Harnessing Fuzzy Cognitive Maps for Advancing AI with Hybrid Interpretability and Learning Solutions

Maikel Leon

Department of Business Technology, Miami Herbert Business School,
University of Miami, Florida, USA. E-mail: mleon@miami.edu

Abstract. Advancements that merge the clarity of symbolic AI with the adaptive learning traits of sub-symbolic AI show great potential at the intersection of these two AI forms. This study introduces Fuzzy Cognitive Maps (FCMs). This hybrid model integrates the optimal characteristics of both frameworks to address the challenges of interpretability and explainability in artificial intelligence (AI) systems. FCMs provide a robust framework for logically and intuitively supporting decision-making processes and representing causal relationships. Their capacity to handle the inherent vagueness and uncertainty of real-world scenarios enables a more natural and flexible approach to problem-solving. Due to their intrinsic adaptability and learning capabilities derived from sub-symbolic AI, FCMs are particularly suited for applications demanding high levels of interpretability and explainability.

Keywords: Fuzzy Cognitive Maps, Adaptability, Learning, Symbolic AI, and Sub-symbolic AI.

1 Introduction

In the constantly shifting realm of AI, two primary paradigms have surfaced, each characterized by its unique method for modeling intelligence and addressing problems. Symbolic AI, on one end of the spectrum, uses explicit symbols and rules to represent knowledge, mirroring the logical frameworks of human reasoning. Conversely, sub-symbolic AI opts for a less transparent but more potent method, learning directly from data and embodying patterns and statistical relationships that underlie intelligence in ways often elusive to human understanding [1]. Although these paradigms appear contradictory, they represent AI's dual avenues to emulate or exceed human cognitive functions. A deeper exploration into their strengths and weaknesses reveals a compelling story—one indicating that AI's future might not hinge on one paradigm prevailing over the other but rather on their combined strengths [2].

Symbolic AI uses symbol manipulation and logical operations to accomplish tasks, resolve problems, and make decisions. This method, pivotal in the early triumphs of AI research, thrives in areas where rules are explicit and outcomes are foreseeable. Its transparency and traceability, where each decision follows a precise, logical sequence, are precious in fields requiring explainability and adherence to regulatory norms. Yet, symbolic AI's inflexibility, dependence on comprehensive rule sets, and challenges in encoding commonsense knowledge restrict its effectiveness in complex, commonsense scenarios marked by ambiguity and uncertainty [3]. Symbolic AI is fundamentally a logic-based field, traditionally relying on classical (usually monotonic) logic and positing that this logical processing underpins machine intelligence. For example, considering this paradigm, querying "What is an apple?" would yield responses defining an apple as "a fruit," "having red, yellow, or green color," or "bearing a roundish shape." These attributes are termed symbolic because they utilize symbols (like color, shape, and type) to describe the apple.

DOI:10.5121/acij.2024.15501

From the 1950s to the 1980s, this paradigm was the predominant approach in AI. Regarding the implementation of symbolic AI, one of the oldest yet still the most popular logic programming languages is Prolog (its roots are in first-order logic) [4].

In contrast, sub-symbolic AI, encompassing neural networks and deep learning, adopts a markedly different methodology. By learning from extensive datasets, sub-symbolic AI constructs an internal representation of the world, excelling in areas such as pattern recognition, language processing, and predictive modeling, often outperforming human capabilities. However, this efficiency comes at the expense of transparency, leading to the "black box" issue, where the rationale behind a model's decisions remains obscure and challenging to explain. The fundamental premise of the sub-symbolic approach is that a model's success hinges on its ability to derive an influential model from limited data. Instead of using transparent, human-understandable relationships, this paradigm employs complex, less interpretable mathematical formulas to tackle problems. Among the most prevalent sub-symbolic AI models are neural networks, ensemble models, regression models, decision trees, and support vector machines, which are frequently encountered in developing machine learning models. During the 80s, the sub-symbolic AI paradigm took over symbolic AI's position as the leading subfield [5].

The dichotomy between symbolic and sub-symbolic AI has led to a vibrant discourse on the future direction of AI research and application. Within this discourse, Fuzzy Cognitive Maps (FCMs) emerge as a fascinating hybrid technique, combining the explicit knowledge representation of symbolic AI with the adaptability and learning capabilities of sub-symbolic AI [6]. FCMs leverage fuzzy logic to manage ambiguity and map out complex systems using networks of concepts and causal links, effectively bridging the deterministic symbol-based world and the probabilistic realm of sub-symbolic learning. As we approach a new epoch in AI development, synthesizing symbolic and sub-symbolic methodologies can potentially unleash novel capabilities. By combining symbolic AI's clear interpretability and structured knowledge representation with the dynamic learning abilities and adaptability of sub-symbolic AI, we can establish the groundwork for more advanced, flexible, and reliable AI systems [7]. This article delves into the distinct strengths and limitations of both symbolic and sub-symbolic AI, spotlights FCMs as a leading example of hybrid AI approaches, and envisions a future where AI's fullest potential is achieved through the seamless integration of these two paradigms [8]. In doing so, we may find that the future of AI is not a question of either/or but a confluence of both, harnessing the best of what each approach has to offer.

The rest of this paper is organized as follows. Sec. 2 presents the origins and notable cases of this classical approach to AI. Sec. 3 refers to theoretical conceptions in Machine Learning. Sec. 4 presents the idea of the need for suitable explanations offered by these systems. Sec. 5 digs deep into why AI's future should contain more traceable and interpretable models. Sec. 6 holds the idea of merging both symbolic and subsymbolic approaches. Sec. 7 highlights the well-known Artificial Neural Networks' relevance in connectionist computing. Sec. 8 introduces a paradigm aiming to benefit from symbolic and subsymbolic AI. Last, Sec. 9 serves as a reflection and to understand the need for new and more AI models that are solid computationally and transparent to human understanding.

2 The "Good Old-Fashioned" AI

Symbolic AI, or "Good Old-Fashioned Artificial Intelligence," refers to a branch of AI research and development emphasizing symbolic representations of problems, logic, and search. This approach to AI relies on manipulating symbols and expressions to perform tasks, solve problems, and model the world. The following report delves into symbolic AI's origins, notable case studies, advantages, and disadvantages.

2.1 Origins

Symbolic AI traces its roots back to the mid-20th century, with foundational work by figures such as Alan Turing, John McCarthy, and Marvin Minsky. Turing's conceptualization of the Turing machine and the Turing test laid the groundwork for thinking about machines that could simulate human intelligence. In the 1950s and 1960s, John McCarthy, often considered one of the fathers of AI, coined the term "artificial intelligence" and introduced the concept of using symbolic logic to represent and solve problems. Marvin Minsky's work on frames and knowledge representation further advanced the development of symbolic AI. The period from the 1950s to the late 1980s is often considered the golden age of symbolic AI, during which researchers focused on developing systems that could reason about the world using symbolic logic. This era saw the creation of expert systems, among the first commercial applications of AI. These systems used rules and databases of knowledge to make inferences and provide advice in specialized domains such as medicine and engineering. This approach to AI relies on manipulating symbols and using predefined rules to solve problems akin to human deductive reasoning. During this period, symbolic AI achieved significant milestones, such as the development of expert systems that could mimic the decision-making abilities of human experts in specific fields and natural language processing that could interpret and generate human-like text based on structured logic and grammar rules [9].

However, the approach led to the AI winter—a period of reduced funding and interest in AI research—because it struggled to scale with complexity and could not handle real-world ambiguity. Symbolic AI systems required extensive manual labor to create and maintain their rule-based systems, and they were brittle, often failing outside narrowly defined scenarios. They also struggled with learning from data, which became increasingly important as the volume of digital data grew. As a result, the limitations of symbolic AI became apparent, leading to disillusionment and a shift towards other paradigms, such as machine learning and neural networks, which promised greater flexibility and adaptability.

2.2 Notable Case Studies

– MYCIN: Developed in the early 1970s at Stanford University, MYCIN was an expert system designed to diagnose bacterial infections and recommend antibiotics. It was one of the first successful demonstrations of symbolic AI in medicine, using a rule-based system to make decisions [10].
– SHRDLU: Created by Terry Winograd in the 1970s, SHRDLU was a natural language understanding system that could interact with a user in English to move blocks around a virtual world. It demonstrated the potential of symbolic AI for understanding and manipulating language and objects in a constrained environment.
– Deep Blue: Although primarily known for its chess-playing ability, IBM's Deep Blue represents a blend of symbolic AI (in terms of chess strategy and positions represented symbolically) and brute-force computation. In 1997, Deep Blue famously defeated world chess champion Garry Kasparov, showcasing the potential of AI in complex decision-making.

