

FASTER-AI: A Comprehensive Framework for Enhancing the Trustworthiness of Artificial Intelligence in Web Information Systems

Christos Troussas^a, Christos Papakostas^b, Akrivi Krouska^c, Phivos Mylonas^d
and Cleo Sgouropoulou^e

*Department of Informatics and Computer Engineering, University of West Attica, Egaleo, Greece
{ctrouss, cpapakostas, akrouska, mylonasf, csgouro}@uniwa.gr*

Keywords: Trustworthy AI, Web Information Systems, Fairness and Bias Mitigation, Explainability and Transparency, Security and Privacy, Robustness, Ethical AI, AI Implementation Framework, AI Evaluation Roadmap, Case Studies in AI, FASTER-AI.


Abstract: With increasing embedding of artificial intelligence (AI) in web information systems (WIS), the maximum assurance on the reliability of such AI systems is solicited. Although this aspect is gaining importance, no comprehensive framework has yet been developed to ensure AI reliability. This paper aims to bridge that gap by proposing the AI FASTER framework to enhance the reliability of AI in WIS. The key dimensions of concern within the framework are FASTER-AI: Fairness/bias mitigation, explainability/transparency, security/privacy, robustness, and ethical considerations/accountability. Each one guides in precisely the area where trust shall be accomplished: a decrease in bias, model interpretability, protection of data, resilience of models, and ethics in governance. The implementation methodology for these dimensions involves preliminary assessment, planning, integration, testing, and continuous improvement. Validation of proof for FASTER-AI was created based on in-depth case studies across different verticals: e-commerce, finance, health, and fraud detection. This work has demonstrated how FASTER-AI is applied through illustrative case studies showing promising performance. From the initial results of high improvement in terms of fairness, transparency, security, and robustness, it may be effectively inferred that FASTER-AI can be successfully applied.


1 INTRODUCTION


Web Information Systems (WIS) have grown to become the backbone of the digital space, without which the management, storage, and delivery of information over the internet would not be possible. Applications range from e-commerce websites, social media networks, online education, content management systems, etc., basing on seamless data flow and user interaction (Bhutani & Mittal, 2023; Ge et al., 2023; Huang, 2022; Kardaras et al., 2024; Troussas et al., 2015, 2022; Virvou et al., 2012; Zenkert & Fathi, 2023). Efficiency and reliability


become crucial aspects of WIS, being that such issues relate to the experience of users, data accessibility, and overall functionality important to billions of people based on daily internet service.


Artificial Intelligence (AI) being integrated into WIS, does bring a new change in the internal operation methodology. WIS can be extended through AI technologies for improved ability to automate complex tasks and personalize user experience more efficiently in the optimization of content delivery methods (Ganesh & Rastogi, 2023; Krouska et al., 2020). Examples include recommendation systems, e-commerce with the power of AI-driven algorithms in e-commerce,

^a  <https://orcid.org/0000-0002-9604-2015>

^b  <https://orcid.org/0000-0002-5157-347X>

^c  <https://orcid.org/0000-0002-8620-5255>

^d  <https://orcid.org/0000-0002-6916-3129>

^e  <https://orcid.org/0000-0001-8173-2622>

content filtering on social media, and automated customer support on various online services. It enables a system to improve operating efficiency and, more importantly, fosters innovation in driving the empowerment of Web information systems to offer more intelligent, responsive services.

In particular, modern web information systems are increasingly dependent on AI, further complicating the critical concern of trustworthiness. Fairness, transparency, security, robustness, and ethics were assumed within the frame of the AI when the AI systems take decisions based on inputs to directly affect the user's decisions, for instance, product recommendations, content moderation, or management of private sensitive personal data in (Martin, 2022). Only truthful AI can be maintained for user trust, and it will also go on to serve in meeting regulatory requirements to protect against such harms due to biased decision-making, data breach, or actions which are not ethically right. Without a solid foundation of trust, the potential benefits arising from AI within web information systems cannot be realized fully, mainly since the associated risks in decisions taken by AI may also crush the integrity and credibility of such systems.

