Neuro-symbolic AI and Thinking Fast and Slow with SOFAI

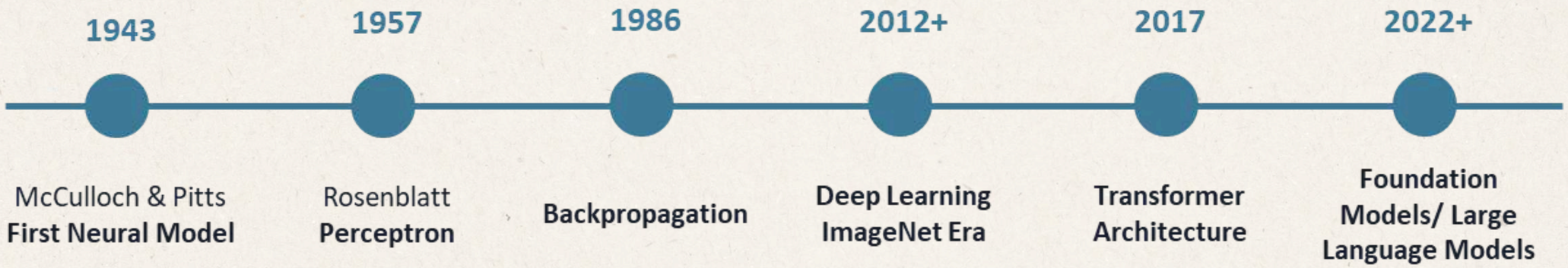
Antonella Coviello - Politecnico di Torino

Mentor: Prof. Zeynep Kiziltan

Outline

1. Neuro-symbolic AI
2. SOFAI (Slow and Fast Artificial Intelligence)
3. Applications of SOFAI
4. AGI and Neuro-symbolic AI

Statistical, Learning-based AI



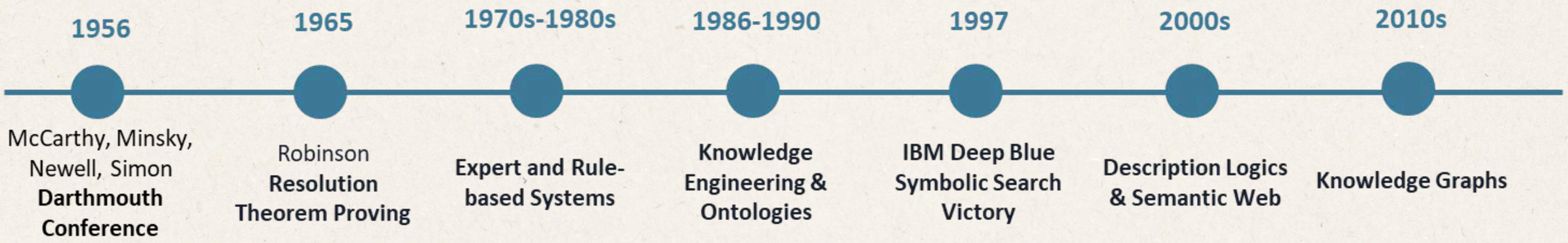
Why ML became dominant?

- Generalises from raw data
- Internet & GPU revolution
- Scalable across domains
- Less hand-crafted rules

Limitations

- Black-box decisions
- Data quality & bias issues
- Lacks true reasoning
- Poor generalisation on Out of Distribution samples

Symbolic AI



What is it?

- Logic & inference rules: Prolog reasoning
- Explicit knowledge representation : Knowledge graphs
- Planning & search : A* algorithm
- Symbolic theorem provers : Coq / Lean

Strenghts

- Interpretable & explainable
- Consistent & verifiable
- No training data needed
- Logical guarantees

Limitations

- Brittle in open environments
- Needs controlled conditions
- Expensive to hand-craft rules

Statistical vs Symbolic AI

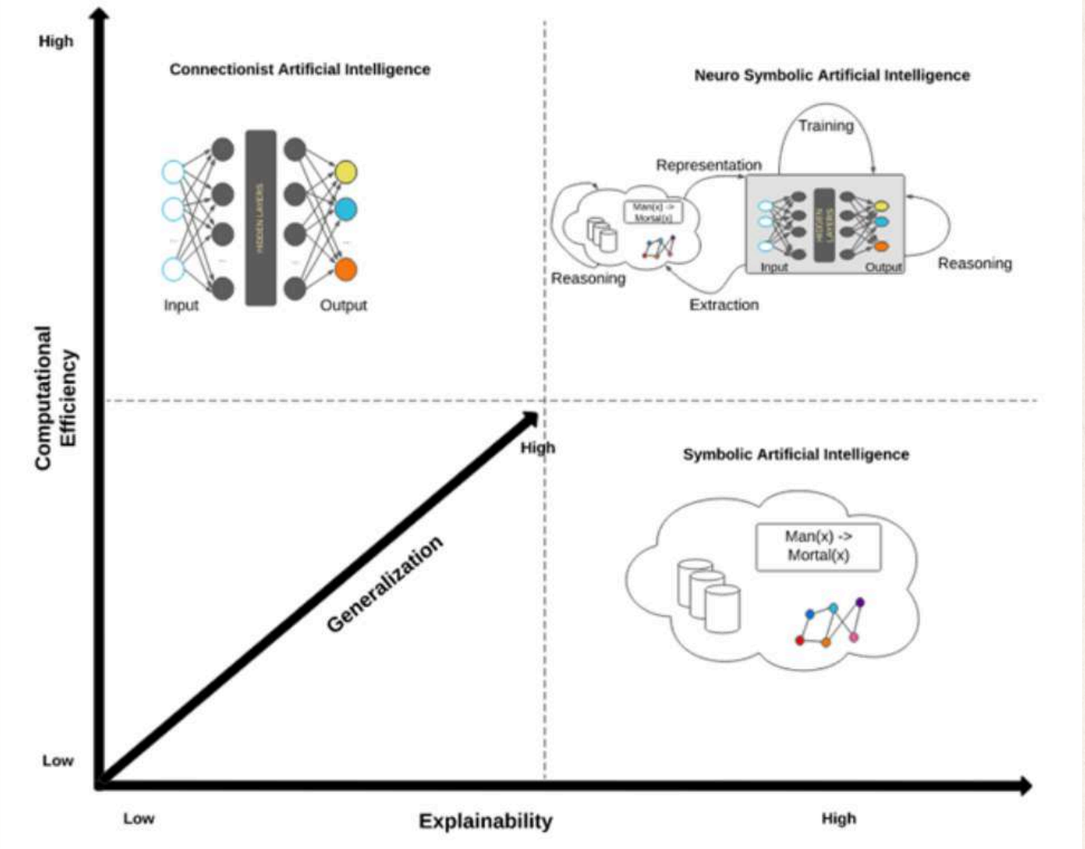
	Statistical, Learning-based AI (ML/DL)	Symbolic AI
Efficiency	⬆ Scalable training and fast inference on large datasets	⬇ Rule evaluation and search can be computationally expensive
Generalization	⬆ Strong statistical generalisation within distribution	⬇ Limited to encoded rules unless extended manually
Data Needs	⬇ Requires large labelled datasets	⬆ Does not require training data (relies on encoded knowledge)
Explainability	⬇ Opaque internal representations (“black box”)	⬆ Traceable logical inference and explicit rules
Robustness	⬇ Sensitive to distribution shift and adversarial noise	⬆ Stable under logical consistency; brittle if knowledge incomplete
Reasoning	⬇ Approximate, emergent reasoning patterns	⬆ Explicit deductive, inductive, abductive reasoning

