Towards Long Context Search for Enhanced Robot Assistance

1. Introduction

The increasing integration of robots into human-centric environments, such as homes and workplaces, necessitates advanced capabilities that enable them to effectively understand and respond to user needs. A crucial aspect of this is the ability for robots to retain and access information about their past operations and observations over extended periods. Consider the scenario where a user asks a robot assistant about the location of a misplaced tool, potentially hours after the tool was last used. To address such queries effectively, the robot requires a system capable of processing and retrieving information from a vast and continuous stream of data, encompassing both textual logs of its activities and visual records of its surroundings. This challenge aligns with the concept of "long context search" in artificial intelligence, which deals with the complexities of handling and extracting relevant information from extensive sequences of data. This report outlines a research proposal focused on exploring potential approaches for developing an effective long context search system tailored for a robot assistant, ultimately aiming to identify key research questions and a methodological framework for such a project. The core objective is to enable a robot to possess a comprehensive "memory" of its interactions and environment, allowing for specific and contextually relevant responses to user inquiries, even those referencing events from the distant past.

The ability of AI models to remember information is often described in terms of their "context window," which refers to the amount of information the model can consider at any given time [1]. A longer context window allows the model to retain more information, leading to improved accuracy and understanding, akin to remembering all episodes of a TV series rather than just the current one [1]. Models with larger context windows, such as Google's Gemini 1.5 Pro with its million-token capacity, can process substantial amounts of data, including lengthy documents and extended conversations [1]. This capability suggests that the technological underpinnings for creating a robot with a substantial memory of its activities are becoming increasingly viable. The analogy between a long context window and human memory, particularly short-term memory [2], provides a useful conceptual framework for understanding the user's requirement for a robot with a persistent and searchable history of its actions. The rapid advancements in context window sizes in state-of-the-art models indicate a clear trend towards enabling AI systems to handle increasingly larger amounts of information within a single session [2]. This progression directly supports the feasibility of developing a system that could allow a robot to effectively recall events from hours of operation.

2. Background and Motivation

Long context search, in the context of this research, can be formally defined as the ability of an intelligent system to efficiently and accurately retrieve specific pieces of information from a continuous and potentially multimodal stream of data accumulated over an extended period. This differs from traditional search methods that typically operate on smaller, discrete datasets. For a robot assistant, this continuous data stream would include textual logs of its actions, interactions with users, and descriptions of its environment, as well as visual data captured by its sensors, such as video recordings or image sequences. Beyond the user's specific example of locating misplaced tools, long context search in robotics holds significant potential for various applications. These include enabling robots to learn complex tasks through continuous observation and instruction, providing detailed reports of their activities over extended missions, and adapting their behavior based on a comprehensive understanding of past interactions and environmental changes. For instance, a robot could learn the nuances of a specific task by remembering a series of demonstrations performed hours or even days prior, or it could provide a detailed account of anomalies it observed during a prolonged monitoring session.

However, current AI models and information retrieval systems face several limitations in effectively handling the type of long-term, multimodal memory envisioned for this robot assistant. One significant challenge is the computational cost and scalability associated with processing extremely long sequences of data [2]. While context windows are expanding, they are not infinite, and managing the vast amounts of data generated over hours of robotic operation requires efficient memory management and information retrieval strategies. Maintaining information retrieval accuracy over such extensive datasets is another hurdle, as models can struggle to pinpoint specific details within a large volume of information [2]. Furthermore, seamlessly integrating and reasoning across textual and visual data streams presents a complex problem, as these modalities often require different processing techniques and operate on different timescales [8]. The "lost in the middle" problem, where information located in the middle portions of long contexts tends to be overlooked or underutilized by models [6], also poses a significant challenge for effectively leveraging very long context. This suggests that simply increasing the context window size might not be sufficient, and the position of relevant information within the sequence can impact retrieval accuracy. Finally, there is a trade-off between the length of the context and the relevance of the information it contains. Including too much irrelevant data can introduce noise and degrade the performance of the search system [6]. The debate surrounding long context models versus Retrieval-Augmented Generation (RAG) [5] highlights a fundamental design choice: should the robot rely on a very large internal memory or on retrieving relevant information from an external store? This decision likely depends on factors such as the dynamic nature of the environment and the computational resources available on the robot.

