Flip Flops in Neural Models

Vineeth Bhat*
IIIT Hyderabad

Mitansh Kayathwal*
IIIT Hyderabad

Abstract

This paper critically evaluates Flip-Flop Neuron (FFN) architectures, comparing their performance and capabilities with Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs). While FFNs require fewer parameters and resist interference, they do not consistently outperform LSTMs or GRUs across tasks. We explore FFN's theoretical basis, linking it to human brain mechanisms, and highlight its potential for improving memory retention and cognitive efficiency. Additionally, we propose new benchmarks to rigorously assess FFN memory performance in cognitive tasks and take initial steps toward understanding its internal dynamics through mechanistic analysis.

1 Introduction

The Flip-Flop Neuron (FFN) architecture (Kumari et al., 2023; Kumar et al., 2021), inspired by bi-stable neurons in the human brain, offers a novel approach to modeling memory and cognition with minimal parameters. Proponents claim FFNs outperform traditional recurrent networks in tasks like sentiment analysis and require less training time than LSTMs, making them ideal for rapid learning and adaptability. By mimicking human neuronal principles, FFNs aim to enhance performance in complex sequence processing and decision-making.

We conduct a detailed replication[1] of the tasks outlined in (Kumari et al., 2023), including signal generation, sentiment analysis, and video frame prediction, while expanding comparisons to include GRUs and RNNs. Our experiments reveal that FFNs can match the performance of larger models while utilizing fewer parameters. However, in certain cases, RNNs also achieve comparable performance with even fewer parameters, challenging the claim of FFN superiority in all contexts.

To deepen our theoretical understanding of FFNs, we extend their analysis beyond the bi-stable neuron analogy (Kumar et al., 2021) by exploring a probabilistic interpretation that highlights analogies between biological neural systems and recurrent models. Additionally, we analyze a bi-directional perspective, demonstrating that neurons can model flip-flop circuits and not just the reverse, broadening the theoretical foundation of FFNs.

Finally, we introduce novel tasks, such as those focused on memory retention, to benchmark FFN performance in previously unexplored contexts. Complementing this, we analyze FFN internals to gain insights into their mechanistic interpretability and operational dynamics.

2 Literature Review

A significant body of research has explored the integration of cognitive insights into machine learning models to enhance their performance, especially in data-scarce environments. Below, we review some of the most notable contributions in this domain.

LSNB and eLSNB Models. The LSNB and eLSNB models embed human cognitive biases into their algorithms, enabling effective generalization from small, noisy, and biased samples (Taniguchi and Others, 2019). This approach makes them well-suited for tasks with limited or costly data, showcasing the value of cognitive principles in data-scarce machine learning.

PPE-Enhanced GBDTs. PPE-Enhanced Gradient-Boosted Decision Trees (Sense and Others, 2022) leverage principles derived from cognitive models of human memory to enhance predictive accuracy, particularly in scenarios with limited data availability. By incorporating memory-inspired mechanisms, these models optimize feature importance and improve generalization, allowing them to perform well despite data

* Equal Contribution (VB: 2021101103, MK: 2021101026)

[1] github.com/flip-flops-in-neural-models/Flip-Flop-Neurons-INCM

scarcity.

Generative Models. Generative models have been increasingly applied to cognitive decision-making tasks (Malloy and Gonzalez, 2024; Wang, 2023), where they combine the strengths of generative AI and cognitive theories to improve memory representation and predict participant behavior. These models are designed to simulate human decision-making processes by representing and retrieving information in ways analogous to human memory systems.

The aforementioned models leverage human cognition but focus on small datasets, lacking recurrent architectures for modeling sequential dependencies and temporal dynamics. They emphasize short-term memory and localized decision-making, with limited support for long-term memory retention. In contrast, recurrent FFNs (Kumari et al., 2023; Kumar et al., 2021) emphasize long-term memory preservation. However, even within this paradigm, the focus often remains narrow of primarily addressing the preservation of long-term memory states without analyzing into broader cognitive functionalities or dynamic adaptability.

3 Replication and Analysis

We replicated the experiments from (Kumari et al., 2023) with a PyTorch-based implementation, ensuring reproducibility through a comprehensive codebase, detailed instructions, and fixed random seed settings for consistent results. The FFN architecture was optimized in PyTorch for better efficiency and performance.