The literature has matured well on the issue of AI trustworthiness, with key dimensions such as fairness, explainability, security, robustness, and ethics. For example, Mehrabi et al. (2021) discussed AI bias and fairness, which proposed methods for bias detection and mitigation. For instance, in the aspect of explainability, some of these advance to better steps of constructing the AI-interpretable models, like LIME and SHAP when announced in 2016 by Ribeiro et al. It is a security loophole, as described in the research topic on adversarial attacks and defenses by Carlini et al., 2019. Another proposition to harden AI models was made by Goodfellow et al. in 2015 and later modified by Madry et al. in 2017, but this adversarial training has remained challenging for real-time application in web information systems.

Several frameworks do attempt at ensuring trustworthiness of AI but are scoped only to narrow levels. Floridi et al. (2018) developed an ethical framework for AI and robotics; however, it is devoid of concrete technical guidelines for WIS. Shahriari & Shahriari (2017) submitted ethical guidelines for AI development; however, these omitted the issues that are specific to WIS. Rieke et al. (2020) proposed a GDPR-compliant, privacy-preserving AI framework dealing largely with data privacy aspects.

The earlier works, while useful, often address individual trust dimensions such as fairness in Agarwal et al. (2018) or transparency in Doshi-Velez

and Kim (2017) without placing them into context with respect to WIS. Works such as Veale & Binns 2017 and Danks & London 2017 discuss fairness, accountability, and social impact brought about by AI but do not go far enough to offer a more holistic and multi-dimensional framework aimed at WIS. Yet other proposals articulate high-level ethical considerations but do not incorporate within the proposals themselves low-level technical safeguards, such as security and robustness specific to WIS.

In turn, corollary to the pervasiveness of AI in web information systems today, there is thus a pressing need for an integrated framework surrounding concerns across all facets of trustworthy AI in such settings. The paper proposes the FASTER-AI framework related to five main dimensions: Fairness and Bias Mitigation, Explainability and Transparency, Security and Privacy, Robustness, and Ethical Considerations and Accountability. FASTER-AI has been designed to populate practical guidelines for the organizations and developers on how to make their AI systems in WIS more trustworthy. The framework provided by FASTER-AI can be implemented to construct AI-driven WIS whose performance is optimal using credible models, and to ensure the best achievable level of trust and dependability.

2 FASTER-AI FRAMEWORK FOR TRUSTWORTHY AI IN WIS

Although significant breakthroughs have been realized in fast-tracking AI integration into the WIS and automation of jobs for improved user experiences and processes, at the same time, it tends to lead to several concerns about dependability of AI systems related to fairness, transparency, security, robustness, and ethics. In view of the above challenges, we propose a framework built upon insights and approaches detailed in the literature presented in Section 2 that could enable further extension of work on development and evaluation of trustworthy AI in WIS. Figure 1 presents the proposed framework over five dimensions: Fairness and Bias Mitigation; Explainability and Transparency; Security and Privacy; Robustness; and Accountability and Ethics. These start with the first letter of the dimensions: Fairness, Accountability, Security, Transparency, Ethical considerations, Robustness-adding AI to close off to the acronym with its full meaning underlining, as said, the focus on fostering

trustworthiness within a given context, in particular that of artificial intelligence systems. Each of these dimensions contributes much to the overall credibility of the AI system and, when integrated, provides pragmatic frameworks for organizations to operationalize AI in online contexts.



Figure 1: FASTER-AI Structure.

The first dimension, Fairness and Bias Mitigation, is about the need for the AI systems to avoid discriminatory outcomes. For WIS, where choices made by AI could involve a diverse user base, fairness will be critical in maintaining public trust and meeting legal obligations. FASTER-AI emphasizes that ongoing surveillance and evaluation of AI models have to be performed in order to continuously identify biases and take action against them accordingly. Some of the strategies ensuring the provision of justice to all groups of users could include fairness-aware learning algorithms, re-weighting the training datasets, and modifications in post-processing. Furthermore, fairness audits should be built into the AI lifecycle of development to continually evaluate and update models with ever-exposed data and scenarios.

The second dimension refers to making clear and explaining how artificial intelligence systems improve the understandability of the latter for the end-users and stakeholders. In the context of WIS, such decisions are likely to have far-reaching repercussions on users, with the transparency and interpretability of AI in such decisions becoming paramount. In general, FASTER-AI encourages the use of interpretable models or tools that explain how the black-box models work, such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP). On the other hand, explanation interfaces should be designed

in a user-centered way because they are insightful interfaces that transform intelligible and understandable forms to the needs of technical experts, end-users, and regulatory entities. Faster-AI makes its software sold transparently, hence gaining user trust and propping more informed decision-making for all stakeholders.

