A Parametric Model of Cognitive Complexity for Symbolic Knowledge-Based Decision Systems

Federico Sabbatini[1,2,3,*]

[1]Department of Computer Science and Engineering, University of Bologna, Viale del Risorgimento 2, 40136, Bologna, Italy
[2]National Institute for Nuclear Physics, Section in Florence, Via Bruno Rossi 1, 50019, Sesto Fiorentino, Italy
[3]Department of Pure and Applied Sciences, University of Urbino, Via Santa Chiara 27, 61029, Urbino, Italy

Abstract

The growing adoption of symbolic knowledge-based decision systems highlights the need of principled methods to quantify their interpretability. While symbolic models are often considered inherently understandable, the cognitive effort required to interpret them varies significantly depending on both their structural properties and on user-specific factors. Existing approaches typically rely on coarse structural proxies, neglect the user perceptions and are often limited to specific representation formats. This work introduces a representation-agnostic and parametric framework for modelling the cognitive complexity of symbolic knowledge bases. First, a taxonomy is proposed to characterise structural and procedural elements of knowledge representations contributing to cognitive complexity, including size, depth, logical composition, and evaluation dynamics. Building on this taxonomy, a unified formal model is defined. The model decomposes cognitive complexity into micro-level components, capturing the internal structure of knowledge units, and macro-level components, accounting for their global organisation and evaluation process. The model also incorporates user-defined tolerance parameters, enabling a profiled assessment of perceived complexity. The proposed framework is validated through an experimental analysis on a set of knowledge bases with different structural properties. Results show that the model captures meaningful differences in cognitive effort across representations and user profiles, providing a flexible tool for designing, evaluating, and comparing interpretable symbolic models.

Keywords

Explainable artificial intelligence, Interpretability, Symbolic knowledge representation, User-centred modelling

1. Introduction

The increasing adoption of machine learning (ML) models in high-stakes domains has amplified the need of outcomes that are not only accurate but also understandable to humans [1, 2, 3, 4]. In this context, symbolic representations as decision rule lists, trees, and other structured knowledge bases are often regarded as inherently interpretable, as they expose the decisional process in a human-readable form [5, 6, 7, 8, 9]. However, interpretability is not a binary property of a model, but rather a matter of degree, strongly influenced by the actual effort required to understand the adopted representation [10].

A central challenge in this area is the lack of a principled and general framework to quantify such cognitive effort. Existing approaches typically rely on simple structural proxies, such as the number of rules, their length, or the depth of a tree [11, 12]. While these measures provide useful heuristics, they capture only limited aspects of the cognitive process involved in understanding symbolic knowledge. In particular, they often neglect the internal structure of rules, the combinatorial complexity induced by logical constructs, and the procedural aspects arising from the organisation of knowledge units. Moreover, most existing metrics are tightly coupled to specific representational forms and do not generalise across different types of symbolic models.

Another fundamental limitation is the lack of explicit modelling of user-dependent factors [13]. The effort required to understand a knowledge representation depends not only on its structural properties, but also on the user background, expertise, and familiarity with specific forms of reasoning [14, 15]. As

Joint Workshop on Statistics and Knowledge Integration for Logic, Learning, Ethical Decisions, and LLMs (SKILLED-LLMs 2026), co-located with FLoC 2026, Lisbon, Portugal
[*]Corresponding author.
f.sabbatini@unibo.it (F. Sabbatini)
0000-0002-0532-6777 (F. Sabbatini)

a result, the same knowledge base may be perceived as simple or complex depending on the observer, a variability that is rarely captured in existing formalisations.

In this work, these limitations are addressed by introducing *(i)* a taxonomy of the cognitive load to understand a knowledge base and *(ii)* a unified framework for modelling the cognitive complexity of symbolic knowledge representations. The proposed taxonomy characterises knowledge bases in terms of structural and procedural dimensions, independently of their external representation. Building on this taxonomy, a parametric model is introduced to decompose cognitive complexity into quantifiable sub-components, capturing the contribution of both the intrinsic structure of knowledge units and their global organisation within the knowledge base. Since the model incorporates user-defined tolerance parameters, a flexible and personalised assessment of cognitive effort is enabled.

The abstraction from specific representational formats in order to focus on the underlying structural properties of knowledge bases allows a representation-agnostic evaluation of cognitive complexity. The integration of user-defined parameters allows the framework to capture variability in how such complexity is perceived. As a consequence, the model makes it possible to analyse and compare different symbolic models within a common formal setting, while accounting for both objective complexity and subjective perception.

2. Related Works

The problem of quantifying interpretability and cognitive complexity has been widely investigated in the context of explainable artificial intelligence (XAI), although a unified and principled framework is still missing. Early work has emphasised that interpretability is inherently user-dependent and cannot be reduced to a purely structural property of a model. For instance, the work in [16] highlights how human understanding is shaped by cognitive and social factors, whereas the findings in [14] advocate for a more rigorous and human-centred evaluation of interpretability.

In the context of symbolic ML models and rule-based decision systems, interpretability is often associated with structural simplicity. Common interpretability proxies include the number of rules, their length, and the depth of decision trees, as discussed in several works on XAI [17, 18, 19, 20]. Similarly, approaches based on the induction of decision sets [21], lists [22] or trees [23] explicitly aim at producing compact and human-readable representations, typically favouring sparsity and limited rule complexity. While these measures are intuitive and easy to compute, they capture only coarse-grained aspects of cognitive effort and do not account for the internal structure of rules or the combinatorial effects induced by logical operators.

More recent work has started to recognise the importance of modelling finer-grained aspects of complexity [24]. For example, different logical constructs, such as conjunctions, disjunctions, and threshold-based conditions, may induce different levels of cognitive load, even when standard metrics such as knowledge size, tree depth, or rule length remain constant. However, these aspects are typically analysed in isolation and lack a unified formalisation that can be consistently applied across different representations. In addition, most existing approaches are tightly coupled to specific model classes, such as decision trees or rule lists, limiting their applicability in more general settings.

Another important line of research concerns the user role in assessing interpretability. Several works have explored personalised or human-in-the-loop approaches, where explanations are adapted to the needs and preferences of users. For instance, the work in [15] proposes the incorporation of human priors into the learning process, whereas the survey in [25] distinguishes between explanation and justification from a user-centric perspective. Despite these advances, the integration of user-dependent factors into formal models of cognitive complexity remains limited, and often treated qualitatively rather than quantitatively.

Despite the breadth of existing work, current approaches exhibit three main limitations. First, most interpretability proxies rely on coarse structural indicators, which do not capture the combinatorial complexity induced by logical constructs nor the internal structure of predicates. Second, existing metrics are typically tied to specific model classes and therefore lack a unified formulation that generalises across heterogeneous symbolic representations. Third, the role of users is rarely formalised quantitatively, even though several works acknowledge that interpretability is inherently user-dependent.