Neuro-symbolic AI

Neuro-symbolic AI combines the **main strengths** of statistical and symbolic AI: **computational efficiency** and the **explainability**

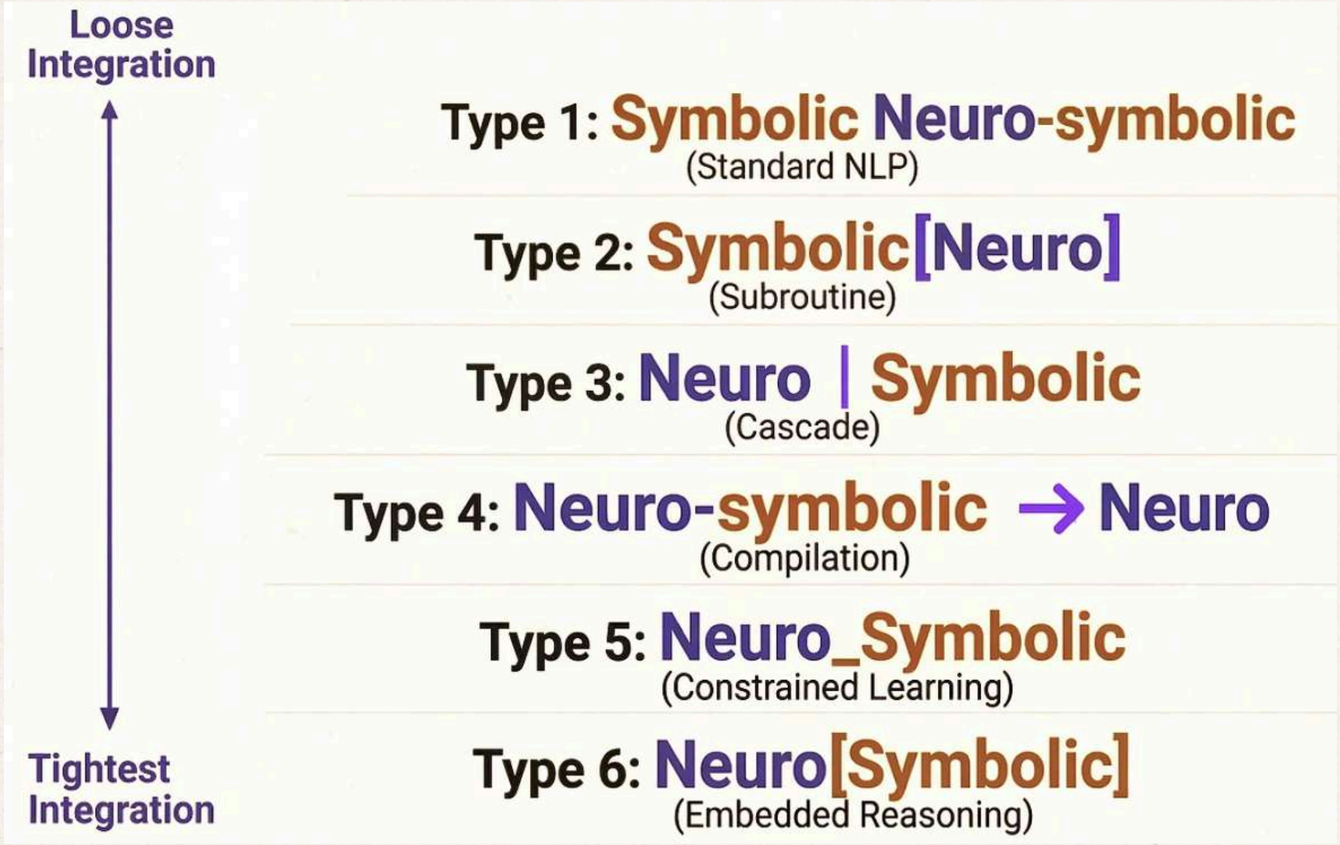
1. **Representation**
   - Distributed embeddings + symbolic structures
2. **Learning**
   - Data-driven induction + rule extraction
3. **Reasoning**
   - Neural pattern recognition + symbolic inference
4. **Decision-making**
   - Fast intuition + structured deliberation

How to best integrate neural & symbolic architectures?

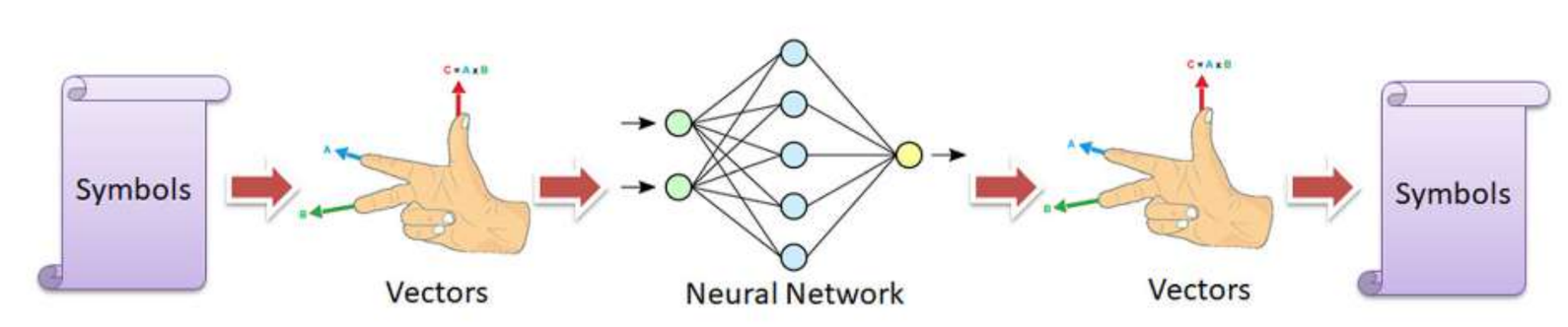


Kautz Taxonomy

Six families of neuro-symbolic integration - H. Kautz, AAAI Robert S. Engelmore Memorial Lecture 2020