2.3 Advantages of Symbolic AI

Exploring symbols and applying logical rules to mimic human reasoning offer several distinct advantages, such as interpretability and transparency. It is well-suited for domains where decision-making processes, such as legal or regulatory settings, must be clear and justifiable. Symbolic AI excels in handling complex problem-solving within constrained parameters, leveraging its rule-based systems to perform tasks that require strict adherence to predefined rules and logical structures. Additionally, because it operates based on explicit rules, it can be easier to debug and modify than other AI paradigms. Let's break it down as follows:

- Explainability: One of the primary advantages of symbolic AI is its inherent explainability. Because decisions are made through explicit logical rules, it is easier to understand and trace symbolic AI systems' reasoning processes than more opaque models like deep neural networks.
- Efficiency in Domain-Specific Knowledge: Symbolic AI systems excel in domains where knowledge can be clearly defined and encoded in rules. This makes them particularly useful for expert medicine, law, and engineering systems.
- Handling Logical Reasoning and Complex Problems: Symbolic AI is well-suited for tasks that involve complex problem-solving and logical reasoning, where clear rules and relationships can be established.

2.4 Disadvantages of Symbolic AI

While symbolic AI has played a pivotal role in the development of artificial intelligence, it also comes with notable drawbacks. One of the primary limitations is its inability to learn from data autonomously. Unlike machine learning models that adapt and improve over time by analyzing vast amounts of data, symbolic AI requires explicit programming of rules and logic, making it less flexible and scalable in dynamic environments. This rigidity often leads to systems that can fail when encountering scenarios not pre-envisaged by the developers, limiting their applicability in complex, real-world situations where unpredictability is typical [11]. Additionally, the maintenance and updating of symbolic AI systems can be labor-intensive, as it involves manual adjustments to the rule base whenever new knowledge or corrections are needed. This makes it less efficient for tasks that require continual learning or adaptation, such as language processing and pattern recognition in continuously evolving datasets. For example:

- Knowledge Acquisition Bottleneck: One of the major challenges of symbolic AI is the knowledge acquisition bottleneck. Encoding expert knowledge into rules and symbols is time-consuming and requires significant expertise. This makes scaling symbolic AI systems difficult.
- Lack of Flexibility: Symbolic AI systems are often criticized for lacking flexibility and adaptability. They struggle with handling uncertainty, learning from new data, and performing in unstructured environments [12].
- Limited Perception and Learning: Unlike their machine learning counterparts, symbolic AI systems have limited abilities to learn from data or perceive complex patterns without explicitly programmed knowledge. This limits their applicability in tasks that require significant generalization or data-driven learning.

2.5 Summing-up

Symbolic AI has been pivotal in shaping the development of the AI field. Its focus on logic, explicit knowledge representation, and symbolic reasoning has driven substantial progress in replicating certain facets of human intelligence. Despite its contributions, the limitations of symbolic AI, particularly its lack of scalability, flexibility, and adaptive learning capabilities, have prompted the exploration of alternative methods, such as machine learning and neural networks. Nonetheless, the advantages of symbolic AI, including its explainability and effectiveness in specific contexts, continue to make it a valuable area for both study and practical application [13]. Emerging hybrid approaches that integrate the strengths of symbolic AI with those of machine learning show promise in addressing both paradigms' drawbacks. As the AI landscape evolves, the principles of symbolic AI are likely to remain influential in shaping the development and understanding of intelligent systems.

3 Sub-symbolic AI

Sub-symbolic AI marks a distinct paradigm in AI research, deviating from the traditional symbolic methodology. While symbolic AI depends on well-defined symbols and rules to represent and process knowledge, sub-symbolic AI delves into the foundational mechanisms of intelligence. This approach seeks to emulate the subconscious patterns and thought processes, frequently taking cues from the operations of the human brain and biological systems. This report investigates the origins, critical case studies, benefits, and drawbacks of sub-symbolic AI, providing a comprehensive overview of its role and impact in the field.

3.1 Origins of Sub-symbolic AI

The origins of sub-symbolic AI can be traced back to the early days of AI research. However, it gained significant momentum in the 1980s with the resurgence of neural networks and the development of algorithms that could learn from data. The limitations of symbolic AI (particularly its inability to handle ambiguous or incomplete information and to learn from raw data) motivated researchers to explore alternative models that could mimic the brain's ability to learn and generalize from experiences. The advent of connectionism, which emphasizes the role of neural networks and parallel distributed processing in cognitive functions, marked a pivotal shift towards sub-symbolic AI.

Sub-symbolic AI gained prominence in the 1980s and continues to thrive today, mainly contributing to the resurgence and expansion of AI research and applications after the AI winters. This shift was driven by the ability of sub-symbolic approaches to learn directly from data, allowing them to adapt to new tasks without requiring explicit programming of rules. Unlike symbolic AI, sub-symbolic AI excels in handling ambiguity, noisy data, and complex pattern recognition, making it ideal for tasks like image and speech recognition, which are prevalent in today's digital landscape. The development of backpropagation and the increase in computational power facilitated the training of deep neural networks, leading to groundbreaking advances in fields such as autonomous driving, language translation, and personalized recommendations. Furthermore, the advent of big data provided the necessary fuel for these algorithms to learn and improve continuously. The success of sub-symbolic AI rekindled interest and investment in AI and broadened its applicability across various sectors, marking a significant leap from the limitations of the earlier AI approaches.

3.2 Notable Case Studies

- Deep Learning for Image Recognition: Convolutional Neural Networks (CNNs), a class of deep neural networks, have revolutionized image recognition. A landmark moment was when AlexNet, a CNN designed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, won the ImageNet Large Scale Visual Recognition Challenge in 2012, significantly outperforming traditional image recognition methods.

- Natural Language Processing (NLP): Sub-symbolic AI has dramatically improved the ability of machines to understand and generate human language. Google's BERT (Bidirectional Encoder Representations from Transformers) and OpenAI's GPT (Generative Pre-trained Transformer) series are prime examples of how deep learning models can grasp complex language patterns, enabling breakthroughs in translation, summarization, and question-answering systems.
- AlphaGo: Developed by DeepMind, AlphaGo is a program that defeated the world champion Go player in 2016. It used deep neural networks and reinforcement learning to master a game known for its complexity and strategic depth, a feat previously thought to be decades away [14].

3.3 Advantages of Sub-symbolic AI

Sub-symbolic AI, primarily represented by machine learning and neural networks, offers several significant benefits that have fueled its widespread adoption. This approach excels in learning from and adapting to large volumes of data without explicit rule-based programming, making it highly effective for complex pattern recognition applications such as image and speech analysis. Sub-symbolic AI can generalize from past experiences to handle novel situations, a critical capability in dynamic environments like financial markets or autonomous vehicle navigation. Its proficiency in processing unstructured data also enables practical applications in natural language processing and personalized user experiences, enhancing technologies like chatbots and recommendation systems. The scalability of sub-symbolic AI systems also stands out, as they can improve continuously with additional data, driving advancements in fields from healthcare diagnostics to customer service automation. This learning capability increases efficiency and fosters innovation by unlocking new data interpretation and utilization possibilities. More in detail, we list the following:

- Learning from Data: One of the most significant advantages of sub-symbolic AI is its ability to learn directly from data without explicit programming. This makes it incredibly powerful in handling complex, high-dimensional data such as images, speech, and text.
- Generalization: Sub-symbolic AI models, particularly deep learning networks, can generalize, meaning they can perform well on unseen data after training on a sufficiently large and representative dataset. This ability to generalize from examples is closer to human learning and is a key strength of sub-symbolic AI.
- Handling Ambiguity and Uncertainty: Unlike symbolic AI, sub-symbolic AI is adept at dealing with ambiguity and incomplete information. Neural networks, for instance, can make probabilistic predictions and decisions even in uncertain or incomplete data [15].