The present work aims to bridge these gaps by introducing a representation-agnostic and parametric framework for modelling cognitive complexity. Unlike existing approaches, the proposed model combines a fine-grained structural analysis of knowledge representations with an explicit parametrisation of user-dependent perception. This allows different sources of complexity (e.g., unit structure, combinatorial expansion, and evaluation dynamics) to be captured within a unified formal setting, while enabling flexible adaptation to different user profiles and contexts.

3. A Cognitive Complexity Taxonomy for Knowledge Representations

To characterise the cognitive complexity of symbolic knowledge base representations, it is necessary to identify the different forms in which a knowledge base can be expressed and the structural properties that influence its interpretability. To this end, a taxonomy is introduced to capture the main dimensions along which knowledge representations vary, with particular attention to those aspects that have a direct or indirect impact on human cognitive effort.

A symbolic knowledge base can be defined as a collection of individual symbolic units, whose specific nature depends on the chosen representation. Units may correspond, for instance, to individual rules in list-based systems, distinct regions in decision plots, or sentences in natural language descriptions. Despite this apparent heterogeneity, these units can be abstracted under a common formal perspective as logical formulæ composed of atomic conditions and their compositions, as formally defined later in this work. This abstraction enables a unified treatment of cognitive cost across representations that are heterogeneous yet logically equivalent or comparable.

It is important to highlight that logical equivalence does not imply cognitive equivalence. Different representational forms impose different perceptual and processing demands on the user, affecting how information is accessed, navigated, and mentally structured. For instance, a set of rules presented as a flat list encourages sequential reading, whereas the same knowledge organised as a tree supports hierarchical exploration and chunking. Similarly, graphical representations may facilitate spatial reasoning while introducing additional visual interpretation effort.

In the proposed framework, distinct representational categories are not explicitly modelled. Rather, they are captured indirectly through the structural and procedural properties that characterise each representation. Aspects such as the number of elements to be combined within a unit, the presence of alternative evaluation paths, and the degree of sequential dependency collectively reflect how a representation is processed by the user. As a result, the model remains representation-agnostic at the formal level, while still accounting for the cognitive effects induced by different representational forms through their underlying structural features.

Knowledge size. A first fundamental aspect of knowledge representational complexity concerns the overall dimension of the knowledge base. Indeed, the number of knowledge units plays a central role in determining mental effort, as larger systems require more extensive exploration and impose a greater memory burden. This notion generalises naturally across representations: for instance, it corresponds to the number of rules in a list, the number of root-to-leaf paths in a decision tree, the number of rows in a decision table, the number of regions in a decision plot, or the number of sentences in a natural language description. While these quantities differ in their surface interpretation, they all capture the same underlying concept of unit multiplicity.

Unit depth. A second, equally important structural aspect concerns the depth of knowledge units, namely the amount of information that must be sequentially combined in order to fully understand a single unit. While size captures how many units must be analysed, depth captures how cognitively demanding each unit is in isolation, thereby affecting working memory load and compositional reasoning effort. Importantly, also this dimension is largely independent of the specific format used to express the unit, and instead captures a more general notion of informational density. For instance, depth corresponds to the number of conditions that must be combined, typically through conjunctions, to evaluate a rule in list-based systems. In decision trees, it can be associated with the number of nodes traversed along a root-to-leaf path. In tabular representations, it reflects the number of attributes that are effectively involved in a row for decision making. In graphical representations such as decision plots, depth can be related to the geometric complexity of a region, e.g., in terms of how many sides or bounding conditions are required to define its decision boundary. In natural language descriptions, it corresponds to the number of distinct pieces of information or constraints that must be integrated within a sentence to convey a decision rationale.

Unit structure. Beyond sheer size and depth, the internal structure of knowledge units introduces additional layers of complexity. Units are composed of two parts, one describing pre-conditions and one expressing the decision to return when the pre-conditions are satisfied. These conditions may be simple, atomic predicates or more general compositions of atomic predicates. Even in the absence of complex logical composition, individual conditions may vary significantly in their interpretability. Simple axis-aligned thresholds, such as comparisons involving a single variable and a scalar constant, are typically easier to process, as they can be readily mapped to intuitive mental representations. In contrast, oblique conditions involving linear combinations of variables, fuzzy predicates defined over linguistic partitions, or constraints expressed over sets and intervals may require additional cognitive effort to interpret and manipulate. It is also important to notice that the cognitive impact of such constructs is not uniform across users. The perceived difficulty of these elements is thus user-dependent and must be explicitly accounted for in a cognitive model.

Combinatorial complexity. A complementary aspect of complexity arises at the level of knowledge units. Logical constructs such as disjunctions and more general constructs (e.g., m-of-n compositions) determine how many alternative evaluation paths must be considered to assess the truth of a unit. This combinatorial aspect is particularly relevant from a cognitive perspective, as it directly affects the number of hypotheses that must be simultaneously entertained during reasoning. Additionally, combinatorial aspects are not evenly perceived across users: while some individuals may find disjunctive or m-of-n rules intuitive, others may perceive them as significantly more demanding. This considerations must be properly encoded when modelling the cognitive effort.

Sequentiality. Another crucial dimension concerns the dynamics of evaluation, that is, the manner in which knowledge units are processed during reasoning. Indeed, the overall cognitive load to understand a knowledge base is not solely determined by its size and the intrinsic complexity of individual units, but also by the way in which they are organised and accessed. In unordered knowledge bases, units can be considered independently, and the effort required to understand each unit does not depend on its position [21]. In contrast, ordered structures impose a sequential evaluation process in which each unit must be interpreted in the context of all preceding ones. In such cases, understanding a given unit implicitly requires considering the negation of all prior unit pre-conditions, leading to an accumulation of cognitive effort along the sequence. Therefore, the average positional cost induced by sequential evaluation must be explicitly accounted for as a distinct source of cognitive effort.

3.1. From Taxonomy to Parametric Modelling of Cognitive Complexity

The taxonomy serves a dual purpose: providing a structured framework to analyse and compare different knowledge bases from a cognitive perspective, and establishing a clear conceptual link between knowledge properties and the components of the cognitive complexity model proposed here.

The central idea underlying the proposed model is that cognitive complexity arises from the interplay of two distinct types of components. On the one hand, structural factors are intrinsic to the knowledge base and depend on how information is organised and represented. On the other hand, perceptual factors relate to how such structural properties are interpreted by the user, reflecting differences in expertise, familiarity, and cognitive preferences.