Our work extended the original study by including RNNs and GRUs for comparative analysis, providing a broader context for FFN performance among recurrent models. Below, we summarize key experiment results and analyze their relevance to cognitive tasks, enhancing the understanding of FFN performance in cognitive modeling.

3.1 Signal Generation

Signal generation involves synthesizing data that mimics natural stimuli to study cognitive processes and model perceptual systems. It reflects the brain's ability to process and predict temporal patterns, aiding in the assessment of models' pattern recognition and predictive reasoning capabilities, crucial for memory and decision-making.

Following (Kumari et al., 2023), we test models on recreating fixed-length signals from a single input. Experiments with sinusoidal signals (Figure 1) show FFNs perform comparably to other recurrent models, though simple RNNs achieve similar results with fewer parameters. For uniform signals (Figure 2), none of the models achieve satisfactory results, but FFNs exhibit less noisy loss curves than other architectures.

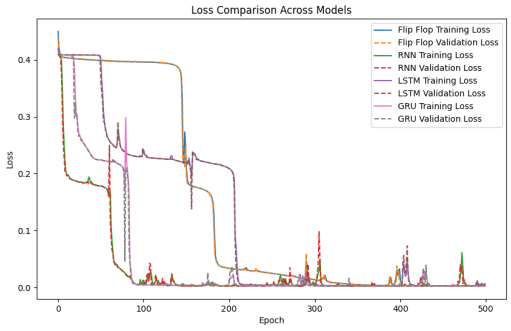

Figure 1: *Signal Generation Results on Sinusoidal Data*. Trained for 500 epochs using a learning rate of 0.001 with 2 sinusoidal classes with 64 signals each. Experiment path: `src/experiments/signal_gen/sinusoidal_0`

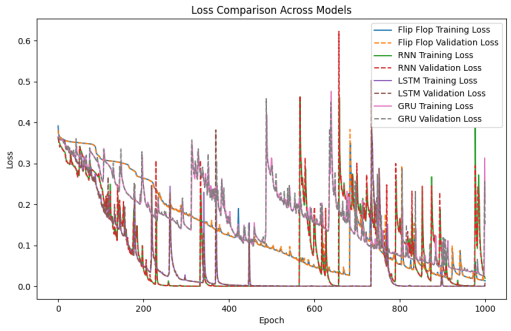

Figure 2: *Signal Generation Results on Uniform Data*. Trained for 1000 epochs using a learning rate of 0.0005 with 2 uniform classes with 64 signals each. Experiment path: `src/experiments/signal_gen/uniform_1`

3.2 Sentiment Analysis

Sentiment analysis models how language conveys emotions and influences decisions by classifying text as positive, negative, or neutral, mimicking human emotional comprehension. Using the IMDB dataset (Maas and Others, 2011), we examine how models emulate cognitive processes like emotional regulation and inference.

Figures 3 and 4 show that FFNs match the performance of larger recurrent models and outperform smaller models like RNNs, consistent with prior studies.

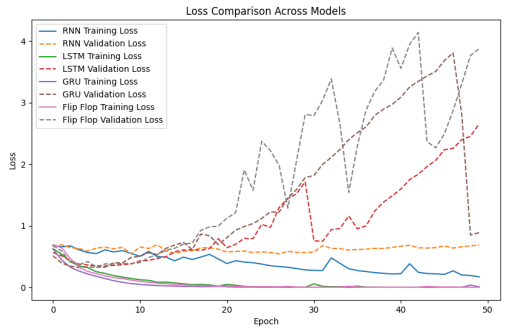

Figure 3: *Sentiment Analysis Loss Curves*. Trained for 50 epochs using a learning rate of 0.001 with a vocabulary size of 20000 and maximum length 200. Experiment path: `src/experiments/sent_analysis/original...`

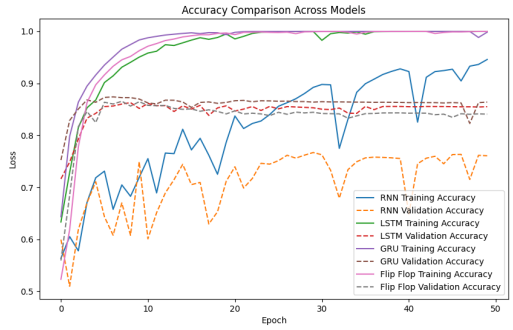

Figure 4: *Sentiment Analysis Loss Curves*. Parameters remain the same as figure 3.

