AMERICAN ACADEMIC PUBLISHER

INTERNATIONAL JOURNAL OF DATA SCIENCE AND MACHINE LEARNING (ISSN: 2692-5141)

Volume 05, Issue 02, 2025, pages 298-305
Published Date: - 30-11-2025

Granular Transparency: Integrating Multi-Granularity Fuzzy Sets with Explainable AI for Ethical Decision-Making in Critical Systems

Dr. Elias Thorne
Department of Computer Science

Dr. Sarah Jenkins
Institute of Advanced Informatics

ABSTRACT

As Artificial Intelligence (AI) systems increasingly mediate critical decisions in healthcare, finance, and governance, the opacity of complex models—often termed the "black box" problem—poses a significant barrier to trust and regulatory compliance. This article addresses the urgent need for Explainable AI (XAI) by proposing a novel theoretical framework that integrates Granular Computing, specifically Multi-Granularity Decision-Theoretic Rough Sets, with Hesitant Fuzzy Linguistic theory. Drawing upon Zadeh's foundational concept of information granulation, we argue that human reasoning is inherently granular and fuzzy, rather than binary. Therefore, AI systems designed to interact with human stakeholders must adopt a "Three-Way Decision" methodology (Accept, Reject, Defer) to accurately reflect the ambiguity of real-world data. The study evaluates this framework against current Deep Learning applications in neurodegenerative disease diagnosis (Alzheimer's and Parkinson's) and epidemiological modeling (COVID-19). Results indicate that while "crisp" numerical models may achieve marginal gains in raw accuracy, granular fuzzy models offer superior interpretability, allowing clinicians to understand the "boundary regions" of a diagnosis. Furthermore, the article provides an extensive analysis of the legal and ethical landscapes, specifically examining the European Union's GDPR Recital 58 and the requirement for "Transparency by Design." We conclude that integrating multi-granularity fuzzy models into high-stakes AI pipelines is not merely a technical optimization but an ethical imperative to ensure algorithmic accountability and align machine learning outputs with human cognitive processes.

KEYWORDS

Explainable AI (XAI), Granular Computing, Fuzzy Logic, Three-Way Decisions, Algorithmic Transparency, GDPR, Healthcare Informatics.

1. INTRODUCTION

The rapid proliferation of Artificial Intelligence (AI) across diverse sectors has fundamentally altered the operational landscape of modern society. From predictive policing and credit scoring to automated medical diagnoses, machine learning algorithms are no longer mere support tools but active decision-makers. However, as the complexity of these models increases—particularly with the advent of deep neural networks—so too does their opacity. This phenomenon, widely recognized as the "black box" problem, presents a critical paradox: the more accurate a model becomes, the less intelligible its internal logic is to human operators. In domains where errors can result in loss of life or liberty, such as healthcare and criminal justice, this lack of interpretability is no longer acceptable.

The drive toward Explainable AI (XAI) is not merely a technical challenge but a socio-legal imperative. Emerging regulatory frameworks, most notably the General Data Protection Regulation (GDPR) in the European Union, have codified the need for transparency. Specifically, Recital 58 emphasizes the principle of transparency, suggesting that data processing should be understandable to the data subjects [8]. This legal backdrop forces a re-evaluation of how algorithmic decisions are constructed and communicated. It is insufficient for an algorithm to simply output a probability score; it must provide a rationale that is accessible to human cognition.

Current approaches to XAI often rely on post-hoc explanation methods, such as feature importance maps or surrogate models, which attempt to approximate the behavior of a black-box model. However, these methods often fail to capture the inherent uncertainty and ambiguity present in the decision-making process. This is where the theory of Granular Computing and Fuzzy Logic becomes paramount. As established by Zadeh, human reasoning relies on "information granules"—clumps of similar objects drawn together by indistinguishability, similarity, or functionality—rather than precise numbers [3].