This distinction is directly grounded in the taxonomy introduced above. Structural dimensions such as size, depth, logical composition, and evaluation dynamics define the objective characteristics of the knowledge base and its units. At the same time, several of these dimensions do not have a fixed cognitive impact, but rather contribute differently depending on the user's ability to process specific forms of

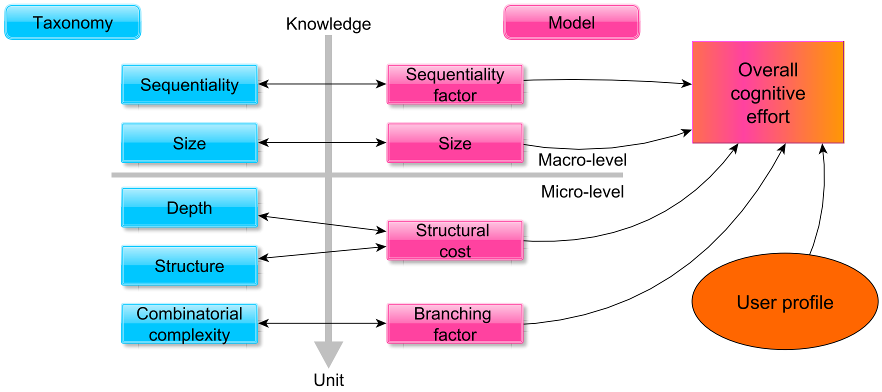


Figure 1: Correspondence between the taxonomy concepts and the modelling framework modules.

complexity. As a result, cognitive complexity cannot be reduced to a purely structural measure, but must account for the interaction between objective properties of the representation and user perception.

Building on this perspective, the proposed model separates the computation of intrinsic complexity (micro-level) from its aggregation and user-dependent modulation (macro-level). This separation enables a representation-agnostic treatment of knowledge bases while allowing the same structure to be perceived with different levels of difficulty across users and contexts.

More in detail, at the micro-level the focus is on the structural complexity of individual knowledge units. This analysis extends to the atomic conditions composing each unit. Dimensions such as logical, semantic, and combinatorial complexity directly contribute to the effort required to interpret a single unit in isolation. In particular, the number of elements that must be jointly considered within a unit, captured by the notion of depth introduced in the taxonomy, plays a central role in determining the amount of information that must be simultaneously processed. These aspects are modelled through a structural cost function, quantifying unit depth and structure, and a branching factor, reflecting combinatorial properties that contribute to cognitive effort.

At the macro-level, the model accounts for global properties of the knowledge base, as well as for the variability in how different users perceive and tolerate the accumulated complexity across units. Dimensions such as size and evaluation dynamics determine the overall burden required to analyse a complete knowledge base, while user-dependent tolerance parameters are introduced to modulate how the aforementioned different sources of complexity are perceived. In this sense, the macro-level integrates both high-level structural aspects of the knowledge base, i.e., its size and a sequentiality factor, and subjective factors related to the user's cognitive profile, allowing the model to represent variability in interpretability across different users and contexts.

Each dimension identified in the taxonomy is encoded in a component of the modelling framework, as depicted in Figure 1, where links between taxonomy and framework are explicitly shown. Micro- and macro-level are depicted separately, but in integration with the user contribution. This multi-level perspective ensures that the proposed model remains both expressive and flexible. As a result, different sources of cognitive complexity are captured in a coherent, unified manner and the model can account for a wide range of cognitive phenomena while preserving a clear and interpretable structure.

3.2. Current Limitations

The proposed taxonomy and modelling framework intentionally focus on a specific subset of factors that contribute most to the cognitive complexity of symbolic knowledge representations. While this choice enables a clear and tractable formalisation, it necessarily leaves out several aspects that may also play a relevant role in human understanding.

A first class of limitations concerns semantic properties of the knowledge base. The current formulation does not explicitly account for factors such as the total number of variables involved, the degree of overlap between knowledge units, the coverage rate of the input space, or the presence of conflicting units. These properties influence how knowledge is organised and interpreted, often affecting the ease with which a user can build a coherent mental model of the system.

A second limitation relates to the expressiveness and heterogeneity of knowledge units. In the present work, the focus is placed on the structure of the conditions that define each unit, abstracting away from potential variability in their internal composition. In practice, knowledge bases may include units of different nature, such as combinations of crisp, fuzzy, or aggregated constructs, which may introduce additional cognitive overhead due to heterogeneity. Similarly, numerical aspects such as the precision and granularity of constants are not explicitly modelled, despite their potential impact on interpretability, especially in domains where fine-grained thresholds are difficult to process mentally.

Furthermore, the proposed framework concentrates on the cognitive cost associated with the body of knowledge units (i.e., the unit pre-conditions), without explicitly modelling the complexity of their heads (i.e., the output decision). While this assumption is reasonable in many classification settings where outputs are simple labels, it becomes more restrictive in scenarios involving structured outputs, continuous predictions, or functional relationships.

Finally, the modelling of graphical and natural language representations are not considered here, as they introduce additional perceptual dimensions beyond the structural abstraction considered here. Nonetheless, the proposed framework has been conceived in order to possibly support future extensions based on the evaluation of representational properties at the stage of micro- and macro-level assessments.

These limitations are not merely omissions, but reflect deliberate design choices aimed at isolating the contribution of structural and combinatorial factors to cognitive complexity.

4. A User-Defined Parametric Model of Cognitive Complexity

4.1. Formalism

Common knowledge unit representations (including conjunctions, disjunctions, m-of-n, oblique and fuzzy predicates) can be captured within a single functional and combinatorial formalism, avoiding *ad-hoc* definitions. To this purpose, an atomic formula φ is defined in this work as follows:

$$\varphi = \sum_{i=1}^{k} c_i X_i \bowtie c, \qquad \text{(atom)} \tag{1}$$

where c_i are constant coefficients, X_i are variables, $\bowtie \in \{<, \leq, >, \geq, =, \in\}$ is an operator, and c is a constant that may be a scalar value, an interval, a finite set, or a fuzzy value. This formulation is naturally suitable to encode oblique decision boundaries (e.g., $2X - Y < 4$), but it may be as well simplified to represent other axis-aligned boundaries (e.g., $X > 6$, $Z = medium$[1]). The $\in$ operator is overloaded to denote both membership in a finite set (for categorical variables) and inclusion within an interval (for numerical expressions).