1 - Symbolic Neuro-symbolic (Standard NLP)



Symbolic input → Neural → Symbolic Output

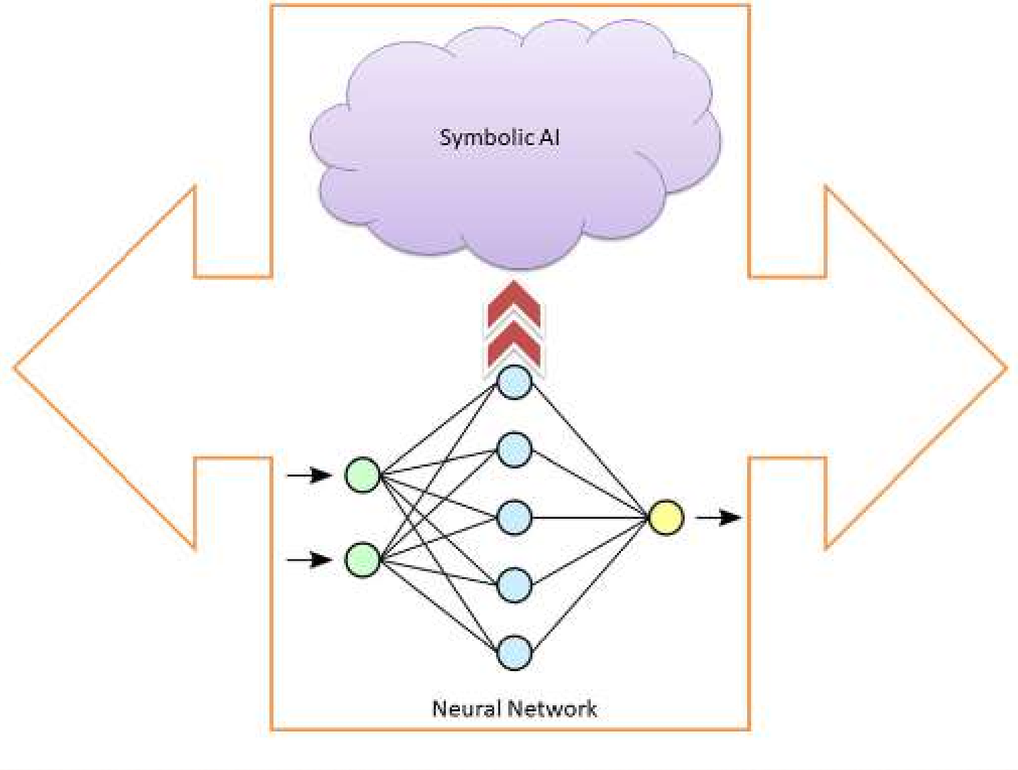
Definition: Symbolic data → Neural processing → Symbolic output

Concrete Example: Word2Vec + neural classifier: text tokens converted to vectors, processed by a network, output mapped back to a symbolic category

Key architectural idea: Standard NLP pipeline: Symbolic input/outputs wrap a neural black-box. The symbolic framing makes I/O interpretable even if the internals are not

2 - Symbolic [Neuro] (Subroutine)

Symbolic solver uses Neural Heuristic



Definition: Neural network as a subroutine inside a symbolic solver

Concrete example: AlphaGo → MCTS (symbolic) uses a neural network as its heuristic evaluation function

Key architectural idea: The symbolic system retains control and logical structure, the neural component provides learned heuristics where explicit rules would be impractical

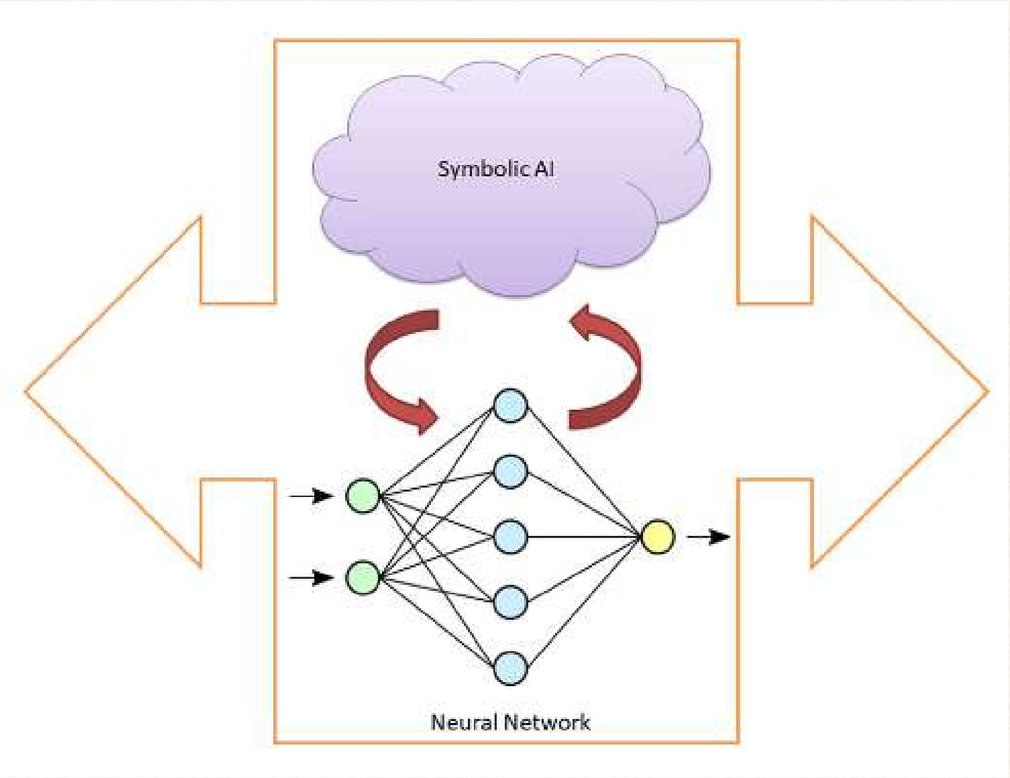
3 - Neuro | Symbolic (Cascade)