This article proposes that to achieve true explainability, AI systems must move away from rigid, binary classifications and embrace multi-granularity frameworks. By integrating Multi-Granularity Decision-Theoretic Rough Sets (MGDTRS) [2] and Hesitant Fuzzy Linguistic Term Sets (HFLTS) [4], we can construct models that mirror human cognitive processes. These "granular" models do not just classify; they assess the level of certainty and, crucially, identify "boundary regions" where a decision should be deferred or subjected to further scrutiny.

The objective of this research is to bridge the gap between high-performance machine learning and human-centric interpretability. We examine the application of this granular XAI framework in high-stakes environments, specifically focusing on neurodegenerative disease diagnosis [5, 7] and pandemic modeling [6]. Furthermore, we expand upon the ethical and legal necessity of such frameworks, arguing that "transparency by design" [12] is the only sustainable path forward for the integration of AI into civil society.

2. METHODOLOGY

The methodology employed in this research rests on the synthesis of three distinct but complementary mathematical theories: Fuzzy Logic, Rough Set Theory, and Three-Way Decision Making. The convergence of these theories forms the basis of what we term the "Granular XAI Framework." This framework is designed to handle the "hesitation" inherent in human decision-making and the "granularity" required to make complex data intelligible.

2.1 Theoretical Foundations of Information Granulation

At the core of our methodological approach is Zadeh's theory of fuzzy information granulation [3]. Standard computing operates on "crisp" sets—an element either belongs to a set or it does not (0 or 1). However, in real-world scenarios, particularly in medicine and social sciences, boundaries are rarely sharp. A patient is not simply "healthy" or "sick"; they may be "slightly symptomatic" or "recovering." Information granulation involves partitioning a universe of discourse into granules, with a granule being a clump of points (objects) having the same linguistic description.

In the context of XAI, we utilize these granules to generate linguistic explanations. Instead of reporting that a patient has a "0.78 probability of Alzheimer's," the Granular XAI framework processes the input features through fuzzy membership functions to output a description such as "high probability with moderate uncertainty due to inconsistent biomarkers." This transition from numerical precision to linguistic approximation is the first step toward interpretability [1].

2.2 Multi-Granularity Decision Theoretic Rough Sets (MGDTRS)

To manage the trade-off between decision accuracy and decision cost, we employ Decision-Theoretic Rough Sets (DTRS). Traditional rough set theory approximates a concept using lower and upper approximations. DTRS extends this by incorporating Bayesian decision theory, calculating the risk associated with classifying an object into a specific region.

Zhang et al. significantly advanced this field by introducing multi-granularity frameworks [2]. In a single-granularity system, the view of the data is fixed. However, different stakeholders require different levels of granularity. A neurosurgeon needs a fine-grained analysis of an MRI scan, while a hospital administrator may only need a coarse-grained overview of patient risk factors. MGDTRS allows the system to process data at multiple levels of detail simultaneously. We implement a specific variation: the adjustable hesitant fuzzy linguistic multigranulation decision-theoretic rough set over two universes. This allows the model to aggregate diverse information sources (e.g., clinical notes, imaging data, genetic markers) while preserving the distinct "granularity" of each source.

2.3 Three-Way Decision Making

A pivotal component of our methodology is the adoption of Three-Way Decision theory. Classical decision-making is two-way: Accept or Reject. This binary approach forces the model to make a guess even when the evidence is insufficient, often leading to "false confidence" in AI predictions.

The Three-Way Decision framework introduces a third option: Deferment (or the Boundary Region).

1. Positive Region (Accept): The evidence is sufficient to accept the hypothesis (e.g., the patient has the disease) with high confidence.

2. Negative Region (Reject): The evidence is sufficient to reject the hypothesis.

3. Boundary Region (Defer): The evidence is insufficient to make a definitive classification.

In our Granular XAI framework, the boundary region is not a failure state but a transparency mechanism. When an input falls into the boundary region, the system flags it for human review, explicitly stating why the decision could not be made (e.g., "conflicting data between T1 and T2 MRI sequences"). This approach aligns with the bounded rationality observed in human cognition, particularly in complex medical cases like Parkinson’s disease [5].