The formalism can be enriched to support the negation and composition of atomic formulæ, expressed as $\neg\varphi$ and $\Diamond^m(\varphi_1, \ldots, \varphi_n)$, respectively. The composition operator is defined as the disjunction of all possible conjunctions of m formulæ φ_i, as follows:

$$\Diamond^m(\varphi_1, \ldots, \varphi_n) = \bigvee_{\substack{S \subseteq \{1,\ldots,n\} \\ |S| = m}} \bigwedge_{i \in S} \varphi_i. \qquad \text{(composition)} \tag{2}$$

4.2. Micro-Level Cognitive Effort

Let φ be a logical formula representing the pre-conditions of a knowledge unit, defined recursively from: (i) atoms, (ii) negations, and (iii) composition of simpler formulæ. Two functions are then defined

[1]Categorical and fuzzy conditions can be seen as particular cases of Equation (1), where the linear combination involves a single variable and the operator $\bowtie$ denotes equality or membership over possibly structured domains (e.g., sets or intervals).

to quantify the effort required by a human to understand the logical formula φ:

$B(\varphi)$. Branching factor, measuring the degree of logical multiplicity of φ, i.e., how many alternative evaluation paths or condition combinations must be considered to determine its truth. It captures the combinatorial expansion induced by disjunctions and constructs such as m-of-n combinations;

$C(\varphi)$. Structural cost, measuring the cognitive effort required to evaluate the atomic components of φ (in terms of variables and constants) and to combine them according to the logical structure of φ.

Branching Factor. The branching factor is recursively defined as follows:

$$B(\varphi) = 1, \qquad \text{(atom)} \tag{3}$$

$$B(\neg\varphi) = B(\varphi), \qquad \text{(negation)} \tag{4}$$

$$B\big(\Diamond^m(\varphi_1, \ldots, \varphi_n)\big) = \sum_{\substack{S \subseteq \{1,\ldots,n\} \\ |S|=m}} \prod_{i \in S} B(\varphi_i). \qquad \text{(composition)} \tag{5}$$

In case of m-of-n predicates over atomic formulæ, the branching factor can be evaluated as $\binom{n}{m}$.

Structural Cost. Let $var(\varphi)$ and $const(\varphi)$ denote the sets of variables and constants appearing in φ, respectively. The structural cost is recursively defined as follows:

$$C(\varphi) = |var(\varphi)|^2 + \sum_{c \in const(\varphi)} cost(c), \qquad \text{(atom)} \tag{6}$$

$$C(\neg\varphi) = C(\varphi) + 1, \qquad \text{(negation)} \tag{7}$$

$$C\big(\Diamond^m(\varphi_1, \ldots, \varphi_n)\big) = \frac{\displaystyle\sum_{\substack{S \subseteq \{1,\ldots,n\} \\ |S|=m}} \left(\prod_{i \in S} B(\varphi_i)\right) \cdot \left(\sum_{i \in S} C(\varphi_i)\right)}{B\big(\Diamond^m(\varphi_1, \ldots, \varphi_n)\big)}, \qquad \text{(composition)} \tag{8}$$

where

$$cost(c) = \begin{cases} 1 & \text{if } c \text{ is a scalar (single numeric or categorical value),} \\ 2 & \text{if } c \text{ is a numerical interval (defined by two bounds),} \\ |c| & \text{if } c \text{ is a finite set of categorical values,} \\ k & \text{if } c \text{ is a fuzzy value associated with a variable partitioned into } k \text{ fuzzy intervals.} \end{cases}$$

Equation (6) explicitly penalises the occurrence of variables symbols in the conditional part of knowledge units, in order to reflect the higher cost of oblique representations. Equation (8) corresponds to the expected cost over all evaluation paths induced by the branching structure.

4.3. Macro-Level Cognitive Effort

As mentioned above, the cognitive effort required to understand a knowledge base depends not only on the average complexity of its units, but also on global properties as its size and structural organisation. In particular, the number of units and their evaluation order introduce additional cognitive burdens that are not captured at the micro-level. Furthermore, the user's cognitive profile must be considered.

The size of a knowledge base can be trivially obtained through the unit count. Conversely, estimating the mental effort due to knowledge structural organisation requires the following assumptions. In unordered knowledge bases, each unit can be analysed independently, and the cognitive effort per unit remains constant. Conversely, in ordered knowledge bases, units must be evaluated sequentially. As a result, the effort required to understand a unit depends on the number of preceding units, since all prior conditions must be evaluated and implicitly considered in negated form. Consequently, for a knowledge base of size s, the average cognitive cost per unit is expected to increase by a factor of $\sum_{i=1}^{s} \frac{i}{s} = \frac{s+1}{2}$.

The cognitive effort required to understand a knowledge base and its units depends on the perspective and background of users. Therefore, quantitative assessments of cognitive costs are parametrised in the proposed model with user-defined tolerance parameters ($\gamma \in \mathbb{R}^+$) reflecting variations in expertise, experience, and comfort with different aspects of cognitive complexity.

Specifically, $\gamma = 1$ represents the baseline scenario, where the perceived cognitive effort matches exactly the computed cost. No user customisation is introduced; $\gamma > 1$ amplifies the perceived cost. For example, the user may be sensitive to particular structural aspects of the units, as depth or combinatorial complexity. In this case, even a moderately complex unit is perceived as cognitively heavy; $\gamma < 1$ attenuates the perceived cost. This choice corresponds to users who are more tolerant of complex representations, typically aligned with expert users or contexts where high complexity is acceptable.

This strategy enables a flexible and user-centred evaluation of the cognitive effort. Furthermore, users can be profiled according to specific preferences, allowing the system to predict how effort-intensive a knowledge base is perceived from different human operators or classes of operators.

γ parameters are applied as exponential scaling factors for the structural micro- and macro-level assessments. This choice is motivated by two complementary considerations. From a cognitive perspective, human perception of complexity is often non-linear and exhibits sensitivity effects consistent with psychophysical principles such as the Weber-Fechner law, where perceived intensity grows logarithmically with respect to physical stimuli. In this context, exponential scaling provides a simple yet effective mechanism to model amplification or attenuation of perceived effort as a function of user sensitivity. From a modelling perspective, exponential factors ensure a monotonic and smooth modulation of the baseline cognitive cost, preserving the relative ordering of knowledge units while allowing controlled emphasis or de-emphasis of specific complexity components. This property is particularly desirable when comparing alternative representations under different user profiles. Therefore, γ parameters act as flexible, dimensionless modifiers that enable the model to capture subjective variability without altering its underlying structural consistency. γ values close to 1 ensure that these modifiers act as subtle adjustments of the computed costs.

Two user-defined tolerance parameters are introduced to adjust the micro-level assessments, namely: γ_C to modulate the structural complexity of units and γ_B to weigh their combinatorial complexity. Two more parameters are introduced for the macro-level measurements: γ_s, to modulate the impact of the knowledge size, and γ_α, representing the user sensitivity to sequential evaluation.

Individual Cognitive Cost. The structural cost $C(\varphi)$ captures the intrinsic effort required to evaluate atomic components, while the branching factor $B(\varphi)$ models combinatorial expansion. γ_C is applied to the structural cost of a formula, enabling users to privilege or not complex representation such as units based on many conjunctions, m-of-n predicates, or oblique decision boundaries, since they usually come with higher structural costs. Analogously, γ_B is applied to the branching factor in order to model the user comfort in managing multiple alternative evaluation paths, as common in units based on many disjunctions or m-of-n constructs. The weighted product of the micro-level assessments defines the individual cognitive cost $W(\varphi)$ for formulæ representing units of the symbolic knowledge base.

