

Article

Not peer-reviewed version

Ethical Decision-Making in Artificial Intelligence: A Logic Programming Approach

[José Machado](#)*, [Regina Sousa](#), [Hugo Peixoto](#), António Abelha

Posted Date: 30 October 2024

doi: 10.20944/preprints202410.2406.v1

Keywords: n/a; Computational Ethics; Logic Programming; Artificial Moral Ethics; Moral Agents




Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Ethical Decision-Making in Artificial Intelligence: A Logic Programming Approach

José Machado , Regina Sousa, Hugo Peixoto and António Abelha

ALGORITMI Centre/LASI, University of Minho, Braga, Portugal; jmac@di.umimnho.pt

Abstract: This paper explores the incorporation of ethical considerations into computational systems, emphasizing the integration of ethical analysis in the design, development, and deployment of technologies such as Artificial Intelligence. Critical topics discussed encompass algorithmic bias, data privacy, autonomous systems, and AI-facilitated ethical decision-making. The paper underscores the necessity of interdisciplinary collaboration and delineates strategies including ethical frameworks, transparency, and stakeholder engagement to ensure that technological advancements correspond with human values. The paper also looks at the concept of Artificial Moral Agents and their role in ethical decision-making in Artificial Intelligence systems. Some specific applications, including healthcare decision support systems and autonomous vehicles, are examined to demonstrate how Artificial Intelligence can reconcile operational efficiency and ethical principles, such as privacy protection and harm reduction. The main contribution of this work is the introduction of Continuous Logic Programming as a method for modelling moral reasoning in AI, providing a transparent and explainable approach to ethical decision-making. The article also addresses the difficulties in evaluating ethical outcomes, especially in situations with incomplete, ambiguous or contradictory data, and highlights the need for continuous supervision and strengthening of ethical frameworks.

Keywords: computational ethics; logic programming; artificial moral ethics; moral agents

1. Introduction

Computational ethics refers to the field of study that integrates ethics with computing technologies, focusing on the ethical implications and responsibilities associated with the design, development, deployment, and use of computational systems [1]. This interdisciplinary area encompasses a broad range of issues, including but not limited to algorithmic bias and fairness; privacy and data protection; autonomous systems and robotics; Artificial Intelligence (AI) and moral decision-making; ethical design and human-computer interaction; and digital accessibility and inclusivity [2]. Computational ethics seeks to proactively address these and other ethical challenges by integrating ethical analysis and considerations into the entire lifecycle of technological development, from initial design to deployment and beyond. It involves collaboration among ethicists, computer scientists, legal scholars, policymakers, and other stakeholders to ensure that technology advances in a way that respects and promotes human values and well-being [3].

Controlling and ensuring the application of ethics in computational systems involves a multi-faceted approach that integrates ethical considerations throughout the design, development, and deployment processes. The several key strategies for embedding ethics into computational systems are ethical guidelines and frameworks; ethical impact assessments; diverse and inclusive design teams; transparency and explainability; public engagement and stakeholder participation; regulation and policy development; professional education and training; continuous monitoring and evaluation; and accountability mechanisms. By incorporating these strategies, organizations and individuals can work towards ensuring that computational systems are developed and used in a manner that aligns with ethical principles and promotes the well-being of all stakeholders involved [4][5].

A particular field of Computational Ethics is Artificial Moral Ethics [6], which aims to ensure that AI systems operate within ethical boundaries, balancing efficiency with moral responsibility, and addressing the potential for unintended consequences of their actions. Some very special fields of

applications are Healthcare (AI making treatment decisions based on the patient's best interests, while respecting privacy and consent); Autonomous Vehicles (making ethical decisions during potential accidents, such as deciding how to minimize harm) or AI in Warfare (following ethical rules of engagement, distinguishing combatants from civilians, and making proportional responses).

Artificial Moral Ethics should deal with Intelligent Agents, with ethical principles that require integrating moral reasoning and ethical decision-making frameworks into their design [7]. This process involves combining technology, philosophy, and cognitive science to ensure these agents can make choices that align with ethical standards while considering the potential impact on human well-being.

Designing agents with ethical principles should involve:

- Define the ethical framework, deciding which ethical framework the agents will follow. The common frameworks include utilitarianism (maximizing overall happiness or reducing harm) [8]; deontology (following moral rules or duties and respecting rights or autonomy) [9]; virtue ethics (emphasizing traits like fairness, honesty and empathy) [10]; and ethical pluralism (combining ethical theories) [11].
- Design ethical decision-making algorithms from moral decision models means developing algorithms that can process ethical rules and principles, creating models that identify stakeholders, evaluate consequences and resolve conflicts.
- Implement ethical constraints involves dealing with hard constraints (strict rules that cannot break) and soft constraints (rules that should be strived to follow but can compromise on where necessary).
- Contextual and situational awareness deals with context sensitivity, sensing and interpreting data as well as understanding legal, cultural and social norms.
- Learning and adapting ethical behavior in reinforcement learning (scenarios are based on feedback (reward or penalty) for ethical or unethical actions), supervised learning (datasets include ethically annotated situations) and human-in-the-loop (allowing humans to intervene or provide feedback).
- Transparency and explainability means designing systems to be able to explain their ethical reasoning process in human-understandable terms and ensuring that the systems can provide clear justifications for their actions.
- Simulating moral dilemmas is the process of subjecting agents to various moral dilemmas in controlled environments, adjusting the ethical decision-making algorithms based on performance to better align with moral outcomes.
- Incorporate ethical review boards are committees that involves interdisciplinary teams to assess the ethical behavior and ensures that the systems adhere to standards, laws and regulations.
- Continuous monitoring and updates are essential for ongoing ethics supervision and self-correction mechanisms.