3. Literature Review

3.1 Existing Techniques for Long Sequence Handling in NLP and Computer Vision

Several techniques have been developed in the fields of Natural Language Processing (NLP) and Computer Vision to address the challenges of handling long sequences of information. These approaches can be broadly categorized into memory networks, transformers with extended context windows, and hierarchical attention mechanisms.

Memory networks, particularly Long Short-Term Memory (LSTM) networks [13], are a type of recurrent neural network designed to capture long-term dependencies in sequential data. Unlike traditional RNNs, LSTMs incorporate a memory cell and gating mechanisms (input, forget, and output gates) that control the flow of information, allowing the network to selectively retain or discard information over extended periods [13]. This explicit design for handling long-term dependencies makes LSTMs a strong candidate for enabling the robot to remember events over extended durations. More advanced memory network architectures, such as the Hierarchical Memory Transformer (HMT) [17], further enhance long-context processing by mimicking human memorization behavior through memory-augmented segment-level recurrence and a hierarchical memory structure. HMT organizes memory into sensory, short-term, and long-term levels, with a memory retrieval mechanism that can handle multiple topics in a long document [18]. The demonstrated ability of HMT to improve performance on long-context tasks with relatively fewer parameters and lower inference memory requirements [20] suggests its potential for use in resource-constrained robotic platforms.

Transformer models have revolutionized NLP, but their original architecture faces limitations in handling very long sequences due to the quadratic complexity of the self-attention mechanism [9]. This complexity arises from the need for each token in the sequence to attend to every other token. To overcome this, various techniques have been proposed to extend the context window of Transformers. Positional Interpolation (PI) is a straightforward method that scales the token positions to fit a new, longer context length [22]. Rotary Position Embeddings (RoPE) and its extensions, such as NTK-aware scaling, YaRN, and LongRoPE, aim to improve the encoding of relative positional information, allowing for context extension with reduced performance degradation [22]. Other methods like Alibi penalize attention based on the distance between tokens [22]. While these techniques show promise in extending the context window, the fundamental computational demands of processing extremely long sequences with Transformers might still pose limitations, particularly for real-time robotic applications with limited resources. Furthermore, the "lost in the middle" problem [9] remains a challenge for effectively utilizing very long Transformer contexts.

Hierarchical attention mechanisms [26] offer an alternative approach for processing long sequences by dividing the input into blocks and applying attention at different hierarchical levels, such as words, sentences, and paragraphs. This allows the model to focus on relevant parts of the input at different scales, potentially reducing the computational burden and improving efficiency [26]. Examples of architectures utilizing hierarchical attention include Hierarchical Attention Networks (HANs) [27] and Hierarchical Convolutional Attention Networks (HCANs) [29]. By processing information in a hierarchical manner, a robotic system might be able to quickly narrow down relevant timeframes or types of events before performing a more detailed search, thereby enhancing the efficiency of querying its memory.

3.2 Multimodal Approaches for Long Context Understanding

For a robot assistant operating in the real world, the ability to process and understand long sequences of multimodal data, such as text and video, is crucial [3]. This involves challenges such as aligning information across different modalities, maintaining temporal coherence over extended periods, and extracting high-level semantic meaning from the combined data streams [8]. Research in this area is exploring how to extend long context capabilities to handle multiple modalities effectively. Models like Gemini 1.5 have demonstrated the ability to reason and answer questions about long-form video and audio, showcasing use cases such as video question answering and summarizing long audio recordings [3].

Hierarchical attention mechanisms have also been investigated in the context of multimodal data processing [31]. For instance, the Hierarchical Cross Attention Model (HCAM) has been proposed for multimodal emotion recognition, effectively fusing audio and text data using a combination of recurrent and co-attention neural network models [34]. Similarly, multimodal hierarchical attention structures with word-level alignment have been developed for sentiment analysis and emotion recognition from text and audio, addressing the challenges of extracting informative features and handling time-dependent interactions between modalities [35]. These approaches suggest that processing long multimodal sequences can benefit from hierarchical structures that allow the model to focus on relevant information within each modality and across modalities at different levels of abstraction. The complexity of aligning and reasoning over long streams of text and video captured by a robot could potentially be managed more effectively through such hierarchical methods.