Raw Input → Neural → Symbolic Struct → Symbolic Reasoner

Definition: Neural perception layer converts raw input into symbolic data structures for a symbolic reasoner

Concrete example: Neuro-Symbolic Concept Learner: a CNN extracts object representations, then a symbolic reasoner answers relational questions

Key architectural idea: Clean separation of perception (neural) and reasoning (symbolic). Enables interpretable reasoning on complex perceptual inputs like images



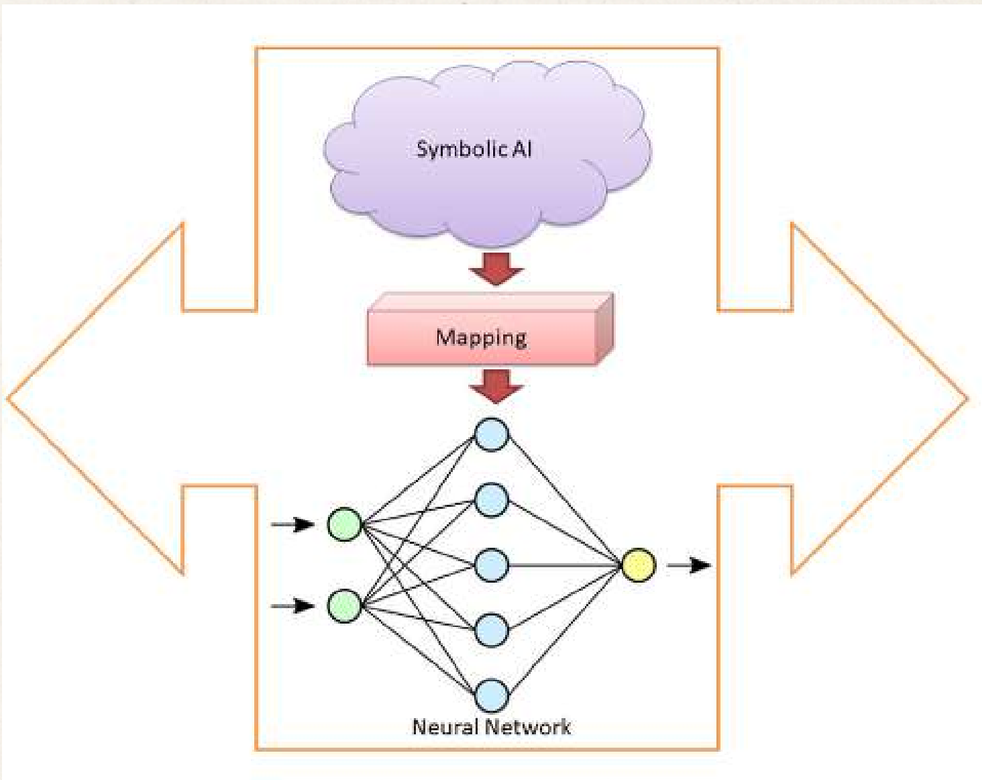
4 - Neuro-symbolic → Neuro (Compilation)

Symbolic Rules → train → Neural Network (inference)

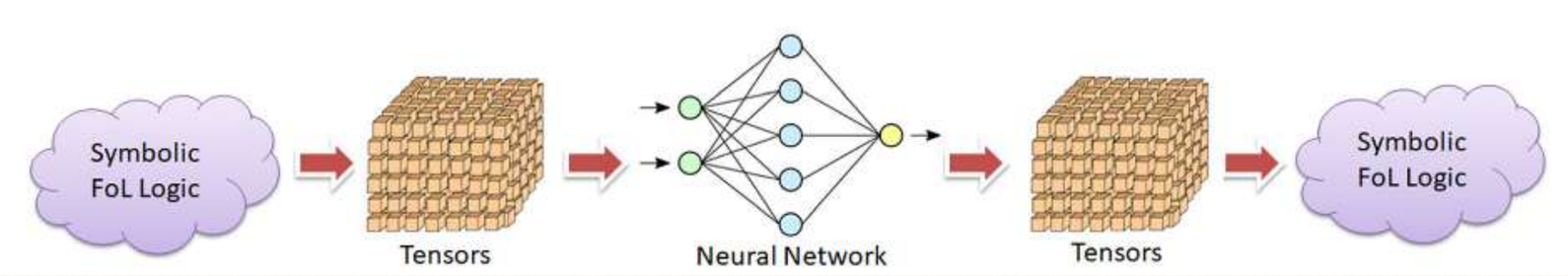
Definition: A special training regime based on symbolic rules guides neural learning

Concrete example: Logic-guided neural networks where symbolic constraints (e.g., physics equations) are injected into the loss function during training

Key architectural idea: Symbolic knowledge steers training without being present at inference. The final model is purely neural but was shaped by symbolic priors



5 - Neuro_Symbolic (Constrained Learning)



Symbolic rules compiled into Neural Architecture (end-to-end)

Definition: Symbolic rules are compiled into the structure of the neural network itself

Concrete example: Tensor Product Representations & Logic Tensor Networks: logical rules become template layers or constraints inside the network architecture

Key architectural idea: Deep fusion: the symbolic structure is baked in. Enables end-to-end differentiable reasoning but makes the architecture less flexible

