Four Freedoms for Deontic Logic: A Framework for Scalable AI Ethics

Ali Farjami
University of Luxembourg
Luxembourg
ali.farjami@uni.lu

The Four Freedoms for Deontic Logic framework is based on input/output logic, introduced by David Makinson and Leon van der Torre [16]. While they did not mathematically define the framework, they provided illustrative examples. Unlike frameworks based on metaphysical assumptions, input/output logic focuses on normative patterns, making it adaptable to various contexts [15]. I have compared the input/output logic framework to Richard Stallman's principles of free software. The first two freedoms relate to openness and transparency, while the last two emphasize distribution and adaptability for normative systems. This framework emphasizes:

- **Freedom to Choose the Logical Base of Normative Systems:** The choice of a base logic plays a crucial role in building normative systems, as different logical frameworks offer distinct advantages depending on the context and application. For example, the use of classical logic in Mimamsa deontic logic [14] aligns with the acceptance of reductio ad absurdum for resolving contradictions, while the intuitionistic logic in Talmudic deontic logic [1] emphasizes conflict resolution through external mechanisms. These examples illustrate how selecting an appropriate base logic allows normative systems to reflect the underlying principles and reasoning structures of the domains they model, highlighting the need for flexibility in foundational choices to ensure adaptability across diverse applications. The input/output logics investigated in the literature are built on top of classical propositional logic [16] and intuitionist propositional logic [18]. It has been shown that the input/output version of any abstract logic can be constructed [12]. We further developed the abstract algebraic logic approach to input/output logic, where the family of selfextensional logics was proposed as a general background environment for input/output logics [11]. We introduced the generalizations of several types of permission (negative, dual negative, static, dynamic), as well as their interactions with normative systems, to various families of selfextensional logics, thereby proposing a systematic approach to the definition of normative and permission systems on nonclassical propositional bases [11].
- **Freedom to Characterize Normative Systems:** In modal logic, there is some degree of freedom for characterizing normative systems through the addition or removal of axioms. However, input/output logic provides even greater flexibility, as it allows the modification of fundamental inference rules to better suit specific normative applications. This freedom is essential for tailoring normative systems to specific applications. For instance, in discursive reasoning, modifying input/output logic by omitting the conjunction rule (AND) allows for more nuanced discourse analysis [12]. Similarly, Alexander Bochman's adaptation of input/output logic (by adding the bottom axiom) for causal reasoning demonstrates how altering logical structures can effectively model causality [7]. However, transitioning between these customized logical systems can be complex. An algebraic characterization offers a granular approach to designing and understanding normative systems, facilitating smoother transitions and more precise modeling across various logical

Submitted to:
FEAR 2025

frameworks. We further developed the algebraic approach to input/output logic initiated in [8], where subordination algebras and a family of their generalizations were proposed as a semantic framework for various input/output logics. In particular, we explored precontact algebras as a suitable algebraic environment for modeling negative permission and characterized the properties of several types of permission (negative, static, dynamic), as well as their interactions with normative systems, using appropriate modal languages to encode outputs [10].

- **Freedom to Combine Normative Systems:** Compositionality is a crucial principle in logic and normative systems, as it ensures that complex structures can be understood and constructed by combining simpler components in a systematic way. Normative systems are complex, and designing normative applications requires combining different normative components with distinct properties. For instance, various types of permissions are defined based on how permissive norms interact with obligatory norms [17]. Furthermore, constitutive norms and regulative norms must be combined to represent social phenomena effectively [19]. Thus, finding a methodological approach to characterize the possibilities for combining normative systems provides significant flexibility in their design. In particular, we have studied a set of first-order formulas, known as Kracht formulas [9], which offer a framework for integrating obligation, permission, and prohibition systems for diverse applications. We characterized the syntactic shape of first-order conditions on algebras endowed with subordination, precontact, and dual precontact relations, ensuring that these conditions correspond to axioms in the aforementioned modal language. Additionally, we introduced algorithms for computing the first-order correspondents of modal axioms on such algebras and, conversely, for computing the modal axioms whose first-order correspondents satisfy the specified syntactic shape [9].

- **Freedom to Implement and Adapt Normative Systems:** The LogiKEy methodology offers a powerful framework for formalizing and automating complex normative reasoning by embedding deontic logics into classical higher-order logic (HOL), making it compatible with off-the-shelf theorem provers such as Isabelle/HOL [6]. While effective for several established deontic logics, including Åqvist's dyadic logic [5] and the more intricate logic by Carmo and Jones [4], LogiKEy faces challenges when applied to input/output (I/O) logic due to its operational semantics, which do not align well with shallow semantic embeddings. Existing translations of I/O logic into modal logic [3] capture only part of the picture. To address this, two complementary methods have been proposed for mathematically formulating I/O logic. The first builds on its operational semantics by introducing algebraic operators to define sound and faithful embeddings into HOL [13], with ongoing efforts exploring relational semantics to address computational complexity. The second starts from the proof system of I/O logic—using derivation rules such as AND, OR, and CT—and aligns it with subordination algebras [8], enabling straightforward implementation of modal algebras for obligations and permissions. The resulting soundness results demonstrate computational efficiency comparable to LogiKEy benchmarks [2].[1] By systematically connecting I/O logic with modal algebras, this approach enhances the scalability and expressivity of the LogiKEy framework for a wider array of logical formalisms with greater computational efficiency for normative reasoning.