4. Proposed Research Directions

4.1 Continuous Information Retention (Textual Logs and Visual Data)

Developing a long context search system for a robot assistant necessitates effective methods for continuously capturing and storing the robot's operational data, encompassing both textual logs and visual data [39]. For textual information, this would involve logging robot actions, user interactions, and environmental descriptions in a structured format suitable for long-term storage and efficient retrieval. For visual data, this could include video feeds from the robot's cameras or sequences of captured images. The ReMEmbR (Retrieval-augmented Memory for Embodied Robots) project [40] provides a compelling example of how continuous visual data can be processed and stored for later querying by a robot. ReMEmbR uses video captioning to describe short segments of what the robot sees and stores these captions, along with location and time data, in a searchable vector database [40]. This approach suggests a viable method for the visual aspect of the user's request, allowing the robot to "remember" what it has seen over extended periods.

Efficient data representation formats will be crucial for managing the potentially vast volumes of textual and visual data generated over hours of robot operation. For text, using embeddings to capture the semantic meaning of words and sentences could facilitate more effective semantic search [50]. For video data, extracting key features from video frames using techniques like convolutional neural networks (CNNs) and storing these features as embeddings could provide a compact representation suitable for similarity search. Given the potential for significant storage requirements, investigating data compression techniques and strategies for selective storage of the most relevant information will be essential. Furthermore, exploring the use of episodic memory systems [42] for robots, which are designed to store and replay past experiences, could offer a valuable mechanism for long context search by allowing the robot to recall specific events or sequences of actions. Tools for visual data logging, such as AdvantageScope and Foxglove [44], could be valuable for recording and analyzing the robot's visual experiences during the data capture phase.

4.2 Efficient Querying and Retrieval (Semantic Search, Object Recognition, Temporal Reasoning)

Once the robot's operational data is continuously captured and stored, the next critical challenge is to enable efficient querying and retrieval of specific information from this long-term memory store, which contains both textual and visual data. Semantic search techniques [50] are likely to be more effective than traditional keyword-based search in this scenario. Semantic search aims to understand the meaning behind user queries and match them to relevant information in the memory, even if the exact keywords are not present [50]. The user's example query, "where did I put this tool," highlights the need for the system to understand the concept of a "tool" and potentially relate it to visual information about the object.

Integrating object recognition capabilities into the long context search system would significantly enhance its ability to answer visually grounded queries [50]. This would allow the robot to retrieve information based on the visual content of its memory, for example, by identifying the location of a specific object like a "red screwdriver" if the user asks. Furthermore, temporal reasoning [53] will be essential for answering queries that involve the timing or sequence of events, such as "when did I last use this tool?" or "what happened before I placed this object here?". The system needs to maintain and understand the temporal relationships within its long-term memory to accurately respond to such questions. The ReMEmbR project [46] demonstrates an approach to querying long-term robot memory by using an LLM agent to generate text-based, spatial, and temporal queries to a vector database. Vector databases [4] are particularly well-suited for efficient similarity search over embeddings of both textual and visual data, enabling the retrieval of semantically similar information based on user queries.

4.3 Robot Assistant Architecture Utilizing Long Context Search

To effectively implement long context search capabilities, a well-defined architectural design for the robot assistant is necessary [49]. A modular architecture, where separate components handle data acquisition, storage, indexing, querying, and response generation, could provide a flexible and scalable framework. A central "memory" module could be responsible for integrating both textual and visual information and providing a unified interface for querying. The ReMEmbR system [49] employs a two-phase architecture involving a memory building phase and a querying phase, which appears to be a logical approach. In the memory building phase, the robot continuously records and processes information from its sensors, storing it in a structured format. In the querying phase, it receives user queries and retrieves relevant information from its memory to formulate a response.