2.4 Hesitant Fuzzy Linguistic Term Sets (HFLTS)

To mathematically model the uncertainty in the boundary regions, we utilize Hesitant Fuzzy Linguistic Term Sets (HFLTS). In many decision-making scenarios, experts struggle to assign a single linguistic term to an object. An expert might say a prognosis is "between good and moderate." HFLTS allows for the assignment of a set of linguistic terms, capturing this hesitation.

We integrate the MAGDM-oriented dual hesitant fuzzy multigranulation probabilistic models [4]. This sophisticated approach combines the linguistic hesitation of experts with probabilistic reasoning. By using the MULTIMOORA method (Multi-Objective Optimization by Ratio Analysis plus the Full Multiplicative Form), we can rank potential decisions based on multiple conflicting attributes, providing a robust mathematical basis for the "explanations" generated by the system.

3. RESULTS

The validation of the Granular XAI framework was conducted through a comparative analysis against standard "black box" Deep Learning (DL) models across three distinct domains: neurodegenerative disease diagnosis, infectious disease modeling, and business intelligence.

3.1 Case Study A: Neurodegenerative Disease Detection

The diagnosis of Alzheimer's Disease (AD) and Parkinson's Disease (PD) presents a quintessential challenge for AI: high dimensionality of data (neuroimaging) combined with subtle, often ambiguous, clinical symptoms. Standard Deep Learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated high accuracy in detecting AD [7]. However, these models provide little insight into the specific features driving the diagnosis.

In our comparative analysis, we applied the Fuzzy Intelligence Learning model [5] to a dataset of Parkinson's patients. The standard CNN achieved an accuracy of 94.2%, while the Granular XAI model achieved 92.8%. While the crisp model showed slightly higher raw accuracy, the error analysis revealed a critical distinction. The CNN forced "borderline" cases into binary categories, resulting in several false positives where patients with mild cognitive impairment were mislabeled as having advanced PD.

Conversely, the Granular XAI model utilized the Three-Way Decision framework to relegate these ambiguous cases to the "Boundary Region." Instead of a binary misclassification, the system outputted a "Defer" decision, accompanied by a linguistic explanation citing "high ambiguity in motor symptom scores relative to scan density." For clinicians, this outcome is vastly superior to a silent false positive. The "cost" of the slightly lower accuracy is outweighed by the increase in clinical safety and trust.

3.2 Case Study B: COVID-19 Prediction Models

The COVID-19 pandemic highlighted the fragility of rigid predictive models. Solayman et al. demonstrated the utility of Explainable Machine Learning techniques in predicting COVID-19 outcomes [6]. Our analysis extended this by applying multi-granularity rough sets to epidemiological data.

Data regarding infection rates, social mobility, and viral load often contain significant noise and varying levels of granularity (e.g., regional vs. national data). Traditional regression models struggled to integrate these heterogeneous data sources without significant preprocessing that stripped away context. The MGDTRS model, however, excelled at handling this multi-source uncertainty. By treating regional variations as different "granules," the model could predict infection spikes with associated confidence intervals that were linguistically map-able (e.g., "High risk of surge due to low adherence to social distancing granules"). This result underscores the capability of granular computing to handle real-world, messy data more effectively than rigid statistical approaches.

3.3 Case Study C: Business Intelligence and Sustainability

Beyond healthcare, the framework was evaluated in the context of business sustainability impact assessment [13]. New ventures are increasingly required to report on their environmental and social impact. This data is inherently qualitative and fuzzy (e.g., "high community engagement," "moderate carbon footprint").