3.3 Action Recognition

Action recognition simulates cognitive processes of visual perception and motion understanding, crucial for identifying activities in real-world scenarios. Models classify actions from visual input by leveraging sequential video frames, mimicking human recognition of actions like walking or running. We use the UCF11 dataset (Liu and Others, 2009), following the original study's approach.

Our results, in figure 5, contradict the claims of (Kumari et al., 2023) by showing that the LSTM architecture outperforms the FFN's convultional variant.

3.4 Remarks

Our replication of (Kumari et al., 2023) compares FFN performance across tasks with established recurrent architectures. FFNs show competitive results in signal generation but are matched by simpler RNNs with fewer parameters, while LSTMs outperform FFNs in action recognition. Additional ablations are included in our released codebase, focusing here on novel analyses and their implications for cognitive modeling.

4 Theoretical Grounding

4.1 Sequential Decision Models and the Brain

The exploration of Sequential Decision Models in relation to brain function provides a fascinating insight into how cognitive processes can be mirrored in computational frameworks. These models are predicated on the understanding that the brain operates as a probabilistic system (Rao and Others, 2002; Pouget and Others, 2013), continuously updating its beliefs based on incoming sensory information (Mazziotta and Others, 2001). This dynamic nature of neural computation is reflected in Long Short-Term Memory (LSTM) networks, which adapt their internal states to refine predictions over time as new data is received.

In essence, the brain encodes sensory information using probabilistic representations, enabling it to make decisions under uncertainty. This mirrors how LSTMs leverage historical context to inform current outputs (Hashemzadeh and Others, 2020; Jain and Huth, 2018), thereby managing uncertainty and enhancing decision-making efficiency. The parallels drawn between neural circuits and LSTM architectures suggest that both systems utilize a form of probabilistic inference, allowing for nuanced processing of information and adaptive learning capabilities.

4.2 From Neurons to Flip Flops

FFNs are inspired by the bi-stable nature of neurons in the human brain, particularly in the prefrontal cortex (Kumar et al., 2021), drawing parallels to flip-flops in electronic circuits by maintaining states through excitation and inhibition.

Neural Flip-Flops (NFFs) (Yoder, 2024) bridge biological neural functions and electronic logic circuits by mimicking neuron dynamics, enabling the modeling of cognitive processes like memory retention and oscillatory activity. NFFs operate by modulating neuron activity levels rather than altering structural networks, reflecting the adaptive nature of biological systems. This stability supports tasks requiring sustained attention and memory recall, akin to flip-flops.

A notable insight is the capability of individual neurons to serve as basic logic primitives due to their intrinsic properties of excitation and inhibition. The memory mechanisms of NFFs have

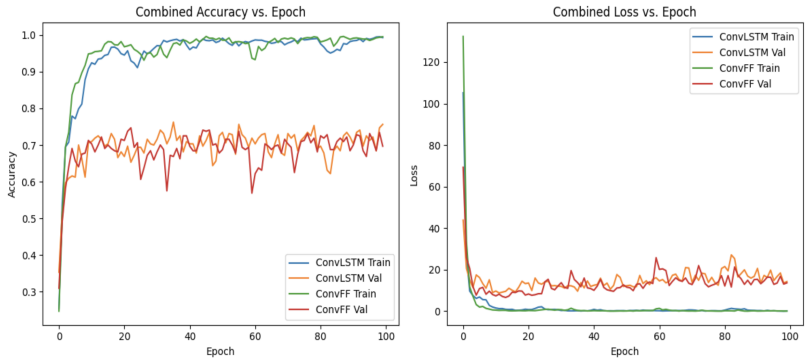

Figure 5: *Results on Action Recognition Task*. Experiment Path: `src/incm-action-recognition.ipynb`

been shown to exhibit phenomena associated with short-term memory, with research suggesting that these networks can store and retrieve information through specific firing patterns, paralleling biological short-term memory systems (Jarne, 2024; Yoder, 2024).