By combining technical design with philosophical insights and human oversight, we can build AI systems capable of acting ethically in complex, real-world environments.

While extensive research has been conducted on integrating ethics into AI systems, much of the existing work focuses on theoretical models or non-symbolic AI approaches such as machine learning, which often lack transparency and explainability in ethical decision-making. For example, traditional AI systems like neural networks (often referred to as black-box models) offer limited insight into how decisions are made, which poses significant ethical challenges in terms of accountability and trust [12].

On the other hand, symbolic approaches such as rule-based systems and logic programming have been explored for their potential to enhance transparency and provide structured ethical frameworks. However, many of these approaches have yet to address the full complexity of moral reasoning in real-world applications. There is also a notable gap in integrating Continuous Logic Programming (CLP) [13] into practical AI systems for ethical decision-making, particularly in areas like healthcare and autonomous systems, where the stakes are high.

This paper aims to fill that gap by proposing the use of CLP as a formal method for modeling ethical decision-making in AI. By focusing on the transparency, explainability, and traceability of these logic-based systems, the paper introduces a structured approach to embedding moral reasoning into AI. Additionally, the paper presents novel applications of these frameworks in healthcare, where ethical decision-making is critical, providing examples of how CLP can resolve ethical conflicts and improve AI accountability.

2. Literature Review

2.1. Overview Of Existing Work

Research on computational ethics has seen significant growth in recent years, particularly with the increasing role of AI in decision-making processes. Studies have focused on various approaches to embed ethics in AI systems, ranging from non-symbolic methods, such as machine learning, to symbolic approaches like rule-based systems and logic programming.

Non-symbolic approaches, particularly machine learning (ML), have gained widespread attention for their ability to handle large datasets and make predictions. However, ML algorithms often function as black-box systems, where the decision-making process is not transparent or explainable to users. For example, research by [14] highlighted the use of LIME (Local Interpretable Model-agnostic Explanations) to provide some interpretability to ML models, but this approach still faces challenges in offering full transparency, particularly in complex ethical decision-making scenarios.

On the other hand, symbolic approaches such as rule-based systems and logic programming offer a clearer, more explainable framework for embedding ethical reasoning into AI. Russell and Norvig [15] explored the potential of rule-based systems to integrate ethical principles into decision-making processes. However, their work primarily focused on predefined, rigid rule sets, which may not adapt well to the fluid and often ambiguous nature of real-world ethical dilemmas.

In recent years, studies have explored the integration of formal logic into ethical AI. CLP are examples of symbolic methods that offer both transparency and traceability, as they rely on well-defined rules and logical reasoning. Work by [16] has demonstrated how logic programming can be used to simulate moral reasoning and provide justifications for AI decisions. However, much of the research remains theoretical, with limited application to real-world ethical challenges.

Specific application domains, such as healthcare and autonomous systems, have also been the subject of ethical AI research. In healthcare, AI systems have been developed to assist in decision-making processes, such as patient diagnosis or treatment recommendations. Studies like those by [17] emphasized the importance of transparency and patient trust, but there remains a gap in how to integrate continuous moral reasoning into these systems. Similarly, research on autonomous vehicles by [18] explored how AI could make ethical decisions in critical situations, but most solutions were based on simplified moral dilemmas and did not account for the full complexity of ethical frameworks.

2.2. Theoretical Framework

The theoretical framework of this paper draws on several ethical theories that have been foundational in both philosophy and AI research. Three key ethical approaches are relevant to computational ethics: utilitarianism, deontological ethics, virtue ethics and ethical pluralism:

Utilitarianism, as originally proposed by Jeremy Bentham [8] and later expanded by John Stuart Mill [19], advocates for decisions that maximize overall happiness or minimize harm. This approach has been integrated into AI systems where the objective is to calculate and optimize outcomes based on predicted consequences. However, one challenge with utilitarian approaches in AI is that they often fail to capture the complexities of individual rights and justice.

Deontological ethics, based on Immanuel Kant's work [9], focuses on the adherence to moral rules or duties. AI systems that follow deontological frameworks are programmed to prioritize rules

over outcomes. This has been explored in logic programming, as demonstrated by [16], where rules-based systems are designed to follow strict ethical guidelines. The limitation here is that such systems might struggle in scenarios where strict rule-following conflicts with other moral considerations, such as context-sensitive judgment.

Virtue ethics, which stems from Aristotle's philosophy [10], focuses on the development of moral character and emphasizes virtues like fairness, honesty, and empathy. While less commonly implemented in AI, this framework is critical in understanding how AI systems should act in ways that mirror human ethical behavior. Integrating virtue ethics into logic programming remains an open challenge due to the difficulty in codifying abstract traits like empathy into logical rules. In modern applications, virtue ethics is used to encourage individuals to develop good character traits and think about ethical behavior as part of their overall moral growth, rather than as isolated acts of right or wrong.

Ethical pluralism, by combining different ethical theories, such as utilitarianism, deontology, and virtue ethics. The framework, known as principlism, offers a pluralistic approach to ethical decision-making in healthcare by balancing multiple ethical principles like autonomy, beneficence, non-maleficence, and justice [11].

CLP, the core approaches explored in this paper, offer a middle ground between rigid rule-following (as in deontology) and flexible outcome-based approaches (as in utilitarianism). These methods enable AI systems to operate using both strict rules and defeasible rules (i.e., rules that can be overridden in certain contexts), providing a flexible ethical framework that can adapt to complex moral dilemmas.

2.3. Identification of Gaps

Despite the advancements in both non-symbolic and symbolic AI methods, several key gaps remain in the current research:

Lack of Transparency in Non-Symbolic Methods: Machine learning systems, though powerful, often lack the transparency needed for ethical decision-making, especially in high-stakes areas like healthcare and autonomous systems. Current explainability tools, like LIME, are limited in their ability to provide comprehensive insight into the decision-making process. This paper addresses this gap by proposing CLP as more transparent alternatives that offer traceability in ethical reasoning.