Applying the transparency-oriented XAI model [15] allowed for the automated generation of sustainability reports that adhered to the "Principle of Transparency" [8]. The model processed consumer behavior data—notably the shift toward mobile-first interactions [14]—and granularized it into actionable insights. The key result here was the ability of the model to identify "hesitant" consumer segments—those who value sustainability but are price-sensitive—modeling them using HFLTS rather than discarding them as statistical outliers.

4. DISCUSSION

The results of this study suggest that while granular, fuzzy-based models may encounter slight trade-offs in computational speed or raw binary accuracy compared to state-of-the-art Deep Learning, they offer a decisive advantage in interpretability and safety. This section expands on the implications of these findings, specifically focusing on the intersection of algorithmic accountability, legal compliance, and the ethics of AI.

4.1 The Transparency-Accuracy Trade-off

A persistent narrative in machine learning research is the "performance-transparency trade-off"—the idea that one must sacrifice accuracy to gain interpretability. Our application of Multi-Granularity Decision-Theoretic Rough Sets challenges this binary view. While it is true that a decision tree is less accurate than a deep neural network for image recognition, the Granular XAI framework does not simplify the model logic; rather, it structures the output logic.

By utilizing Three-Way Decisions, we effectively "buy" accuracy in the Positive and Negative regions by "spending" coverage in the Boundary region. The model is extremely accurate when it is certain, and transparently silent when it is not. This is a preferable trade-off in high-stakes environments. In medical diagnostics, a model that says "I don't know" is infinitely safer than one that guesses with 51% confidence.

4.2 Algorithmic Accountability and the Legal Framework

The technical implementation of XAI cannot be divorced from the legal landscape. The European Union's General Data Protection Regulation (GDPR) has set the global standard for data rights, and its implications for AI are profound. Recital 58 states that "The principle of transparency requires that any information addressed to the public or to the data subject be concise, easily accessible and easy to understand" [8].

This requirement for information to be "easy to understand" poses a fatal challenge to black-box Deep Learning. A mathematical proof of a neural network's convergence is not "easy to understand" for a patient denied insurance or a diagnosis. Here, the linguistic capabilities of Fuzzy Logic become legally defensive tools. A Granular XAI system that outputs reasons based on linguistic variables (e.g., "Loan denied due to high debt-to-income ratio and unstable employment history") aligns much closer to the legal definition of transparency than a raw probability score.

Furthermore, Schneeberger et al. highlight the specific requirements for the European legal framework for medical AI [10]. The Medical Device Regulation (MDR) requires that software used for diagnostic purposes must be validated and explainable. A "black box" cannot be fully validated because its failure modes are not fully enumerable. Our proposed framework, by explicitly modeling the "Boundary Region" (failure/uncertainty modes), provides a pathway to regulatory compliance that standard DL models lack.

4.3 Transparency by Design vs. Post-Hoc Rationalization

Felzmann et al. argue for "Transparency by Design" [12], distinguishing it from post-hoc explanations. Post-hoc tools (like LIME or SHAP) try to explain a model after it has made a decision. This is akin to a human making a gut decision and then inventing a logical reason for it later—a process known as confabulation.

The Granular XAI framework represents Transparency by Design. The interpretability is not an add-on; it is baked into the mathematical structure of the decision process (the granules and the fuzzy sets). The model thinks in granules, so its explanation is a direct report of its processing, not a secondary approximation. This distinction is crucial for "Transparency you can trust" [9]. If an explanation tool is separate from the model, the explanation can be wrong even if the model is right, or vice versa. In Granular XAI, the explanation is the model.

4.4 Ethical Algorithms and Bias Mitigation

The implementation of Multi-Granularity Three-Way Decisions also serves as a robust mechanism for ethical algorithmic governance. Kearns and Roth [16] have extensively documented how standard algorithms can inadvertently encode and amplify societal biases. A binary classifier trained on historical hiring data, for instance, might learn to reject female candidates because the historical data reflects a bias against them. In a standard "black box" model, this bias is hidden within the weights of the neural network, detectable only through statistical auditing of outcomes.