Let K be a knowledge base with size s. The individual cognitive cost of a knowledge unit, represented by a formula φ, is defined as follows:

$$W(\varphi) = C(\varphi)^{\gamma_C} \cdot B(\varphi)^{\gamma_B}. \tag{9}$$

This formulation reflects the intuition that structural and combinatorial aspects contribute jointly to the perceived cognitive effort, and that their relative importance may vary depending on user sensitivity.

The average cognitive cost for knowledge units is computed as follows:

$$\overline{W}(K) = \sum_{\varphi \in K} \frac{W(\varphi)}{s}. \tag{10}$$

Overall Cognitive Effort. To quantify the total effort perceived to understand a knowledge base, γ_s and γ_α are applied to its global properties. γ_s increases or decreases the perceived cognitive effort based on the knowledge size, whereas γ_α reflects the extent to which sequential reasoning is cognitively heavy for human users evaluating the knowledge base. For instance, users favouring compact and parallel representations will set higher γ_s and γ_α values, whereas users accustomed to structured or procedural evaluation may exhibit lower sensitivity to these aspects.

Let K be a knowledge base composed of s units having average cognitive cost $\overline{W}(K)$. Its overall cognitive effort E is defined as follows:

$$E(K) = s^{\gamma_s} \cdot \overline{W}(K) \cdot \alpha^{\gamma_\alpha}, \tag{11}$$

where the sequentiality factor α models the average positional cost per unit induced by sequential evaluation:

$$\alpha = \begin{cases} 1 & \text{if } K \text{ is not ordered,} \\ \dfrac{s+1}{2} & \text{if } K \text{ is ordered.} \end{cases}$$

5. Experimental Evaluation

The proposed framework has been validated in a simple but significant setting, by using the Iris data set. Several knowledge bases have been selected, with identical or similar decision boundaries defined through different structures of varying cognitive complexity (Section 5.1). A non-exhaustive pool of user profiles has been defined, with the aim of identifying a minimal set of clusters of human cognitive preferences (Section 5.2). The cognitive effort model introduced in this work is then applied to analyse the perceived impact of each knowledge base across all user profiles (Section 5.3).

5.1. Illustrative Knowledge Bases

The Iris data set is a well-known and simple task characterised by the classification of flower specimens amongst three distinct classes. Four input features are available to this purpose, i.e., the length and width of flower sepals and petals. It is possible to build infinite classification models for this data set. In this work the focus is on symbolic models making decisions only based on the petal width and length. In particular, six knowledge bases are defined according to different paradigms and structures, such as decision trees, unordered sets, and ordered lists composed of fuzzy, oblique, and m-of-n rules.

KB1. An ordered list of simple rules possibly containing conjunctions or disjunctions. The knowledge units must be evaluated in order and the last unit acts as a default rule. The decision boundaries corresponding to this knowledge base are shown in Appendix A;

KB2. An unordered set of rules yielding the same decision boundaries of KB1. Units may be evaluated independently and no branching is allowed within units. Multiple conditions on the same variable are collapsed into single interval-inclusion constraints. Units U2.2 and U2.3 are partially overlapping;

KB3. An ordered rule list obtained from a decision tree (shown in Appendix A), logically equivalent to (yet cognitively different from) KB1;

KB4. An ordered rule list with oblique rules. Units contain no conjunctions or disjunctions. The corresponding decision boundaries, shown in Appendix A, are not axis-aligned;

KB5. An unordered set of rules of m-of-n predicates, preserving logical equivalence with KB1;

KB6. An unordered rule set with fuzzy predicates, where each fuzzy variable is partitioned into three trapezoidal fuzzy intervals, as depicted in Appendix A. The corresponding decision boundaries are logically equivalent to those of KB1. The notions of 'being low or $medium$' (cf. U6.3) and 'being not $high$' (cf. U6.4) are logically equivalent but exhibit a different cognitive impact.

The adopted knowledge bases are formally listed in Table 1. Computed micro- and macro-level assessments as branching factors, structural costs, knowledge base sizes and sequentiality factors are reported in the same. It can be noticed from the table that branching factors $B(\varphi)$ are equal to 1 if no disjunctions or m-of-n constructs are present in the knowledge base units. Analogously, the sequentiality factors α are set to 1 for unordered knowledge bases. The size s always corresponds to the number of units displayed in the table. Finally, the structural cost computation reveals some

	Knowledge bases (KB) and units (U)	$B(\varphi)$	$C(\varphi)$	s	α
KB1	Ordered rule list with conjunctions and disjunctions (cf. Figure 2a)			3	2
U1.1	IF PetalLength $\leq$ 2.9 AND PetalWidth $\leq$ 0.7 THEN setosa	1	4		
U1.2	ELSE IF PetalLength $>$ 5.0 OR PetalWidth $>$ 1.7 THEN virginica	2	2		
U1.3	ELSE versicolor	1	0		
KB2	Unordered rule set without disjunctions			5	1
U2.1	IF PetalLength $\leq$ 2.9 AND PetalWidth $\leq$ 0.7 THEN setosa	1	4		
U2.2	IF PetalLength $>$ 5.0 THEN virginica	1	2		
U2.3	IF PetalWidth $>$ 1.7 THEN virginica	1	2		
U2.4	IF PetalLength $\in$ [2.9, 4.5] AND PetalWidth $\leq$ 1.7 THEN versicolor	1	5		
U2.5	IF PetalLength $\leq$ 5.0 AND PetalWidth $\in$ [0.7, 1.7] THEN versicolor	1	5		
KB3	Ordered rule list from a decision tree (cf. Figure 3)			5	3
U3.1	IF PetalLength $>$ 5.0 THEN virginica	1	2		
U3.2	ELSE IF PetalWidth $>$ 1.7 THEN virginica	1	2		
U3.3	ELSE IF PetalLength $>$ 2.9 THEN versicolor	1	2		
U3.4	ELSE IF PetalWidth $>$ 0.7 THEN versicolor	1	2		
U3.5	ELSE setosa	1	0		
KB4	Ordered rule list with oblique predicates (cf. Figure 2b)			3	2
U4.1	IF 0.4 $\cdot$ PetalLength + PetalWidth $\leq$ 1.7 THEN setosa	1	6		
U4.2	ELSE IF 1.5 $\cdot$ PetalLength + PetalWidth $>$ 8.9 THEN virginica	1	6		
U4.3	ELSE versicolor	1	0		
KB5	Unordered rule set with m-of-n predicates			3	1
U5.1	IF 2-of-{PetalLength $\leq$ 2.9, PetalWidth $\leq$ 0.7} THEN setosa	1	4		
U5.2	IF 1-of-{PetalLength $>$ 5.0, PetalWidth $>$ 1.7} THEN virginica	2	2		
U5.3	IF PetalLength $\leq$ 5.0 AND PetalWidth $\leq$ 1.7 AND 1-of-{PetalLength $>$ 2.9, PetalWidth $>$ 0.7} THEN versicolor	2	6		
KB6	Unordered rule set with fuzzy predicates (cf. Figure 4)			4	1
U6.1	IF PetalLength IS low AND PetalWidth IS low THEN setosa	1	8		
U6.2	IF PetalLength IS $high$ OR PetalWidth IS $high$ THEN virginica	2	4		
U6.3	IF PetalLength IS med AND PetalWidth $\in \{low, med\}$ THEN versicolor	1	11		
U6.4	IF PetalLength IS NOT $high$ AND PetalWidth IS med THEN versicolor	1	9		