Insufficient Application of Symbolic Methods in Real-World Scenarios: While logic programming has been widely discussed as a theoretical framework for ethical AI, there is a notable gap in practical applications. Much of the existing research, such as the work of [16], has focused on hypothetical moral dilemmas or simplified simulations, but few studies have applied these models to real-world systems. This paper bridges this gap by applying CLP to concrete examples in healthcare, showcasing how these methods can be operationalized in practice.

Over-reliance on Rigid Ethical Frameworks: Many AI systems that employ symbolic methods, such as rule-based systems, rely heavily on rigid ethical frameworks that may not adapt well to complex, real-world situations where exceptions or contextual factors need to be considered. This paper addresses this limitation by proposing the use of defeasible rules in CLP, allowing for more context-sensitive decision-making that can better handle ambiguous or conflicting ethical scenarios.

Limited Exploration of Multi-disciplinary Collaboration: While there is widespread acknowledgment of the importance of interdisciplinary collaboration in computational ethics, many studies still adopt a siloed approach, focusing on either the technical or philosophical aspects of the problem. This paper emphasizes the integration of philosophy, computer science, and real-world stakeholder engagement, advocating for a more holistic approach to ethical AI development.

While this paper aims to address key gaps in computational ethics, it acknowledges that full solutions remain elusive due to the complexity of the field. While CLP provide more transparency and flexibility than existing methods, achieving complete transparency, real-world adaptability, and nuanced moral reasoning in AI remains a challenge. Additionally, interdisciplinary collaboration, though emphasized, requires further integration across fields. This paper offers meaningful advancements, but recognizes that continuous effort and research will be necessary to fully address these ethical challenges.

This work touches on moral dilemmas but simplifies the resolution process into rule-based models without delving deeply into the nuances of ethical conflicts. Real-world ethical decisions often involve trade-offs that are not easily captured by rule-based systems, and a deeper analysis of these complexities would be welcomed [20].

The theoretical models proposed, like the use of CLP for ethical decision-making, lack empirical validation or examples of successful implementation. Pilot studies or simulations will be object of future work, in order to strengthen the argument that these approaches are feasible and effective in real-world applications.

We are aware that there are significant difficulties in consistently implementing moral frameworks in AI systems, as real-world ethical situations often require nuanced judgment that can be challenging to codify. Additionally, relying too heavily on automated ethical reasoning poses risks, particularly without sufficient human oversight. While CLP offer structured ethical frameworks, they cannot fully capture the complexity of human moral reasoning. As a result, human involvement remains crucial to ensure accountability and to handle exceptions or moral dilemmas that AI systems may struggle to resolve on their own.

3. Methodology

Strict rules are also referred as Exact Rules, and Soft Rules are Defeasible Rules. Some particular application fields of Artificial Moral Ethics are healthcare and autonomous vehicles. Here are some examples of application in Healthcare and in Autonomous Cars .

In Healthcare:

$$\begin{aligned} s_1: & \rightarrow \text{protect_privacy}(a). \\ r_1: & \text{medical_emergency}(a). \\ r_2: & \text{protect_privacy}(X) \neg \text{disclose_data}(X). \\ s_2: & \text{medical_emergency}(X) \rightarrow \text{disclose_data}(X). \end{aligned}$$

In Autonomous Cars:

$$\begin{aligned} s_1: & \rightarrow \text{traffic_law}(b). \\ r_1: & \text{avoid_accident}(b). \\ r_2: & \text{traffic_law}(X) \text{obey}(X). \\ s_2: & \text{avoid_accident}(X) \rightarrow \neg \text{obey}(X). \end{aligned}$$

In the two examples, s_i are strict rules and r_j are defeasible rules. The two programs are contradictory because they manage to demonstrate at the same time $\text{disclose_data}(a)$ and $\neg \text{disclose_data}(a)$ using $[s_1, r_2]$ or using $[r_1, s_2]$. Given that defeasible rules are used in both cases, there is clearly a conflict that needs to be resolved. The same happens in the second example for autonomous cars, where using defeasible rules it is possible to demonstrate at the same time $\text{obey}(b)$ and $\neg \text{obey}(b)$.

Several questions remain open regarding information and its quality. The same can be applied to evaluating decisions and assessing whether a decision is morally acceptable.

Information: Information has quality, how to select information, how to evaluate the quality of information and overstep the problems with information that is incomplete, ambiguous, contradictory or nebulous?

Knowledge: Is knowledge complete?

Decision: Is decision morally acceptable?

These questions are all related to define how to reduce costs and risks in systems with default reasoning.

They are many attributes or dimensions related to quality, in particular inclusion, objectivity, accessibility, actuality, confidence, precision or validity. The quality of information (QoI) is related with the user point of view and emotions more than with technology. The main challenge is how to attend moral and legal requisites, as well as data quality, information security, access control and privacy. On the other hands, some ethical attributes are among autonomy, beneficence, non-maleficence, dignity, justice and truthfulness. Those attributes must be quantified and its importance can not sub-estimated in the decision making process [21].

Being:

Γ a Program

AS_i and AS_j two different answer sets of Γ

EAS_i and EAS_j the extensions of predicates p in AS_i and AS_j

An action set AS_i is considered morally preferable to another action set AS_j if $AS_i \prec AS_j$, where \prec represents the moral preference relation. This implies that for each predicate p_1 , there exists a predicate p_2 such that $p_1 \prec p_2$, and the intersection $EAS_i \cap EAS_j$ is not empty [22]. By incorporating defeasible rules, AI systems can make decisions that are more ethically aligned, better capturing the complexities and subtleties of real-world scenarios.