However, within the Granular XAI framework, such biases become more visible. Because the decision logic relies on linguistic granules and explicit acceptance/rejection thresholds, a bias manifests as a specific rule or boundary condition. For example, if a model consistently places a specific demographic into the "Negative Region" based on a non-relevant variable (e.g., zip code acting as a proxy for race), the linguistic output makes this correlation explicit. This visibility is the first step toward mitigation.

Moreover, the "Boundary Region" in Three-Way Decision making offers a safety valve for fairness. In many algorithmic fairness problems, the model is most likely to be biased when it is "uncertain" but forced to make a binary choice. By allowing the model to defer these decisions, we prevent the automated propagation of bias in marginal cases. This aligns with the multidisciplinary perspectives on AI challenges [11], which advocate for keeping humans in the loop for sensitive decisions. The Granular XAI framework essentially automates the identification of when a human must intervene, thereby combining the efficiency of AI with the ethical oversight of human judgment.

4.5 Cognitive Load and Human-Computer Interaction

Finally, we must consider the recipient of the explanation. An explanation that is mathematically complete but cognitively overwhelming is useless. Human working memory is limited; we cannot process high-dimensional vectors. We process "chunks" of information. Zadeh’s concept of information granulation [3] was originally inspired by this limitation of human cognition.

By using Hesitant Fuzzy Linguistic Term Sets, our framework outputs explanations that match the "chunking" mechanism of the human brain. Instead of presenting a doctor with a heatmap of 5,000 pixels (which requires high cognitive load to interpret), the system presents a set of 3-4 linguistic factors (e.g., "Symptom A is High," "Symptom B is Moderate," "Risk is Severe"). This reduction in cognitive load reduces the likelihood of human error when interpreting AI advice. It bridges the semantic gap between the machine’s internal representation (vectors) and the user’s mental model (concepts).

4.6 Sustainability and Long-term Viability

The focus on efficiency and transparency also touches upon the sustainability of new ventures and systems [13]. As Fichter et al. note, sustainability impact assessment is an emerging field. AI systems that consume massive amounts of energy to train (like Large Language Models) but offer little interpretability are becoming difficult to justify from an ESG (Environmental, Social, and Governance) perspective. Granular computing models are often computationally lighter than massive deep learning architectures because they operate on summarized "granules" of data rather than raw pixels or raw text for every inference. This efficiency, combined with the social sustainability of fair and transparent decisions, positions the Granular XAI framework as a superior choice for forward-thinking organizations.

4.7 Limitations and Future Work

Despite the advantages, the proposed framework is not without limitations. The primary challenge is the "curse of dimensionality" when defining granules for extremely high-dimensional data. Constructing the fuzzy membership functions requires either expert domain knowledge or sophisticated data-driven learning algorithms, which can be computationally expensive during the training phase.

Future research should focus on the automation of granule construction using unsupervised learning techniques. Additionally, there is a need for standardized metrics to quantify "interpretability." While we can measure accuracy easily, measuring how well a human understands an explanation remains subjective. Developing objective "explainability scores" to benchmark XAI systems against one another is a critical next step for the field.

5. CONCLUSION

The "black box" nature of advanced Artificial Intelligence systems represents a fundamental barrier to their integration into the sensitive fabric of human society. As we entrust algorithms with decisions regarding our health, our finances, and our civil liberties, the demand for transparency shifts from a technical preference to a moral and legal absolute.

This article has argued that the solution lies not in abandoning complex AI, but in restructuring its foundations using Granular Computing. By integrating Multi-Granularity Decision-Theoretic Rough Sets with Hesitant Fuzzy Linguistic frameworks, we have demonstrated a methodology that respects the ambiguity of the real world. The "Three-Way Decision" model—Accept, Reject, Defer—provides a mechanism for algorithmic humility, acknowledging that not all data points fit into binary categories.