Large language models (LLMs) are likely to play a crucial role as the core reasoning engine for processing user queries and generating answers based on the retrieved long context [49]. Their ability to understand natural language and perform complex reasoning makes them well-suited for interpreting the user's intent and synthesizing information from the robot's memory. The long context search system would need to be tightly integrated with the robot's action planning and execution capabilities to enable it to not only answer questions but also potentially act based on its memory. For instance, if the user asks the robot to retrieve a tool, the system would need to locate the tool in its memory and then plan and execute the physical action of retrieving it. The concept of a coordinator and worker LLM, as explored in research on multitasking robots [57], could also be relevant, where one LLM manages high-level decisions and task prioritization, while another executes specific actions and updates the robot's memory.

5. Research Questions

This research proposal aims to address the following key questions:

1. What are the most effective data representation formats for long-term storage and efficient retrieval of continuous textual and visual data generated by a robot assistant operating over extended periods?
2. How can semantic search techniques be adapted and optimized for querying a long-term, multimodal memory store in a robotic system to retrieve information based on conceptual meaning rather than just keywords?
3. What architectural design patterns are most suitable for implementing a long context search system in a robot assistant, considering the need for continuous data capture, efficient storage and retrieval, and integration with the robot's reasoning and action capabilities?
4. How can object recognition capabilities be effectively integrated into a long context search system for a robot to enable retrieval of information based on visual content and answer queries about the location and history of specific objects?
5. What strategies are required to incorporate robust temporal reasoning into a long context search system for a robot assistant to accurately answer queries about the sequence and timing of past actions and observations?
6. How does the "lost in the middle" problem, observed in long context language models, manifest in a robotic long context search scenario involving multimodal data, and what techniques can be employed to mitigate its impact on retrieval accuracy?
7. What are the most appropriate evaluation metrics for assessing the performance of a long context search system for a robot assistant, considering both the accuracy and efficiency of retrieving textual and visual information, as well as the

overall responsiveness of the system to user queries?

8. What are the computational trade-offs associated with different approaches to long context search (e.g., very long internal context vs. retrieval-augmented generation) in the context of a resource-constrained robot platform?

These questions cover various critical aspects of developing an effective long context search system for a robot assistant, including data management, retrieval techniques, system architecture, reasoning capabilities, evaluation methodologies, and the practical constraints of robotic platforms.

6. Evaluation Plan

To assess the effectiveness of a developed long context search system for a robot assistant, a comprehensive evaluation plan is essential. This plan should include key performance indicators (KPIs) and evaluation metrics that measure both the accuracy and efficiency of the system. For textual and visual information retrieval accuracy, metrics such as recall (the fraction of relevant information retrieved), precision (the fraction of retrieved information that is relevant), F1-score (the harmonic mean of precision and recall), and mean average precision (MAP) can be used [7]. Recall measures the system's ability to find all relevant information, while precision measures its ability to avoid retrieving irrelevant information. The F1-score provides a balanced measure, and MAP is useful for evaluating the ranking of retrieved results.

Retrieval efficiency can be evaluated using metrics like query latency (the time taken to process a query and retrieve results) and throughput (the number of queries the system can handle per unit of time) [7]. These metrics are crucial for ensuring that the robot assistant can provide timely responses to user queries. If the system involves generating answers or taking actions based on the retrieved information, metrics for the quality and correctness of these outputs will also be necessary. For instance, answer correctness can be evaluated using methods that compare the generated answer to a ground truth or rely on LLMs as judges to assess relevance and factual accuracy [7].

A suitable dataset will be required for evaluating the system. This might involve collecting real-world data from a robot assistant operating in a specific environment over an extended period, capturing both textual logs and visual recordings. Alternatively, simulated data that mimics the continuous stream of information a robot would encounter could be used. The dataset should include a variety of scenarios and user queries that test different aspects of the long context search capabilities, including queries about object locations, past actions, and temporal relationships between events. Datasets like LongBench and L-Eval [60] used for evaluating long context language models could provide inspiration for creating a relevant evaluation dataset for the robotic scenario. Leveraging LLMs as judges [7] can provide a scalable and robust way to evaluate the correctness and relevance of the system's responses to user queries.

