Exploring Neuro-Symbolic AI
for Facial Emotion Recognition

Jens Gebele[a, b, *]**, Anne Vetter**[b]**, Philipp Brune**[a]**, Frank Schwab**[b] **and Sebastian von Mammen**[b]

[a]Neu-Ulm University of Applied Sciences
[b]University of Würzburg
ORCID (Jens Gebele): https://orcid.org/0009-0009-5879-4409, ORCID (Philipp Brune):
https://orcid.org/0000-0002-2570-7462, ORCID (Sebastian von Mammen): https://orcid.org/0000-0002-7866-5788

Abstract. Facial Emotion Recognition (FER) aims at interpreting emotional states from facial behaviors. Deep Learning (DL) models have achieved notable successes in FER through inductive pattern recognition, yet their real-world effectiveness remains limited. This limitation stems from difficulties in capturing the multifaceted and nuanced range of facial behavior in both theoretical models and datasets. To address these issues, this paper proposes adopting Neuro-Symbolic AI (N-SAI), i.e. approaches that combine the rule-based strengths of symbolic AI with the numerical power of DL. We explore various N-SAI strategies, with a particular focus on abductive learning, which interprets sub-symbolic data into logical facts and uses logical abduction to correct misconceptions. This approach not only improves the adaptability of FER systems, but also fosters new insights into the relationship between facial behaviors and emotional states, substantially enhancing the practical utility and effectiveness of FER technologies. Additionally, the analysis relates N-SAI reasoning to human cognition, deduction, induction, and abduction, as a conceptual lens on hybrid intelligence.

1 Introduction

Facial Emotion Recognition (FER) entails the analysis and interpretation of emotional states based on observable facial behavior. This technology enhances human-machine interactions [7, 19], and plays a crucial role in supporting interpersonal understanding [31, 35]. The current state-of-the-art in FER predominantly relies on Deep Learning (DL) techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks, and Generative Adversarial Networks [23].

Despite these advances, FER systems face multiple challenges. Key issues include the complex and variable relationship between facial expressions and emotional states [3], the scarcity of diverse training data [23], and different environmental factors such as lighting, background and occlusions [25]. Inconsistencies in data annotations further complicate matters [12]. Moreover, the field also contends with different emotional theories that propose either a continuous or a discrete model concept of emotional states [3], such as the well-known theory of seven universal basic emotional states [9]. The ongoing debate about the universal applicability of facial expressions to infer emotional states highlights the variability and complexity inherent in human expressions [2]. Consequently, while FER systems

exhibit robust performance on controlled datasets, their effectiveness in real-world scenarios remains limited [37, 29].

To overcome these limitations, we propose Neuro-Symbolic AI (N-SAI) architectures for FER. N-SAI merges the rule-based precision of symbolic AI with the adaptive, data-driven capabilities of neural approaches, encompassing both symbolic reasoning and neural inductive reasoning [16]. This innovative approach is designed to address both data-related and theoretical challenges in emotion recognition. This paper conceptually explores various N-SAI architectures for FER, with a particular focus on abductive learning, to reveal new insights into the correlations between facial expressions and emotional states across different contexts and conditions. The structure of this paper is as follows: Section 2 examines current FER research. This is followed by an examination of six N-SAI design patterns. Section 4 highlights the most promising N-SAI design for FER, focusing on abductive learning and drawing conceptual parallels to human cognitive processes, namely, deduction, induction, and abduction.

2 Facial Emotion Recognition

In the field of FER, there are two main approaches that researchers are using. The first approach involves directly recognizing an emotional state from facial expressions in a single step. This method typically utilizes DL models trained on a large corpus of annotated data [30]. In contrast, the second approach starts with the detection of facial movements, which are categorized as Action Units (AUs) based on the Facial Action Coding System (FACS) [11] standard. These AUs are then mapped to corresponding emotional states [30]. Ultimately, both approaches rely on annotations of emotional states grounded in emotional theories, using either discrete [9] or continuous label spaces [14, 28]. In particular, the discrete model of the seven universal basic emotions is predominantly used.

A significant challenge across both approaches is the accurate mapping of facial behavior to emotional states, especially as the relationship between these behaviors and states is recognized to be more complex than previously understood, thus complicating the fidelity of data annotations based on discrete and continuous emotional models [3]. This complexity remains whether the mapping is performed directly through features learned by DL models or indirectly through the detection of Action Units by certified FACS coders [11] or semi-automated tools [1]. In all cases, a relationship to emotional states

---

must be established at the end.