Table 1
Examples of knowledge bases for the Iris data set. Objective micro- and macro-level assessments are reported.

interesting patterns. *(i)* Units associated with default rules come along with no cognitive cost (e.g., U1.3). *(ii)* Units with the same number of atoms having the same structure, but differing in the composition operator (e.g., units U1.1 and U1.2) exhibit the same $B(\varphi) \cdot C(\varphi)$ product. This is the expected behaviour of the framework, since conjunction acts as an accumulator of complexity and disjunction acts as a distributor of complexity. *(iii)* The conciseness of atoms expressing inclusion within numeric intervals corresponds to a slightly lower structural cost than the analogous logic construct expressed as the conjunction of two inequalities (e.g., units of KB2). *(iv)* Oblique constructs increase the structural cost of knowledge units, but have no impact on combinatorial complexity (cf. KB4). *(v)* m-of-n constructs increase both the structural cost and the combinatorial complexity of knowledge units, since represent a concise formalism to express a disjunction of conjunctions (see U5.3). This consideration does not apply to m-of-n atoms degenerated into conjunctions ($m = n$, U5.1) or disjunctions ($m = 1$, U5.2). *(vi)* Knowledge units based on fuzzy atoms are the most demanding in terms of structural cost. This is

Profile name	Parameters γ_B	γ_C	γ_s	γ_α	Cognitive preferences
Neutral	1.0	1.0	1.0	1.0	Baseline profile, no strong preferences
Branching-averse	1.3	0.7	1.3	0.7	Against branching and large-sized knowledge bases
Depth-averse	0.7	1.3	0.7	1.3	Privileging branching and simple, independent units
Domain user	1.3	1.3	0.7	0.7	Non-expert user, favourable to (possibly many) simple units
Domain expert	0.7	0.7	1.3	1.3	Privileging conciseness against redundancy and sequentiality

Table 2
Examples of different user profiles encoding varying cognitive preferences.

ascribable to the cognitive effort to associate linguistic values to fuzzy intervals and to the fact that multiple linguistic labels are more difficult to be merged and simplified than numeric intervals and inequalities (see U6.3 and U6.4).

5.2. Illustrative User Profiles

A set of illustrative user profiles is defined by varying the tolerance parameters associated with different sources of cognitive complexity. Each profile emphasises or attenuates specific dimensions, to enable a controlled analysis of their individual and combined effects.

Neutral. Average user with no particular requirements. All elements of cognitive complexity are perceived to require the same effort to be understood;

Branching-averse. User privileging small-sized knowledge bases where individual units may be structurally complex. Sequentiality is tolerated and preferred over combinatorial complexity;

Depth-averse. User privileging independent, structurally simple units. Combinatorial complexity and knowledge size are tolerated;

Domain user. User preferring micro-level simplicity and accepting macro-level complexity. Simple units with low combinatorial complexity are preferred and large-sized knowledge bases, possibly ordered, are tolerated;

Domain expert. User requesting macro-level simplicity and accepting micro-level complexity. Structural and combinatorial complexities are tolerated. Sequentiality and knowledge size are penalised.

The pool of user profiles is formally detailed in Table 2. γ values equal to 0.7 represent tolerance with respect to micro- and macro-level assessments, whereas values equal to 1.3 correspond to penalties.

5.3. Quantifying the Overall Cognitive Effort

Table 3 summarises the experimental validation of the framework, listing the average cognitive cost (per knowledge unit) and the overall cognitive effort of each knowledeg base as it is perceived by each user profile. The results reported in the table provide several insights on the behaviour of the proposed model and its ability to capture meaningful variations in perceived cognitive effort.

First, the model clearly differentiates knowledge bases that are logically equivalent but structurally distinct. For instance, KB1, KB3, and KB5 induce different cognitive costs across all profiles, despite encoding similar decision boundaries. This confirms that the framework effectively captures the impact of structural organisation and evaluation dynamics beyond mere functional equivalence.

Second, the role of user-dependent parameters emerges as a key factor in shaping the perceived complexity. No single knowledge base is uniformly optimal across all profiles. Instead, each representation is favoured under specific cognitive preferences. For example, KB4 is preferred by branching-averse users, as it eliminates combinatorial structures at the cost of higher structural complexity, while KB5 becomes optimal for domain experts, who tolerate micro-level complexity but penalise redundancy and size. Conversely, KB2 and KB3 are favoured by profiles that prioritise simplicity at the unit level, even when this results in larger or sequentially organised knowledge bases.

Profiles $\rightarrow$	Neutral		Branching-averse		Depth-averse		Domain user		Domain expert	
$K \downarrow$	$\overline{W}(K)$	$E(K)$	$\overline{W}(K)$	$E(K)$	$\overline{W}(K)$	$E(K)$	$\overline{W}(K)$	$E(K)$	$\overline{W}(K)$	$E(K)$
KB1	2.67	**16.00**	2.21	15.00	3.35	17.82	4.04	14.17	1.76	18.07
KB2	3.60	18.00	2.41	19.54	5.44	**16.78**	5.44	16.79	2.41	19.54
KB3	1.60	24.00	1.30	22.72	1.97	25.35	1.97	**13.11**	1.30	43.93
KB4	4.00	24.00	2.34	**14.83**	6.85	36.38	6.85	24.00	2.34	24.00
KB5	6.67	20.00	5.09	21.23	8.92	19.24	12.47	26.91	3.66	**15.26**
KB6	9.00	36.00	5.20	31.52	16.16	42.73	17.46	46.08	4.65	28.17

Table 3
Cognitive effort computed for the knowledge bases appearing in Table 1 for the different user profiles listed in Table 2. Bold values highlight the most understandable knowledge base for each user profile.

Third, results highlight the non-trivial interaction between micro- and macro-level components. Knowledge bases with low average unit cost (e.g., KB3) may still result in high overall effort due to size and sequentiality effects. Similarly, compact representations (e.g., KB4) may be penalised when structural complexity is amplified by user sensitivity. This confirms that neither micro-level nor macro-level factors alone are sufficient to explain perceived interpretability.