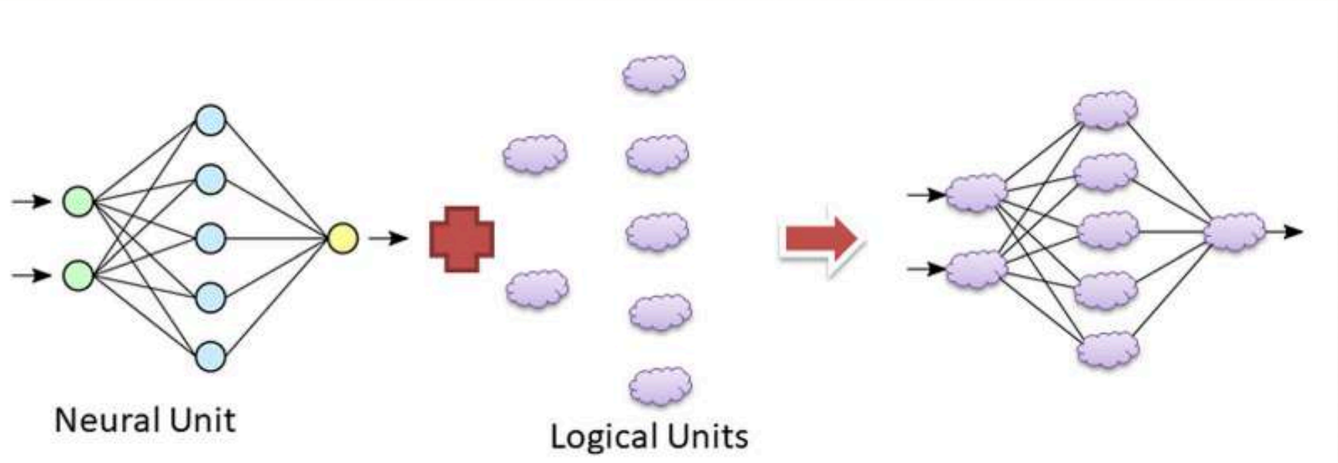
6 - Neuro[Symbolic] (Embedded Reasoning)

Neural (S1) <-> Meta (Governs switching)→ Symbolic (S2)

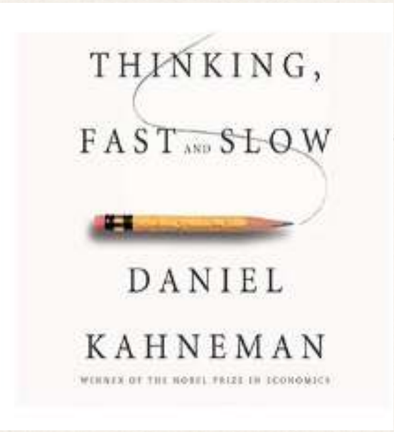
Definition: A symbolic reasoning engine is embedded inside a neural engine

Concrete example: SOFAI architecture: LLM or RL agent (System 1, neural) orchestrated by a metacognitive layer that can invoke a symbolic solver (System 2)

Key architectural idea: Inspired by Kahneman's dual-process theory. The neural component handles fast intuitive reasoning; the symbolic component handles deliberate, correct reasoning. Metacognition governs when to switch



Khaneman Thinking Fast and Slow



Daniel Kahneman , Nobel Prize-winning psychologist and pioneer of **behavioral economics**, was the **author** of **Thinking, Fast and Slow**.

System 1 can perform cognitively easy tasks:

- Local and parallel
- Handles and exploits causality to build approximate views of the world

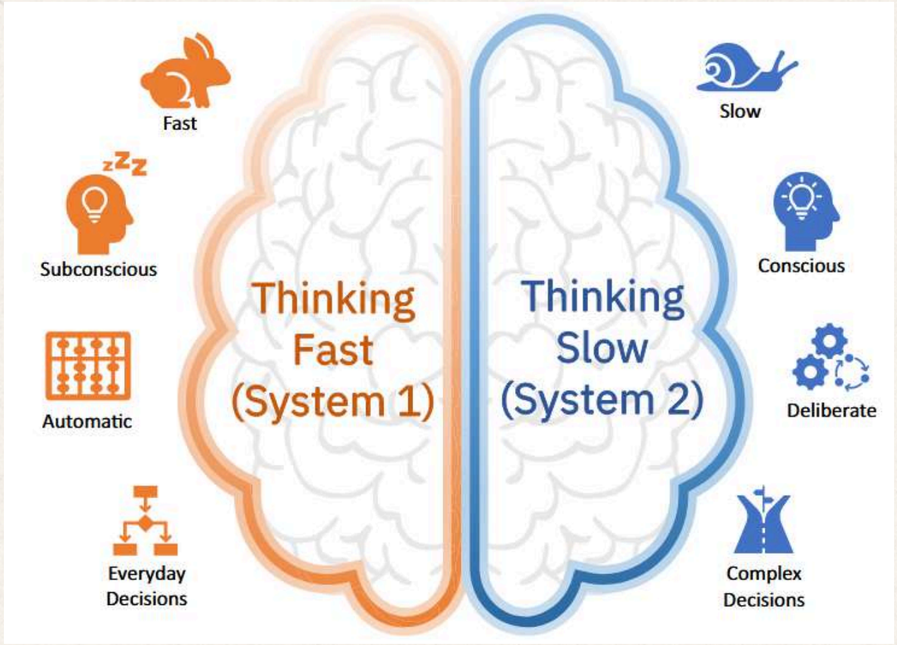
System 2 handles more complex tasks:

- Requires all our attention
- Global and sequential

System 2's **search for solution** is usually supported by System 1's **heuristics**.

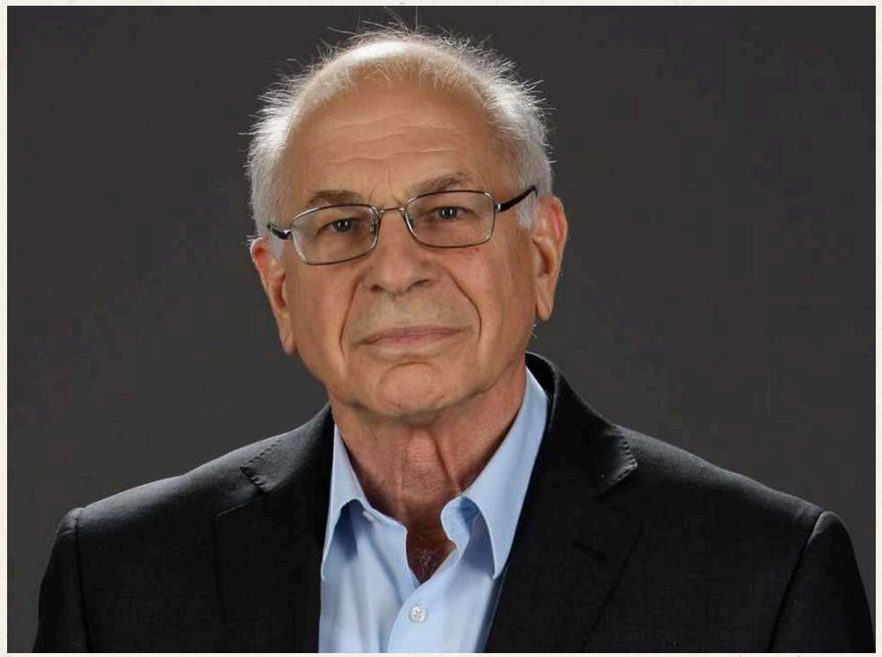
Skill learning : some tasks transfer from System 2 to System 1 over time.