Given this complexity, it is not surprising that while DL models achieve impressive results in controlled environments, their performance in real-world settings often falls short [23]. This limitation is primarily due to the models' dependence on approximating the distribution of the training data. Capturing the nuanced and multifaceted nature of facial expressions of emotions remains a significant challenge, necessitating considerable efforts to ensure datasets are diverse and of high quality [13]. Despite growing awareness of the foundational issues related to emotion theory concepts, validated approaches that address these underlying problems are still relatively rare.

The AI community is exploring solutions through training on data of higher quantity and quality, and by employing more complex DL strategies, such as multitask networks and transfer learning [23]. However, even without the data quantity challenge, the effectiveness of purely inductive DL approaches remains questionable. The complexity and variability of the relationship between facial expressions and emotional states, further influenced by cultural and age-related differences, make this task particularly challenging. Consequently, the mapping of facial behavior to emotional states, a critical foundation for annotating data of facial expressions, continues to pose a major issue.

3 Neuro-Symbolic AI (N-SAI)

Given the complex challenges identified in the field of FER, particularly in accurately mapping facial behaviors to emotional states across different contexts and demographics, we see N-SAI as a promising solution. N-SAI combines the rule-based strengths of symbolic AI (symbolism), characterized by deductive reasoning, with the numerical power of sub-symbolic DL, known for its inductive learning capabilities. This hybrid approach is designed to leverage the precision of symbolic rules and the adaptability of DL (connectionism) to effectively tackle the nuanced complexity of FER.

In this context, the foundational concept of symbols, fundamental to symbolic AI, is beneficial. There is an ongoing philosophical debate about symbolic versus non-symbolic or sub-symbolic data [34]. To simplify this issue, we adopt Berkeley's interpretation [4], which defines a symbol as meeting the following criteria:

1. It represents an object, category, or relationship.
2. It can be either simple or composite, consisting of other symbols.
3. It requires a defined process for creating new symbols from existing ones, ensuring that each new symbol also represents something.

Recent research interest in N-SAI has surged, reflecting a growing recognition of its potential. However, the existing literature remains diverse and predominantly empirical, making it challenging to navigate. To provide clarity, we draw on the seminal works of ten Teije and van Harmelen [34] and Kautz [21], which offer complementary perspectives on N-SAI architecture designs. While ten Teije and van Harmelen focus on detailed architectural analysis, Kautz provides high-level design patterns that serve as the foundation for our systematic overview. In this paper, we adopt Kautz's classification of six key N-SAI design patterns, illustrating each with concrete examples, advantages, and disadvantages, before demonstrating which of these frameworks address key challenges in FER.

3.1 Symbolic-Neuro-Symbolic

The *Symbolic-Neuro-Symbolic* pattern describes an approach where both the input and output are symbolic, while the processing in between is handled numerically by a neural architecture. For example, words can be converted to numerical vectors, e.g. by means of GloVe [26], which are processed by DL models to yield outputs in the form of sequences or categories represented again as symbols [34, 21]. This architecture offers the advantage of inductive learning from large amounts of data, enabling the system to discover patterns and generalize effectively. The lack of explainability is a key drawback, as the neural component operates as a black box, making it difficult to interpret decisions. The system is also highly dependent on data quality and quantity. For FER, this design pattern is ineffective because it requires processing sub-symbolic data, like image pixels, which symbolic rules alone cannot adequately represent.

3.2 Symbolic[Neuro]

The *Symbolic[Neuro]* pattern focuses on a symbolic problem-solving method, enhanced by a neural network acting as pattern recognition subroutine. In this setup, the neural network's numerical capabilities support the decision-making of the symbolic core. A notable example is AlphaGo [32], where the symbolic problem-solving core is implemented using Monte Carlo Tree Search [5], guided by a neural network as the evaluation subroutine. This architecture is particularly effective in scenarios requiring complex decision-making, such as autonomous driving [34, 21].

A key advantage of this pattern is its effective search process, where the symbolic system explores possible decisions guided by inductively learned representations from the neural network. This combination provides strong generalizability, allowing it to adapt to similar tasks, and easy transferability to other structured environments (e.g., different games) without requiring extensive domain-specific knowledge. However, the pattern has limitations. Its explainability is reduced when sub-symbolic (e.g., pixel-based) input is fed into the symbolic module without interpretable representations. Additionally, its transferability is limited in tasks lacking well-defined rules, particularly in settings characterized by continuous or ambiguous inputs. Moreover, the lack of abstract reasoning capabilities can hinder performance in tasks requiring logical generalization beyond learned patterns [32]. In the context of FER, this design pattern leverages the numerical power of DL and symbolic logic, but its one-sided flow of information - from the neural network to the symbolic system - prevents the emergence of higher reasoning capabilities through dynamic interaction.