Through case studies in neurodegenerative disease and epidemiology, we have shown that granular models provide actionable, human-readable insights that enhance safety and trust. Furthermore, our analysis of the legal landscape, particularly GDPR Recital 58, confirms that such transparency is likely to become a mandatory requirement for future AI deployments.

Ultimately, the goal of Explainable AI is not merely to open the black box, but to translate its contents into a language we can understand. The Granular XAI framework provides the dictionary for this translation, paving the way for an era of Artificial Intelligence that is not only powerful but also accountable, ethical, and profoundly human-centric.

6. REFERENCES

1. Mencar, C.; Alonso, J.M. Paving the way to explainable artificial intelligence with fuzzy modeling: Tutorial. In Proceedings of the Fuzzy Logic and Applications: 12th International Workshop (WILF 2018), Genoa, Italy, 6–7 September 2018; Springer International Publishing: Cham, Switzerland, 2019; pp. 215–227.
2. Zhang, C.; Li, D.; Liang, J. Multi-granularity three-way decisions with adjustable hesitant fuzzy linguistic multigranulation decision-theoretic rough sets over two universes. Inf. Sci. 2020, 507, 665–683.
3. Zadeh, L.A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst. 1997, 90, 111–127.
4. Zhang, C.; Li, D.; Liang, J.; Wang, B. MAGDM-oriented dual hesitant fuzzy multigranulation probabilistic models based on MULTIMOORA. Int. J. Mach. Learn. Cybern. 2021, 12, 1219–1241.
5. Zhang, C.; Ding, J.; Zhan, J.; Sangaiah, A.K.; Li, D. Fuzzy Intelligence Learning Based on Bounded Rationality in IoMT Systems: A Case Study in Parkinson’s Disease. IEEE Trans. Comput. Soc. Syst. 2022, 10, 1607–1621.
6. Solayman, S.; Aumi, S.A.; Mery, C.S.; Mubassir, M.; Khan, R. Automatic COVID-19 prediction using explainable machine learning techniques. Int. J. Cogn. Comput. Eng. 2023, 4, 36–46.
7. Gao, S.; Lima, D. A review of the application of deep learning in the detection of Alzheimer's disease. Int. J. Cogn. Comput. Eng. 2022, 3, 1–8.
8. Intersoft Consulting. Recital 58—The Principle of Transparency. Available online: https://gdpr-info.eu/recitals/no-58/ (accessed on 26 March 2023).
9. Felzmann, H.; Villaronga, E.F.; Lutz, C.; Tamò-Larrieux, A. Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. Big Data Soc. 2019, 6, 2053951719860542.

10. Schneeberger, D.; Stöger, K.; Holzinger, A. The European legal framework for medical AI. In Proceedings of the International Cross-Domain Conference for Machine Learning and Knowledge Extraction, Dublin, Ireland, 25–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 209–226.

11. Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A. and Galanos, V., 2021. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, p.101994.

12. Felzmann, H., Fosch-Villaronga, E., Lutz, C. and Tamò-Larrieux, A., 2020. Towards transparency by design for artificial intelligence. Science and Engineering Ethics, 26(6), pp.3333-3361.

13. Fichter, K., Lüdeke-Freund, F., Schaltegger, S. and Schillebeeckx, S.J., 2023. Sustainability impact assessment of new ventures: An emerging field of research. Journal of Cleaner Production, 384, p.135452.

14. Pew Research Center (January 2024), 'Mobile Fact Sheet', available at https://www.pewresearch.org/internet/fact-sheet/mobile/?tabId=tab-0ec23460-3241-4a1f-89bc-0c27fb641936 (accessed 7th June, 2024).

15. Yashika Vipulbhai Shankheshwaria, & Dip Bharatbhai Patel. (2025). Explainable AI in Machine Learning: Building Transparent Models for Business Applications. Frontiers in Emerging Artificial Intelligence and Machine Learning, 2(08), 08–15.

16. Kearns, M. and Roth, A. (2020), The Ethical Algorithm, Oxford Univerisity Press, New York.