, 2017, p. 4768-4777.

[6] M. T. Ribeiro, S. Singh, et C. Guestrin, « " Why should i trust you?" Explaining the predictions of any classifier », in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, p. 1135-1144.

[7] E. Doumard, J. Aligon, E. Escriva, J.-B. Excoffier, P. Monsarrat, et C. Soulé-Dupuy, « A comparative study of additive local explanation methods based on feature influences », in *24th International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data ((DOLAP 2022)*, CEUR-WS. org, 2022, p. 31-40. [En ligne]. Disponible sur: https://hal.science/hal-03687554/

[8] A. B. Arrieta *et al.*, « Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI », *Information Fusion*, vol. 58, p. 82-115, 2020.

[9] L. Longo, R. Goebel, F. Lecue, P. Kieseberg, et A. Holzinger, « Explainable artificial intelligence: Concepts, applications, research challenges and visions », in *International Cross-Domain Conference for ML and Knowledge Extraction*, Springer, 2020, p. 1-16.

[10] A. Adadi et M. Berrada, « Peeking inside the black-box: a survey on explainable artificial intelligence (XAI) », *IEEE access*, vol. 6, p. 52138-52160, 2018.

[11] P. Hase et M. Bansal, « Evaluating explainable AI: Which algorithmic explanations help users predict model behavior? », *arXiv preprint arXiv:2005.01831*, 2020.

[12] Z. C. Lipton, « The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. », *Queue*, vol. 16, nº 3, p. 31-57, 2018.

[13] T. Miller, « Explanation in artificial intelligence: Insights from the social sciences », *Artificial intelligence*, vol. 267, p. 1-38, 2019.

[14] F. Doshi-Velez et B. Kim, « Towards a rigorous science of interpretable machine learning », *arXiv preprint arXiv:1702.08608*, 2017.

[15] C. Rudin, « Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead », *Nature Machine Intelligence*, vol. 1, nº 5, p. 206-215, 2019.

[16] V. Sumanasena *et al.*, « Artificial Intelligence for Electric Vehicle Infrastructure: Demand Profiling, Data Augmentation, Demand Forecasting, Demand Explainability and Charge Optimisation », *Energies*, vol. 16, nº 5, p. 2245, 2023.

[17] R. Yan et S. Wang, « Ship detention prediction using anomaly detection in port state control: model and explanation », *Electron. Res. Arch*, vol. 30, p. 3679-3691, 2022.

[18] J. Hangl, S. Krause, et V. J. Behrens, « Drivers, barriers and social considerations for AI adoption in SCM », *Technology in Society*, vol. 74, p. 102299, 2023.

[19] E. Rajabi, S. Nowaczyk, S. Pashami, M. Bergquist, G. S. Ebby, et S. Wajid, « A Knowledge-Based AI Framework for Mobility as a Service », *Sustainability*, vol. 15, nº 3, p. 2717, févr. 2023, doi: 10.3390/su15032717.

[20] F. Olan, K. Spanaki, W. Ahmed, et G. Zhao, « Enabling explainable artificial intelligence capabilities in supply chain decision support making », *Production Planning & Control*, p. 1-12, févr. 2024, doi: 10.1080/09537287.2024.2313514.

[21] S. Bhatia et A. S. Albarrak, « A blockchain-driven food supply chain management using QR code and XAI-faster RCNN architecture », *Sustainability*, vol. 15, nº 3, p. 2579, 2023.

[22] S. Verma, J. Dickerson, et K. Hines, « Counterfactual explanations for machine learning: A review », *arXiv preprint arXiv:2010.10596*, vol. 2, 2020, [En ligne]. Disponible sur: https://ml-retrospectives.github.io/neurips2020/camera_ready/5.pdf

[23] J. van der Waa, E. Nieuwburg, A. Cremers, et M. Neerincx, « Evaluating XAI: A comparison of rule-based and example-based explanations », *Artificial intelligence*, vol. 291, p. 103404, 2021.