Finally, the framework exhibits consistent and interpretable behaviour with respect to the variation of tolerance parameters. Changes in γ_B, γ_C, γ_s, and γ_α produce predictable shifts in the knowledge base ranking, without altering the internal coherence of the model. This property is particularly relevant in practice, as it enables controlled adaptation of interpretability assessments to different user profiles.

Overall, the experimental analysis supports the validity of the proposed model as a flexible and expressive tool for quantifying cognitive complexity. While the experimental setting is intentionally simplified, the results demonstrate that the framework is able to capture both structural and user-dependent aspects of interpretability in a unified and principled manner.

6. Conclusions

This work focuses on quantifying the cognitive complexity of symbolic knowledge-based decision systems and highlights the limitations of existing approaches, generally based on coarse structural proxies and representation-specific metrics. Accordingly, a taxonomy capturing both structural and procedural dimensions of knowledge representations is introduced to support representation-agnostic assessments of cognitive effort.

Building on this taxonomy, a unified model is proposed to decompose cognitive complexity into micro-level components, accounting for the internal structure and combinatorial properties of individual knowledge units, and macro-level components, capturing the global organisation and evaluation dynamics of knowledge bases. A key aspect of the framework is the explicit integration of user-defined tolerance parameters, enabling a flexible and personalised assessment of perceived cognitive effort.

The joint contribution of taxonomy and framework provides a principled and flexible foundation for modelling cognitive complexity in symbolic systems, supporting the design, evaluation, and comparison of interpretable models in a user-centred perspective.

The experimental evaluation, conducted on a set of logically equivalent yet structurally diverse knowledge bases, demonstrates that the proposed model effectively captures meaningful differences in cognitive complexity. Results highlight how structurally distinct representations lead to different cognitive costs, and how these differences are further modulated by user-specific preferences. This confirms that interpretability cannot be reduced to purely structural measures, but must be understood as the result of an interaction between representation properties and user perception.

The proposed framework focuses on a subset of structural and combinatorial factors, currently neglecting other potential relevant aspects. Future work will therefore investigate the integration of these dimensions, as well as the empirical validation of the model through user studies and its application to more complex and heterogeneous real-world knowledge bases.

Declaration on Generative AI

During the preparation of this work, the author used Generative AI tools for grammar and spell checking. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] R. Andrews, J. Diederich, A. B. Tickle, Survey and critique of techniques for extracting rules from trained artificial neural networks, Knowledge-Based Systems 8 (1995) 373–389. doi:10.1016/0950-7051(96)81920-4.

[2] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Computing Surveys 51 (2018) 1–42. doi:10.1145/3236009.

[3] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence 1 (2019) 206–215. doi:10.1038/s42256-019-0048-x.

[4] L. Scorzato, Reliability and interpretability in science and deep learning, Minds and Machines 34 (2024) 27. URL: https://doi.org/10.1007/s11023-024-09682-0. doi:10.1007/S11023-024-09682-0.

[5] J. Huysmans, K. Dejaeger, C. Mues, J. Vanthienen, B. Baesens, An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models, Decision Support Systems 51 (2011) 141–154. doi:10.1016/j.dss.2010.12.003.

[6] P. M. Murphy, M. J. Pazzani, Id2-of-3: Constructive induction of m-of-n concepts for discriminators in decision trees, in: Machine Learning Proceedings 1991, Elsevier, 1991, pp. 183–187.

[7] J. R. Quinlan, Simplifying decision trees, International Journal of Man-Machine Studies 27 (1987) 221–234. doi:10.1016/S0020-7373(87)80053-6.

[8] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kauffmann, San Mateo, CA, USA, 1993. URL: https://dl.acm.org/doi/10.5555/152181.

[9] F. Sabbatini, Four decades of symbolic knowledge extraction from sub-symbolic predictors. a survey, ACM Computing Surveys 58 (2025). URL: https://doi.org/10.1145/3749097. doi:10.1145/3749097.

[10] F. Sabbatini, R. Calegari, Symbolic knowledge-Extraction evaluation metrics: The FiRe score, in: K. Gal, A. Nowé, G. J. Nalepa, R. Fairstein, R. Radulescu (Eds.), ECAI 2023 - 26th European Conference on Artificial Intelligence, September 30 – October 4, 2023, Kraków, Poland – Including 12th Conference on Prestigious Applications of Intelligent Systems (PAIS 2023), volume 372 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2023, pp. 2033–2040. doi:10.3233/FAIA230496.

[11] J. Huysmans, B. Baesens, J. Vanthienen, ITER: An algorithm for predictive regression rule extraction, in: Data Warehousing and Knowledge Discovery (DaWaK 2006), Springer, 2006, pp. 270–279. doi:10.1007/11823728_26.

[12] F. Sabbatini, R. Calegari, Symbolic knowledge extraction from opaque machine learning predictors: GridREx & PEDRO, in: G. Kern-Isberner, G. Lakemeyer, T. Meyer (Eds.), Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning, KR 2022, Haifa, Israel. July 31 - August 5, 2022, 2022. URL: https://proceedings.kr.org/2022/57/. doi:10.24963/kr.2022/57.

[13] F. Sabbatini, R. Calegari, On the evaluation of the symbolic knowledge extracted from black boxes, AI and Ethics 4 (2024) 65–74. doi:https://doi.org/10.1007/s43681-023-00406-1.

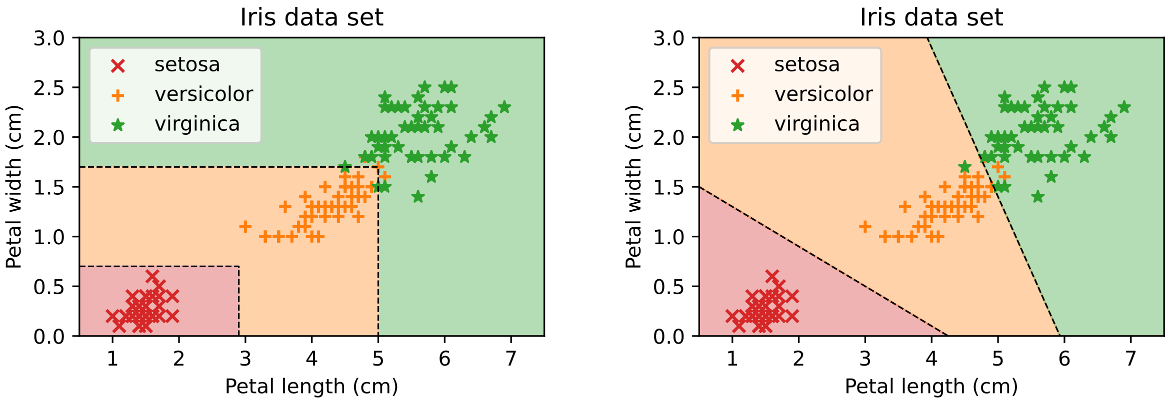
[14] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning, arXiv preprint arXiv:1702.08608 (2017).

