Slide 1 – Title

Thanks for giving me this opportunity to tell you about my proposed Pioneer Project.

Slide 2

I want to explore higher cognitive function in brains by asking this question:

How do brains and large language models integrate sensory inputs, remember, plan and act over hours?

Slide 3

Let's look into the brain of a monkey hunting for food in the wild.

Each of these circles is a brain area involved in finding and capturing food.

There is self-generated planning activity throughout the brain while the monkey searches for food.

A prey appears.

A rapid sequence of activity up the visual cortical hierarchy identifies the prey.

A decision is made to reach out and catch the prey, all in less than a second.

The neurons in this sensorimotor chain have been well studied.

The brain then returns to self-generated activity on what to do next.

What happens during planning and thinking is more difficult to study than the fast sensorimotor response and is poorly understood.

This self-generated activity is what I want to understand.

Slide 4

Memory mechanisms in brains span from seconds to your lifetime.

The mechanisms for fast signaling at synapses lasting a few seconds are well understood.

Short–term working memory of items like a telephone number is maintained by rehearsal on a time scale of minutes.

Long-term working memory is what happens when you go to a talk and listen to word after word and see slide after slide, and you can remember enough to ask a question an hour later.

This is different from long-term memory because it is only temporary, and the next day, you only remember the gist of what you heard and a few salient details that you paid attention to.

Slide 5

It's generally believed that memories are formed by changing the strengths of synapses.

There are a million billion synapses in your brain.

Many genes associated with mental disorders affect synaptic signaling.

This diagram shows the range of time scales for synaptic mechanisms on a log scale.

Fast synaptic signaling lasts milliseconds, and short-term synaptic plasticity lasts seconds.

Short-term working memory can be sustained by circulating electrical activity for minutes during rehearsal.

Long-term changes in the structures of synapses can last a lifetime.

Many cortical synapses are labile and turn over every day.

What neural mechanism can fill the gap between minutes and hours needed for long-term working memory?

Slide 6

ChatGPT has a long-term working memory that lasts as long as you are engaged in a dialog.

I was curious about how this was accomplished.

As you input your query into ChatGPT, the output words are looped back to the input.

The long input vector keeps getting longer and longer during your dialog.

ChatGPT transforms temporal sequences of words into spatial sequences of words.

Is there any evidence that brains can do something similar?

Slide 7

The earliest electrical signals recorded from the scalp were 10 Hz brain oscillations.

Since then, direct recordings from the cortex have revealed oscillations in frequency bands from 4 Hz to 80 Hz.

Here is what an oscillation looks like in the gamma band.

Power is maximum at 40 Hz and lasts for a few hundred milliseconds.

This burst of spikes is triggered in visual cortex by attending to an object.

Power in the gamma band is reduced in schizophrenia patients and may be related to their thought disorder.

Slide 8

However, this oscillation is not synchronous.

It’s a slow wave that travels across the cortex.

Here are recordings at three sites in the cortex separated by half a mm with offset peaks.

Like the transformer, the traveling wave converts temporal sequences into long spatial sequences.

However, for long-term working memory, we need a neural mechanism that can rapidly change the strength of synapses.

Slide 9

A candidate is **spike-timing dependent plasticity** -- **STDP** that occurs when a presynaptic input is paired within 10 ms of a postsynaptic spike, as shown here.

Repeated pairing by bursts at 10-40 Hz can increase or decrease synaptic strength by a factor of 2 depending on the order of the synaptic input and output spike.

The change in strength is rapid and lasts for around 20 minutes after a few bursts and a few hours after 5-10 bursts.

The wavefronts of traveling waves, which are constantly generated in the cortex, provide precisely timed bursts of inputs needed to induce STDP.

Slide 10

Here is my Working Hypothesis: Sequences of inputs are encoded by traveling waves and temporarily stored for hours by STDP.

Which synapses are involved? Perhaps these are the small, labile synapses that account for 90% of all excitatory synapses onto pyramidal neurons in the cortex.

If so, the potential capacity for long-term working memory is vast – most cortical synapses.

Slide 11

This slide demonstrates a simple ring model that supports sparse traveling waves like the cortex.

The spiking neurons are connected along the ring with the same Mexican hat connectivity found in the cortex: local reciprocal excitatory connections and more distant inhibitory influence.

The video shows the input current to each model neuron along the ring.

The waves go in both directions and can pass through each other because only a small fraction of model neurons are active when a wave passes, as observed in the cortex.

DALL-E illustrates a 2D cortical network in which only a handful of glowing blue neurons are active.

Our ring model has 100,000 spiking neurons, about the number beneath a square mm of cortex.

This project aims to simulate long-term working memory and replicate working memory tasks at a biological scale.

A mouse has 3 square cm’s of cortex with 30 million neurons, thousands of times larger than our current model.

The amount of computation needed far exceeds the budget of an R21.

I will collaborate with experimental labs to test the predictions of our model.

Slide 12

The Global Workspace is a concept in cognitive psychology that assumes information is shared across cortical regions during cognition.

My conceptual framework for long-term working memory has the long-term scale and high capacity needed to support information sharing across the cortex.

Can ChatGPT think?

When your dialog with ChatGPT ends, activity stops, and there is no internal monologue.

Hence, there is no self-generated activity, which is the hallmark of thinking.

Fast weight changes like STDP in the cortex could be implemented in ChatGPT to make them self-generative and more like us.

Slide 13

Does my research project merit a Pioneer award?

The payoff for studying the mechanisms underlying long-term working memory is potentially high and could give us a path for understanding human cognition.

I pioneered computational neuroscience in the 20th century, which today is helping us understanding brain function.

I am now a leader in NeuroAI, a new field at the nexus between neuroscience and AI that is broadening our horizons.

Slide 14

Thank you for your attention and your long-term working memory.