Conversely, our approach posits that CLP offers advantages that can address the limitations of other methods, such as those relying on black-box reasoning or case-based reasoning. A key feature of using CLP to model morality lies in its reliance on deontological principles as a trustworthy knowledge source and its focus on expert evaluation. Modeling morality through principles, rules, and exceptions or abducibles is inherently understandable for experts, allowing for traceability through proof trees, with a processing method that is transparent, predictable, and adaptable.

The primary aim of using CLP is to support decision-making architectures that consider the moral context, rather than to replicate moral reasoning itself. In Healthcare, CLP provides the ability to justify moral decisions and address uncertainties in real-time for clinical staff, while clearly presenting the reasoning process underlying the recommendations made by the system.

On the other hand, Explainable AI (XAI) aligns with the broader goals of the XAI field, emphasizing interpretability, user comprehension, and the role of explanations as interfaces that improve AI transparency. Researchers in XAI are deeply focused on bridging the interpretation gap, especially in applications where transparency is paramount for end users, such as in healthcare and finance. XAI techniques are designed to provide insights into how models make decisions, which can be complex with deep learning and other opaque models. Tools like Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) are often used to attribute decisions to specific input features, making models more understandable by highlighting what influences their decisions [23]. The broader XAI research focus on creating models that balance interpretability with performance. There is also a push for methods that inherently integrate explainability, such as rule-based approaches, which allow stakeholders to understand decisions at a high level, supporting more transparent AI systems that align with principles of Responsible AI.

The use of CLP to model ethical decisions has several advantages. Firstly, its clarity and simplicity in defining rules and facts makes the modelling of moral dilemmas transparent. Rules can be explicitly defined and easily adjusted for different ethical systems, such as utilitarianism or deontology, making it easier to adapt to multiple scenarios. Another benefit is that CLP allows moral reasoning to be automated. Instead of relying exclusively on manual programming for each decision, the system can automatically deduce the most ethical action, based on logical rules and contextual data. However, there are also limitations. The main one is that complex moral situations often involve subjective and

emotional factors, which are difficult to quantify or represent logically. Also, while CLP is effective in rule-based systems, it often lacks adaptive learning capabilities, such as in machine learning systems, which could adjust their rules based on new information or historical data following hybrid methods.

Although CLP offers a powerful framework for modelling ethical decisions, questions arise about accountability and transparency. Who should be held responsible for the decisions made by a system programmed in CLP? In addition, programmed rules may not capture the moral complexity of certain situations, raising doubts about whether such systems can make truly ethical decisions without human intervention. Another ethical challenge is the bias in the programmed rules. Modelling decisions can reflect prejudices or unfair assumptions. It is essential that logical rules are reviewed and validated by experts to ensure that the system operates fairly.

Various approaches to Knowledge Representation and Reasoning have been developed using the Logic Programming (LP) framework from Mathematical Logic, especially in the domains of Model Theory [24,25] and Proof Theory [26]. This research follows a Proof Theoretical approach, extending the LP language. An Continuous Logic Program, often referred to simply as a Logic Program, is defined as a finite set of clauses, structured as follows:

$$p_1 \wedge \dots \wedge p_m \wedge \text{not } q_{m+1} \wedge \dots \wedge \text{not } q_{m+n} \rightarrow q; \text{ and}$$

$$?p_1 \wedge \dots \wedge p_m \wedge \text{not } q_{m+1} \wedge \dots \wedge \text{not } q_{m+n}$$

where $?$ is a domain atom representing falsity, and q and each p_i are literals, meaning formulas such as a or $\neg a$, where a is an atom, for $m, n \in \mathbb{N}_0$. In this context, the CLP introduces a different type of negation: strong negation, denoted by the classical negation symbol \neg . In many cases, it is useful to represent a as a literal, provided that a can be proven. In CLP, the expressions a and $\text{not } a$, where a is a literal, are considered extended literals, while a or $\neg a$ are referred to as simple literals. Intuitively, $\text{not } a$ holds true when there is no justification to believe a , whereas $\neg a$ requires a proof for the negation of a .

Additionally, every CLP program is associated with a set of abducibles [27]. Abducibles serve as hypotheses that offer potential solutions or explanations to given queries, typically framed as exceptions to the extension of predicates within the program. To reason about the knowledge base encapsulated in a specific program or theory, based on the aforementioned formalism, let us consider a procedure defined through the extension of a predicate called *demo*, using CLP. This predicate enables reasoning about the knowledge base in a specific domain, structured according to the formalism described earlier. Defeasible rules are represented as abducibles. Given a query, it returns a solution based on a particular set of assumptions.

The use of languages based on CLP can not only preserve the power of first-order logic, but also help to describe forms of incompleteness that occur in databases and overcome the drawbacks mentioned above. Firstly, it has to be verified that CLP is capable of doing at least what first-order logic allows, as well as being able to represent databases with null values and abducibles. Furthermore, the computational and conceptual costs are modest.