3.3 Neuro | Symbolic

In the *Neuro | Symbolic* design, a neural network takes sub-symbolic (or non-symbolic) inputs, such as image pixels, and converts them into a format that a symbolic reasoning system can understand and process. An example of this is the Neuro-Symbolic Concept Learner [24], which integrates object-based scene representations and symbolic program execution to perform tasks like visual question answering and semantic parsing without direct supervision. In the *Symbolic[Neuro]* design, the neural module serves as a secondary subroutine, whereas in the *Neuro | Symbolic* approach it functions as a co-routine, working in parallel with the symbolic system [34, 21].

This design offers several advantages. It requires only weak data supervision, eliminating the need for pixel-level annotations. The architecture can learn novel visual and linguistic concepts, enabling

strong generalization across tasks. Additionally, the symbolic reasoning component improves interpretability compared to purely neural approaches by offering logical, structured explanations of decisions. However, the design faces challenges. It has a high dependency on the sub-symbolic perception module, as the symbolic system cannot correct errors made by the neural classifier. This limits performance in real-world scenarios with ambiguous object boundaries or new, unseen categories. Furthermore, the lack of end-to-end differentiability complicates training, as symbolic reasoning does not support backpropagation of gradients. The system also struggles with abstract reasoning, especially when dealing with high-level concepts that are not directly captured by visual perception [24].

For FER, this pattern shows potential due to its ability to process sub-symbolic data and provide interpretable outputs, but its inability to correct neural classifier errors makes it highly limited for facial behavior, which can be highly ambiguous. Additionally, as the knowledge describing AU-emotion relationships is often debated in emotion psychology, this pattern does not support the necessary refinement or adaptation of symbolic knowledge.

3.4 Neuro: Symbolic → Neuro

The *Neuro: Symbolic → Neuro* approach adheres to the architecture outlined in 3.1, with the distinction that training leverages symbolic rules instead of textual data. A notable implementation is the 2020 study by Lample and Charton [22], focusing on symbolic mathematics. They developed a transformer model trained to simplify mathematical expressions from one form (A) to another (B). Post-training, the model demonstrated the ability to accurately simplify new, previously unseen expressions, providing correct solutions directly without step-by-step derivations [34, 21].

This approach has several advantages. It exhibits high generalization power, allowing it to solve new, unseen problems. The end-to-end nature of training and inference eliminates the need for symbolic reasoning during prediction, making it computationally efficient. Studies by Lample and Charton [22] show that such models can outperform traditional symbolic systems, and the architecture is transferable to other symbolic tasks by adjusting the training data. However, there are challenges. The lack of explainability is a key drawback, as the system does not produce step-by-step derivations, making it difficult to trace how solutions are reached. The model is highly dependent on the quality and quantity of training data, and new symbolic expressions require retraining with updated data due to its inductive reasoning-only approach. Additionally, this method does not validate predictions, potentially reducing reliability. For FER, this pattern is similarly ineffective, like 3.1, because it solely relies on symbolic rules without the ability to process sub-symbolic data such as facial images.

3.5 Neuro_{Symbolic}

The *Neuro_{Symbolic}* architecture incorporates symbolic representations (rules) as templates to structure neural networks. Logic tensor networks [8] and tensor product representations [33] have successfully integrated hierarchical and abstract concepts within these neural networks. By encoding disjunctive rules (*OR*), this approach could facilitate combinatorial reasoning, enabling the system to manage multiple scenarios simultaneously [34, 21].

This architecture offers the advantage of combining inductive DL with symbolic reasoning, allowing systems to benefit from both data-driven pattern discovery and structured logic. The integration of sym-bolic logic enhances robustness, as it provides logical constraints or guard rails that can enforce requirements, such as those mandated by regulations. This dual nature improves generalization in tasks where structured reasoning and pattern recognition are both required. These advantages are accompanied by notable challenges. Both training and inference become more complex. The performance is highly dependent on the quality of the symbolic component; poorly defined symbolic rules can limit generalization and lead to errors when examples fall outside predefined logic. Additionally, encoding symbolic knowledge into neural networks is nontrivial and can require significant domain expertise. Explainability can be an advantage or disadvantage: well-defined symbolic knowledge enhances transparency, but complex encodings can reduce interpretability [8, 33]. For FER, this architecture is problematic because it relies on symbolic representations (concepts) embedded in neural networks, which are heavily debated in emotion psychology, making it unsuitable as an N-SAI architecture.

