Mechanism-Based AGI Architectures

Amar Nath Thakur
Associate Professor
Department of Information Technology
ABES Engineering College
***Email ID**: amarnath14thakur@gmail.com*

ABSTRACT

Artificial General Intelligence (AGI) aims to develop machines that can perform any intellectual task that a human can do. Most present AI systems are task-specific and rely heavily on data-driven learning without deep understanding of internal mechanisms. Recently, researchers are focusing on mechanism-based architectures where intelligence emerges from structured cognitive components such as memory, reasoning, perception, planning, and learning modules that interact in systematic way. These architectures attempt to mimic how human cognition works rather than only optimizing statistical performance. This paper reviews mechanism-based AGI architectures, their design principles, cognitive inspirations, and implementation strategies. Various architectural models like symbolic-subsymbolic hybrids, cognitive architectures, neurosymbolic models, and world-model based agents are discussed. Advantages, challenges, and future research directions are also presented. Tables and figures are included to show architectural comparisons and module interactions.

KEYWORDS: *Artificial General Intelligence, Cognitive Architecture, Mechanism-based Design, Neurosymbolic AI, World Models, Memory Systems, Reasoning Engine.*

INTRODUCTION

Artificial Intelligence today is dominated by narrow systems such as deep learning networks, large language models, and reinforcement learning agents. These systems show impressive

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

performance but lack general reasoning, adaptability, and causal understanding. AGI research is therefore shifting from only data-driven models toward architectures inspired by cognitive mechanisms of human brain.

Mechanism-based AGI architectures focus on *how intelligence operates internally*. Instead of treating intelligence as a black box, these architectures define modules for perception, memory, reasoning, decision making, and learning. Interaction among these modules forms a complete intelligent agent. This approach is inspired from psychology, neuroscience, and cognitive science.

The objective of this paper is to review such architectures and analyze how mechanism-level design can contribute to building general intelligence.

CONCEPT OF MECHANISM-BASED ARCHITECTURE

A mechanism-based architecture for AGI is built by decomposing intelligence into **functional cognitive mechanisms** that resemble how human cognition is organized. Instead of relying on a single, end-to-end neural model, intelligence is distributed across specialized modules, each responsible for a well-defined role. These modules cooperate through structured interfaces, shared representations, and feedback loops. The goal is not only to perform tasks, but to **understand, reason, remember, plan, and adapt** across domains.

This design philosophy emphasizes **functional decomposition**, **inter-module communication**, and **explainability**. Each mechanism can be improved, replaced, or scaled independently without disturbing the entire system.

1. Sensory Perception Mechanism

The perception module acts as the **input gateway** of the AGI system. It processes raw sensory data such as text, images, and speech, video, or sensor streams and converts them into structured internal representations.

Key roles:

- Feature extraction from raw signals
- Multi-modal fusion (e.g., combining vision and language)

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

- Object, event, and pattern recognition
- Context detection from environment

Typical techniques:

- Convolutional Neural Networks (CNNs) for images
- Transformers for text and multi-modal learning
- Speech encoders for audio understanding

The output of perception is not the final decision, but a **symbolic or vector representation** passed to memory and reasoning modules.

2. Working and Long-Term Memory

Memory is central to mechanism-based AGI. Unlike standard AI models that "forget" between tasks, AGI systems require persistent knowledge.

Table 1: Types of memory

Memory Type	Function	AGI Equivalent
Working Memory	Temporary context for current task	Context buffer, attention window
Long-Term Memory	Stored facts, rules, experiences	Knowledge base, vector DB
Episodic Memory	Past experiences and events	Logs, interaction history
Semantic Memory	General knowledge of world	Knowledge graphs, ontologies

Memory allows the agent to recall previous situations, learn continuously, and personalize responses.

3. Knowledge Representation

For reasoning to happen, knowledge must be stored in structured format.

Forms of representation:

- Knowledge graphs
- Ontologies
- Semantic networks
- Rule bases
- Embedding spaces

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

Knowledge representation acts as a bridge between memory and reasoning. It provides **meaning** to the stored data rather than raw numbers.

4. Logical and Causal Reasoning

Reasoning is what makes AGI different from pattern-matching AI. This module performs inference, deduction, and causal analysis.

Types of reasoning:

- Deductive reasoning (logic rules)
- Inductive reasoning (learning patterns)
- Abductive reasoning (best explanation)
- Causal reasoning (cause-effect relationships)
- Probabilistic reasoning (Bayesian inference)

The reasoning engine uses knowledge from memory and inputs from perception to draw conclusions.

5. Planning and Decision Making

This mechanism determines **what action to take** based on goals and reasoning outcomes.

Functions:

- Goal prioritization
- Action sequence generation
- Risk and reward evaluation
- Handling uncertainties

Techniques used:

- Markov Decision Processes (MDP)
- Monte Carlo Tree Search (MCTS)
- Heuristic search algorithms (A*)
- Utility-based decision frameworks

Planning ensures that AGI behaves purposefully rather than reactively.