The meta-predicate *demo* is defined as a meta theorem-solver designed to handle incomplete information, and it is represented by the signature [28]

$$\text{demo}: T, V \rightarrow \{\text{true}, \text{false}, \text{unknown}\}.$$

In other words, it determines the valuation V of a theorem T based on the truth values *false* (or 0),

$$T \rightarrow \text{demo}(T, \text{true}).$$

$$\neg T \rightarrow \text{demo}(T, \text{false}).$$

$$\text{not } T, \text{not } \neg T \rightarrow \text{demo}(T, \text{unknown}).$$

As a basic example, we can represent the knowledge using CLP by considering the extensions of predicates that describe certain aspects of medical ethical modeling. This can be formally expressed using predicates such as *survival_rate*, *quality_of_life*, *futility_of_treatment*, and *consent*. The formalization is as follows:

survival_rate: Patient x Intervention x Rate

quality_of_life: Patient x Intervention x Measurement

futility_of_treatment: Patient x Intervention x Degree_of_Utility

consent: Patient x Intervention x Value

where the first argument indicates the identification of the patient, the second refers to the intervention being analyzed, and the third represents the value of the predicate's attribute, specifically the survival rate, a quantification of the expected quality of life for the patient, the degree of utility of a procedure, and the particulars of informed consent. For instance, let us assume that the knowledge is represented in terms of the logic program:

Program 1 - The extended logic program for predicate *survival_rate*

```
{
    not survival_rate(X,Y,Z), not_abducible_survival_rate(X,Y,Z) →
    survival_rate(X,Y,Z).

    → survival_rate(Andrew, procedure_B, 0.3).
    → survival_rate(Marie, procedure_C, 0.86).
    → abducible(survival_rate(James, procedure_A, 0.4)).
    → abducible(survival_rate(James, procedure_A, 0.7)).
    → abducible(survival_rate(James, procedure_A, 0.75)). }
```

In Program 1, the first clause signifies the closure of the predicate *survival_rate*. The second and third clauses indicate that the survival rate values for patients named *Andrew* and *Marie* for the procedures *procedure_B* and *procedure_C* are 0.3 and 0.86, respectively. The fourth, fifth, and sixth clauses specify that the value for the patient named *James* undergoing *procedure_A* is unknown, but the possible values belong to the set {0.4, 0.7, 0.75}.

Program 2 - program for predicate *quality_of_life*

```
{
    not quality_of_life(X,Y,Z), not_abducible(quality_of_life((X,Y,Z)) →
    quality_of_life(X,Y,Z).

    quality_of_life(X,Y,φ) → abducible(quality_of_life(X,Y,Z)).
    → quality_of_life(Andrew, procedure_B, φ).
    → quality_of_life(Marie, procedure_A, 0.25).
    → quality_of_life(James, procedure_C, 0.43).
}
```

In Program 2, the first clause represents the closure of the predicate *quality_of_life*. In the second and third clauses, the symbol ϕ denotes a null value, indicating that the variable *Z* can take on any value within its domain. The fourth and fifth clauses specify that the *quality_of_life* values for the

patients named *Marie* and *James* undergoing the procedures *procedure_A* and *procedure_C* are 0.25 and 0.43.

Program 3 - The extended logic program for predicate *futility_of_treatment*

```
{
    not      futility_of_treatment(X,Y,Z),not      abduciblefutility_of_treatment(X,Y,Z).      →
    futility_of_treatment(X,Y,Z).

    → futility_of_treatment(Andrew,procedure_A,0.6).
    → futility_of_treatment(Marie,procedure_B,0.4).
    → abducible(futility_of_treatment(James,procedure_C,0.7)).
    → abducible(futility_of_treatment(James,procedure_A,0.6)).
}
```

In Program 3, the first clause indicates the closure of the predicate *futility_of_treatment*. The second and third clauses specify that the values of *futility_of_treatment* for the patients named *Andrew* and *Marie*, and for the procedures *procedure_A* and *procedure_B*, are 0.6 and 0.4, respectively. The value for the patient named *James* undergoing *procedure_C* or *procedure_A* is unknown, but the possible values belong to the set {0.7, 0.6} or possibly both.

Program 4 - The extended logic program for predicate *consent*

```
{

    not consent(X,Y,Z), not abducibleconsent(X,Y,Z). → ¬consent(X,Y,Z).

    → consent(Andrew,procedure_A,0.55).
    → consent(Marie,procedure_C,0.66).
    → abducible(consent(James,procedure_B,0.33)).
    → abducible(consent(James,procedure_B,0.42)).
    ?((abducible(consent(X1,Y1,Z1))∨
    abducible(consent(X2,Y2,Z2)))∧
    ¬(abducible(consent(X1,Y1,Z1))∧
    abducible(consent(X2,Y2,Z2))
    ))
}
```

In Program 4, the fourth and fifth clauses indicate that the consent values for the patient named *James* undergoing the procedure *procedure_B* can be either 0.33 or 0.42, or possibly both. However, the sixth clause introduces an invariant that implements the XOR operator, meaning it asserts that the consent value can be either 0.33 or 0.42, but not both simultaneously.

It is now feasible to generate all possible scenarios to represent the universe of discourse based on the information provided in Logic Programs 1, 2, 3, and 4.

The objective of the *QoI* [29] based approach to program assessment is to develop a quantification process derived from logic programs or theories created through an evolutionary process aimed at addressing problems in environments with incomplete information. The *QoI* associated with a generic predicate *p* can be analyzed and measured as a truth value within the interval [0-1], where 0 represents the truth value false and 1 corresponds to the truth value *true*. In cases where the value is unknown,

the QoI is defined as:

$$QoI_p = \lim_{N \rightarrow \infty} \frac{1}{N} = 0 \quad (N \gg 0)$$

In situations where the information is unknown but can be inferred from a set of values, the QoI is expressed as $QoI = \frac{1}{Card}$, where $Card$ denotes the cardinality of the set of abducibles for p , assuming the set of abducibles is disjoint. If the set of abducibles is not disjoint, the QoI is defined as:

$$QoI_p = \frac{1}{C_1^{Card} + \dots + C_{Card}^{Card}}$$

where C_{Card}^{Card} is a card-combination subset, with $Card$ elements.