In conclusion, the exploration of Sequential Decision Models and the development of Neural Flip-Flops illustrate the synergy between cognitive neuroscience and computational modeling. Drawing parallels between brain function and LSTM networks reveals how complex decision-making can be modeled, while NFFs highlight the potential of biologically inspired architectures to mimic cognitive functions.

5 Novel Experiments

We describe novel experiments on the FFN below.

5.1 Memory Retention

We evaluate the FFN architecture on memory retention to explore how models mimic the way neurons retain information over time. This is important as biological systems can forget due to interference, unlike machine learning models, which do not naturally forget. By introducing interference effects, we assess how the FFN handles memory decay and the ability to retain or overwrite information.

This experiment follows a signal generation approach, where recurrent architectures learn a target signal while facing interference from lower-magnitude signals. The number of signals and degree of interference are varied to assess how the FFN manages memory retention, decay, and information overwriting.

5.2 Working Memory

We analyze the flip-flop architecture's ability to replicate human short-term memory by evaluating the recurrent model's working memory capacity. This involves varying the length and complexity of uniform random input sequences to assess how well the model retains and manipulates information over short durations. This experiment offers insights into the model's ability to maintain relevant data, manage cognitive load, and adapt to input complexity, aligning with human working memory characteristics.

5.3 Biologically Inspired Learning Rules

To enhance the FFN model's learning, we incorporated biologically inspired rules like Hebbian learning and Spike-Timing-Dependent Plasticity (STDP) (Markram and Others, 1997). Hebbian learning, based on the principle that neurons firing together strengthen their connection, helps encode associative patterns vital for learning and memory. STDP, a temporal rule, adjusts connection strength based on spike timing; a pre-synaptic spike before a post-synaptic one strengthens the connection (potentiation), while the reverse weakens it (depression).

We hypothesize that integrating Hebbian learning and STDP into the FFN model would create adaptive memory states, enabling the network to retain sequential patterns more reliably, even in noisy conditions. We modified the architecture to include Hebbian learning, aligning input and output activations to reinforce associations, and simulated random spikes for STDP to adjust connection strengths based on timing.

5.4 Model Internals

To gain insights into the FFN model's internal workings, we visualize the dynamics of the J and K gates across training epochs to understand their evolution and adaptation. We also examine activation distributions to see how information is represented and propagated, focusing on variations in gate activations.

Additionally, we compare output trajectories over multiple timesteps of signal prediction to observe the model's coherence and stability. Finally, we analyze gate activations and hidden state trajectories in FFN models versus LSTMs to identify similarities and differences in processing sequential data, and assess output-target cross-correlation to measure the alignment between model outputs and expected targets.

5.5 Other Experiments

We also analyze noise robustness and task parametrization. The results of these experiments are available in our released repository.

6 Results and Discussion

Building on the results presented in section 3, we present the results of our novel tasks in this section.

6.1 Memory Retention

The results of this experiment are shown in Figure 6. Based on 11 distinct trials, we found that Flip-Flop neurons demonstrated greater robustness to interference effects compared to other recurrent network architectures. Specifically, the loss curves for the FFN model exhibited a more stable and uniform decrease, indicating consistent performance and the ability to maintain information over longer periods. This aligns with cognitive theories suggesting that the brain's memory system, particularly in regions such as the prefrontal cortex, relies on mechanisms that can adaptively suppress interference to preserve relevant information.

In contrast, the loss curves for other recurrent networks showed higher variability, with instances of sudden increases, which could imply less effective mechanisms for mitigating interference and managing the decay of working memory. This comparison suggests that Flip-Flop neurons are better suited for simulating aspects of human-like memory retention, which requires the ability to handle competing information while maintaining focus on the task at hand.