3.4 Disadvantages of Sub-symbolic AI

Despite its impressive capabilities, sub-symbolic AI also presents several drawbacks. One of the most significant issues is the "black box" nature of these systems, where the decision-making processes are often opaque, making it difficult to trace how conclusions are drawn. This lack of transparency can be problematic in critical applications such as medical diagnosis or judicial decisions where accountability is essential. Additionally, sub-symbolic AI requires vast data to train effectively, which can introduce biases if the data is not carefully curated. These biases can perpetuate and even amplify existing prejudices in automated decisions. Sub-symbolic AI systems are also computationally intensive, requiring significant resources for training and operation, which can limit their accessibility and increase environmental impacts due to high energy consumption. Lastly, these systems are susceptible to adversarial attacks where slight, often imperceptible, inputs can deceive the AI into making erroneous decisions, posing security risks, especially in security-sensitive areas. Let's wrap up some of them:

- Opacity (Black-Box Problem): A significant drawback of sub-symbolic AI, especially deep neural networks, is its lack of transparency. These models are often described as "black boxes" because it is difficult to understand how they arrive at specific decisions or predictions, complicating efforts to debug or explain their behavior [16].
- Data and Computational Requirements: Training sub-symbolic AI models, particularly deep learning networks, requires vast data and significant computational resources. This can make cutting-edge AI research and applications inaccessible to organizations with limited resources [17].
- Overfitting and Generalization Issues: While sub-symbolic AI models are good at generalizing from data, they can also be prone to overfitting, where they perform well on training data but poorly on new, unseen data. To mitigate this risk, careful design, regularization techniques, and validation strategies are required [18].

3.5 Summing-up

Sub-symbolic AI has emerged as a powerful approach to AI, offering capabilities that surpass traditional symbolic methods in many areas, particularly those involving complex pattern recognition, learning from data, and generalization [19]. The achievements of deep learning and neural networks have highlighted sub-symbolic AI's capacity to address issues that were once deemed unsolvable. Nevertheless, the difficulties associated with interpretability, the substantial data and resource demands, and the risk of overfitting underscore the necessity for continuous research and development in this area. The future of AI seems to be leaning towards a hybrid methodology that integrates the strengths of both symbolic and sub-symbolic AI. Such an approach would combine symbolic systems' clear transparency and structured knowledge representation with sub-symbolic models' robust learning abilities and flexibility.

4 Explainable AI

Explainable AI (XAI) encompasses methodologies and techniques designed to make AI system outputs transparent and understandable to human users. XAI strives to develop a collection of machine learning models that not only maintain high-performance levels in terms of accuracy but also enhance their explainability. This allows human users to understand, trust, and manage artificial intelligence systems more effectively. The goal is to foster an environment where AI tools can be reliable and transparent partners in various applications. This report delves into the origins, significant case studies, benefits, and drawbacks of explainable AI. It also highlights how the symbolic AI paradigm, known for its clear interpretability and robust reasoning capabilities, facilitates the easy tracing of the logic behind specific outcomes, thereby supporting the principles of XAI. Yet, expressing the entire relation structure, even in a particular domain, is difficult [20]. Symbolic AI models often struggle to encompass all possibilities without considerable effort. In contrast, sub-symbolic AI paradigms yield highly effective models that can be developed and trained with less effort relative to their accuracy performance. Nonetheless, a significant limitation of sub-symbolic models lies in the transparency of their decision-making processes. The reliance on high-performing yet opaque models is problematic, particularly in sensitive areas where clear reasoning is crucial for outcomes, such as judicial decisions, military operations, and financial loan approvals. This highlights a critical need for models that perform well and provide understandable and traceable decision paths [21].

4.1 Origins of Explainable AI

The idea of XAI isn't a recent innovation but dates back to the early stages of AI research. Interest in XAI has surged recently as increasingly complex machine learning models, like deep learning, have become widespread. These models are often described as "black boxes" because their decision-making processes are not transparent. The growing demand for explainability is driven by concerns about accountability, fairness, transparency, and the need to meet regulatory requirements, such as those specified in the European Union's General Data Protection Regulation, which mandates a right to explanation [9]. Historically, AI systems were more interpretable because they predominantly employed symbolic AI methods, such as rule-based systems, where the rationale behind decisions could be quickly followed and comprehended. As the field shifted towards more powerful but less interpretable models, the demand for techniques to make these models explainable grew [22].

4.2 Notable Case Studies

1. Healthcare Diagnosis: AI models are increasingly used to diagnose diseases from medical imaging. Researchers have developed XAI systems that can identify specific features in imaging data that lead to their diagnosis, providing doctors with insights into why the AI system made a particular diagnosis. This not only aids in validating the AI's conclusions but also enhances the doctor's understanding and trust in the tool. 2. Financial Services for Loan Approval: AI models evaluate loan applications in the financial sector. XAI can be crucial in explaining why a loan was approved or denied, ensuring compliance with regulations against discriminatory practices, and helping applicants understand what factors influenced the decision. 3. Criminal Justice Risk Assessment Tools: Tools like COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) have been used to assess the likelihood of reoffending. XAI methods can help uncover, explain, and correct biases in such predictive models, ensuring fair and transparent decision-making.

4.3 Advantages of Explainable AI

- Increased Trust and Confidence: Explainability builds trust among users and stakeholders by transparentizing decision-making. When users understand how an AI system arrives at its conclusions, they are more likely to trust it.
- Improved Model Debugging and Validation: XAI techniques enable developers to identify and correct errors or biases in AI models. Developers can make targeted adjustments to improve performance and fairness by understanding the factors influencing model decisions.
- Regulatory Compliance: Many industries are subject to regulations that require decisions made by automated systems to be explainable. XAI facilitates compliance with such regulations, enabling AI solutions deployed in highly regulated sectors like finance and healthcare.

- Ethical and Fair Decision-Making: Explainable AI can help identify and mitigate biases in AI models, promoting more ethical and fair decision-making processes. This is particularly important in applications with significant social implications, such as criminal justice and employment.

4.4 Disadvantages of Explainable AI

- Potential Reduction in Model Performance: Sometimes, making a model more explainable may require simplifying its architecture or using less complex algorithms, which can reduce accuracy or performance [23].
- Complexity and Resource Requirements: Developing explainable AI models can be more complex and resource-intensive than traditional models. It requires additional efforts in design, implementation, and validation to ensure that explanations are meaningful and accurate.
- Risk of Oversimplification: There is a risk that the explanations provided by XAI systems might oversimplify the underlying processes, potentially leading to misunderstandings or misplaced trust in the AI system's capabilities.
- Security and Privacy Concerns: Explaining how AI systems work might inadvertently reveal sensitive information about the data or the model itself, posing security and privacy risks [24].

4.5 Summing-up

XAI marks a pivotal development in meeting AI systems' transparency, trust, and comprehension needs. As AI increasingly permeates critical sectors of society, the significance of explainability is set to escalate. The key challenge involves balancing the demand for complex, high-performing AI models and the necessity for transparency and understandability. While symbolic AI models inherently offer explainability, sub-symbolic AI models typically do not. Addressing this issue involves two main approaches: XAI, which focuses on enhancing model explainability by creating inherently more comprehensible models for human users, and Neuro-Symbolic Computation (NSC), which seeks to merge sub-symbolic learning algorithms with symbolic reasoning techniques to develop high-performing AI models that also possess reasoning capabilities [25]. This dual strategy is essential for advancing AI in a way that aligns with ethical standards and societal expectations. Future developments in XAI will likely focus on innovative approaches to maintaining or enhancing model performance while providing clear, accurate, and helpful explanations [26]. As the field evolves, it will also be essential to develop standardized metrics for explainability and ensure that explanations are accessible and understandable to all users, regardless of their technical background. Ultimately, the success of explainable AI will depend on its ability to foster trust and collaboration between humans and machines, enable more informed decision-making, and ensure that AI systems align with societal values and ethical principles [27].

5 Interpretable AI

Interpretable AI is dedicated to creating comprehensible models and algorithms for human users. This approach ensures that individuals can follow and understand an AI system's decisions, predictions, or classifications. Interpretable AI is vital in sensitive and critical areas where grasping the logic behind AI decisions is fundamental for building trust, ensuring compliance, and facilitating continual improvement. This report explores the origins, significant case studies, benefits, and drawbacks of interpretable AI, highlighting its importance and the challenges it faces in integrating transparency with advanced AI functionality [10].

Interpretable AI is considered more advanced than explainable AI due to its inherent transparency and clarity in decision-making processes. Interpretable AI models are designed to be inherently understandable, often using more straightforward or structured frameworks that allow direct insight into how inputs are transformed into outputs. This contrasts with explainable AI, which typically involves complex models like deep neural networks that require additional explanation layers or techniques to make their operations understandable. The critical advantage of interpretable AI is its ability to provide intuitive explanations directly from the model's structure. It ensures stakeholders can trust and verify the AI's decisions without needing auxiliary tools or methods. This intrinsic understandability is crucial in fields where decisions must be accurate and justifiable, such as healthcare and finance, making interpretable AI more transparent and potentially more reliable in sensitive applications [28].