The next aspect of the model to consider is the relative importance assigned to each attribute under observation by a predicate, denoted as w_{ij} , which represents the relevance of attribute j for predicate i . It is also assumed that the weights of all predicates are normalized, meaning that:

$$\sum_{j=1}^n w_{ij} = 1, \exists i$$

It is now possible to define a scoring function $V_i(x)$ that, given a value $x = (x_1, \dots, x_n)$ in a multidimensional space, represents all the domains of the attributes. This scoring function is expressed in the following form [22]:

$$V_i(x) = \sum_{1 \leq j \leq n} w_{ij} \times V_{ij}(x_j)$$

It quantifies the QoI resulting from invoking a logic program to prove a theorem; specifically, this is achieved by placing the $V_i(x)$ values into a multidimensional space and projecting them onto a two-dimensional plane. Using this procedure, it is defined a circle. Here, the dashed n-slices of the circle (in this example built on the extensions of four predicates (named as survival_rate, quality_of_file, futility_of_treatment, consent) denote the QoI that is associated with each of the predicate extensions that make the logic program, being respectively 0.1925, 0.165, 0.1925 and 0.2075. For the particular case of Andrew, Marie and James, the QoI is respectively 0.25, 0.33 and 0.18. The more complete predicate is consent and the more complete general information is about Marie.

The selection of the optimal theory or logic program will be determined by a relational order applied to the QoI values of all possible scenarios. In practical terms, at the conclusion of this process, we will obtain a set of theories (or scenarios) that represent the best models of the universe of discourse, which will provide the data to be input into a possible connectionist module.

The extensions of the predicates that constitute the universe of discourse must be generated and incorporated into the CLP program construction to establish a foundation for decision-making. This process is bi-directional, as it must address not only organizational, functional, technical, and scientific requirements but also ethical and legal considerations, along with data quality, information security, access control, and privacy. In healthcare, this generation is derived from Electronic Health Records (EHR) [30]. EHR serves as a core application that horizontally integrates across healthcare units, facilitating a comprehensive analysis of medical records across various services, units, or treated conditions. It brings computational models, technologies, and tools to healthcare settings, based on data, agents, multi-agent systems, and ambient intelligence.

An EHR consists of standardized, ordered, and concise documents aimed at recording actions and medical procedures; it is a compilation of information gathered by physicians and other healthcare professionals, encompassing all relevant patient health data and a follow-up on risk values and clinical profiles. The primary objective is to enhance data processing while reducing time and costs, resulting in more effective and quicker patient assistance, thereby improving overall quality.

Any conceptualization of an information society related to healthcare is built on three fundamental components: raw medical data, reconstructed medical data, and derived medical data [22]. Clinical research and practice involve a systematic data collection process to organize knowledge about patients, their health status, and the reasons for their healthcare admission. Concurrently, the data must be recorded in an organized manner to facilitate effective automation and support through Information Technologies. For instance, patient data collected from an information repository should be registered in an efficient, consistent, clear, and manageable manner to enhance understanding of diseases and therapies. The medical processes for data registration are complemented by the information exchange among different physicians involved in the patient's care, ensuring that clinical data recording is preserved within the EHR application and procedural framework. Interoperability further enables the sharing of information across multiple information systems.

The data collection process originates from a clinical recording format that includes a problem list, a database containing the patient history along with physical examination and clinical findings, diagnostic, therapeutic, and educational plans, and daily SOAP (Subjective, Objective, Assessment, and Plan) progress notes[31]. The problem list acts as an index for the reader, with each problem tracked until resolution. This system significantly influences note-taking by recognizing the four distinct phases of the decision-making process: data collection, problem formulation, development of a management plan, and evaluation of the situation with necessary plan revisions.

The following two cases, drawn from [28], illustrate ethical dilemmas in medical decision-making, in the Intensive Care Unit (ICU) [32]:

Case 1: Mr. PD, an 81-year-old man with a history of cardiopathy and diabetes, is admitted to ICU with Acute Respiratory Distress Syndrome (ARDS). Despite advancements, his chances of survival are low, and his quality-adjusted life expectancy post-ARDS is expected to be poor. During a medical meeting, the assistant physician asks whether ICU resources should continue to be used on Mr. PD, given the survival rates, treatment costs, and expected quality of life.

Case 2: Mrs. GB, a 36-year-old woman, is hospitalized after a car accident and diagnosed with sepsis, Acute Lung Injury (ALI), and a Glasgow coma scale of 3. She requires ICU care, but there are limited beds, meaning Mr. PD would need to be transferred. While moving Mr. PD poses risks due to his fragile state, Mrs. GB's younger age and better prognosis suggest a higher likelihood of recovery with better quality of life.

The assistant physician must decide how to allocate the ICU resources between the two patients.

These cases highlight the complexities of ethical decision-making in healthcare, where resource allocation, survival probabilities, and quality of life must be carefully weighed.

The continuous logic program for agent survival-rate:

```
{ (not survival-rate(X, Y) and not abducible(survival-rate(X, Y)) →
¬ survival-rate(X, Y)),
(survival-rate(X, unknown-survival-rate) → abducible(survival-rate(X, Y))),
(ards(X) and pao2(X, low) and evaluate(X,Y) → survival-rate(X, Y)),
(abducible(survival-rate(gb, 0.5))),
?((abducible(survival-rate(X,Y)) or abducible(survival-rate(X,Z))) and
¬ (abducible(survival-rate(X,Y)) and
abducible(survival-rate(X, Z)))
/This invariant states that the exceptions to the predicate survival-rate follow an exclusive or/
}agsurvival-rate
```