Metric	Type	Description	Calculation	Relevant Snippets
Recall	Accuracy	Fraction of relevant information retrieved.	(Number of relevant items retrieved) / (Total number of relevant items)	[7]
Precision	Accuracy	Fraction of retrieved information that is relevant.	(Number of relevant items retrieved) / (Total number of items retrieved)	[7]
F1-score	Accuracy	Harmonic mean of precision and recall.	2 * (Precision * Recall) / (Precision + Recall)	[7]
MAP	Accuracy	Mean Average Precision; evaluates the ranking of retrieved results.	Average of precision at each relevant item's rank.	[7]
Query Latency	Efficiency	Time taken to process a query and retrieve results.	Time from query submission to receiving results.	[7]
Throughput	Efficiency	Number of queries the system can handle per unit of time.	(Number of queries) / (Time taken)	[59]

Answer Correctness	Accuracy	Measures how correct the generated answer is to the ground truth.	Evaluated by comparing to ground truth or using LLM as judge.	[7]
Answer Relevance	Accuracy	Degree to which a response directly addresses and is appropriate for a query.	Evaluated by assessing the presence of redundant information or incomplete answers.	[59]

7. Conclusion and Expected Contributions

This research proposal outlines the need for and potential approaches to developing a long context search system for a robot assistant capable of remembering and reasoning over extended periods of operation. The user's vision of a robot that can recall past actions and observations, such as the location of misplaced tools, highlights a significant challenge in robotics that requires advancements in continuous information retention, efficient querying and retrieval, and suitable robot assistant architectures. The literature review indicates that techniques like memory networks (especially LSTMs and HMT), transformers with extended context windows, and hierarchical attention mechanisms offer promising avenues for addressing these challenges. Furthermore, the extension of long context capabilities to multimodal data, such as text and video, is crucial for a robot operating in real-world environments.

The proposed research directions focus on investigating methods for continuous information retention of both textual and visual data, exploring efficient querying and retrieval techniques like semantic search and object recognition, and designing a suitable robot assistant architecture that can effectively utilize long context search. The formulated research questions aim to delve into specific aspects of these directions, including data representation, retrieval optimization, system architecture, temporal reasoning, evaluation methodologies, and the mitigation of challenges like the "lost in the middle" problem. The evaluation plan emphasizes the importance of using appropriate metrics to assess both the accuracy and efficiency of the developed system.

The expected contributions of this research include:

- A novel architecture or methodology for implementing long context search in a robot assistant that can effectively handle continuous, multimodal data streams.
- Improved techniques for integrating and querying textual and visual information within a long-term robot memory system, potentially leveraging semantic search, object recognition, and temporal reasoning.
- A comprehensive evaluation of the performance of the proposed long context search system using relevant metrics and datasets.
- Valuable insights into the trade-offs between different approaches to long context search in the context of resource-constrained robotic platforms.

Ultimately, the successful development of an effective long context search system for a robot assistant would represent a significant advancement in the field of human-robot interaction. It would pave the way for more capable, helpful, and user-friendly robots that can truly learn from their experiences and provide nuanced, contextually relevant assistance to humans in a variety of environments.

Πηγές αναφοράς

1. What does a long context window mean for an AI model, like Gemini? - ZDNET, πρόσβαση Μαρτίου 23, 2025, https://www.zdnet.com/article/what-does-a-long-context-window-mean-for-an-ai-model-like-gemini/
2. What is long context and why does it matter for AI? | Google Cloud Blog, πρόσβαση Μαρτίου 23, 2025, https://cloud.google.com/transform/the-prompt-what-are-long-context-windows-and-why-do-they-matter
3. Long context | Generative AI - Google Cloud, πρόσβαση Μαρτίου 23, 2025, https://cloud.google.com/vertex-ai/generative-ai/docs/long-context
4. Long context | Gemini API | Google AI for Developers, πρόσβαση Μαρτίου 23, 2025, https://ai.google.dev/gemini-api/docs/long-context
5. How Long-Context LLMs are Challenging Traditional RAG Pipelines - Medium, πρόσβαση Μαρτίου 23, 2025, https://medium.com/@jagadeesan.ganesh/how-long-context-llms-are-challenging-traditional-rag-pipelines-93d6eb45398a
6. RAG vs Long-Context LLMs: Approaches for Real-World Applications - Prem, πρόσβαση Μαρτίου 23, 2025, https://blog.premai.io/rag-vs-long-context-llms-which-approach-excels-in-real-world-applications/
7. Long Context RAG Performance of LLMs | Databricks Blog, πρόσβαση Μαρτίου 23, 2025, https://www.databricks.com/blog/long-context-rag-performance-llms
8. What are the limitations of current multimodal AI models? - Milvus, πρόσβαση Μαρτίου 23, 2025, https://milvus.io/ai-quick-reference/what-are-the-limitations-of-current-multimo