Cognitive control: system 1 reacts while system 2 may override it.



Khaneman on Current AI (2020)

“Since the **triumphs** of **machine learning** in **2012**, artificial intelligence produces solutions by the **opaque operations** of large **neural net**, which shares many features with the **automatic intuitions** of **System 1**. The **integration** of **intuition and reasoning** in artificial intelligence **was not achieved** during the **first** decade of the **machine-learning era**."



SOFAI Overview

A multi-agent architecture (Thinking Fast and Slow in AI, AAAI 2021) where:

System 1 solvers (statistical/learning-based):

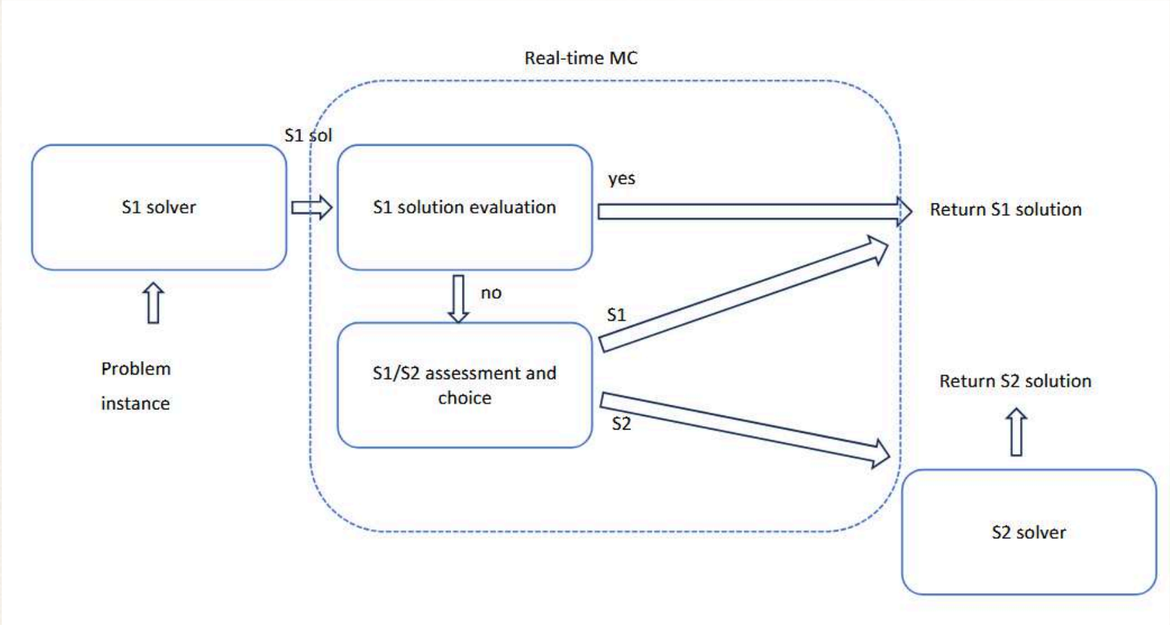
- Rely on past experience and model of the world
- Not reason on the problem
- React to arrival of new problem instance
- Generate a candidate solution

System 2 solvers (symbolic):

- Reason on the given problem
- Computational complexity dependent on the input size
- Activated by meta-cognition
- Generate solution

Model/solver updater

- Acts in background to update models of world, self, and others



SOFAI Architecture

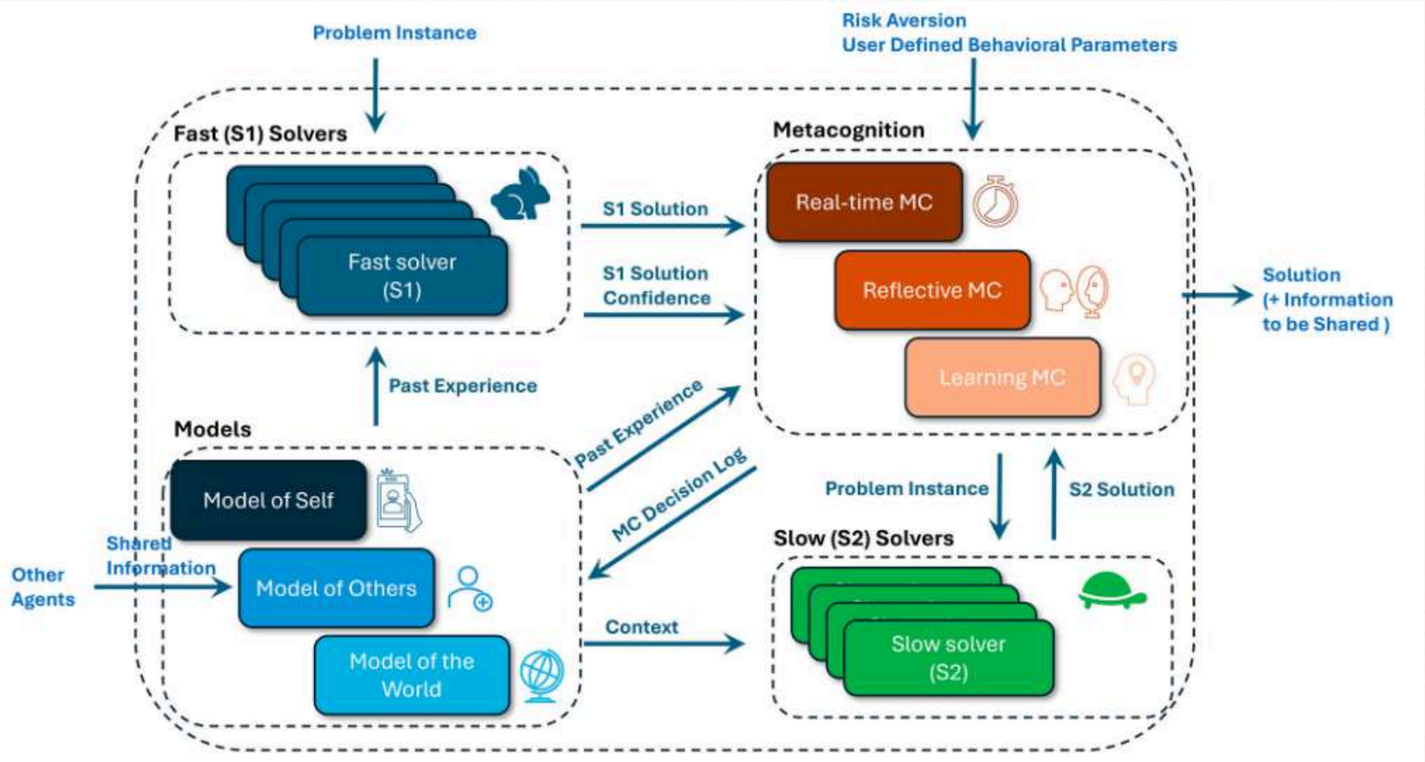
Fast solvers: Only based on past experience