3.6 Neuro[Symbolic]

The *Neuro[Symbolic]* architecture combines symbolic reasoning and neural processing by embedding a symbolic engine within a neural engine to enhance "superneuro" and combinatorial reasoning capabilities. Inspired by Daniel Kahneman's dual-process theory from his seminal work, "Thinking, Fast and Slow" [18], this architecture harmonizes the rapid, intuitive operations of neural networks (System 1) with the methodical, thoughtful processes of symbolic reasoning (System 2). This configuration allows for both fast pattern recognition and thoughtful, detailed analysis within the same AI system [34, 21]. A key feature of this architecture is the dynamic interaction between the two systems, where one subsystem can activate the other, enabling a bidirectional flow of information. Insights from symbolic reasoning (System 2) can refine and enhance the pattern recognition capabilities of the neural network (System 1), while System 1 can provide data-driven insights that improve symbolic reasoning.

In our view, a concrete example of this principle is offered by research on abductive learning, a topic notably absent from the foundational works of ten Teije and van Harmelen [34] as well as Kautz [21]. Abductive Learning provides a powerful framework that integrates Machine Learning (ML) with logical reasoning [6]. It leverages abductive reasoning, a cognitive process central to hypothesis generation, creative problem-solving, and the generation of plausible explanations. As such, it constitutes a crucial component within the broader Neuro[Symbolic] pattern [20].

The advantages of *Neuro[Symbolic]* architecture are numerous. They offer increased explainability due to the symbolic reasoning component. The architecture enables reasoning capabilities that go beyond the inductive reasoning of DL and the deductive reasoning of symbolic logic, supporting more abstract and flexible forms of reasoning. Furthermore, the system benefits from improved generalization through logical constraints, which can help to reduce overfitting. The bidirectional flow of information between the neural and symbolic components enables the system to validate predictions, correct errors, and detect instances of new and unseen classes, thereby improving learning and adaptability over time. Most importantly, support for abductive reasoning equips the system with the ability to generate hypotheses and explanations based on incomplete information, making it effective in tasks involving uncertainty [38, 17].

These benefits come with certain challenges. Integrating the two systems is complex, particularly when resolving conflicts between

Table 1. N-SAI Patterns with Their Advantages and Disadvantages

Pattern	Advantages	Disadvantages
A	Generalization	Limited explainability; heavily dependent on data quality and quantity
B	Generalization; transferability	Lacks explainability; limited transferability in absence of well-defined rules; lacks abstract reasoning
C	Effective with weak supervision; strong generalization	Still limited explainability; dependent on neural classifier output; no end-to-end differentiable training; lacks abstract reasoning
D	Strong generalization; transferability	Limited explainability; dependent on data quality and quantity; requires extensive re-training
E	Enhanced generalization; potential for greater transparency	Complex training and inference; depends on symbolic logic quality; encoding requires expertise
F	Improved explainability; advanced reasoning; better generalization; error correction; new concept detection	Complex integration; symbolic components may limit transferability; lacks end-to-end differentiable training

the neural and symbolic parts to ensure consistency. The symbolic reasoning component can also limit the system's transferability to other domains, particularly to those requiring learning or tasks outside the defined symbolic knowledge base. Additionally, the architecture does not fully support end-to-end training using back-propagation, as the symbolic reasoning component introduces non-differentiable operations that complicate optimization [6].

In contrast to the previously discussed N-SAI architectures, the *Neuro[Symbolic]* architecture enables mutual improvement between the neural and symbolic components. This integration provides reasoning capabilities beyond traditional inductive and deductive approaches. Such advanced reasoning capabilities are particularly well-suited to the domain of FER, where the mapping between facial expressions and emotional states is often ambiguous, context-dependent, and highly variable. Given these strengths, we consider the *Neuro[Symbolic]* pattern to be the most promising architectural approach for addressing the challenges inherent in FER.

A particularly compelling instantiation of this pattern is abductive learning, which we examine in greater detail in the following section. To provide a structured comparison, Table 1 summarizes the key advantages and disadvantages of the six N-SAI patterns, highlighting their respective trade-offs and underscoring the strengths of the *Neuro[Symbolic]* approach.

4 The Role of Abductive Learning in FER

Building upon the limitations of existing FER approaches and the capabilities of various N-SAI design patterns outlined in Section 3, we propose a novel conceptual integration of abductive learning into FER. Unlike existing models that rely mostly on inductive learning via deep neural networks, our approach extends them via abductive reasoning, a cognitive process central to hypothesis generation and creative problem solving. This integration draws parallels to human cognitive reasoning processes, specifically, deduction, induction, and abduction. Deductive reasoning involves deriving specific conclusions from general principles; inductive reasoning entails identifying general patterns based on specific observations; and abductive reasoning involves hypothesizing plausible explanations based on incomplete or ambiguous information [20].