dal-ai-models

9. De-Coded: Understanding Context Windows for Transformer Models - Medium, πρόσβαση Μαρτίου 23, 2025, https://medium.com/towards-data-science/de-coded-understanding-context-windows-for-transformer-models-cd1baca6427e
10. Long-Context LLMs Meet RAG: Overcoming Challenges for Long Inputs in RAG | OpenReview, πρόσβαση Μαρτίου 23, 2025, https://openreview.net/forum?id=oU3tpaR8fm¬eId=8X6xAgSGa2
11. Long Context Models Explained: Do We Still Need RAG?, πρόσβαση Μαρτίου 23, 2025, https://www.louisbouchard.ai/long-context-vs-rag/
12. Long Context vs. RAG for LLMs: An Evaluation and Revisits - arXiv, πρόσβαση Μαρτίου 23, 2025, https://arxiv.org/html/2501.01880v1
13. What is LSTM - Long Short Term Memory? - GeeksforGeeks, πρόσβαση Μαρτίου 23, 2025, https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/
14. What is LSTM? Introduction to Long Short-Term Memory - Analytics Vidhya, πρόσβαση Μαρτίου 23, 2025, https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/
15. Unlocking the Power of Long Short-Term Memory (LSTM) Networks - Medium, πρόσβαση Μαρτίου 23, 2025, https://medium.com/@sachinsoni600517/unlocking-the-power-of-long-short-term-memory-lstm-networks-7a02da292d7a
16. Long short-term memory - Wikipedia, πρόσβαση Μαρτίου 23, 2025, https://en.wikipedia.org/wiki/Long_short-term_memory
17. HMT: Hierarchical Memory Transformer for Long Context Language Processing - arXiv, πρόσβαση Μαρτίου 23, 2025, https://arxiv.org/html/2405.06067v1
18. OswaldHe/HMT-pytorch: [NAACL 2025] Official ... - GitHub, πρόσβαση Μαρτίου 23, 2025, https://github.com/OswaldHe/HMT-pytorch
19. HMT: Hierarchical Memory Transformer for Long Context Language Processing | AI Research Paper Details - AIModels.fyi, πρόσβαση Μαρτίου 23, 2025, https://www.aimodels.fyi/papers/arxiv/hmt-hierarchical-memory-transformer-long-context-language
20. HMT: Hierarchical Memory Transformer for Efficient Long Context Language Processing - arXiv, πρόσβαση Μαρτίου 23, 2025, https://arxiv.org/html/2405.06067v3
21. Recent Advances in In-Memory Prompting for AI: Extending Context, Memory, and Reasoning | by Jose F. Sosa - Medium, πρόσβαση Μαρτίου 23, 2025, https://medium.com/@josefsosa/recent-advances-in-in-memory-prompting-for-ai-extending-context-memory-and-reasoning-f38cff8bf7ec
22. Why and How to Achieve Longer Context Windows for LLMs ..., πρόσβαση Μαρτίου 23, 2025, https://towardsdatascience.com/why-and-how-to-achieve-longer-context-windows-for-llms-5f76f8656ea9/