Slow solvers: Reasoning about the problem istance

Metacognition:

- **Real-time**: decides which solver to use
- **Offline- reflective** : counterfactual **S2** comparison and possible adjustment of MC parameters
- **Offline-learning**: solvers/models update

Fast solvers act indipendently, slow solvers need to be triggered by MC



Metacognition

Online Metacognition (Real - time)

- **Phase 1** : Rapid **S1** Assessment:
  - Is **S1** confidence above the threshold?
  - Is experience sufficient?
  - Are resources available?
  - → If YES : adopt **S1** solution

- **Phase 2** : Cost benefit Analysis
  - Expected reward gain from **S2**
  - vs. cost of invoking **S2**
  - → if gain > cost : activate **S2**

Offline Metacognition:

- **Learning MC:**
  - Stores action, reward, solver ID
  - Updates **S1** solvers models
  - Improves Model of Self/World
  - → Enables improvement over time

- **Reflective MC:**
  - Counterfactual **S2** simulations
  - Compare past solutions vs **S2**-only
  - Adjust MC parameters if **S2** better
  - → Adapts meta-policy over time

Emerging Capabilities

Three capabilities emerge from the SOFAI architecture:

- **Skill learning**: Initially SOFAI relies mostly on **S2**. As experience accumulates, reflective MC, shift decisions toward **S1**, mirroring how humans move from deliberate to automatic processing
- **Adaptability** : When multiple **S1** solvers with different competence levels are available, MC learns to exploit the best one, maintaining solution quality while minimising compute, adapts to the solver's actual capabilities
- **Cognitive control**: In high-risk or highly constrained scenarios, SOFAI becomes more risk-averse and systematically invokes **S2**, reducing constraint violations, similar to how humans engage slow reasoning in high-stakes contexts

Grid Navigation Example

Constraints over cells and moves (actions)
Goals:

- **Maximize reward**, minimize time and length
- **High risk aversion**: minimize constraint violation

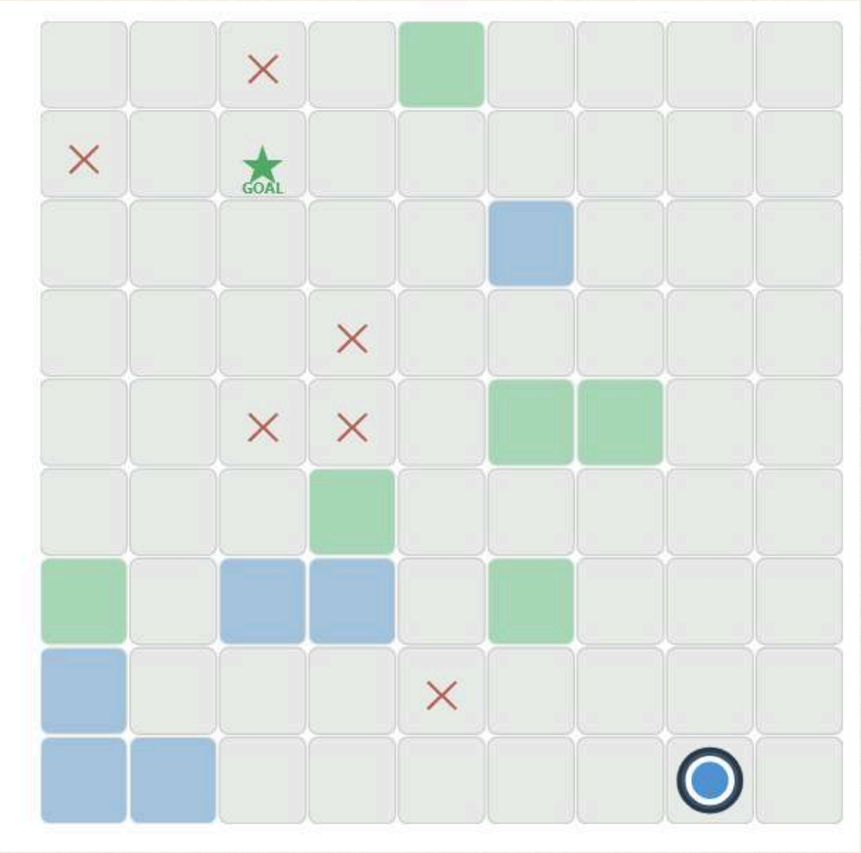
Solvers:

- **S1 solver**: RL agent, trained on past trajectories and their reward
- **S2 solver**: MDFT (Multi-attribute Decision Field Theory)

Meta-cognition:

- **Real-time MC:** two-phases choice between **S1** and **S2**
- **Reflective (counterfactual) MC**: compare past trajectories with simulated **S2**-only ones, and adjust MC parameter
- **Learning MC:** update model of self and **S1** solver