AI's historical evolution began with the symbolic era, noted for its deductive reasoning capabilities. This phase was followed by the sub-symbolic era, which highlighted inductive learning through so-

phisticated DL models [16, 15]. Building on this evolution, we propose that a promising next step in AI development may lie in the incorporation of abductive learning, which enables AI to formulate hypotheses and insights beyond the limitations of existing symbolic and sub-symbolic systems. This transition is essential for expanding AI's problem-solving capabilities beyond traditional deductive and inductive reasoning frameworks. Incorporating abductive learning aligns seamlessly with the *Neuro[Symbolic]* design pattern described in subsection 3.6. Additionally, we advocate for the synthesis of symbolic and sub-symbolic AI components in such a way that they mutually enhance the capabilities of each other.

Depending on the specific operational task or field, the choice of the most suitable N-SAI design pattern can vary. Given the unique challenges of FER, particularly the ambiguity and variability in the relationship between facial expressions and emotional states, we find *Neuro[Symbolic]* architectures to be particularly relevant. Unlike other designs, *Neuro[Symbolic]* integrates symbolic and sub-symbolic systems as mutually interacting routines. This capability allows the FER system to generate new hypotheses and insights beyond the available data and symbolic rules, effectively "thinking" outside the conventional datasets. This approach is essential because not all facial expression emotion relationships are universally applicable or adequately represented by data alone, making *Neuro[Symbolic]* uniquely equipped to navigate these complexities.

To this end, the research on abductive learning by Dai et al. [6] seems highly relevant, demonstrating how ML models can detect basic logical facts and use symbolic reasoning to correct errors and refine predictions. This method is exemplified in their work on decoding Mayan hieroglyphs, which involves recognizing numbers visually from the glyphs and using knowledge of mathematics and calendars to interpret them symbolically. Follow-up research has shown how symbolic knowledge can be refined if it is incomplete (i.e. new concepts can be detected) or inaccurate [38, 17].

For FER, this approach is particularly advantageous because it aligns DL model predictions with theoretical knowledge about the relationship between emotional states and facial behavior. The key feature of this approach is its reliance on high-quality AU annotations, enabling the DL model and symbolic reasoning system to mutually refine and enhance each other's outputs. Practically, this involves a (pre-trained) DL Model (perception model) predicting an emotional state (ground truth label) from a facial input image. Additionally, the facial images include AU annotations, ideally from certified FACS
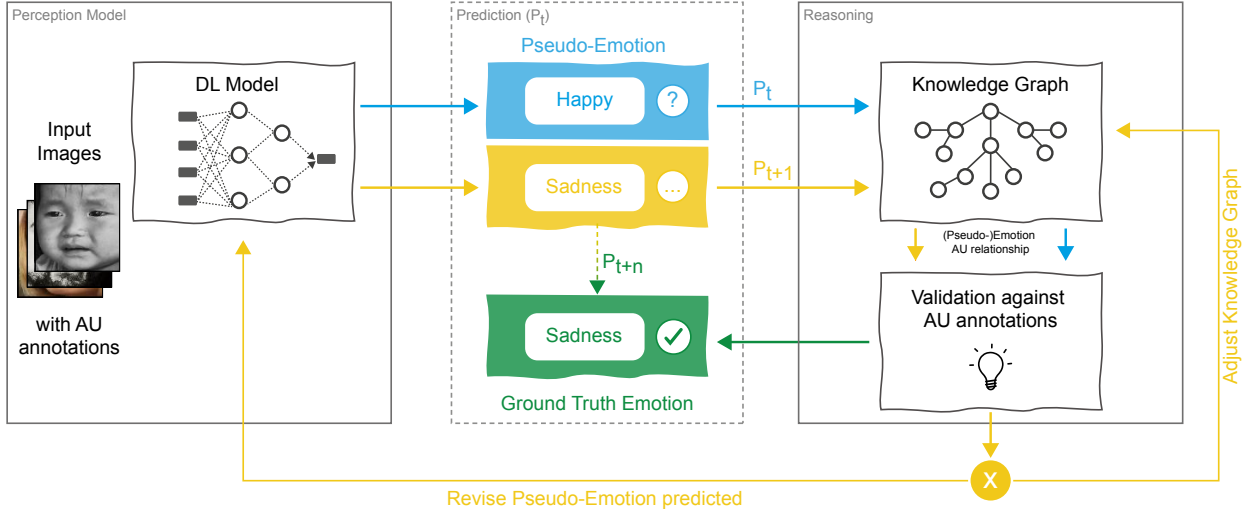
Figure 1. Illustration of abductive learning through mutual interaction between a DL model and a symbolic reasoning part.

coders. A knowledge graph (reasoning), built on expert knowledge of AU-emotion relationships, is used to deduce the AUs associated with the predicted emotion.