23. Exploring Ways to Extend Context Length in Transformers - GitHub Pages, πρόσβαση Μαρτίου 23, 2025, https://muhtasham.github.io/blog/posts/explore-context/
24. Aren't context lengths for transformers an artificial restriction? - AI Stack Exchange, πρόσβαση Μαρτίου 23, 2025, https://ai.stackexchange.com/questions/42313/arent-context-lengths-for-transformers-an-artificial-restriction
25. Extending context length with Hugging Face's transformers | by Leanne Tan - Medium, πρόσβαση Μαρτίου 23, 2025, https://medium.com/@leannetan/extending-context-length-with-hugging-faces-transformers-6b04db05b39a
26. The big picture: Transformers for long sequences | by Lukas Nöbauer - Medium, πρόσβαση Μαρτίου 23, 2025, https://medium.com/@lukas.noebauer/the-big-picture-transformers-for-long-sequences-890cc0e7613b
27. Hierarchical Attention Neural Networks: New Approaches for Text Classification - CloudSEK, πρόσβαση Μαρτίου 23, 2025, https://www.cloudsek.com/blog/hierarchical-attention-neural-networks-beyond-the-traditional-approaches-for-text-classification
28. Advanced Attention Mechanisms for Long Sequence Transformers | by Freedom Preetham | Autonomous Agents | Medium, πρόσβαση Μαρτίου 23, 2025, https://medium.com/autonomous-agents/advanced-attention-mechanisms-for-long-sequence-transformers-6c88b2b41514
29. Hierarchical Convolutional Attention Networks for Text Classification - OSTI, πρόσβαση Μαρτίου 23, 2025, https://www.osti.gov/servlets/purl/1471854
30. Hierarchical Attention Encoder Decoder - arXiv, πρόσβαση Μαρτίου 23, 2025, https://arxiv.org/pdf/2306.01070
31. Hierarchical Self-Attention: Generalizing Neural Attention Mechanics to Hierarchy | OpenReview, πρόσβαση Μαρτίου 23, 2025, https://openreview.net/forum?id=qODJnX99hi
32. Hierarchical multiples self-attention mechanism for multi-modal analysis - ResearchGate, πρόσβαση Μαρτίου 23, 2025, https://www.researchgate.net/publication/372527161_Hierarchical_multiples_self-attention_mechanism_for_multi-modal_analysis
33. Hierarchical Attention Learning for Multimodal Classification | Request PDF - ResearchGate, πρόσβαση Μαρτίου 23, 2025, https://www.researchgate.net/publication/373414561_Hierarchical_Attention_Learning_for_Multimodal_Classification
34. HCAM - Hierarchical Cross Attention Model for Multi-modal Emotion Recognition - arXiv, πρόσβαση Μαρτίου 23, 2025, https://arxiv.org/html/2304.06910v2
35. Multimodal Affective Analysis Using Hierarchical Attention Strategy ..., πρόσβαση Μαρτίου 23, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC6261375/
36. Towards mental time travel: a hierarchical memory for reinforcement learning agents - NeurIPS, πρόσβαση Μαρτίου 23, 2025, https://proceedings.neurips.cc/paper_files/paper/2021/file/ed519dacc89b2bead3f