The third dimension of our framework is Security and Privacy. This imperative relates to protection from AI systems against intrusion of the destructive nature that may affect user data confidentiality. Since the information being handled in WIS is sensitive, strict security mechanisms need to be implemented to ensure security. Because of this, FASTER-AI encourages adversarial training for safeguarding AI models from bad inputs performing privacy-preserving operations such as differential privacy, federated learning, and homomorphic encryption. It accomplishes these through methods that ensure AI systems operate securely and privately in decentralized settings or where there is a high risk of breaching the data. Further, FASTER-AI identifies encryption mechanisms and access controls that impede illegitimate persons from accessing sensitive information, as well as compliance with the concerned protection regulations such as GDPR or CCPA.

Robustness, being the fourth dimension, aims at assuring an AI system's reliability and effectiveness, regardless of the range of conditions in WIS. These models should be robust to distributional shifts, outliers, and adversarial circumstances that could render them unpredictably harmful. FASTER-AI advocates for increased testing on edge cases, which the methodologies of cross-domain generalization and transfer learning can help with. Similarly, the constant monitoring and retraining of artificial intelligence models to adapt them to shifting data distributions and operational environments are recommended so that such systems remain robust and trustworthy throughout their life cycle.

The fifth dimension, Ethical Considerations and Accountability, is a very important parameter with respect to the integrity and social acceptability of AI systems. FASTER-AI embeds ethical considerations such as fairness, justice, and respect for individual autonomy into the design and operation of AI models. It also places great emphasis on accountability structures such as full documentation, audit trails, and organizational oversight bodies. Such initiatives ensure that artificial intelligence systems are in bounds of fixed parameters of ethics, and any problem or non-conformity may quickly be brought under the control of the concerned authority.

Additionally, FASTER-AI also recommends the artificial intelligence be governed by frameworks created through ethics committees, which can continuously provide oversight and assurance that AI systems conform to the moral values of the communities they serve.

Faster-AI embeds multiple dimensions of AI trustworthiness, including fairness, transparency, security, and robustness, into one framework that is important for WIS. It also underlines the fact that an end which is expected to be fair needs transparency and robust security protocols relevant to privacy protection. With this approach, it is possible for organizations to ensure state-of-the-art standards for AI systems while at the same time aligning with citizen values.

With such a wide variety of users and dynamic data, WIS introduces a very special set of challenges into the world of AI systems, which should be proficient at decision-making and trustworthy. FASTER-AI addresses this issue by developing appropriate metrics for each of these dimensions—fairness, transparency, and security—which permit periodic assessments and their alignment with ethical standards. This framework should be agile and adaptable to the needs in various sectors, like healthcare or e-commerce, and scalable for organizations. In this regard, FASTER-AI will enable organizations to build trust and credibility in their AI systems by being compliant with regulations and societal expectations.

The major steps of the FASTER-AI implementation methodology include: an assessment phase that audits the existing AI systems for their fairness, transparency, and security by aligning with the dimensions of FASTER-AI. The planning and design phase is next, during which the tools required, AI Fairness 360 and SHAP, would be identified in a manner to ensure the involvement of stakeholders.

Next comes the integration phase, where cross-functional teams—data scientists and lawyers—come together to apply these tools. A wide range of datasets and scenarios during the testing and validation phase ensure that AI models are nondiscriminatory, robust, and transparent. In this iteration process, users perform ongoing monitoring and improvement through KPIs and real-time dashboards; issues that emerge from it are identified and fixed.

The last step involves reporting and communication, through which updates would continue to come in via reports, audits, and communication strategies, thereby building trust in users for the long term about the technical and ethical standards followed.

3 ROADMAP FOR EVALUATION AND PRELIMINARY RESULTS

In the operation to assess how well FASTER-AI is able to enhance the trustworthiness of AI systems in WIS, there needs to be an efficient evaluation strategy in place. This section presents a general evaluation framework for FASTER-AI and the central dimensions: fairness, transparency, security, robustness, and ethical accountability. Finally, the interim outcomes of some first case studies are analyzed to provide preliminary views concerning the practical impact of FASTER-AI.

3.1 Evaluation Roadmap

However, the value of FASTER-AI has to be tested within several real-world settings where the AI systems are crucial. It measures best the impact caused by the implementation of FASTER-AI within each one of these dimensions.