5.1 Origins of Interpretable AI

The roots of interpretable AI stretch back to the field's inception when more straightforward, rule-based systems were standard. These systems were inherently interpretable, allowing users to trace the AI's logical steps to a conclusion. However, as AI research evolved, particularly with the development of more complex models like deep neural networks, the emphasis shifted toward enhancing performance, often at the expense of interpretability. The increasing use of AI systems in critical domains such as healthcare, finance, and criminal justice has reemphasized the need for interpretability. Stakeholders in these areas require AI to make decisions and provide explanations that are understandable to humans. This necessity has catalyzed the advancement of new techniques and spurred research efforts to make even the most sophisticated models interpretable [29].

5.2 Notable Case Studies

- Healthcare Diagnosis and Treatment: AI systems are increasingly used to diagnose diseases and recommend treatments. For instance, models that predict cardiovascular diseases based on patient data must be interpretable so that healthcare providers can understand the reasoning behind the predictions. This ensures trust and allows healthcare professionals to make informed decisions.
- Financial Services Compliance and Decision-Making: In finance, AI models are used for credit scoring, fraud detection, and automated trading. Interpretability in these models helps users understand the factors influencing decisions, ensuring compliance with regulatory standards and building customer trust.
- Criminal Justice and Bail Decisions: AI is used to assess the risk of recidivism and inform bail and sentencing decisions. Using interpretable AI models in this context is crucial for fairness, transparency, and accountability, allowing for scrutinizing decisions that significantly impact individuals' lives.

5.3 Advantages of Interpretable AI

- Trust and Transparency: Interpretable AI fosters trust from users by making the decision-making process transparent. When stakeholders understand how decisions are made, they are more likely to trust and accept AI solutions.

- Improved Decision-Making: Interpretability allows users to verify the correctness of the AI's reasoning, leading to more informed and better decision-making. This is especially important in domains where decisions have significant consequences [30].
- Regulatory Compliance: Many industries are subject to regulations that require decisions to be explainable. Interpretable AI facilitates compliance with such regulations, avoiding potential legal and financial penalties.
- Error Detection and Model Improvement: By understanding how an AI system makes decisions, developers and users can identify errors or biases in the model, leading to continuous improvement of AI systems.
- Ethical Considerations: Interpretable AI can help identify and mitigate biases in AI systems, promoting fairness and ethical decision-making.

5.4 Disadvantages of Interpretable AI

- Potential Trade-off Between Interpretability and Performance: Sometimes, making a model more interpretable may require simplifying its architecture or using less complex algorithms, potentially leading to decreased accuracy or performance.
- Complexity in Interpretation: Achieving true interpretability can be challenging for complex models. Even when interpretations are provided, they may be difficult for non-experts to understand, limiting their usefulness.
- Risk of Misinterpretation: There's a risk that interpretations provided by AI systems might be misunderstood by users, leading to incorrect conclusions or decisions based on those interpretations [31].
- Time and Resource Intensive: Developing interpretable AI models can require additional time and resources. Designing models that balance interpretability and performance involves extra effort in model selection, development, and validation [32].

5.5 Summing-up

Interpretable AI is essential for the responsible implementation of AI technologies, particularly in sensitive and high-stakes areas. It fulfills the critical need for transparency, trust, and ethical practices within AI systems. As AI progresses and becomes more ingrained in various societal aspects, the demand for interpretable models is expected to grow, pushing the limits of current research and development efforts [33]. Future progress in interpretable AI will focus on resolving the existing compromises between performance and interpretability, establishing standardized interpretation criteria, and generating more accessible explanations. These advancements will help ensure that AI systems are robust, effective, and congruent with societal values and ethical standards, thereby enhancing the acceptance and integration of AI technologies across diverse sectors [34].

6 The merge of both approaches

The intersection of symbolic and sub-symbolic AI presents a compelling and promising research area within AI. It merges symbolic AI's explicit reasoning and interpretability with the learning capabilities and adaptability of sub-symbolic AI, particularly neural networks. This hybrid approach aims to address the limitations that each method faces when used alone, facilitating the development of AI systems that are both powerful and understandable. This report delves into the origins, significant case studies, benefits, and drawbacks of blending symbolic and sub-symbolic AI highlighting how this intersection drives the evolution of more effective and user-friendly AI technologies [35].

6.1 Origins

The dichotomy between symbolic and sub-symbolic AI traces back to the early phases of AI research. Initially, symbolic AI was predominant, focusing on logic and rule-based systems. In contrast, sub-symbolic AI, which became prominent with machine learning and neural networks, emphasizes learning directly from data and recognizing patterns. The concept of merging these two approaches emerged from recognizing their complementary strengths and weaknesses. Symbolic AI excels in complex reasoning and explicit knowledge representation, while sub-symbolic AI is adept at processing raw data and learning from experience. This synergy presented a strong case for their integration, suggesting a unified approach that could leverage both advantages to overcome their limitations [31].

6.2 Notable Case Studies

- Neuro-Symbolic AI for Visual Question Answering (VQA): Research projects have combined neural networks with symbolic reasoning to improve VQA systems, which answer questions about images. These hybrid systems use neural networks to interpret visual data and symbolic systems to reason about the content, enabling more accurate and interpretable answers.
- Commonsense Reasoning: Projects like OpenAI's GPT-3 have integCommonsenselic reasoning to enhance the model's ability to perform commonsense reasoning tasks. These systems can better understand human-like responses by embedding symbolic representations within a neural framework.
- Robotics and Planning: Combining symbolic AI for high-level planning and decision-making with sub-symbolic AI for perception and motion control has led to more versatile and efficient robots. This approach allows robots to navigate and interact with their environment in a more human-like manner, adapting to new tasks and environments through learning.

6.3 Advantages

- Enhanced Reasoning and Generalization: Integrating symbolic and sub-symbolic AI can lead to systems that learn from data and apply logical reasoning to generalize beyond their training data. This results in more flexible and capable AI systems [36].
- Improved Interpretability and Transparency: Symbolic components can provide clear explanations for the decisions made by sub-symbolic models, addressing one of the major drawbacks of purely sub-symbolic AI systems.
- Efficient Learning and Knowledge Representation: Symbolic AI can encode domain knowledge that guides the learning process of sub-symbolic models, making them more efficient and effective in learning from data. Conversely, sub-symbolic models can discover patterns and relationships that can be formalized into symbolic knowledge.
- Flexibility and Adaptability: Hybrid systems can adapt to new tasks and environments more readily by leveraging the learning capabilities of sub-symbolic AI with the structured knowledge representation of symbolic AI.

6.4 Disadvantages

- Complexity in Integration: Combining symbolic and sub-symbolic AI involves significant challenges, including integrating disparate representations and reasoning mechanisms. This complexity can make the development of hybrid systems more challenging and resource-intensive.

- Scalability Issues: The scalability of hybrid AI systems can be limited by the symbolic component, which may not easily handle the vast amounts of data that sub-symbolic models can process [37].
- Limited Understanding of Integration Mechanisms: The field is still exploring the most effective ways to integrate symbolic and sub-symbolic AI. This includes challenges in combining learning and reasoning, representing knowledge, and ensuring that the systems are robust and reliable.

6.5 Summing-up

The intersection of symbolic and sub-symbolic AI offers significant potential for advancing the field of AI. By melding the strengths of both approaches, researchers and practitioners are working to develop AI systems that not only have the power to learn from extensive data sets but also possess the capability to reason, generalize, and articulate their decisions in ways that are understandable to humans. This hybrid approach aims to produce AI that is both robust in its analytical abilities and transparent enough to ensure trust and accountability, marking a pivotal evolution in the development of intelligent systems. This hybrid approach represents a step towards more sophisticated, versatile, and trustworthy AI systems that can be effectively applied in various domains, from healthcare and finance to autonomous systems [15]. Realizing the full potential of the intersection between symbolic and sub-symbolic AI involves surmounting significant challenges, such as integrating diverse AI paradigms, ensuring scalability, and devising effective mechanisms to meld learning with reasoning. Continued research and experimentation are essential for pushing the boundaries of what's currently possible and achieving the aim of creating AI systems that are both intelligent and interpretable. As the field progresses, the fusion of symbolic and sub-symbolic AI is anticipated to be instrumental in developing next-generation AI systems. These systems are expected to tackle complex problems with unparalleled efficiency and effectiveness, marking a crucial advancement in AI technology.

7 From ANN (sub-symbolic) to Rules (symbolic)

Humans excel at understanding and applying rules due to our cognitive ability to process abstract concepts, reason deductively, and learn from specific examples. This proficiency is deeply rooted in our linguistic and social development, where understanding and following rules are essential for communication and societal functioning. In contrast, networks and other decision-making models often rely on statistical patterns and data-driven learning, which can obscure decisions' underlying logic and rationale. While these models excel at identifying patterns and making predictions from large datasets, they lack the human-like capacity to grasp and reason through abstract rules and principles intuitively. This fundamental difference makes rule-based systems more aligned with human logic, facilitating easier comprehension, troubleshooting, and modification by human operators.