453b0b05a4a8b-Paper.pdf

37. HCAM-CL: A Novel Method Integrating a Hierarchical Cross-Attention Mechanism with CNN-LSTM for Hierarchical Image Classification - MDPI, πρόσβαση Μαρτίου 23, 2025, https://www.mdpi.com/2073-8994/16/9/1231
38. (PDF) HCAM-CL: A Novel Method Integrating a Hierarchical Cross-Attention Mechanism with CNN-LSTM for Hierarchical Image Classification - ResearchGate, πρόσβαση Μαρτίου 23, 2025, https://www.researchgate.net/publication/384150880_HCAM-CL_A_Novel_Method_Integrating_a_Hierarchical_Cross-Attention_Mechanism_with_CNN-LSTM_for_Hierarchical_Image_Classification
39. Automated Data Capture: Here's All You Need to Know - Docsumo, πρόσβαση Μαρτίου 23, 2025, https://www.docsumo.com/blog/automated-data-capture-detailed-guide
40. ReMEmbR: Building and Reasoning Over Long-Horizon Spatio-Temporal Memory for Robot Navigation | PromptLayer, πρόσβαση Μαρτίου 23, 2025, https://www.promptlayer.com/research-papers/remembr-building-and-reasoning-over-long-horizon-spatio-temporal-memory-for-robot-navigation
41. DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation, πρόσβαση Μαρτίου 23, 2025, https://arxiv.org/html/2403.07788v1
42. Task-Unaware Lifelong Robot Learning with Retrieval-based Weighted Local Adaptation, πρόσβαση Μαρτίου 23, 2025, https://arxiv.org/html/2410.02995v2
43. Data Logging | Virtual Robotics Toolkit, πρόσβαση Μαρτίου 23, 2025, https://www.virtualroboticstoolkit.com/documentation/sections/7/articles/117
44. AdvantageScope — FIRST Robotics Competition documentation - WPILib Docs, πρόσβαση Μαρτίου 23, 2025, https://docs.wpilib.org/en/stable/docs/software/dashboards/advantagescope.html
45. Foxglove - Visualization and observability for robotics developers., πρόσβαση Μαρτίου 23, 2025, https://foxglove.dev/
46. A Trip to ReMEmbR - Hackster.io, πρόσβαση Μαρτίου 23, 2025, https://www.hackster.io/news/a-trip-to-remembr-c58f284ff58d
47. ReMEmbR: Building and Reasoning Over Long-Horizon Spatio-Temporal Memory for Robot Navigation - arXiv, πρόσβαση Μαρτίου 23, 2025, https://arxiv.org/html/2409.13682v1
48. ReMEmbR shows how generative AI can help robots reason and act, says NVIDIA, πρόσβαση Μαρτίου 23, 2025, https://www.therobotreport.com/remembr-generative-ai-enables-robots-reason-act-says-nvidia/
49. Using Generative AI to Enable Robots to Reason and Act with ReMEmbR | NVIDIA Technical Blog, πρόσβαση Μαρτίου 23, 2025, https://developer.nvidia.com/blog/using-generative-ai-to-enable-robots-to-reason-and-act-with-remembr/
50. Semantic Video Search: Workflow, Examples & Use Cases - FastPix, πρόσβαση Μαρτίου 23, 2025, https://www.fastpix.io/blog/search-videos-semantically-workflow-and-applications

51. Video semantic search with AI on AWS | AWS for M&E Blog, πρόσβαση Μαρτίου 23, 2025, https://aws.amazon.com/blogs/media/video-semantic-search-with-ai-on-aws/
52. Semantic Video Search - Multimodal Makers | Mixpeek, πρόσβαση Μαρτίου 23, 2025, https://blog.mixpeek.com/what-is-semantic-video-search/
53. Enhancing Math Understanding with Spatial-Temporal Models: A Visual Learning Approach, πρόσβαση Μαρτίου 23, 2025, https://blog.mindresearch.org/blog/enhancing-math-understanding-with-spatial-temporal-models-a-visual-learning-approach
54. Spatial-Temporal Reasoning: The Importance of Touch - Blog, πρόσβαση Μαρτίου 23, 2025, https://blog.mindresearch.org/blog/bid/295457/spatial-temporal-reasoning-the-importance-of-touch
55. Differential temporal dynamics during visual imagery and perception - PMC, πρόσβαση Μαρτίου 23, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC5973830/
56. Understanding Autonomous Agent Architecture - SmythOS, πρόσβαση Μαρτίου 23, 2025, https://smythos.com/ai-agents/agent-architectures/autonomous-agent-architecture/
57. Robots Can Multitask Too: Integrating a Memory Architecture and LLMs for Enhanced Cross-Task Robot Action Generation | PromptLayer, πρόσβαση Μαρτίου 23, 2025, https://www.promptlayer.com/research-papers/robots-can-multitask-too-integrating-a-memory-architecture-and-llms-for-enhanced-cross-task-robot-action-generation
58. Long Context RAG Performance of LLMs - YouTube, πρόσβαση Μαρτίου 23, 2025, https://www.youtube.com/watch?v=jheFz4kL07o
59. A list of metrics for evaluating LLM-generated content - Microsoft Learn, πρόσβαση Μαρτίου 23, 2025, https://learn.microsoft.com/en-us/ai/playbook/technology-guidance/generative-ai/working-with-llms/evaluation/list-of-eval-metrics
60. Long Context Evaluation Guidance - OpenCompass' documentation! - Read the Docs, πρόσβαση Μαρτίου 23, 2025, https://opencompass.readthedocs.io/en/latest/advanced_guides/longeval.html