Fairness is evident in the AI-driven decisions when they affect people from different demographic groups. For example, the fairness evaluation of an e-commerce recommendation system can be done based on the distribution of recommendations in different user segments, using Disparate Impact Ratio and Demographic Parity metrics. It shall be able to demonstrate if FASTER-AI efficiently removed biases and promoted fairness.

The transparency domain in financial services will be a focal point in this research where the interpretability of the decision by AI can either be important or very important. This will consist of deploying mechanisms that provide transparency behind the AI decisions and improve the AI decisions' interpretability to end-users. Associated domains with such measures may include customer satisfaction ratings, disputes, regulatory compliance, among other indicators of performance.

This would imply security and privacy tests under environments dealing with sensitive information; say, healthcare institutions. In this case, evaluation will be done on how much potential FASTER-AI has in protecting data, factoring in protection measures, including encryption, differential privacy, and federated learning. This would have been evaluated by the number of data breaches, testing robustness of the security protocols against simulated attacks, and also gauging data privacy stakeholder confidence.

Robustness will be studied through controlled experiments that will expose the AI models to adversarial conditions, such as data corruption or malicious attacks. Comparison of error rate and

resilience of the models before and after applying the strategies for improving robustness recommended by FASTER-AI will enable insight into the framework's capacity for improving stability and reliability under stressful conditions.

Governance frameworks will help organizational ethical accountability with the constitution of AI ethics boards. The effectiveness of governance frameworks will be measured on the extent to which they solve ethical dilemmas, introduce transparency in the decision-making process, and maintain compliance with set ethical standards.

3.2 Preliminary Results

Preliminary case study-based applications of FASTER-AI give initial evidence on its potential impact, and the case studies show exactly how, where, and what benefits FASTER-AI is likely to work in practice. First, these cases apply to the very nature of the different industries; hence applicability and the effectiveness of the proposed framework will be different in some characteristics. Specifically, the electronic commerce site at stake is one of the world's largest international retailers, with an extremely varied client base, and it runs a state-of-the-art recommendation engine. In the financial case study, FASTER-AI was piloted on a long-established institution that is highly regulatory inquisitive and services a wide and diverse customer base. Finally, the health-related case study was based on an average-sized care provider heavily invested in AI-supported clinical decision support systems. The final assessed dimensions were the robustness and ethical accountability for the AI-driven fraud detection system of a financial institution.

3.2.1 Fairness in E-Commerce Recommendation Systems

In an e-commerce platform's recommendation system, the integration of fairness measures aimed at improving the equitable distribution of recommendations was tested. Over a three-month period, the system was monitored using demographic parity and disparate impact ratio metrics. As shown in Table 1, the demographic parity improved by 20%, suggesting a more balanced representation of different user groups in the recommendations. Additionally, customer feedback indicated a 15% reduction in complaints related to perceived biases, signalling an increase in user satisfaction with the fairness of the system.

Table 1: Fairness Evaluation in E-Commerce Recommendation System.

Metric	Before Implementation	After Implementation	% Change
Demographic parity	0.65	0.85	+20%
Customer Complaints (monthly)	120	102	-15%

3.2.2 Transparency in Financial AI Systems

Within finance, protocols for explainability made by FASTER-AI were used on a credit scoring system applied at a leading financial institution. This was included for the combination of explainability tools that would enable customers to fathom factors affecting their credit score. Evaluation to include customer satisfaction via surveys and disputes on loan decisions. As would be seen in Table 2, these initial results amount to a 15% reduction in disputes. This also tended to coincide with improved levels of customer satisfaction ratings. These would therefore suggest that transparency initiatives increased user confidence and contributed toward better regulatory outcomes.

Table 2: Transparency Evaluation in Credit Scoring System.

Metric	Before Implementation	After Implementation	% Change
Customer Disputes (monthly)	200	170	-15%
Customer Satisfaction Score	7.2/10	8.4/10	+1.2%

3.2.3 Security and Privacy in Healthcare AI

The security aspect of FASTER-AI was evaluated within the context of healthcare, where artificial intelligence models are used for clinical decision-making support. The tests included state-of-the-art encryption methods that are differential privacy mechanisms for the protection of patient data. There were no instances of leakage within the test period. Therefore, this is the representation of how effective the security methods adopted are. According to medical professionals and patients through interviews and surveying, increase in trust was realized when sensitive protection information was catered to. Therefore, the measures undertaken, though resource-intensive as shown in Table 3, paid off-for

there were no incidents of security, and stakeholders gave favourable assessments.