[15] I. Lage, A. S. Ross, S. J. Gershman, B. Kim, F. Doshi-Velez, Human-in-the-loop interpretability prior, in: S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems 31: Annual Conference

on Neural Information Processing Systems 2018, NeurIPS 2018, December 3–8, 2018, Montréal, Canada, 2018, pp. 10180–10189. URL: https://proceedings.neurips.cc/paper/2018/hash/0a7d83f084ec258aefd128569dda03d7-Abstract.html.
[16] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence 267 (2019) 1–38. URL: https://doi.org/10.1016/j.artint.2018.07.007. doi:10.1016/j.artint.2018.07.007.
[17] A. A. Freitas, Comprehensible classification models: a position paper, ACM SIGKDD Explorations Newsletter 15 (2014) 1–10. doi:10.1145/2594473.2594475.
[18] R. Setiono, W. K. Leow, FERNN: An algorithm for fast extraction of rules from neural networks, Appl. Intell. 12 (2000) 15–25. URL: https://doi.org/10.1023/A:1008307919726. doi:10.1023/A:1008307919726.
[19] X. Fu, C. Ong, S. Keerthi, G. G. Hung, L. Goh, Extracting the knowledge embedded in support vector machines, in: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), volume 1, 2004, pp. 291–296. doi:10.1109/IJCNN.2004.1379916.
[20] N. H. Barakat, A. P. Bradley, Rule extraction from support vector machines: A sequential covering approach, IEEE Trans. Knowl. Data Eng. 19 (2007) 729–741. URL: https://doi.org/10.1109/TKDE.2007.190610. doi:10.1109/TKDE.2007.190610.
[21] H. Lakkaraju, S. H. Bach, J. Leskovec, Interpretable decision sets: A joint framework for description and prediction, in: B. Krishnapuram, M. Shah, A. J. Smola, C. C. Aggarwal, D. Shen, R. Rastogi (Eds.), Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13–17, 2016, ACM, 2016, pp. 1675–1684. URL: https://doi.org/10.1145/2939672.2939874. doi:10.1145/2939672.2939874.
[22] E. Angelino, N. Larus-Stone, D. Alabi, M. I. Seltzer, C. Rudin, Learning certifiably optimal rule lists for categorical data, Journal of Machine Learning Research 18 (2017) 234:1–234:78. URL: https://jmlr.org/papers/v18/17-716.html.
[23] L. Breiman, J. Friedman, C. J. Stone, R. A. Olshen, Classification and Regression Trees, CRC Press, 1984.
[24] F. Sabbatini, R. Calegari, ICE: An evaluation metric to assess symbolic knowledge quality, in: A. Artale, G. Cortellessa, M. Montali (Eds.), AIxIA 2024 – Advances in Artificial Intelligence, volume 15450 of *Lecture Notes in Computer Science*, Springer, Cham, Switzerland, 2025, pp. 241–256. doi:10.1007/978-3-031-80607-0_19, XXIII International Conference of the Italian Association for Artificial Intelligence, AIxIA 2024, Bolzano, Italy, November 25–28, 2024, Proceedings.
[25] O. Biran, C. Cotton, Explanation and justification in machine learning: A survey, in: IJCAI 2017 Workshop on Explainable AI (XAI), volume 8, 2017, pp. 8–13.

A. Details on the Illustrative Knowledge Bases

Figure 2 shows the decision boundaries for KB1 (Figure 2a) and KB4 (Figure 2b). Figure 3 depicts the KB3 decision tree. Figure 4 represents the fuzzy membership functions adopted for KB6.



(a) Axis-aligned.

(b) Oblique.

Figure 2: Decision boundaries for the Iris data set.

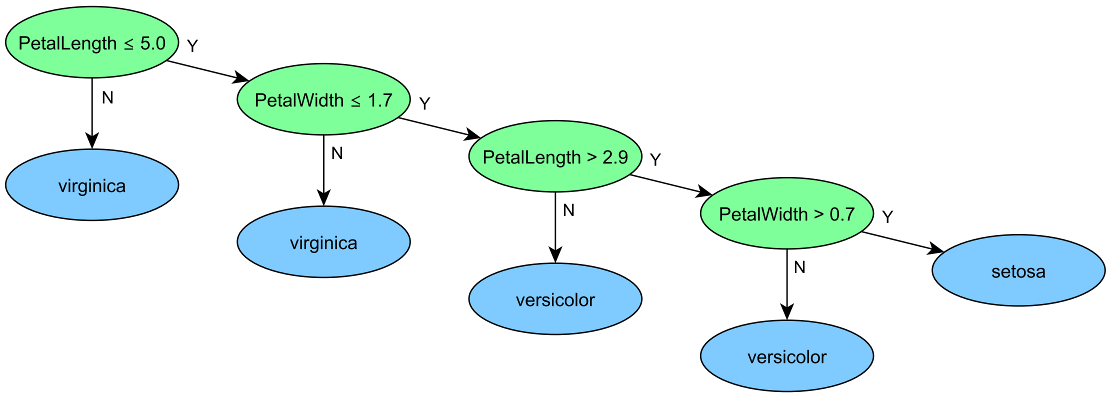


Figure 3: Decision tree providing the same outcomes displayed in Figure 2a.

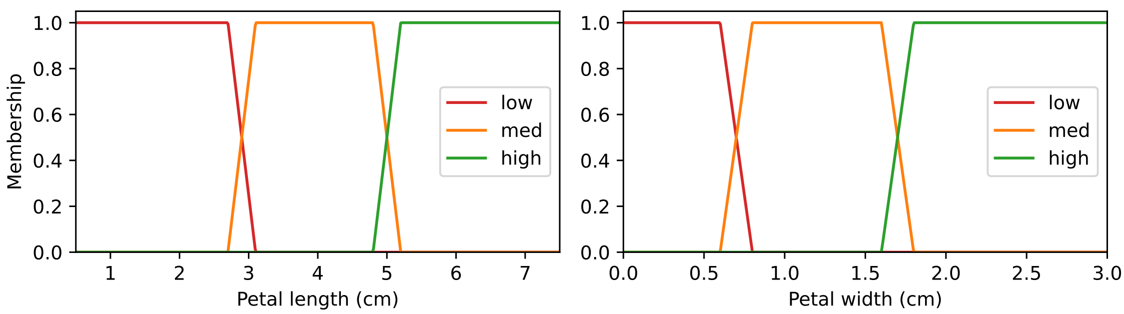


Figure 4: Fuzzy membership functions for petal length and width.