Extracting rules from Artificial Neural Networks (ANNs) is essential for making these models more transparent and their decisions more understandable to humans. This involves converting the complex, non-linear relationships the network has learned into rules that people can easily understand. We will look at a detailed example showing how to derive rules from an ANN trained on a primary dataset to predict loan approvals based on various applicant features.

7.1 Background

Consider a financial institution that has developed an ANN to assess loan applications. The ANN inputs include applicant features, such as Age, Income, Credit Score, and Employment Status, and it outputs a binary decision: Approve or Deny. Despite the ANN's high accuracy, the decision-making process is opaque, making it difficult for loan officers to justify decisions to applicants or to ensure compliance with regulatory standards. The institution seeks to extract interpretable rules from the ANN to address this.

7.2 ANN Architecture

The ANN in this example is a simple feedforward network with one hidden layer. The input layer has four neurons corresponding to the applicant features. The hidden layer has a few neurons (say five for simplicity) using ReLU (Rectified Linear Unit) as the activation function [20]. The output layer has one neuron and uses a sigmoid activation function to output a probability of loan approval.

7.3 Rule Extraction Process

The rule extraction process involves several steps designed to translate the ANN's learned weights and biases into a set of if-then rules that replicate the network's decision-making process as closely as possible:

– Simplification: The first step involves simplifying the ANN to make the rule extraction more manageable. This could include pruning insignificant weights (shallows values) and neurons that have little impact on the output based on sensitivity analysis.
– Discretization: Since ANNs deal with continuous inputs and hidden layer activations, a discretization process is applied to convert these continuous values into categorical ranges. For instance, age might be categorized into 'Young', 'Middle-aged', and 'Old'; Income into 'Low', 'Medium', and 'High'; Credit Score into 'Poor', 'Fair', 'Good', and 'Excellent'; and Employment Status into 'Unemployed' and 'Employed'.
– Activation Pattern Analysis: Next, the activation patterns of the neurons in the hidden layer are analyzed for each input pattern. This involves feeding various combinations of the discretized input variables into the simplified network and observing which neurons in the hidden layer are activated for each combination. An activation threshold is defined to determine whether a neuron is considered activated.
– Rule Generation: Based on the activation patterns observed, rules are generated to replicate the ANN's decision process. Each rule corresponds to a path from the input layer through the activated hidden neurons to the output decision. For example:
  - If (Age is Young) and (Income is High) and (Credit Score is Good) and (Employment Status is Employed), then Approve Loan.
  - If (Age is Middle-aged) and (Credit Score is Poor), then Deny Loan.

  This step involves identifying which combinations of input features and hidden neuron activations lead to loan approval or denial, effectively translating the ANN's complex decision boundaries into more interpretable formats.
– Rule Refinement and Validation: The initial set of rules may be too complex or too numerous for practical use. Rule refinement techniques simplify and consolidate the rules without significantly reducing their accuracy in replicating the ANN's decisions. The refined rules are then validated against a test dataset to accurately reflect the ANN's behavior. This may involve adjusting the regulations based on misclassifications or applying techniques to handle exceptions and edge cases.

After applying the rule extraction process to our hypothetical ANN, we might end up with a set of simplified, human-readable rules such as:

- Rule 1: If (Income is High) and (Credit Score is Excellent), then Approve Loan.
- Rule 2: If (Employment Status is Unemployed) and (Credit Score is Poor or Fair), then Deny Loan.
- Rule 3: If (Age is Old) and (Income is Low) and (Employment Status is Employed), then Deny Loan.

These rules provide clear criteria derived from the ANN's learned patterns, making the decision-making process transparent and justifiable.

7.4 Advantages and Challenges

Advantages:

- Transparency: The extracted rules make the ANN's decisions transparent and understandable to humans.
- Compliance: Clear rules can help ensure compliance with regulatory requirements for explainable AI.
- Trust: Understanding how decisions are made can increase user trust in the AI system.

Challenges:

- Complexity: The rule extraction process can be complex, especially for deep or highly non-linear networks [38].
- Approximation: The extracted rules approximate the ANN's decision process and may not capture all nuances.
- Scalability: Extracting rules from large, deep neural networks with many inputs and hidden layers can be challenging and may result in many complex rules [39].

7.5 Summing-up

Modern large language models (LLMs) [40], such as those based on the Transformer architecture, implement ANNs in their backend primarily through deep learning techniques. These models consist of layers of interconnected neurons, each capable of performing calculations using input data and generating output that feeds into subsequent layers. The Transformer models, a subset of ANNs, utilize attention mechanisms that allow the model to weigh the importance of different words in a sentence regardless of their distance from each other in the text. This ability to manage long-range dependencies within text is crucial for understanding and generating human-like language. The training of these models involves backpropagation, where errors are used to adjust the weights of the connections between neurons across many layers, optimizing the model's performance on language tasks. This architecture enables LLMs to excel in various language processing tasks by capturing complex patterns in large volumes of text data.

Extracting rules from ANNs provides a method for rendering AI decisions more transparent, understandable, and justifiable. Although the process poses challenges, especially in complex networks, it is a vital move toward AI's responsible and ethical use. Making AI systems more interpretable helps build trust with users, ensures adherence to regulatory standards, and offers important insights into how decisions are made.

8 Fuzzy Cognitive Maps

The pendulum in AI is swinging back from purely statistical approaches toward integrating structured knowledge. FCMs emerge as a compelling solution that bridges the gap between symbolic AI's highly structured, rule-based reasoning and the pattern-driven, data-intensive approaches of sub-symbolic AI. FCMs incorporate elements of both paradigms, utilizing a graph-based representation to model complex systems and their behaviors through concepts and causal relationships akin to symbolic AI [41]. Yet, they also integrate aspects of fuzzy logic, allowing for handling uncertainty and imprecision in a more characteristic of sub-symbolic AI. This hybrid approach enables FCMs to model systems dynamically and adaptively, capturing both the structured knowledge of expert systems and the adaptive learning capabilities of neural networks [42]. The need for FCMs arises from the challenges faced by purely symbolic or sub-symbolic systems when dealing with real-world applications that require both interpretability and flexibility. In domains like healthcare, environmental management, and strategic planning, decisions must be made with a clear understanding of causal relationships and an accommodation for uncertainties and ambiguities inherent in real-world data. FCMs address these needs by providing a cognitively intuitive framework capable of learning from empirical data. This dual strength makes FCMs particularly valuable for scenarios where stakeholders must navigate complex decision-making environments. This allows for a balance of precise rule-based logic and adaptive, data-driven inference [43].

8.1 Origins

Bart Kosko introduced the concept of FCMs in the 1980s as an extension of cognitive maps. Cognitive maps, developed by Axelrod, were diagrams that represented beliefs and their interconnections [44]. Kosko's introduction of fuzziness to these maps allowed for the representation of causal reasoning with degrees of truth rather than binary true/false values, thus capturing the uncertain and imprecise nature of human knowledge and decision-making processes. FCMs combine elements from fuzzy logic, introduced by Lotfi A. Zadeh, with the structure of cognitive maps to model complex systems.

8.2 Structure and Functionality

FCMs are graph-based representations where nodes represent concepts or entities within a system, and directed edges depict the causal relationships between these concepts. Each edge is assigned a weight that indicates the relationship's strength and direction (positive or negative) [45]. This structure closely mirrors that of artificial neural networks, particularly in how information flows through the network and how activation levels of concepts are updated based on the input they receive, akin to the weighted connections between neurons in neural networks [46].

However, unlike typical neural networks that learn from data through backpropagation or other learning algorithms, the weights in FCMs are often determined by experts or derived from data using specific algorithms designed for FCMs. The concepts in FCMs can be activated like neurons, with their states updated based on fuzzy causal relations, allowing for dynamic modeling of system behavior over time. Integrating structured knowledge graphs with distributed neural network representations offers a promising path to augmented intelligence. We get the flexible statistical power of neural networks that predict, classify, and generate based on patterns—combined with the formalized curated knowledge encoding facts, logic, and semantics via knowledge graphs [47]. Table 1 provides a high-level comparison of both approaches.

While both serve as models for simulating complex systems and decision-making processes, they differ significantly in their structure and functionality. FCMs utilize a graph-based approach with nodes representing concepts and edges depicting causal relationships, which lends itself to high interpretability and the ability to handle uncertainty through fuzzy logic. This suits FCMs for scenarios requiring transparent reasoning and adaptability to nuanced changes. In contrast, ANNs comprise layered nodes connected by weights, focusing on pattern recognition and classification through a data-driven, often opaque process known as backpropagation. While ANNs excel in tasks involving large datasets and require the identification of patterns or trends, their "black box" nature can make them less suitable for applications where understanding the basis of decisions is crucial. Despite these differences, both models leverage connectivity and iterative learning, underscoring their utility in dynamic and complex problem-solving environments [48].