The continuous logic program for predicate survival-quality: { (not survival-quality(X, Y) and not exception(survival-quality(X, Y)) →
¬survival-quality(X, Y)),

```

(survival-rate(X, unknown-survival-quality) → abducible(survival-quality(X, Y))),
(survival-quality(gb, 0.8)),
(abducible(survival-quality(pd, 0.1))),
(?((exception(survival-quality(X,Y)) or exception(survival-quality(X,Z)))
and ¬(exception(survival-quality(X,Y)) and exception(survival-quality(X, Z)))))
}agsurvival-quality The continuous logic program for predicate cost:

```

```

{ (not cost(X, Y) and not abducible(cost(X, Y)) →
¬ cost(X,Y)),
(abducible(cost(X, Y)) ← cost(X, unknown-cost)),
(cost(gb, unknown-cost)),
(cost(pd, unknown-cost)),
(?((exception(cost(X,Y)) or exception(cost(X,Z))) and
¬(exception(cost(X,Y)) and exception(cost(X, Z)))))
}agcost

```

4. Ethical Challenges In Logic Programming

While CLP provides a comprehensive framework for modelling ethical decision-making, it has specific limitations that require further analysis, especially with regard to bias and human oversight.

The ethical guidelines incorporated into AI systems that utilise CLP are impartial, as are the individuals who create them. Rule-based systems can unintentionally incorporate the prejudices, cultural norms or implicit biases of their creators. For example, prioritising specific ethical principles, such as utilitarianism or deontology, can set aside other legitimate ethical viewpoints. In addition, the data used to formulate these regulations can be biased, especially if it is incomplete or indicates historical disparities. This can result in unethical consequences, especially in critical fields such as healthcare or criminal justice. It is therefore imperative to establish rigorous analysis procedures to identify and mitigate potential biases before implementing CLP-based systems.

Although CLP facilitates rules-based decision-making, numerous real-world ethical dilemmas encompass context-dependent and complex elements that may not be easily encapsulated by rigid regulations. Moral decisions often require emotional intelligence, empathy and awareness of cultural nuances - factors that are difficult to reproduce through logic alone. Ethical frameworks such as virtue ethics and care ethics, which prioritise character traits and relational obligations, can be ignored in favour of more rule-based methodologies such as utilitarianism or deontology. This limits the system's ability to cover the full range of ethical considerations.

Although CLP can automate specific aspects of moral reasoning, it must not completely override human judgement. Human supervision is essential to ensure that the AI system's decisions are in line with social values and legal norms. Humans have enhanced abilities to manage exceptions, interpret ambiguous information and make ethical decisions in different or unexpected situations. Human-in-the-loop models, where AI systems make recommendations while humans retain final decision-making authority, are crucial to ensuring accountability and ethical integrity in AI systems using CLP.

Another challenge concerns the transparency of the decision-making process. Although CLP offers a certain degree of explainability through explicit symbolic and logical rules, it can be difficult for non-experts to understand the application of specific ethical rules in complex scenarios. Ensuring that CLP-based systems are explainable and checkable is essential to preserving trust, as transparency allows stakeholders to understand the logic behind AI decisions. In the absence of transparency, it becomes difficult to hold systems accountable for their actions, especially when decisions lead to adverse outcomes.

However, it is crucial to recognise that CLP systems are usually more interpretable than non-symbolic or non-transparent systems, such as those based on deep learning or alternative machine learning algorithms. Black box models generally lack transparency, complicating the interpretation of how input data results in a particular outcome or decision. In contrast, the fact that CLP relies on clearly structured rules makes it easier to follow the logical progression of the decision-making process. This confers an advantage in high-risk contexts, where accountability and explicit reasoning are essential for ethical decision-making. However, despite CLP, ongoing efforts are essential to ensure that the explanations offered are understandable and meaningful to a diverse audience, including non-specialist users and stakeholders.

Unlike machine learning models that can evolve with new data, rule-based systems such as CLP have limited flexibility. Once programmed, CLP systems generally adhere to established rules and can only progress if they are manually revised. This static feature can restrict the system's ability to solve emerging ethical dilemmas or evolving social norms. Consequently, integrating CLP with more adaptive machine learning methodologies can mitigate this limitation, allowing AI systems to learn from historical data while maintaining compliance with a structured ethical framework.

5. Conclusions

The effective integration of health data through EHRs increases operational efficiency, but raises important ethical issues that need to be addressed. Collecting, storing and sharing sensitive medical data requires careful consideration of the privacy and consent implications. As healthcare professionals access and utilise these sources, transparency becomes a crucial process. Patients must be informed about how their data is being used and have the option to consent to or refuse the use of their personal information.

Although the adoption of EHRs brings many benefits, implementation is not without its challenges. Obstacles include resistance to change on the part of some professionals, concerns about the user-friendliness of the systems and the need for adequate training to ensure that healthcare professionals can use the tools effectively. In addition, data security is a key concern, as health data breaches can have serious consequences for patients and healthcare organisations.

AI is a powerful tool in analysing healthcare data, enabling professionals to identify patterns, predict outcomes and make more informed decisions. AI-based systems can analyse large amounts of data from electronic health records to provide information that supports personalised diagnoses and treatments. However, the use of AI also requires a careful analysis of ethical aspects, such as algorithmic bias and the need for human supervision.

To summarise, the evolution of EHRs and the application of artificial technologies have a lot of potential to transform healthcare delivery. However, it is essential that this progress is accompanied by a solid ethical framework that protects patient privacy and guarantees data quality. The intersection between data, ethics and technology is a dynamic area that requires continuous reflection and adaptation from those interested and involved in healthcare provision.