Table 3: Security Evaluation in Healthcare AI System.

Metric	Before Implementation	After Implementation	% Change
Data Breaches (reported)	2	0	-100%
Stakeholder Confidence Level	6.8/10	8.5/10	+1.7

3.2.4 Robustness in Fraud Detection Systems

A controlled experimental framework is developed to investigate the robustness of the AI model applied in fraud detection. Adversarial training recommendations, with respect to FASTER-AI, are translated into practice under iterative changes, considering multiple attack scenarios.

The results in Table 4 clearly indicated that the model was indeed more resilient by 30%, with lower error rates in an attack environment. These results suggest FASTER-AI might be useful to improve significantly the robustness of artificial intelligence systems. Further testing in a variety of applications is required to confirm these findings.

Table 4: Robustness Evaluation in Fraud Detection System.

Metric	Before Implementation	After Implementation	% Change
Error Rate Under Attack	12%	8%	-30%
Recovery Time (minutes)	15	10	-33%

3.2.5 Ethical Accountability in Financial Institutions

Ethics accountability was done by developing an AI ethics committee in a financial organization. This committee's job was essentially to second-guess AI-based decisions, especially the very sensitive ones such as loan approvals. Early indications were that input from the committee resulted in a more consistent and explainable decision-making process. As seen in Table 5, some of the ethical quagmires were more easily avoided, and the organization had far fewer external complaints about the appropriateness of AI-led decisions. But the actual effectiveness of those governance frameworks over

time will depend on their absorption into the greater organizational culture.

Table 5: Ethical Accountability Evaluation in Financial Institution.

Metric	Before Implementation	After Implementation	% Change
Ethical Dilemma Revolution Time	5 days	3 days	-40%
External Complaints (monthly)	15	9	-40%

Preliminary results from the deep-dive case studies have given a few important pointers about how effective FASTER-AI is for each of the aspects of AI trustworthiness. In the case of an e-commerce platform, improved demographic parity combined with reduced customer complaints points to improved fairness in AI-based recommendation systems brought about by FASTER-AI. This implies that by lowering the intrinsic biases, businesses could increase consumer satisfaction while possibly increasing participation by a wider consumer base.

The reduction in customer complaints after transparency tools were implemented as part of the credit scoring financial institutions mechanism demonstrates well how FASTER-AI makes decisions by artificial intelligence explainable and acceptable. This has remarkable importance within the regulatory framework, since often the requirements involve transparency—the fact that FASTER-AI might help financial institutions get closer to compliance with the obligations while building customers' trust.

As the healthcare case study, the security protocols recommended by FASTER-AI were able to remain stable against data breaches and increase stakeholder confidence. Even though implementation required high resource involvement, the lack of adverse security incidences together with positive feedback from the stakeholders underpins the role which such holistic security measures play in sensitive industries such as healthcare. In this vein, the system's frailty analyses evidenced very low error rates after simulated attacks on fraud detection, suggesting that FASTER-AI recommendations for adversarial training of the poison or trigger models are much more likely to result in stronger AI.

Overall, these results should be particularly encouraging in high-stakes settings where AI models are often under attack by adversaries.

Finally, the setting-up of an AI ethics committee within the bank supported these drivers because it

allowed more standardized and transparent implementation on ethical conflict resolution practices. It showed thus the possibility of FASTER-AI to institutionalize ethical responsibility. The decrease in public grievances finally signals that governance mechanisms along the lines should/can be crucial to build confidence in AI-supported rule-deliberation.

4 CONCLUSIONS AND FUTURE WORK

This paper proposes the FASTER-AI framework in order to enhance AI trustworthiness in WIS along the dimensions of fairness, transparency, security, robustness, and ethical accountability. First case studies conducted in various sectors have demonstrated that the adoption of FASTER-AI enhances the reliability of AI, since it provides higher fairness, explainability, security, and ethical trust. However, these initial findings relate to small samples, and further research is expected to validate the adaptation of FASTER-AI into larger and more complex settings.