Table 1. Simple comparison of FCMs ANNs

Feature	**Fuzzy Cognitive Maps (FCMs)**	**Artificial Neural Networks (ANNs)**
Structure	Graph-based, nodes represent concepts	Layered nodes (neurons) connected by weights
Data Handling	Manages uncertainty and imprecision with fuzzy logic	Processes numerical inputs using activation functions
Learning Method	Heuristic updates based on expert knowledge and data	Backpropagation and other gradient-based methods
Interpretability	High, due to transparent causal relationships	Low, often considered "black boxes"
Application Areas	Complex decision-making, strategic planning	Pattern recognition, classification tasks
Adaptability	Adjusts to new information through causal reinforcement	Learns from large datasets to improve accuracy
Rule Integration	Integrates explicit rules and relationships	Learns rules implicitly through training data

8.3 Case Studies

FCMs have been applied across various domains, demonstrating their versatility and effectiveness as a hybrid AI tool:

- Decision Support Systems: FCMs model complex decision-making processes, integrating expert knowledge and data-driven insights to support decisions in healthcare, environmental management, and business strategy.
- Predictive Modeling: In healthcare, FCMs model the progression of diseases or the impact of treatments, incorporating medical expertise and patient data to predict outcomes and support personalized medicine [49].
- System Analysis and Design: FCMs help analyze and design complex systems, such as socio-economic systems or ecosystems, by modeling the interactions between various factors and predicting the impact of changes or interventions.
- Healthcare Management: FCMs have been employed to model and predict patient outcomes in healthcare settings. For example, an FCM can be developed to understand the complex interplay between patient symptoms, treatment options, and possible outcomes, aiding medical professionals in decision-making [50].

- Environmental and Ecological Systems: In environmental studies, FCMs have been used to model the impact of human activities on ecosystems, allowing for the simulation of various scenarios based on different policies or interventions. This application showcases the strength of FCMs in handling systems where data may be scarce or imprecise [51].
- Business and Strategic Planning: FCMs assist in strategic planning and decision-making within business contexts by modeling the relationships between market forces, company policies, and financial outcomes, offering a tool for scenario analysis and strategy development [52].

8.4 Advantages

The hybrid nature of FCMs offers several advantages:

- Interpretability and Transparency: The symbolic representation of concepts and causal relationships in FCMs provides clarity and understandability, facilitating communication with experts and stakeholders and supporting explainable AI.
- Flexibility and Adaptability: FCMs can be easily updated with new knowledge or data, allowing them to adapt to changing conditions or insights. This makes them particularly valuable in fields where knowledge evolves rapidly.
- Handling of Uncertainty: Using fuzzy values to represent causal strengths enables FCMs to deal effectively with uncertainty and ambiguity, providing more nuanced and realistic modeling of complex systems [53].
- Integration of Expert Knowledge and Data-Driven Insights: FCMs uniquely combine expert domain knowledge with learning from data, bridging the gap between purely knowledge-driven and purely data-driven approaches.
- Interpretability: The graphical representation of FCMs, similar to semantic nets, allows for straightforward interpretation and understanding of the modeled system, making it accessible to experts and stakeholders without deep technical knowledge of AI.
- Flexibility: FCMs can incorporate quantitative and qualitative data, effectively handling uncertainty and imprecision through fuzzy logic. This flexibility makes them suitable for a wide range of applications.
- Dynamic Modeling Capability: FCMs can simulate the dynamic behavior of systems over time, providing valuable insights into potential future states based on different inputs or changes in the system [54].

8.5 Limitations

Despite their advantages, FCMs also face several challenges:

- Complexity with Large Maps: As the number of concepts and relationships in an FCM increases, the map can become complex and challenging to manage, analyze, and interpret [32].
- Learning and Optimization: While FCMs can learn from data, adjusting the fuzzy values of causal relationships can be computationally intensive and require sophisticated optimization techniques, especially for large and complex maps [55].
- Quantification of Expert Knowledge: Translating expert knowledge into precise fuzzy values for causal relationships can be challenging and may introduce subjectivity, requiring careful validation and sensitivity analysis [56].
- Subjectivity in Model Construction: The reliance on expert knowledge for constructing FCMs can introduce subjectivity, especially in determining the strength and direction of causal relationships between concepts.

– Complexity with Large Maps: As the number of concepts increases, the FCM can become complex and challenging to manage and interpret, potentially requiring sophisticated computational tools for simulation and analysis.
– Limited Learning Capability: While FCMs can be adjusted or trained based on data to some extent, they lack the deep learning capabilities of more advanced neural networks, which can autonomously learn complex patterns from large datasets [57].

8.6 FCMs as a Hybrid AI Approach

There is momentum toward hybridizing connectionism and symbolic approaches to AI to unlock potential opportunities for an intelligent system to make decisions. This hybrid approach is gaining ground; FCMs embody a hybrid AI approach through their integration of symbolic and sub-symbolic elements:

– Symbolic Components: The concepts and causal connections in FCMs are symbolic, explicitly representing entities and their interrelations. This aligns with the symbolic AI paradigm, where knowledge is structured and interpretable, allowing for reasoning and inference based on explicit rules and relationships [58].
– Sub-symbolic Components: The strengths of the causal relationships in FCMs are represented by fuzzy values, which are learned and adjusted based on data or expert input, much like the weights in neural networks. This learning capability and the use of fuzzy logic to handle uncertainty and ambiguity mirror the characteristics of sub-symbolic AI, which learns from patterns in data without requiring explicit programming.

FCMs offer a compelling hybrid approach to AI, combining the symbolic representation of knowledge with sub-symbolic learning and reasoning; they bridge a crucial gap between symbolic AI's interpretability and structured knowledge representation and the adaptability and data-driven learning of sub-symbolic AI. Their applications across diverse domains underscore their versatility and potential to address complex problems by integrating human-like reasoning with machine learning. The challenges FCMs face, including complexity management and the quantification of expert knowledge, highlight areas for further research and development. As AI continues to evolve towards more integrated and versatile models, FCMs stand as a testament to the potential of hybrid approaches to combine the strengths of symbolic and sub-symbolic AI, offering a pathway to more intelligent, understandable, and adaptable AI systems [59].

8.7 Summing-up

FCMs offer a robust framework for modeling and analyzing complex systems, blending the best symbolic and sub-symbolic AI by integrating fuzzy logic, graph theory, and neural network-like dynamics. While FCMs provide a powerful tool for understanding system behaviors and decision-making processes, their effectiveness is contingent upon accurately representing causal relationships and managing map complexity. Future developments in FCMs aim to enhance their learning capabilities, reduce subjectivity in their construction, and improve scalability, further solidifying their role as a valuable tool in complex system analysis and decision support across various domains [60].

9 Conclusion and reflection

Both symbolic and sub-symbolic AI approaches effectively address complex problems, but each excels in different contexts. Symbolic AI is particularly well-suited for expert or knowledge-based systems that require human input and domain-specific expertise. In contrast, sub-symbolic AI is ideal for tasks requiring continuous learning, such as natural language processing, speech recognition, and image recognition. Therefore, evaluating the specific requirements and constraints of the problem is crucial before choosing an approach. The success of an AI application largely depends on selecting a strategy that aligns best with the problem's demands. The ongoing debate between proponents of symbolic AI and sub-symbolic AI reflects deep-rooted differences in their perceived capabilities. Advocates for symbolic AI argue that it is the only path to accurate intelligence and understanding, emphasizing its reliance on human-like cognitive processes such as reasoning and logic. They contend that sub-symbolic AI, primarily based on statistical algorithms, falls short in its ability to reason abstractly and produce genuinely intelligent behavior. Conversely, supporters of sub-symbolic AI believe it provides a more adaptable and robust framework for achieving intelligence, drawing on its ability to mimic brain processes and handle the intricacies of real-world scenarios. Sub-symbolic AI's capacity to learn from experience and adapt over time, without extensive hand-coding, is a significant advantage.