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

6. Learning and Adaptation

Learning allows the system to improve with experience. This mechanism updates perception models, memory content, and reasoning rules.

Learning modes:

- Supervised learning from data
- Reinforcement learning from feedback
- Self-supervised learning from environment
- Continual learning without forgetting
- Meta-learning (learning how to learn)

Adaptation ensures flexibility across tasks and environments.

7. Action Execution

The final mechanism translates decisions into actions.

Examples:

- Generating text or speech
- Controlling robots or software agents
- Sending commands to external systems
- Interacting with users

This module closes the perception-action loop.

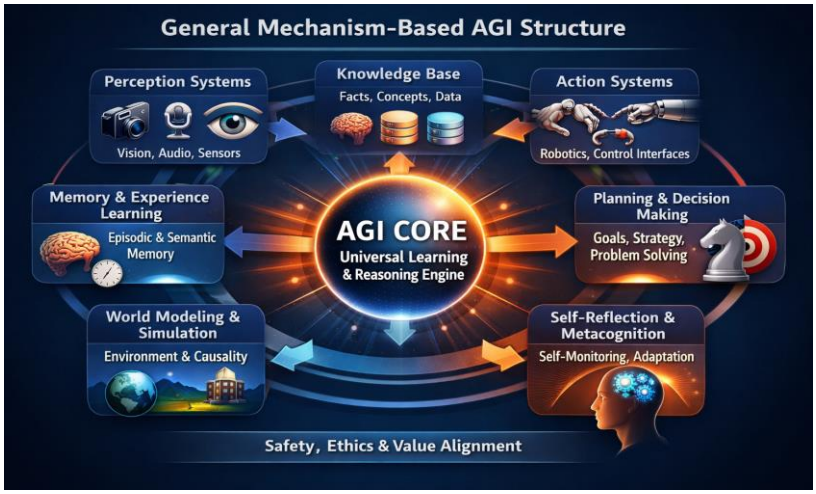


Figure 1: General Mechanism-Based AGI Structure

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

COGNITIVE INSPIRATION FROM HUMAN BRAIN

Mechanism-based AGI architectures are strongly influenced by findings from neuroscience and cognitive psychology which show that **human intelligence emerges from cooperation among many specialized brain subsystems**. The brain does not rely on a single uniform process for thinking. Instead, perception, memory, reasoning, attention, learning, emotion, and action are handled by different regions that continuously exchange information.

This biological organization provides an important blueprint for designing AGI systems where intelligence is distributed across modules rather than embedded in one large network.

1. Modularity of the Brain

The brain is functionally modular:

- Occipital lobe processes vision
- Temporal lobe handles language and memory
- Parietal lobe manages spatial awareness
- Frontal lobe performs reasoning and planning
- Subcortical regions assist in learning and decisions

No part works in isolation. Intelligence arises due to **interaction** among these parts. Mechanism-based AGI adopts this same principle by separating perception, memory, reasoning, learning, and action into modules.

2. Perception Systems: Sensory Cortices

Humans understand the world through specialized sensory systems:

- Visual cortex interprets images and motion
- Auditory cortex interprets sound and speech
- Somatosensory cortex processes touch

These systems transform raw sensory signals into meaningful internal patterns.

AGI inspiration: Dedicated perception models (vision models, speech models, text encoders) convert raw data into structured representations before passing to reasoning modules.

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

3. Memory Systems: Hippocampus and Cortex

Table 1: Human memory is layered and categorized

Type of Memory	Brain Region	Role	AGI Analogy
Working memory	Prefrontal cortex	Temporary task context	Active context window
Episodic memory	Hippocampus	Personal experiences	Interaction logs
Semantic memory	Temporal cortex	General world knowledge	Knowledge graphs
Procedural memory	Basal ganglia	Skills and habits	Learned policies

This separation inspires AGI systems to maintain multiple memory types rather than a single data store.

4. Reasoning and Planning: Prefrontal Cortex

The prefrontal cortex is responsible for:

- Logical thinking
- Planning steps ahead
- Evaluating consequences
- Problem solving

This region operates slower but more carefully than reflex actions.

AGI parallel: A reasoning engine that uses symbolic rules, causal models, and planning algorithms separate from perception networks.

5. Learning and Neural Plasticity

The brain adapts continuously through **neural plasticity**. Humans learn from:

- Experience
- Observation
- Feedback
- Repetition

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

AGI inspiration:

- Continual learning systems
- Reinforcement learning from feedback
- Self-supervised learning from environment
- Updating knowledge without full retraining

6. Attention Mechanisms

Humans cannot process everything simultaneously. Attention mechanisms select what is important.

AGI parallel:

- Attention layers in transformer models
- Memory retrieval based on relevance
- Focused processing in working memory

Attention controls information flow between modules.

MAJOR MECHANISM-BASED AGI ARCHITECTURAL MODELS

1. Cognitive Architectures (SOAR, ACT-R)

These architectures model cognition through symbolic rules and memory chunks. They include:

- Declarative memory
- Procedural memory
- Goal stack
- Rule-based reasoning

They are good for reasoning but weak in perception tasks.