The symbolic reasoning part then validates whether the data-driven prediction, complemented by the corresponding AUs deduced from the knowledge graph, matches the original AU annotations from the facial images. If both the perception model and the reasoning part concur, the output of the DL model (ground truth prediction) is likely correct. However, if they diverge, the inconsistency may stem from either the DL model's prediction errors or an incomplete or inaccurate knowledge graph. This discrepancy can be resolved either by retraining the DL model with a revised prediction or by updating the knowledge graph with refined AU-emotion mappings and retesting it against the model.

In the example illustrated in Figure 1, the process begins with a perception model (e.g. a pre-trained DL model) analyzing an image of a sad-looking child. Despite the visual cues, the model initially incorrectly predicts a pseudo-emotion label "happiness". The image also contains AU annotations, manually provided by FACS experts, which in this case likely include AU 1 (Inner Brow Raiser), AU 4 (Brow Lowerer), and AU 15 (Lip Corner Depressor), typically indicative of "sadness". The predicted pseudo-emotion is then passed to the reasoning module, which infers the corresponding AUs for "happiness" using a knowledge graph. In this context, the graph suggests AU 6 (Cheek Raiser) and AU 12 (Lip Corner Puller), based on established mappings from Ekman's FACS Investigator Guide [10].

These inferred AUs are compared against the expert-annotated AUs present in the image. The mismatch between the deduced and observed AUs initiates a revision process: the system reconsiders the initial pseudo-emotion and iteratively refines its prediction, potentially through classifier retraining with the revised prediction. Ultimately, the process converges on "Sadness" as the final emotion label, whose knowledge-graph derived AUs align with those annotated in the image.

This adaptive feedback loop promotes coherence across three layers: the model's perceptual prediction, symbolic reasoning based on AU-emotion relationships, and empirical AU observations. The system will iteratively refine its predictions and symbolic knowledge, leading to consistency between DL and symbolic reasoning output. This process effectively manages ambiguity and variability in facial expressions. It enables the generation of new hypotheses such as novel labels or AU-emotion combinations. Additionally, human expert feedback can be incorporated into this mechanism [38].

A critical component of abductive learning in FER is the consistency optimization between the perception model and the symbolic reasoning module. This involves dynamically adjusting pseudo-labels predicted by an undertrained DL model, often inaccurate in the early training phases, so that they align with domain knowledge encoded in a symbolic system. While Dai et al. [6] rely on derivative-free optimization (RACOS) [39], alternative strategies such as evolutionary algorithms [27] or LLM-based optimizers [36] may offer greater flexibility, particularly for complex or semantically ambiguous correction tasks. This iterative process refines both the symbolic knowledge graphs and the neural model predictions, promoting coherence across perception and reasoning. Over successive cycles, relational features extracted from consistent hypotheses serve as feedback, enabling the DL model to generalize better, distinguish ambiguous features, and ultimately converge toward symbolically grounded, high-fidelity outputs.

We anticipate that abductive learning will offer significant advantages for FER. By integrating both sub-symbolic data and expert knowledge, this approach is expected to handle the inherent ambiguity and variability in facial expressions more effectively than conventional methods. It also offers enhanced adaptability through the continuous refinement of DL predictions and symbolic knowledge, enabling the system to apply learned patterns to novel, unseen sce-

narios, such as new AU-emotion combinations. Its capacity to reason under uncertainty makes abductive learning well-suited for robust, accurate emotion recognition, even with sparse or noisy data. Additionally, this reasoning process is expected to yield deeper insights into AU-emotion relationships by refining the underlying symbolic knowledge base and enabling a more systematic evaluation of the quality and consistency of research data, both of which are crucial for advancing emotion psychology.

This approach allows FER systems to not only detect patterns in facial expressions but also to evaluate these patterns against symbolic expert knowledge, such as the theory of seven universal emotional states. Furthermore, abductive learning enables the possibility to refine the symbolic knowledge base, ensuring that FER systems adapt to new findings and remain effective in different contexts. This adaptability is particularly valuable in environments where expressions may vary significantly, helping to mitigate biases and enhance the reliability of emotion recognition technologies. In sum, our proposed abductive learning framework within the *Neuro[Symbolic]* pattern offers a novel, cognitively inspired, and technically robust path forward for FER.

5 Conclusion

In this work, we have examined the fundamental challenges faced by current FER systems and systematically evaluated six N-SAI architectures as potential solutions. Among these, we propose the *Neuro[Symbolic]* design pattern as the most suitable framework for addressing the ambiguity, variability, and contextual dependence inherent in facial emotional expression.