The debate between symbolic and sub-symbolic AI is unlikely to be resolved swiftly, as both paradigms offer distinct advantages and limitations. The choice between them depends on the particular needs of the application. However, a deeper understanding of the arguments for each approach can fuel progress in developing more sophisticated and effective AI systems. Integrating symbolic and sub-symbolic approaches is crucial when creating AI systems. Symbolic AI shines in scenarios that demand logical reasoning and problem-solving capabilities that can be articulated through rules and symbols. On the other hand, sub-symbolic AI is superior in domains where pattern recognition and experiential learning are critical, such as speech recognition, image recognition, and natural language processing. One significant benefit of leveraging both approaches is their potential to create hybrid models, which enhance the efficacy of solving complex problems. For example, sub-symbolic AI can be utilized for feature extraction in image recognition tasks, while symbolic AI can handle classification duties.

Understanding the strengths and limitations of each approach and using them in tandem can forge more comprehensive and intelligent AI systems. These systems can surmount the challenges posed by relying solely on one AI approach. While symbolic AI, with its rule-based systems, is effective in environments where problem-solving processes are rule-intensive and well-defined, sub-symbolic AI thrives in more ambiguous settings where patterns and data-driven insights come into play. Symbolic AI requires expert input to craft precise rules, whereas sub-symbolic AI leverages vast datasets to learn and adapt. Moreover, symbolic AI's deterministic and transparent nature allows for accurate tracking of decisions, an advantage in applications where explainability is critical. In contrast, sub-symbolic AI often operates as a "black box," where the internal workings and decision-making processes are not easily discernible, posing challenges in situations requiring clear audit trails. Ultimately, the choice between symbolic and sub-symbolic AI hinges on the specific needs of the application and the project's goals. Employing both complementary can address a broader range of challenges and enhance AI systems' overall functionality and adaptability.

References

1. F. Vanhoenshoven, G. Nápoles, M. Creemers, M. Leon, and K. Vanhoof, "Analyzing the impact of the adaptive clearing mechanism on algorithm accuracy in variable mesh optimization," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8, IEEE, 2016.
2. M. Leon, R. Bello, and K. Vanhoof, "Cognitive maps in transport behavior," in *2009 Eighth Mexican International Conference on Artificial Intelligence*, pp. 179–184, IEEE, 2009.

3. M. Leon, G. Nápoles, C. Rodriguez, M. M. García, R. Bello, and K. Vanhoof, "A fuzzy cognitive maps modeling, learning and simulation framework for studying complex system," in *New Challenges on Bioinspired Applications: 4th International Work-conference on the Interplay Between Natural and Artificial Computation, IWINAC 2011, La Palma, Canary Islands, Spain, May 30-June 3, 2011. Proceedings, Part II 4*, pp. 243–256, Springer Berlin Heidelberg, 2011.
4. G. Nápoles, F. Hoitsma, A. Knoben, A. Jastrzebska, and M. Leon, "Prolog-based agnostic explanation module for structured pattern classification," *Information Sciences*, vol. 622, p. 1196–1227, Apr. 2023.
5. W. Saeed and C. Omlin, "Explainable ai (xai): A systematic meta-survey of current challenges and future opportunities," *Knowledge-Based Systems*, vol. 263, p. 110273, 2023.
6. C. Borrero-Domínguez and T. Escobar-Rodríguez, "Decision support systems in crowdfunding: A fuzzy cognitive maps (fcm) approach," *Decision Support Systems*, vol. 173, p. 114000, 2023.
7. M. Leon, G. Nápoles, R. Bello, L. Mkrtchyan, B. Depaire, and K. Vanhoof, "Tackling travel behaviour: an approach based on fuzzy cognitive maps," *International Journal of Computational Intelligence Systems*, vol. 6, no. 6, pp. 1012–1039, 2013.
8. O. Emir and Şule Önsel Ekici, "An integrated assessment of food waste model through intuitionistic fuzzy cognitive maps," *Journal of Cleaner Production*, vol. 418, p. 138061, 2023.
9. F. Hoitsma, A. Knoben, M. Leon, and G. Nápoles, "Symbolic explanation module for fuzzy cognitive map-based reasoning models," in *Artificial Intelligence XXXVII: 40th SGAI International Conference on Artificial Intelligence, AI 2020, Cambridge, UK, December 15–17, 2020, Proceedings 40*, pp. 21–34, Springer International Publishing, 2020.
10. G. Nápoles, F. Hoitsma, A. Knoben, A. Jastrzebska, and M. Leon, "Prolog-based agnostic explanation module for structured pattern classification," *Information Sciences*, vol. 622, pp. 1196–1227, 2023.
11. M. Leon, G. Nápoles, M. M. García, R. Bello, and K. Vanhoof, "A revision and experience using cognitive mapping and knowledge engineering in travel behavior sciences," *Polibits*, no. 42, pp. 43–50, 2010.
12. M. Leon, B. Depaire, and K. Vanhoof, "Fuzzy cognitive maps with rough concepts," in *Artificial Intelligence Applications and Innovations: 9th IFIP WG 12.5 International Conference, AIAI 2013, Paphos, Cyprus, September 30–October 2, 2013, Proceedings 9*, pp. 527–536, Springer Berlin Heidelberg, 2013.
13. M. Leon, N. M. Sanchez, Z. G. Valdivia, and R. B. Perez, "Concept maps combined with case-based reasoning in order to elaborate intelligent teaching/learning systems," in *Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)*, pp. 205–210, IEEE, 2007.
14. E. Struble, M. Leon, and E. Skordilis, "Intelligent prevention of ddos attacks using reinforcement learning and smart contracts," *The International FLAIRS Conference Proceedings*, vol. 37, May 2024.
15. G. Nápoles, M. L. Espinosa, I. Grau, K. Vanhoof, and R. Bello, *Fuzzy cognitive maps based models for pattern classification: Advances and challenges*, vol. 360, pp. 83–98. Springer Verlag, 2018.
16. M. Zhang, J. Wu, A. Belatreche, Z. Pan, X. Xie, Y. Chua, G. Li, H. Qu, and H. Li, "Supervised learning in spiking neural networks with synaptic delay-weight plasticity," *Neurocomputing*, vol. 409, pp. 103–118, 2020.
17. Y. Xia, J. Wang, Z. Zhang, D. Wei, and L. Yin, "Short-term pv power forecasting based on time series expansion and high-order fuzzy cognitive maps," *Applied Soft Computing*, vol. 135, p. 110037, 2023.
18. H. Wang, Y. Wu, Y. Liu, and W. Liu, "A cross-layer framework for lpwan management based on fuzzy cognitive maps with adaptive glowworm swarm optimization," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2023.
19. M. Leon, L. Mkrtchyan, B. Depaire, D. Ruan, R. Bello, and K. Vanhoof, "Learning method inspired on swarm intelligence for fuzzy cognitive maps: Travel behaviour modelling," in *Artificial Neural Networks and Machine Learning – ICANN 2012* (A. E. P. Villa, W. Duch, P. Érdi, F. Masulli, and G. Palm, eds.), (Berlin, Heidelberg), pp. 718–725, Springer Berlin Heidelberg, 2012.
20. G. Nápoles, Y. Salgueiro, I. Grau, and M. Leon, "Recurrence-aware long-term cognitive network for explainable pattern classification," *IEEE Transactions on Cybernetics*, vol. 53, no. 10, pp. 6083–6094, 2023.
21. G. Nápoles, J. L. Salmeron, W. Froelich, R. Falcon, M. Leon, F. Vanhoenshoven, R. Bello, and K. Vanhoof, "Fuzzy cognitive modeling: Theoretical and practical considerations," in *Intelligent Decision Technologies 2019: Proceedings of the 11th KES International Conference on Intelligent Decision Technologies (KES-IDT 2019), Volume 1*, pp. 77–87, Springer Singapore, 2020.
22. S. S Band, A. Yarahmadi, C.-C. Hsu, M. Biyari, M. Sookhak, R. Ameri, I. Dehzangi, A. T. Chronopoulos, and H.-W. Liang, "Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods," *Informatics in Medicine Unlocked*, vol. 40, p. 101286, 2023.
23. J. Wang, Z. Peng, X. Wang, C. Li, and J. Wu, "Deep fuzzy cognitive maps for interpretable multivariate time series prediction," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 9, pp. 2647–2660, 2021.
24. N. Martinez, M. Leon, and Z. Garcia, "Features selection through fs-testors in case-based systems of teaching-learning," in *Mexican International Conference on Artificial Intelligence*, pp. 1206–1217, Springer, 2007.