2. Hybrid Symbolic-Subsymbolic Architectures

Combines neural networks with symbolic reasoning.

Examples:

- Neural perception + symbolic reasoning engine
- Knowledge graph integrated with LLMs

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

3. Neurosymbolic Architectures

These models integrate logic with learning. Neural nets learn representations, while symbolic modules enforce logical constraints.

4. World Model Based Agents

Agents create internal model of environment and simulate outcomes before acting.

Used in:

- Model-based reinforcement learning
- Predictive coding systems

KEY MECHANISMS IN AGI DESIGN

1. Perception Module

Handles multi-modal inputs like text, image, and audio. Deep learning is commonly used here.

2. Memory System

Memory is divided into:

- Short-term memory (context buffer)
- Long-term memory (knowledge base)
- Episodic memory (experience logs)

3. Reasoning Engine

Uses:

- Logical inference
- Causal reasoning
- Probabilistic reasoning

4. Learning Mechanism

Combination of:

- Supervised learning
- Reinforcement learning
- Self-supervised learning

5. Planning and Decision Module

Uses goal-oriented planning and utility-based decisions.

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

Table 2: Functional Role of Mechanisms

Mechanism	Role	Typical Techniques
Perception	Input processing	CNN, Transformers
Memory	Store knowledge	Vector DB, Knowledge graphs
Reasoning	Inference	Logic rules, Bayesian nets
Learning	Adaptation	RL, Gradient descent
Planning	Goal execution	Search, MDP

INTERACTION BETWEEN MODULES

AGI is achieved when these modules operate in coordination.

Example flow:

- Perception captures environment.
- Memory retrieves relevant knowledge.
- Reasoning evaluates situation.
- Planning decides action.
- Learning updates model after feedback.

ADVANTAGES OF MECHANISM-BASED ARCHITECTURES

- Transparency in decision making
- Better reasoning capability
- Modular upgradability
- Interpretability
- Reduced data dependency

CHALLENGES IN IMPLEMENTATION

- Integration complexity
- High computational cost
- Designing common representation
- Synchronization of modules
- Scalability issues

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

Table 3: Comparison with Pure Deep Learning Models

Feature	Deep Learning	Mechanism-Based AGI
Interpretability	Low	High
Generalization	Limited	Broad
Reasoning	Weak	Strong
Data Requirement	High	Moderate
Modularity	No	Yes

CASE STUDIES AND RESEARCH EXAMPLES

- OpenCog Hyperon cognitive architecture
- DeepMind's Gato and world model agents
- IBM neurosymbolic research
- LLM + Knowledge graph systems

ROLE OF KNOWLEDGE REPRESENTATION

Structured knowledge is essential:

- Ontologies
- Knowledge graphs
- Semantic networks

They help reasoning module to operate logically.

MEMORY-CENTRIC AGI MODELS

Recent research shows memory as central component. Long context memory enables:

- Continual learning
- Personalization
- Experience based reasoning

PLANNING AND GOAL MANAGEMENT

AGI must manage multiple goals. Planning algorithms like A* search, MCTS, and hierarchical planning are integrated.

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved

LEARNING ACROSS DOMAINS

Mechanism-based systems allow transfer learning because knowledge is stored in structured way.

FUTURE TRENDS

- Self-reflective architectures
- Autonomous knowledge building
- Better neurosymbolic integration
- Cognitive digital twins

CONCLUSION

Mechanism-based AGI architectures represent a promising direction toward true general intelligence. By decomposing intelligence into cognitive mechanisms such as memory, reasoning, learning, and planning, these architectures offer transparency, adaptability, and strong reasoning ability. While integration challenges exist, continued research in neurosymbolic models, cognitive architectures, and world-model agents will make AGI more practical. Future systems will not rely only on large data but on structured cognitive processes similar to human thinking.

REFERENCES

1. Newell, A. “Unified Theories of Cognition.” Harvard University Press, 1990.
2. Laird, J. “The SOAR Cognitive Architecture.” MIT Press, 2012.
3. Anderson, J. R. “ACT-R: A Theory of Higher Level Cognition.” Psychological Review, 1996.
4. Garcez, A., Lamb, L. “Neurosymbolic AI: The 3rd Wave.” AI Magazine, 2020.
5. Hassabis, D. et al. “Neuroscience-Inspired AI.” Neuron, 2017.
6. Ha, D., Schmidhuber, J. “World Models.” NIPS, 2018.
7. Lake, B. et al. “Building Machines That Learn and Think like People.” Behavioral and Brain Sciences, 2017.
8. Lenat, D. “Cyc: A Large-Scale Investment in Knowledge Infrastructure.” CACM, 1995.
9. Silver, D. et al. “Mastering the Game of Go with Deep Neural Networks.” Nature, 2016.

Page 122-133 © MANTECH PUBLICATIONS 2025. All Rights Reserved