As AI evolves, so does FASTER-AI, bound to proliferate with emerging challenges. Above all, an effective collaboration between academic institutions, industry players, and the regulators is very instrumental in establishing common metrics and standards for its evaluation. Long-term implications of deploying FASTER-AI on organizational change, user trust, and regulatory compliance would, therefore, be an area of future research, possibly through longitudinal studies.

Concluding, FASTER-AI contributes to the debate on trustworthy AI by delivering a real-world framework for WIS and hence laying the foundation for creating and maintaining trust in AI systems; trust will increasingly be necessary for efficacy and societal acceptance of AI technologies.

REFERENCES

- Agarwal, A., Beygelzimer, A., Dudík, M., Langford, J., & Wallach, H. (2018). A reductions approach to fair classification. *International Conference on Machine Learning*, 60–69.
- Bhutani, S., & Mittal, R. (2023). Blockchain Based Ecommerce Warranty System Using NFT (Non-fungible Token). *2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE)*, 374–377.
- Carlini, N., Liu, C., Erlingsson, Ú., Kos, J., & Song, D. (2019). The secret sharer: Evaluating and testing unintended memorization in neural networks. *28th USENIX Security Symposium (USENIX Security 19)*, 267–284.
- Danks, D., & London, A. J. (2017). Algorithmic Bias in Autonomous Systems. *Ijcai*, 17(2017), 4691–4697.
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *ArXiv Preprint ArXiv:1702.08608*.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., & Rossi, F. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689–707.
- Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics*, 26(3), 1771–1796. <https://doi.org/10.1007/s11948-020-00213-5>
- Ganesh, D., & Rastogi, A. (2023). *Integrating Web 3.0 Technologies for Advanced Information Retrieval Systems in the Educational Sector*. <https://doi.org/10.1109/ICCSAI59793.2023.10421149>
- Ge, M., Pilato, G., Persia, F., & D’Auria, D. (2023). *Recommender System for Social Media: Research Challenges and Future Applications*. <https://doi.org/10.1109/TransAI60598.2023.00033>
- Goodfellow, I., Shlens, J., & Szegedy, C. (2015). *Explaining and harnessing adversarial examples*.
- Huang, D. (2022). *Design of Integrated English online education system based on Artificial Intelligence*. <https://doi.org/10.1109/ISoIRS57349.2022.00021>
- Kardaras, D. K., Troussas, C., Barbounaki, S. G., Tselenti, P., & Armyras, K. (2024). A Fuzzy Synthetic Evaluation Approach to Assess Usefulness of Tourism Reviews by Considering Bias Identified in Sentiments and Articulacy. *Information*, 15(4). <https://doi.org/10.3390/info15040236>
- Krouska, A., Troussas, C., & Sgouropoulou, C. (2020). Applying Genetic Algorithms for Recommending Adequate Competitors in Mobile Game-Based Learning Environments. In V. Kumar & C. Troussas (Eds.), *Intelligent Tutoring Systems* (pp. 196–204). Springer International Publishing.
- Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2017). *Towards Deep Learning Models Resistant to Adversarial Attacks*.
- Martin, K. (2022). *Ethics of data and analytics: Concepts and cases*. Auerbach Publications.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6), 1–35.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should i trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
- Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers,

- R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *Npj Digital Medicine*, 3(1), 119. <https://doi.org/10.1038/s41746-020-00323-1>
- Ryan, M., & Stahl, B. C. (2020). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*, 19(1), 61–86.
- Shahriari, K., & Shahriari, M. (2017). IEEE standard review—Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*, 197–201.
- Troussas, C., Krouska, A., & Sgouropoulou, C. (2022). Enriching mobile learning software with interactive activities and motivational feedback for advancing users' high-level cognitive skills. *Computers*, 11(2), 18.
- Troussas, C., Virvou, M., & Espinosa, K. J. (2015). Using visualization algorithms for discovering patterns in groups of users for tutoring multiple languages through social networking. *J. Networks*, 10(12), 668–674.
- Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2), 2053951717743530.
- Virvou, M., Troussas, C., Caro, J., & Espinosa, K. J. (2012). User Modeling for Language Learning in Facebook. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech and Dialogue* (pp. 345–352). Springer Berlin Heidelberg.
- Zenkert, J., & Fathi, M. (2023). Practice-Oriented Approaches for Information and Metadata Management in a Content Management System—Learnings from the Smart City Project LOKAL-digital. *2023 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 1–6.