25. G. Nápoles, Y. Salgueiro, I. Grau, and M. Leon, "Recurrence-aware long-term cognitive network for explainable pattern classification," *IEEE transactions on cybernetics*, vol. 53, no. 10, pp. 6083–6094, 2022.
26. H. DeSimone and M. Leon, "Explainable ai: The quest for transparency in business and beyond," in *2024 7th International Conference on Information and Computer Technologies (ICICT)*, IEEE, Mar. 2024.
27. A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-Lopez, D. Molina, R. Benjamins, *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information fusion*, vol. 58, pp. 82–115, 2020.
28. G. Nápoles, I. Grau, R. Bello, M. Leon, K. Vahoof, and E. Papageorgiou, "A computational tool for simulation and learning of fuzzy cognitive maps," in *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–8, IEEE, 2015.
29. G. Nápoles and M. Leon, "Métodos para la construccion de mapas cognitivos difusos y sus aplicaciones," *Anales de la Academia de Ciencias de Cuba*, 2016.
30. G. Nápoles, I. Grau, K. Vanhoof, R. Bello, and M. Leon, "Soft computing based optimization and decision models-fuzzy cognitive maps based models for pattern classification: Advances and challenges," *Soft Computing Based Optimization and Decision Models*, pp. 83–98, 2017.
31. M. Leon, "Fuzzy cognitive maps as a bridge between symbolic and sub-symbolic artificial intelligence," *International Journal on Cybernetics & Informatics*, vol. 13, no. 4, pp. 57–75, 2024.
32. R. Schuerkamp and P. J. Giabbanelli, "Extensions of fuzzy cognitive maps: A systematic review," *ACM Comput. Surv.*, vol. 56, no. 2, 2023.
33. M. Leon, I. Bonet, M. M. Garcia, and Z. Garcia, "Combining concept maps and petri nets to generate intelligent tutoring systems: A possible approach," in *Mexican International Conference on Artificial Intelligence*, pp. 797–805, Springer, 2008.
34. M. Leon, "Comparing llms using a unified performance ranking system," *International Journal of Artificial Intelligence and Applications*, vol. 15, no. 4, pp. 33–46, 2024.
35. H. DeSimone and M. Leon, "Leveraging explainable ai in business and further," in *3rd IEEE Opportunity Research Scholars Symposium*, 2024.
36. Y. Xixi, D. Fengqian, and L. Chao, "Time series prediction based on high-order intuitionistic fuzzy cognitive maps with variational mode decomposition," *Soft Computing*, vol. 26, pp. 189–201, 2022.
37. M. Wrzesień, M. Wrzesień, and W. Homenda, "Time series processing with cognitive maps. the case of general forecast modeling for time series of similar nature," in *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–8, 2022.
38. Z. Yang, J. Liu, and K. Wu, "Learning of boosting fuzzy cognitive maps using a real-coded genetic algorithm," in *2019 IEEE Congress on Evolutionary Computation (CEC)*, pp. 966–973, 2019.
39. W. Liang, Y. Zhang, X. Liu, H. Yin, J. Wang, and Y. Yang, "Towards improved multifactorial particle swarm optimization learning of fuzzy cognitive maps: A case study on air quality prediction," *Applied Soft Computing*, vol. 130, p. 109708, 2022.
40. M. Leon, "Benchmarking large language models with a unified performance ranking metric," *International Journal on Foundations of Computer Science & Technology*, vol. 14, no. 4, pp. 15–27, 2024.
41. M. Leon, L. Mkrtchyan, B. Depaire, D. Ruan, and K. Vanhoof, "Learning and clustering of fuzzy cognitive maps for travel behaviour analysis," *Knowledge and information systems*, vol. 39, pp. 435–462, 2014.
42. M. Leon, N. Martinez, Z. Garcia, and R. Bello, "Concept maps combined with case-based reasoning in order to elaborate intelligent teaching/learning systems," in *Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)*, pp. 205–210, 2007.
43. Y. Hu, Y. Guo, and R. Fu, "A novel wind speed forecasting combined model using variational mode decomposition, sparse auto-encoder and optimized fuzzy cognitive mapping network," *Energy*, vol. 278, p. 127926, 2023.
44. M. Leon, L. Mkrtchyan, B. Depaire, D. Ruan, and K. Vanhoof, "A travel behaviour study through learning and clustering of fuzzy cognitive maps," *Decision Aid Models for Disaster Management and Emergencies*, pp. 259–293, 2013.
45. M. Leon Espinosa and G. Napoles Ruiz, "Modeling and experimentation framework for fuzzy cognitive maps," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, Mar. 2016.
46. W. Hoyos, J. Aguilar, and M. Toro, "A clinical decision-support system for dengue based on fuzzy cognitive maps," *Health Care Management Science*, vol. 25, no. 4, pp. 666–681, 2022.
47. W. Hoyos, J. Aguilar, and M. Toro, "Prv-fcm: An extension of fuzzy cognitive maps for prescriptive modeling," *Expert Systems with Applications*, vol. 231, p. 120729, 2023.
48. M. Leon, "Fuzzy cognitive maps as a bridge between symbolic and sub-symbolic artificial intelligence," *International Journal on Cybernetics & Informatics (IJCI). 3rd International Conference on Artificial Intelligence Advances (AIAD 2024)*, vol. 13, pp. 57–75, Aug. 2024.

49. K. Poczeta and E. I. Papageorgiou, "Energy use forecasting with the use of a nested structure based on fuzzy cognitive maps and artificial neural networks," *Energies*, vol. 15, no. 20, 2022.
50. G. D. Karatzinis, N. A. Apostolikas, Y. S. Boutalis, and G. A. Papakostas, "Fuzzy cognitive networks in diverse applications using hybrid representative structures," *International Journal of Fuzzy Systems*, vol. 25, no. 7, pp. 2534–2554, 2023.
51. O. Orang, P. C. de Lima e Silva, and F. G. Guimarães, "Time series forecasting using fuzzy cognitive maps: a survey," *Artificial Intelligence Review*, vol. 56, p. 7733–7794, 2023.
52. M. Leon, "Business technology and innovation through problem-based learning," in *Canada International Conference on Education (CICE-2023) and World Congress on Education (WCE-2023)*, CICE-2023, Infonomics Society, July 2023.
53. M. Leon, "Aggregating procedure for fuzzy cognitive maps," *The International FLAIRS Conference Proceedings*, vol. 36, no. 1, 2023.
54. E. Jiya, O. Georgina, and A. O., "A review of fuzzy cognitive maps extensions and learning," *Journal of Information Systems and Informatics*, vol. 5, no. 1, pp. 300–323, 2023.
55. S. Yang and J. Liu, "Time-series forecasting based on high-order fuzzy cognitive maps and wavelet transform," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3391–3402, 2018.
56. T. Koutsellis, G. Xexakis, K. Koasidis, N. Frilingou, A. Karamaneas, A. Nikas, and H. Doukas, "In-cognitive: A web-based python application for fuzzy cognitive map design, simulation, and uncertainty analysis based on the monte carlo method," *SoftwareX*, vol. 23, 2023.
57. D. Qin, Z. Peng, and L. Wu, "Deep attention fuzzy cognitive maps for interpretable multivariate time series prediction," *Knowledge-Based Systems*, vol. 275, p. 110700, 2023.
58. M. Leon, L. Mkrtchyan, B. Depaire, D. Ruan, and K. Vanhoof, "Learning and clustering of fuzzy cognitive maps for travel behaviour analysis," *Knowledge and Information Systems*, vol. 39, no. 2, pp. 435–462, 2013.
59. G. Nápoles, J. L. Salmeron, W. Froelich, R. Falcon, M. Leon, F. Vanhoenshoven, R. Bello, and K. Vanhoof, *Fuzzy Cognitive Modeling: Theoretical and Practical Considerations*, p. 77–87. Springer Singapore, July 2019.
60. G. Nápoles, M. Leon, I. Grau, and K. Vanhoof, "FCM expert: Software tool for scenario analysis and pattern classification based on fuzzy cognitive maps," *International Journal on Artificial Intelligence Tools*, vol. 27, no. 07, p. 1860010, 2018.

Author

Dr. Maikel Leon is interested in applying AI/ML techniques to modeling real-world problems using knowledge engineering, representation, and data mining methods. His recent research focuses on XAI and has been featured in high-impact factor journals, such as Information Sciences and IEEE Transactions on Cybernetics. Dr. Leon is a reviewer for the International Journal of Knowledge and Information Systems, Journal of Experimental and Theoretical Artificial Intelligence, Soft Computing, and IEEE Transactions on Fuzzy Systems. He is a Committee Member of the Florida Artificial Intelligence Research Society. He is a frequent contributor on technology topics for CNN en Español TV and the winner of the Cuban Academy of Sciences National Award for the Most Relevant Research in Computer Science. Dr. Leon obtained his PhD in Computer Science at Hasselt University, Belgium, previously having studied computation (Master of Science and Bachelor of Science) at Central University of Las Villas, Cuba.