[1] See the GitHub link (direct implementation): https://github.com/farjami110/AlgebriacInputOutput

References

[1] Michael Abraham, Dov M Gabbay & Uri Schild (2011): *Obligations and prohibitions in Talmudic deontic logic*. Artificial Intelligence and Law 19, pp. 117–148.

[2] Christoph Benzmüller, Ali Farjami, David Fuenmayor, Paul Meder, Xavier Parent, Alexander Steen, Leendert van der Torre & Valeria Zahoransky (2020): *LogiKEy Workbench: Deontic Logics, Logic Combinations and Expressive Ethical and Legal Reasoning (Isabelle/HOL Dataset)*. Data in Brief 33, p. 106409.

[3] Christoph Benzmüller, Ali Farjami, Paul Meder & Xavier Parent (2019): *I/O Logic in HOL*. Journal of Applied Logics – IfCoLoG Journal of Logics and their Applications (Special Issue on Reasoning for Legal AI) 6(5), pp. 715–732.

[4] Christoph Benzmüller, Ali Farjami & Xavier Parent (2018): *A Dyadic Deontic Logic in HOL*. In Jan Broersen, Cleo Condoravdi, Shyam Nair & Gabriella Pigozzi, editors: Deontic Logic and Normative Systems — 14th International Conference, DEON 2018, Utrecht, The Netherlands, 3-6 July, 2018, College Publications, pp. 33–50. John-Jules Meyer Best Paper Award.

[5] Christoph Benzmüller, Ali Farjami & Xavier Parent (2019): *Åqvist's Dyadic Deontic Logic E in HOL*. Journal of Applied Logics – IfCoLoG Journal of Logics and their Applications (Special Issue on Reasoning for Legal AI) 6(5), pp. 733–755.

[6] Christoph Benzmüller, Xavier Parent & Leendert van der Torre (2020): *Designing Normative Theories for Ethical and Legal Reasoning: LogiKEy Framework, Methodology, and Tool Support*. Artificial Intelligence 237, p. 103348.

[7] Alexander Bochman (2021): *A logical theory of causality*. MIT Press.

[8] Andrea De Domenico, Ali Farjami, Krishna Manoorkar, Alessandra Palmigiano, Mattia Panettiere & Xiaolong Wang (2022): *Subordination Algebras as Semantic Environment of Input/Output Logic*. In: International Workshop on Logic, Language, Information, and Computation, Springer, pp. 326–343.

[9] Andrea De Domenico, Ali Farjami, Krishna Manoorkar, Alessandra Palmigiano, Mattia Panettiere & Xiaolong Wang (2024): *Correspondence and Inverse Correspondence for Input/Output Logic and Region-Based Theories of Space*. arXiv preprint arXiv:2412.01722.

[10] Andrea De Domenico, Ali Farjami, Krishna Manoorkar, Alessandra Palmigiano, Mattia Panettiere & Xiaolong Wang (2024): *Obligations and permissions, algebraically*. arXiv preprint arXiv:2403.03148.

[11] Andrea De Domenico, Ali Farjami, Krishna Manoorkar, Alessandra Palmigiano, Mattia Panettiere & Xiaolong Wang (2024): *Obligations and permissions on selfextensional logics*. arXiv preprint arXiv:2402.07336. Accepted for Synthese.

[12] Ali Farjami (2020): *Discursive Input/Output Logic: Deontic Modals, and Computation*. Ph.D. thesis, University of Luxembourg, Luxembourg.

[13] Ali Farjami (2021): *New Algebraic Normative Theories for Ethical and Legal Reasoning in the LogiKEy Framework*. arXiv preprint arXiv:2107.11838.

[14] Björn Lellmann, Francesca Gulisano & Agata Ciabattoni (2021): *Mīmāṃsā deontic reasoning using specificity: a proof theoretic approach*. Artificial Intelligence and Law 29(3), pp. 351–394.

[15] David Makinson (1999): *On a fundamental problem of deontic logic*. Norms, Logics and Information Systems. New Studies on Deontic Logic and Computer Science, pp. 29–54.

[16] David Makinson & Leendert van der Torre (2000): *Input/output logics*. Journal of Philosophical Logic 29(4), pp. 383–408.

[17] David Makinson & Leendert van der Torre (2003): *Permission from an input/output perspective*. Journal of Philosophical Logic 32(4), pp. 391–416.

[18] Xavier Parent, Dov Gabbay & Leendert van der Torre (2014): *Intuitionistic basis for input/output logic*. In: David Makinson on Classical Methods for Non-Classical Problems, Springer, pp. 263–286.

[19] Xin Sun & Leendert van der Torre (2014): *Combining constitutive and regulative norms in input/output logic*. In Fabrizio Cariani, Davide Grossi, Joke Meheus & Xavier Parent, editors: *Deontic Logic and Normative Systems — 12th International Conference, DEON 2014, Ghent, Belgium, July 12-15, 2014*, Springer, pp. 241–257.