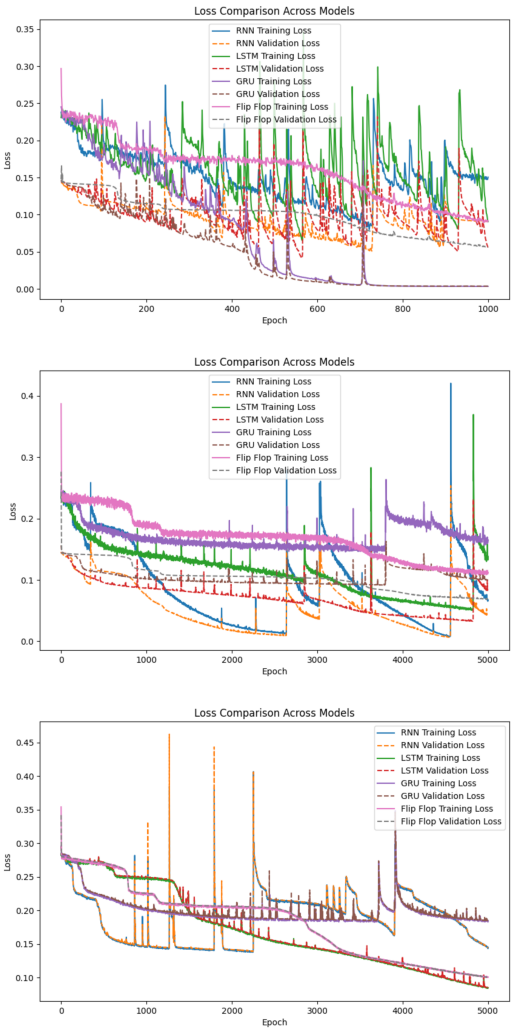

Figure 6: *Memory Retention Experiments*. Loss curves over three selected ablations of the memory retention task with varying percentage of interference effects (>10% and < 30%). Experiments path: `src/experiments/mem_retention`

6.2 Working Memory

Results from 9 experiments led to the conclusion that Flip-Flop neurons exhibit slower learning compared to other models (figure 7). However, their loss curves show significantly less fluctuation, suggesting better pattern recognition and improved generalization with the same input batches. This stability indicates that Flip-Flop neurons are effectively capturing and retaining sequential patterns over time, which is reminiscent of cognitive processes in the brain associated with long-term memory consolidation and the ability to discern patterns in complex information. The less variable learning curves suggest that the model might be better at avoiding overfitting to noise and instead focusing on the underlying structure of the input, similar to

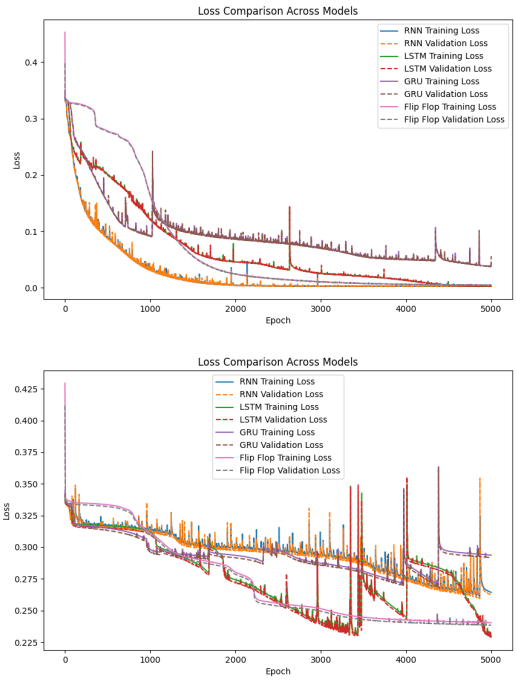

Figure 7: *Working Memory Experiments*. Loss curves over two selected ablations of the working memory task with varying signal lengths and noise. The curves for the FFN are less noisy possibly indicating better generalization. Experiments path: `src/experiments/working_mem`

how the brain generalizes knowledge.

6.3 Biologically Inspired Learning Rules

Integrating biologically inspired learning rules into the Flip-Flop neuron model did not yield significant results, as shown in Figure 8. This suggests that certain biological mechanisms may not be directly applicable within conventional machine learning (ML) architectures without significant adaptation.

Biological systems excel in learning and memory due to complex neural dynamics and contextual interactions, including metabolic and neuromodulatory influences, which ML models lack. We hypothesize that the simplified nature of artificial neurons, with static weight updates and mathematical functions, cannot replicate real-time adaptation and nuanced learning. Additionally, the use of spike-based simulations may have influenced the results.

6.4 Model Internals

Visualizing Weights over Epochs. Figure 9 illustrates how the weights evolve over multiple epochs. We observe rapid convergence initially, followed by minimal fluctuations. This aligns with our previous findings on working memory, suggesting that

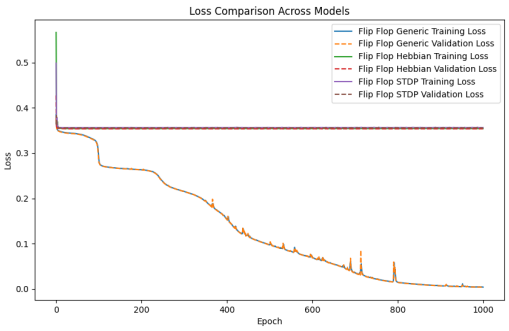

Figure 8: *Adding Hebbian and STDP*. We observe that adding Hebbian and STDP as mechanisms provide diminished returns. Experiment path: `src/experiments/learning_rules/exp_2`

the model generalizes more effectively, leading to a smoother and lower loss curve compared to other recurrent models, which exhibit steeper and noisier loss patterns. After more epochs, the weights maintain a consistent general pattern, becoming more distinct relative to each other and preserving a clear distribution post-convergence.

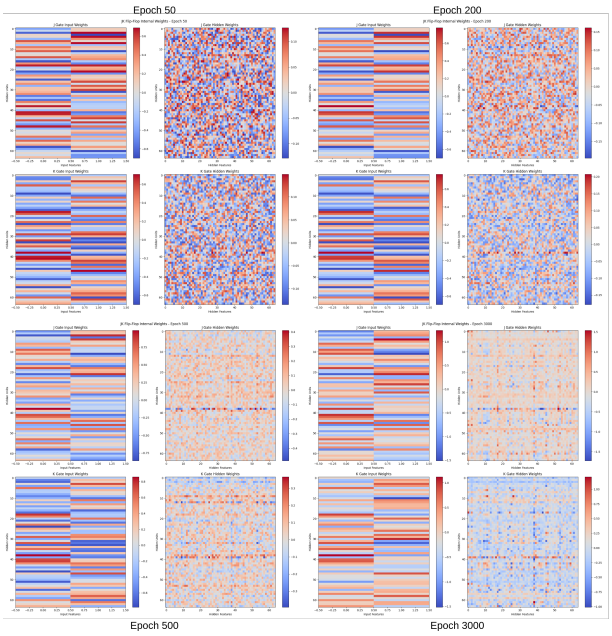

Figure 9: *Visualizing JK weights in the FFN over multiple epochs*. We observe the weights of the FCNs between hidden and input tensor interactions over multiple epochs. Experiments Path: `src/experiments/visualization_data_task`

Evolution of Gate Activations. The evolution of gate activations, in figure 10, supports our analysis on the evolution of gate weights.

Analysis of Output Trajectories. Figure 11 compares the output trajectories of FFNs and LSTMs.

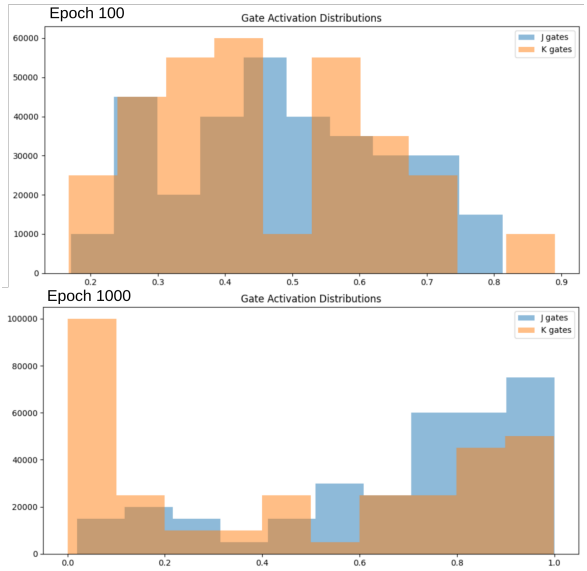

Figure 10: *JK Gate Activations in the FFN*. Experiments Path: `src/experiments/visualization_data_task`

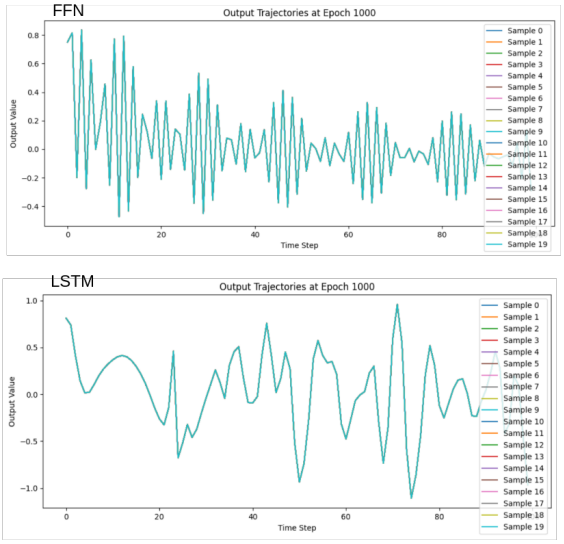

Figure 11: *Comparison of output trajectories between FFN and LSMT*. Experiments Path: `src/experiments/visualization_data_task`

Both models effectively filter out noise, ensuring that samples from the same class produce overlapping signals. However, FFNs demonstrate the ability to model more complex patterns, achieving a nuanced representation of the uniform data, as evidenced by the richer structure in the output graph.

Average Activation Comparison. As shown in Figure 12, activations in the FFN are more nuanced and broadly representative, whereas LSTM activations are dominated by a smaller subset of parameters. This suggests that Flip-Flop neurons support more holistic modeling, potentially explaining their superior robustness to interference effects.

Output-Target Cross Correlation. The correlation, shown in figure 13, supports the argument from previous experiments about the FFN having more nuanced representations.

7 Conclusion

Our work extends prior studies on Flip-Flop Neurons (FFNs) by evaluating their performance across various tasks and introducing novel, cognition-aligned experiments. FFNs show robustness in signal generation and memory retention under interference, while simple RNNs achieve similar results with fewer parameters, and LSTMs excel in tasks like action recognition, highlighting the importance of task-specific architecture selection.

Our experiments on memory retention and working memory demonstrate FFNs' potential to mimic human cognitive functions, such as handling interference and maintaining information over time. Visualizing model internals and comparing FFNs with LSTMs improved our understanding of sequential data processing.

We also provided a theoretical framework positioning FFNs within neural computation and probabilistic decision-making, drawing parallels to biological systems. This work highlights the relevance of FFNs for modeling cognitive processes and bridging neuroscience and machine learning.

Future Work

Mechanistic Interpretability. A deeper analysis to identify neuron activation patterns and their impact on task performance.
Image Task Internals. Exploring how FFNs encode spatial and temporal features in visual data through activation visualization and feature map analysis.
Longitudinal Studies. Assessing FFNs over extended periods and different training conditions to understand their adaptability and stability in real-world, continuous learning scenarios.

8 Addendum

Acknowledgments

The authors are grateful for the opportunity to work on this project as a part of a course on neural and cognitive modeling taken by Prof. Bapi Raju S.

Large Language Model Usage

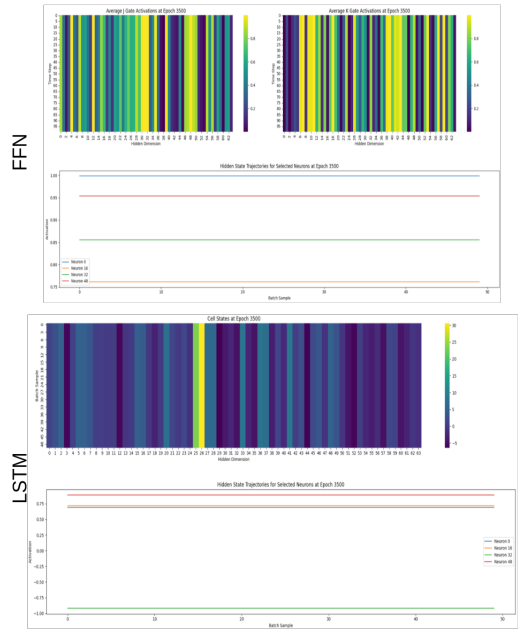

Figure 12: *Comparison of average activations between FFN and LSTM*. Activations in the FFN are more nuanced and representative while the ones in the LTSM are dominated by a few parameters. Experiments Path: `src/experiments/visualization_data_task`

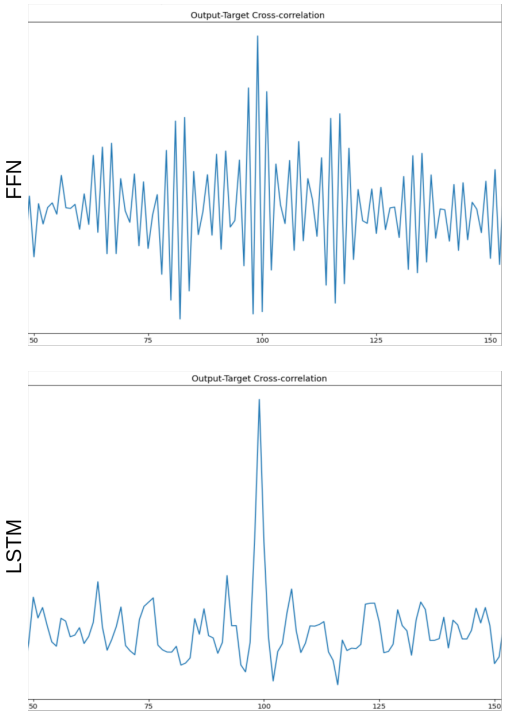

Figure 13: *Comparison of output-target cross-correlation between FFN and LSTM*.

In our implementation, we utilized `Claude-3.5-Sonnet` for aiding us in development and `GPT-4o-mini` for refining this paper, including grammatical corrections and rephrasing.

References

Maryam Hashemzadeh and Others. 2020. From language to language-ish: How brain-like is an lstm's representation of nonsensical language stimuli? *arXiv:2010.07435*.

Shailee Jain and Alexander Huth. 2018. Incorporating context into language encoding models for fmri. *Advances in NIPS*.

Cecilia Jarne. 2024. Exploring flip flop memories and beyond: training recurrent neural networks with key insights. *Frontiers in Systems Neuroscience*.

S Kumar, C Vigneswaran, and V Srinivasa Chakravarthy. 2021. Flip flop neural networks: Modelling memory for efficient forecasting. In *Proceedings of MDCWC 2020*.

Sweta Kumari, Vigneswaran Chandrasekaran, and V Srinivasa Chakravarthy. 2023. The flip-flop neuron: a memory efficient alternative for solving challenging sequence processing and decision-making problems. *Neural Computing and Applications*.

Jingen Liu and Others. 2009. Recognizing realistic actions from videos "in the wild". In *CVPR*, pages 1996–2003. IEEE.

Andrew Maas and Others. 2011. Learning word vectors for sentiment analysis. In *ACL HLT*.

Tyler Malloy and Cleotilde Gonzalez. 2024. Applying generative artificial intelligence to cognitive models of decision making. *Frontiers in Psychology*.

Henry Markram and Others. 1997. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*.

John Mazziotta and Others. 2001. A four-dimensional probabilistic atlas of the human brain. *Journal of the American Medical Informatics Association*.

Alexandre Pouget and Others. 2013. Probabilistic brains: knowns and unknowns. *Nature Neuroscience*.

Rajesh PN Rao and Others. 2002. *Probabilistic models of the brain: Perception and neural function*. MIT press.

Florian Sense and Others. 2022. Cognition-enhanced machine learning for better predictions with limited data. *Topics in cognitive science*.

Hidetaka Taniguchi and Others. 2019. Implementation of human cognitive bias on neural network and its application to breast cancer diagnosis. *SICE JCMSI*.

Shu-Qiang Others Wang. 2023. Generative ai for brain imaging and brain network construction.

Lane Yoder. 2024. Neural flip-flops i: Short-term memory. *Plos one*.