As a key contribution, we identify abductive learning as a novel and underexplored instantiation within the *Neuro[Symbolic]* paradigm, uniquely capable of integrating symbolic (deductive) and sub-symbolic (inductive) reasoning. Inspired by human cognitive processes, particularly the ability to generate plausible explanations from incomplete observations, abductive learning enables FER systems to not only align data-driven predictions with expert emotional theories, but also to iteratively resolve inconsistencies between neural outputs and symbolic knowledge. This integration promotes a more adaptable, interpretable, and conceptually grounded approach to emotion recognition.

We believe this framework opens promising directions for FER research by enabling systems to hypothesize, adapt to uncertainty, and support a deeper understanding of emotional behavior across diverse real-world contexts. Future work should focus on empirically validating the proposed architecture, exploring advanced consistency optimization techniques, and incorporating human-in-the-loop feedback to enhance interpretability and ensure the ethical and trustworthy deployment of affective technologies.

References

[1] M. M. Adnan, M. S. M. Rahim, A. Rehman, Z. Mehmood, T. Saba, and R. A. Naqvi. Automatic Image Annotation Based on Deep Learning Models: A Systematic Review and Future Challenges. *IEEE Access*, 9: 50253–50264, 2021. doi: 10.1109/ACCESS.2021.3068897.

[2] L. F. Barrett. Discrete Emotions or Dimensions? The Role of Valence Focus and Arousal Focus. *Cognition and Emotion*, 12(4):579–599, July 1998. doi: 10.1080/026999398379574.

[3] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements:. *Psychological Science in the Public Interest*, July 2019. doi: 10.1177/1529100619832930.

[4] I. S. N. Berkeley. What the <0.70, 1.17, 0.99, 1.07> is a Symbol? *Minds and Machines*, 18(1):93–105, Mar. 2008. ISSN 1572-8641. doi: 10.1007/s11023-007-9086-y.

[5] G. Chaslot, S. Bakkes, I. Szita, and P. Spronck. Monte-Carlo Tree Search: A New Framework for Game AI. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 4(1):216–217, 2008. ISSN 2334-0924. doi: 10.1609/aiide.v4i1.18700.

[6] W.-Z. Dai, Q. Xu, Y. Yu, and Z.-H. Zhou. Bridging Machine Learning and Logical Reasoning by Abductive Learning. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[7] J. Deng, G. Pang, Z. Zhang, Z. Pang, H. Yang, and G. Yang. cGAN Based Facial Expression Recognition for Human-Robot Interaction. *IEEE Access*, 7:9848–9859, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2891668.

[8] I. Donadello, L. Serafini, and A. d'Avila Garcez. Logic Tensor Networks for Semantic Image Interpretation, May 2017.

[9] P. Ekman. Basic emotions. Handbook of cognition and emotion. *Wiley, New York*, pages 301–320, 1999.

[10] P. Ekman and W. V. Friesen. *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press, 1978.

[11] P. Ekman, W. V. Friesen, and J. C. Hager. *Facial Action Coding Sytem*. A Human Face, Salt Lake City, Utah, 2002.

[12] J. Gebele, P. Brune, and S. Faußer. Face Value: On the Impact of Annotation (In-)Consistencies and Label Ambiguity in Facial Data on Emotion Recognition. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 2597–2604, Aug. 2022. doi: 10.1109/ICPR56361.2022.9956230.

[13] J. Gebele, P. Brune, F. Schwab, and S. Von Mammen. *Assessing Sequential Databases for Spontaneous and Posed Facial Expression Recognition*. Jan. 2025. ISBN 978-0-9981331-8-8.

[14] H. Gunes and B. Schuller. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2):120–136, Feb. 2013. ISSN 0262-8856. doi: 10.1016/j.imavis.2012.06.016.

[15] P. Hitzler, F. Bianchi, M. Ebrahimi, and M. K. Sarker. Neural-symbolic integration and the Semantic Web. *Semantic Web*, 11(1):3–11, Jan. 2020. ISSN 1570-0844. doi: 10.3233/SW-190368.

[16] P. Hitzler, M. K. Sarker, and A. Eberhart, editors. *Compendium of Neurosymbolic Artificial Intelligence*, volume 369 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, July 2023. ISBN 978-1-64368-406-2 978-1-64368-407-9. doi: 10.3233/FAIA369.

[17] Y.-X. Huang, W.-Z. Dai, Y. Jiang, and Z.-H. Zhou. Enabling knowledge refinement upon new concepts in abductive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 7928–7935, 2023.