[24] M. A. Jahin, M. S. H. Shovon, M. S. Islam, J. Shin, M. F. Mridha, et Y. Okuyama, « QAmplifyNet: pushing the boundaries of supply chain backorder prediction using interpretable hybrid quantum-classical neural network », *Scientific Reports*, vol. 13, nº 1, p. 18246, 2023.

[25] S. Laato, M. Tiainen, A. K. M. Najmul Islam, et M. Mäntymäki, « How to explain AI systems to end users: a systematic literature review and research agenda », *Internet Research*, vol. 32, nº 7, p. 1-31, janv. 2022, doi: 10.1108/INTR-08-2021-0600.

[26] W. H. DeLone et E. R. McLean, « Information Systems Success: The Quest for the Dependent Variable », *Information Systems Research*, vol. 3, nº 1, p. 60-95, mars 1992, doi: 10.1287/isre.3.1.60.

[27] « The DeLone and McLean Model of Information Systems Success: A Ten-Year Update », *Journal of Management Information Systems*, vol. 19, nº 4, p. 9-30, avr. 2003, doi: 10.1080/07421222.2003.11045748.

[28] T. El Oualidi et S. Assar, « Does AI Explainability Meet End-Users' Requirements? Insights From A Supply Chain Management Case Study », *ECIS 2024 TREOS*, juin 2024, [En

ligne]. Disponible sur: https://aisel.aisnet.org/treos_ecis2024/43

[29] M. Polanyi, « The tacit dimension », in *Knowledge in organisations*, Routledge, 2009, p. 135-146.

[30] A. J. Karran, T. Demazure, A. Hudon, S. Senecal, et P.-M. Léger, « Designing for Confidence: The Impact of Visualizing Artificial Intelligence Decisions », *Front. Neurosci.*, vol. 16, juin 2022, doi: 10.3389/fnins.2022.883385.

[31] K. Eberhard, « The effects of visualization on judgment and decision-making: a systematic literature review », *Manag Rev Q*, vol. 73, nº 1, p. 167-214, févr. 2023, doi: 10.1007/s11301-021-00235-8.

[32] S. Lechler, A. Canzaniello, B. Roßmann, H. A. von der Gracht, et E. Hartmann, « Real-time data processing in supply chain management: revealing the uncertainty dilemma », *International Journal of Physical Distribution & Logistics Management*, vol. 49, nº 10, p. 1003-1019, janv. 2019, doi: 10.1108/IJPDLM-12-2017-0398.

[33] « One Size Does Not Fit All », Center for Security and Emerging Technology. [En ligne]. Disponible sur: https://cset.georgetown.edu/publication/one-size-does-not-fit-all/

[34] H. A. Simon, « Bounded rationality in social science: Today and tomorrow », *Mind & Society*, vol. 1, nº 1, p. 25-39, mars 2000, doi: 10.1007/BF02512227.

[35] L. Waardenburg et M. Huysman, « From coexistence to co-creation: Blurring boundaries in the age of AI », *Information and Organization*, vol. 32, nº 4, p. 100432, 2022.

[36] « Explicabilité (IA) ». [En ligne]. Disponible sur: https://www.cnil.fr/fr/definition/explicabilite-ia

[37] O. Biran et C. Cotton, « Explanation and justification in machine learning: A survey », in *IJCAI-17 workshop on explainable AI (XAI)*, 2017, p. 8-13.

[38] U. Ehsan, Q. V. Liao, M. Muller, M. O. Riedl, et J. D. Weisz, « Expanding explainability: Towards social transparency in ai systems », in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, p. 1-19.

[39] Y. Li et J. Hahn, « Review of Research on Human Trust in Artificial Intelligence », 2022, Consulté le: 9 décembre 2023. [En ligne]. Disponible sur: https://aisel.aisnet.org/icis2022/ai_business/ai_business/8/

[40] K. Bauer, M. von Zahn, et O. Hinz, « Expl(AI)ned: The Impact of Explainable Artificial Intelligence on Users' Information Processing », *Information Systems Research*, vol. 34, nº 4, p. 1582-1602, déc. 2023, doi: 10.1287/isre.2023.1199.

[41] « AI Act | Shaping Europe's digital future ». [En ligne]. Disponible sur : https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai