

Alex Graves - Neural Turing Machines (2014)

Tom Rochette <tom.rochette@coreteks.org>

August 30, 2025 — [861fb9d0](#)

0.1 Context

0.2 Learned in this study

0.3 Things to explore

1 Overview

1.1 1 Introduction

- Computer programs make use of three fundamental mechanisms: elementary operations (e.g., arithmetic operations), logical flow control (branching), and external memory, which can be written to and read from in the course of computation
- Modern machine learning has largely neglected the use of logical flow control and external memory

1.2 2 Foundational Research

1.3 2.1 Psychology and Neuroscience

- A “central executive” focuses attention and performs operations on data in a memory buffer
- The limitations of working memory is often quantified by the number of “chunks” of information that can be readily recalled

1.4 2.2 Cognitive Science and Linguistics

- In the early 1980s, both fields considered recursive or procedural (rule-based) symbol-processing to be the highest mark of cognition
- The Parallel Distributed Processing (PDP) or connectionist revolution cast aside the symbol-processing metaphor in favour of a so-called “sub-symbolic” description of thought processes
- Fodor and Pylyshyn claimed that connectionist theories were incapable of variable-binding, or the assignment of a particular datum to a particular slot in a data structure
- They also claimed that neural networks with fixed-length input domains could not reproduce human capabilities in tasks that involve processing variable-length structures

1.5 2.3 Recurrent Neural Networks

- A broad class of machines with dynamic state
- Their state depends on both the input to the system and on the current state
- A crucial innovation to recurrent networks was the Long Short-Term Memory (LSTM)
- This architecture was developed to address the “vanishing and exploding gradient” problem, which we might relabel as “vanishing and exploding sensitivity”
- LSTM ameliorates the problem by embedding perfect integrators for memory storage in the network

- With a mechanism that allows an enclosing network to choose when the integrator listens to inputs, we can selectively store information for an indefinite length of time

1.6 3 Neural Turing Machines

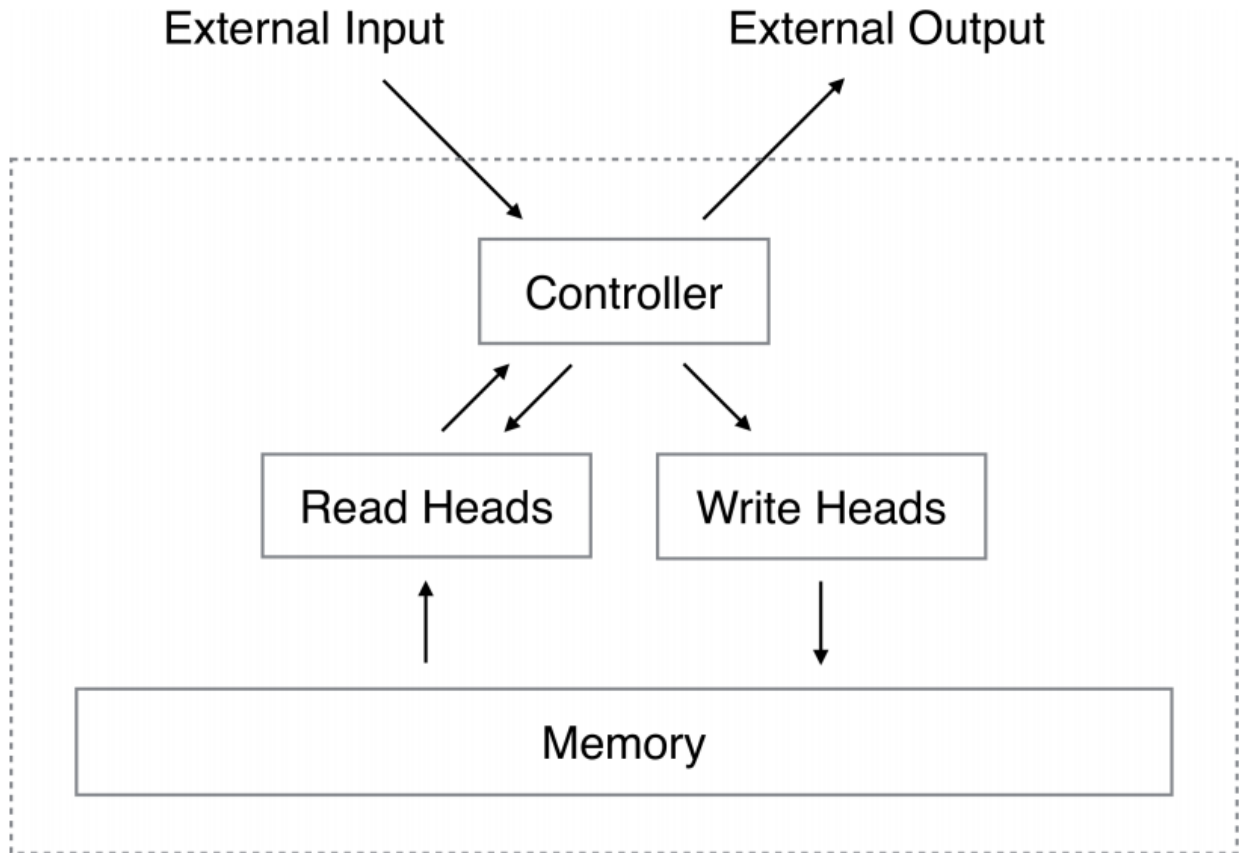


Figure 1: Neural Turing Machine

- Contains two basic components
 - a neural network controller
 - a memory bank
- The controller interacts with the external world via input and output vectors
- The controller also interacts with a memory matrix using selective read and write operations
- Every component of the architecture is differentiable, making it straightforward to train with gradient descent
- “Blurry” read and write operations that interact to a greater or lesser degree with all the elements in memory
 - The degree of blurriness is determined by an attentional “focus” mechanism that constraints each read and write operation to interact with a small portion of the memory, while ignoring the rest

2 See also

3 References

- [arXiv:1410.5401v2](#) [cs.NE]