[18] D. Kahneman. *Thinking, Fast and Slow*. macmillan, 2011.

[19] S. Kamal, F. Sayeed, and M. Rafeeq. Facial emotion recognition for Human-Computer Interactions using hybrid feature extraction technique. In *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, pages 180–184, Mar. 2016. doi: 10.1109/SAPIENCE.2016.7684129.

[20] T. Kapitan. Peirce and the autonomy of abductive reasoning. *Erkenntnis*, 37(1):1–26, July 1992. ISSN 1572-8420. doi: 10.1007/BF00220630.

[21] H. Kautz. The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture. *AI Magazine*, 43(1):105–125, Mar. 2022. ISSN 2371-9621. doi: 10.1002/aaai.12036.

[22] G. Lample and F. Charton. Deep Learning for Symbolic Mathematics, Dec. 2019.

[23] S. Li and W. Deng. Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 13(3):1195–1215, July 2022. ISSN 1949-3045. doi: 10.1109/TAFFC.2020.2981446.

[24] J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, and J. Wu. The Neuro-

Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision, Apr. 2019.

[25] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. Web-Based Database for Facial Expression Analysis. In *2005 IEEE International Conference on Multimedia and Expo*, pages 317–321, Amsterdam, The Netherlands, 2005. IEEE. doi: 10.1109/ICME.2005.1521424.

[26] J. Pennington, R. Socher, and C. Manning. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1162.

[27] P. Rakshit and A. Konar. Foundation in Evolutionary Optimization. In P. Rakshit and A. Konar, editors, *Principles in Noisy Optimization: Applied to Multi-agent Coordination*, pages 1–56. Springer, Singapore, 2018. ISBN 978-981-10-8642-7. doi: 10.1007/978-981-10-8642-7_1.

[28] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 1980. doi: 10.1037/h0077714.

[29] N. Samadiani, G. Huang, B. Cai, W. Luo, C.-H. Chi, Y. Xiang, and J. He. A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data. *Sensors*, 19(8):1863, Jan. 2019. ISSN 1424-8220. doi: 10.3390/s19081863.

[30] D. Seuss, T. Hassan, A. Dieckmann, M. Unfried, K. R. Scherer, M. Mortillaro, and J. Garbas. Automatic Estimation of Action Unit Intensities and Inference of Emotional Appraisals. *IEEE Transactions on Affective Computing*, 14(2):1188–1200, Apr. 2023. ISSN 1949-3045. doi: 10.1109/TAFFC.2021.3077590.

[31] Shan Jia, S. Jia, S. Wang, C. Hu, P. J. Webster, X. Li, X. Li, Xin Li, and Xin Li. Detection of Genuine and Posed Facial Expressions of Emotion: Databases and Methods. *Frontiers in Psychology*, 11, Jan. 2021. doi: 10.3389/fpsyg.2020.580287.

[32] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, Jan. 2016. ISSN 1476-4687. doi: 10.1038/nature16961.

[33] P. Smolensky, M. Lee, X. He, W.-t. Yih, J. Gao, and L. Deng. Basic Reasoning with Tensor Product Representations, Jan. 2016.

[34] A. ten Teije and F. van Harmelen. Architectural patterns for neurosymbolic AI. In P. Hitzler, A. Eberhart, and M. K. Sarker, editors, *Compendium of Neurosymbolic Artificial Intelligence*, Frontiers in Artificial Intelligence and Applications, pages 64–76. IOS Press, 2023. ISBN 978-1-64368-406-2. doi: 10.3233/FAIA230135.

[35] P. Werner, A. Al-Hamadi, R. Niese, S. Walter, S. Gruss, and H. Traue. *Automatic Pain Recognition from Video and Biomedical Signals*. Aug. 2014. doi: 10.1109/ICPR.2014.784.

[36] C. Yang, X. Wang, Y. Lu, H. Liu, Q. V. Le, D. Zhou, and X. Chen. Large Language Models as Optimizers, Sept. 2023.

[37] K. Yang, C. Wang, Z. Sarsenbayeva, B. Tag, T. Dingler, G. Wadley, and J. Goncalves. Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets. *The Visual Computer*, 37(6):1447–1466, June 2021. ISSN 1432-2315. doi: 10.1007/s00371-020-01881-x.

[38] X.-W. Yang, J.-J. Shao, W.-W. Tu, Y.-F. Li, W.-Z. Dai, and Z.-H. Zhou. Safe Abductive Learning in the Presence of Inaccurate Rules. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16361–16369, 2024.

[39] Y. Yu, H. Qian, and Y.-Q. Hu. Derivative-Free Optimization via Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), Mar. 2016. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v30i1.10289.