• Five versions of **S1** solver, with varying degrees of randomness



Demo Link

15-puzzle

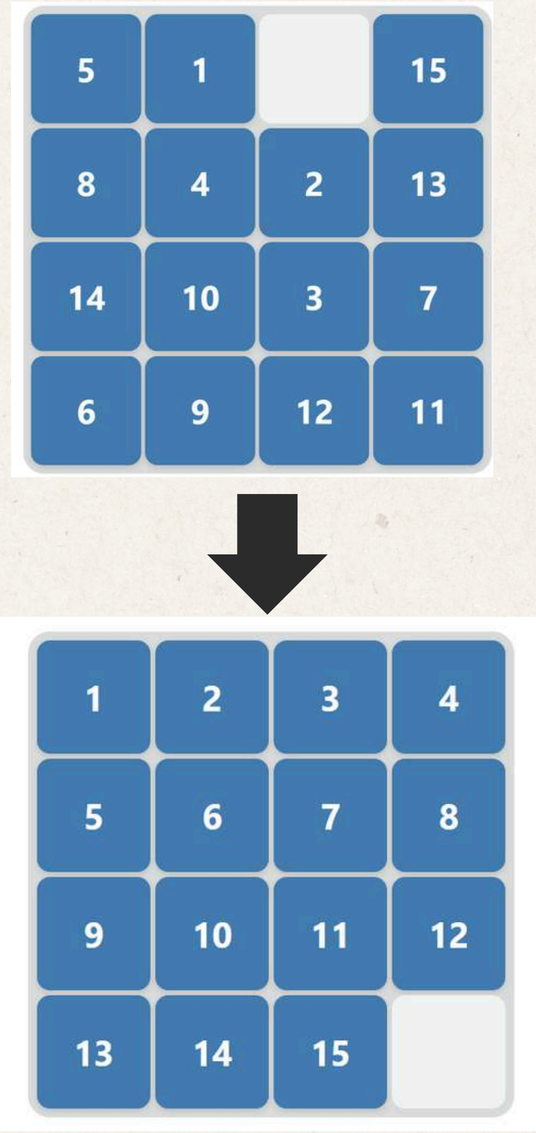
Environment: 4×4 sliding puzzle, goal [1,…,15,0]
Cost: −1 per move
Objective: Find a sequence of moves that solves the sliding tile puzzle
Solvers:

- **S2 solver**: IDA* (Iterative Deepening A*)
- **S1 solver**: Policy trained via imitation learning on backward-generated moves and IDA*

Meta-cognition:

- **Real-time MC:** decides between **S1** policy and **S2** IDA* search
- **Reflective MC:** compares hybrid trajectories with **S2** optimal solutions
- **Learning MC:** update model of self and **S1** solver

Demo Link



Road to Artificial General Intelligence

What would an Artificial General Intelligence (AGI) need? AI that replicates broad human cognitive abilities.

1. Perception	2. Learning	3. Interaction	4. Creation
5. Quantitative Skills	6. Partial Reasoning	7. Persistent Memory	8. Reliable Planning
9. Embodied Action	10. Motivation	11. Emotion	

Light blue = partially demonstrated by LLMs
Dark blue = still major bottlenecks for AGI

Current LLMs show early signs of AGI but lack of full breadth, control, alignment and embodiment remain open challenges.

Yann LeCun (2025, Big Technology Podcast) : "We are **not** going to get to human-level AI by just scaling up LLMs"

How Neuro-symbolic AI contributes

By bridging the gap between narrow AI and AGI

1.Perception + Interaction + Creation

Neural components provide strong representation learning from raw data and support flexible multimodal generation

2.Reasoning + Planning

Symbolic rules, search, and constraints help make reasoning and planning more explicit, controllable, and logically consistent

3.Memory + Knowledge

Structured external memory and knowledge representations can complement limited parametric memory

4.Safety + Interpretability

Rules, constraints, and explicit intermediate representations can make decisions more auditable and easier to inspect

5. Transfer + Compositionality

Symbolic abstractions can support concept reuse and cross-task generalization, while neural modules retain flexibility on messy real-world inputs

6.Not a full solution

Neuro-symbolic AI is promising, but surveys still identify gaps

Limitations of Neuro-symbolic AI

1. **Integration complexity:** Combining neural learning with symbolic reasoning remains technically challenging
2. **Scalability issues:** Symbolic reasoning components may struggle to scale to large, real-world environments
3. **Knowledge engineering cost:** Symbolic representations and rules often require manual design or expert knowledge
4. **Learning–reasoning coordination:** Efficient interaction between learning modules and reasoning modules is still an open research problem

Neuro-symbolic AI vs Large Reasoning Models (LRMs)

Recent **Large Reasoning Models (LRM)** integrate reasoning capabilities directly into LLMs via RL-based training

This raises an important question: **Are explicit symbolic components still necessary?**

Open challenges **specific to LRMs** remain:

1. Lack of formal guarantees
2. Fragility under distribution shift
3. Poor unsolvability detection
4. Steep cost–accuracy tradeoff

Hybrid approaches: combining LRMs and symbolic solvers and metacognitive governance (e.g., SOFAI-LM) may provide a promising direction.

Takeaways & Conclusions

1. **Neither paradigm is sufficient alone**
   - Pure neural systems lack explicit structure and guarantees
   - Pure symbolic systems lack scalability and perceptual robustness
2. **Human cognition suggests a solution**
   - Kahneman's dual-process theory shows that intelligence emerges from the interaction of:
     - Fast, heuristic processing (System 1)
     - Deliberate, structured reasoning (System 2)
   - AI architectures should reflect this complementarity
3. **SOFAI demonstrates the principle in practice**
   - An explicit S1/S2 separation, governed by metacognition, achieves better performance-resource trade-offs than either system alone
4. **We are entering the era of neuro-symbolic AI**
   - The next leap in AI capability will likely come not from scaling one paradigm, but from integrating learning, reasoning, and governance in coherent hybrid systems
5. **Open questions remain**
   - Scalability and integration challenges persist
   - How to integrate LLMs with symbolic components

Neuro-symbolic AI represents a promising path toward systems that are both powerful and trustworthy

Thank you for your attention

Bibliography and Resources

1. **AAAI 2020 Panel** on Thinking Fast and Slow in AI. Panel discussion, AAAI
2. **Booch, G., et al. (2021).** Thinking Fast and Slow in AI. AAAI.
3. **Kautz, H. A. (2022).** The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture. AI Magazine.
4. **Bhuyan, B. P., et al. (2024).** Neuro-symbolic Artificial Intelligence: A Survey. Neural Computing and Applications.
5. **Bhuyan, B. P., et al. (2025).** Neuro-Symbolic Artificial Intelligence: Bridging Logic and Learning. Springer.
6. **Bergamaschi Ganapini., et al. (2025).** Fast, Slow, and Metacognitive Thinking in AI. npj Artificial Intelligence.
7. **Fabiano, F., et al. (2025).** Thinking Fast and Slow in Human and Machine Intelligence. Communications of the ACM.
8. **Yang, X.-W., et al. (2025).** Neuro-Symbolic Artificial Intelligence: Towards Improving the Reasoning Abilities of Large Language Models. IJCAI (Survey Track).
9. **Khandelwal, V., et al. (2026).** SOFAI-LM: A Cognitive Architecture for Building Efficient and Reliable Reasoning Systems with LLMs. AAAI Tutorial and Lab Forum.
10. **SOFAI website**.