6. Acknowledgements

This work has been supported by FCT-Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

References

1. Bostrom, N.; Yudkowsky, E. The ethics of artificial intelligence. In *The Cambridge Handbook of Artificial Intelligence*; Frankish, K.; Ramsey, W.M., Eds.; Cambridge University Press, 2014; pp. 316–334.
2. Floridi, L.; Cowls, J. A unified framework of five principles for AI in society. *Harvard Data Science Review* **2019**. doi:10.1162/99608f92.8cd550d1.
3. Bryson, J.J.; Diamantis, M.E.; Grant, T.D. Of, for, and by the people: The legal lacuna of synthetic persons. *Artificial Intelligence and Law* **2017**, *25*, 273–291.

4. Asaro, P.M. What should we want from a robot ethic? *International Review of Information Ethics* **2006**, 6, 9–16.
5. Van de Poel, I. Embedding values in artificial intelligence (AI) systems. *Minds and Machines* **2020**, 30, 385–409.
6. Anderson, M.; Anderson, S.L. *Machine Ethics*; Cambridge University Press, 2011.
7. Cavalcante, J.V.; Pereira, L.M. Cognitive agents for machine ethics. Proceedings of the 18th Brazilian Symposium on Artificial Intelligence, 2019, pp. 345–354.
8. Bentham, J. *An Introduction to the Principles of Morals and Legislation*; Clarendon Press, Oxford, 1789.
9. Kant, I. *Groundwork of the Metaphysics of Morals*, revised edition ed.; Cambridge University Press: Cambridge, 1785. First published in 1785.
10. Aristotle. *Nicomachean Ethics*; Cambridge University Press, 2009. Original work published ca. 350 BCE.
11. Beauchamp, T.L.; Childress, J.F. *Principles of Biomedical Ethics*, 7th ed.; Oxford University Press: New York, NY, 2012.
12. Winfield, A.F.; Michael, K.; Pitt, J.; Evers, V. Machine ethics: The design and governance of ethical AI and autonomous systems. *Proceedings of the IEEE* **2019**, 107, 509–517.
13. Neves, J.; Martins, M.R.; Vilhena, J.; Neves, J.; Gomes, S.; Abelha, A.; Machado, J.; Vicente, H. A Soft Computing Approach to Kidney Diseases Evaluation. *Journal of Medical Systems* **2015**, 39.
14. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.
15. Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 3rd ed.; Prentice Hall, 2010.
16. Pereira, L.M.; Saptawijaya, A. *Programming Machine Ethics*; Springer, 2016.
17. Topol, E.J. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*; Basic Books, 2019.
18. Bonnefon, J.F.; Shariff, A.; Rahwan, I. The social dilemma of autonomous vehicles. *Nature* **2016**, 536, 425–427.
19. Mill, J.S. *Utilitarianism*; Parker, Son, and Bourn, London, 1863.
20. Wallach, W.; Allen, C. *Moral Machines: Teaching Robots Right from Wrong*; Oxford University Press, 2009.
21. Kakas, A.C.; Moraitis, P. Argumentation based decision making for autonomous agents **2003**. pp. 883–890.
22. Miranda, M.; Machado, J.; Abelha, A.; Pontes, G.; Neves, J. A step towards medical ethics modeling. *IFIP Advances in Information and Communication Technology* **2010**, 335, 27 – 36.
23. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bannetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; Chatila, R.; Herrera, F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* **2020**, 58, 82 – 115. doi:10.1016/j.inffus.2019.12.012.
24. Kakas, A.C.; Toni, F. Computing argumentation in logic programming. *Journal of Logic and Computation* **1998**, 9, 515–562.
25. Pereira, L.M.; Anh, H.P. Agent morality via counterfactuals in logic programming. *Journal of Applied Logic* **2009**, 7, 523–534.
26. Neves, J. A logic interpreter to handle time and negation in logic data bases. Proceedings of the 1984 ACM Annual Conference on Computer Science: The fifth generation challenge, San Francisco, CA, USA, October 1984; Muller, R.L.; Pottmyer, J.J., Eds. ACM, 1984, pp. 50–54.
27. Kakas, A.C.; Sadri, F., Eds. *Computational Logic: Logic Programming and Beyond: Essays in Honour of Robert A. Kowalski, Part I*; Vol. 2407, *Lecture Notes in Computer Science*, Springer: Berlin, Heidelberg, 2002.
28. Machado, J.; Miranda, M.; Pontes, G.; Abelha, A.; Neves, J. Morality in Group Decision Support Systems in Medicine. Intelligent Distributed Computing IV - Proceedings of the 4th International Symposium on Intelligent Distributed Computing - IDC 2010, Tangier, Morocco, September 2010; Essaaidi, M.; Malgeri, M.; Badica, C., Eds., 2010, Vol. 315, pp. 191–200.
29. Neves, J.; Machado, J.; Analide, C.; Abelha, A.; Brito, L. The Halt Condition in Genetic Programming. Progress in Artificial Intelligence, 13th Portuguese Conference on Artificial Intelligence, EPIA 2007, Guimarães, Portugal, December 3-7, 2007, Proceedings; Neves, J.; Santos, M.F.; Machado, J., Eds. Springer, 2007, Vol. 4874, *Lecture Notes in Computer Science*, pp. 160–169.
30. Oliveira, D.; Ferreira, D.; Abreu, N.; Leuschner, P.; Abelha, A.; Machado, J. Prediction of COVID-19 diagnosis based on openEHR artefacts. *Scientific Reports* **2022**, 12.
31. Bickley, L.S.; Szilagy, P.G. *Bates' Guide to Physical Examination and History Taking*, 11th ed.; Lippincott Williams & Wilkins, 2012. SOAP format is discussed extensively as part of clinical history documentation.

32. Portela, F.; Cabral, A.; Abelha, A.; Salazar, M.; Quintas, C.; Machado, J.; Neves, J.; Santos, M.F. Knowledge acquisition process for intelligent decision support in critical health care. *Information Systems and Technologies for Enhancing Health and Social Care* **2013**, p. 55 – 68.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.