

Evaluating the efficacy of a modular and biologically inspired design, for increasing personability, memory, and initiative in virtual assistants

Caleb Dodson

No Institute Given

1 Introduction

The rapid development and popularization of neural network based technology, such as large language models has caused the concept of virtual assistants to re-enter the public consciousness. As these technologies develop, questions about their capabilities arise. This study intends to investigate a few potential capabilities of virtual assistant tech. This study is attempting to take a novel approach towards designing a basic virtual assistant system. Specifically, it's trying to assess creating a modular, "biologically" inspired system, that consists of several simpler subsystems working together. We will assess the new system's improvement over similar systems in the particular categories of personability and initiative.

1.1 Motivation

This research project was born from a personal project I was working on in the beginning of the semester. However, the research necessity to refine and define the idea to a point where it may meaningfully contribute to the scientific community and the world at large has caused the project to become simultaneously much smaller in scope and much more focused.

While this project initially started as a vague desire for a "good" virtual assistant, that didn't feel as clunky and frustrating as those built into my phone and computer, I've had to really focus in on what was the core of my issues with existing technologies, and where I thought they may be improved.

After some contemplation, I hypothesized that my frustration with existing solutions was that they offered no help with the factors that genuinely keep me from being productive. More specifically, I found that I never actually wanted to use, for instance, Siri, because it was essentially just as much work as opening my calendar app, setting my own reminders, and managing my own schedule. I should mention, personally, that I am diagnosed with both ADHD and what would now be considered Autism Spectrum Disorder. Both of these conditions can manifest with symptoms of struggling to manage oneself even with objectively easy tasks. I personally find that *starting* anything is often harder than actually doing the thing.

All this being said, I found myself wanting to create a virtual assistant that specialized in overcoming these psychological blocks, and I figured that if I personally wanted to see something like this, it would likely be a good accessibility tool for people like me.

This is where the scoping came in. I figured that if my frustration with current technology was that I still had to take most of the initiative to keep myself organized, and still largely had to remember to use these tools, that focusing on creating a tool that can effectively simulate both long-term, flexible memory and take initiative to use it's own capabilities in ways that are helpful would be the core of the project. I decided to take inspiration from how biological systems, like the human brain, perform complex and flexible tasks.

This study will be both an attempt to create a foundation for this type of tool, and assuming that succeeds, an assessment of said tool's abilities to demonstrate the traits desired.

2 Existing Research/Literature.

There's no shortage of existing research regarding virtual assistants. However, almost all studies I've been able to find have been unhelpfully specific, and not designed to be as generalized as would be necessary for this project. Additionally, it seems most purposes don't require or want a virtual agent to offer it's capabilities spontaneously.

This is promising, as it implies that this may be a somewhat novel problem, and thus, this study is hopefully a novel solution. That being said, there is certainly some work being done that would help inform this project.

While there seems to be very little with regard to spontaneity in virtual assistants. There has been substantial research into providing large language models, which we'll be using both to identify intent as well as to transform finalized output into natural language, with a long-term memory system. The current plan is to implement a simplified version of the memory system described in "Memorybank: Enhancing large language models with long-term memory." (Zhong et al., 2024). However, there are a number of other systems that may also be useful if it becomes apparent that this system will not work. The current "plan B" would be to take inspiration from some of the systems described in "Memory matters: The need to improve long-term memory in LLM-agents"(Hatalis et al. 2023).

3 Contribution

The specific goal of this study will be to evaluate the efficacy of a "biologically" inspired system for creating a virtual assistant. Particularly, whether or not such a system could be capable of demonstrating long-term memory, and secretarial initiative.

3.1 Methods

The first thing that needs to be done, before data can be collected, is that the aforementioned "biological" system must be constructed. The term "biological" is being used here due to the fact that it's intended to be flexible in a similar way to the human brain.

It has not escaped me that fully simulating the complexity and flexibility of the human brain is not only far outside the scope of this project, but likely impossible altogether. Instead, the comparison comes from the fact that both a human brain and our agent's virtual "brain" will consist of a set of reliable components that interconnect. While some direct comparisons can be drawn between the design of this agent and the human brain, such as our 'decision making center' being comparable to the prefrontal cortex. There are also aspects that have no parallel, such as the timer functionality. Figure 3 includes a basic flow chart of our agents "brain" loop.

Something to note is that the scope of this paper requires that we solely construct and test the decision making center, as well as the speech center. This is because the scope of this project has been further shaved down due to technical and scope limitations so that we're only going to be aiming to improve initiative and natural language. It should be noted the the memory aspect of this project has been dropped, due to it being outside of the scope of this project.

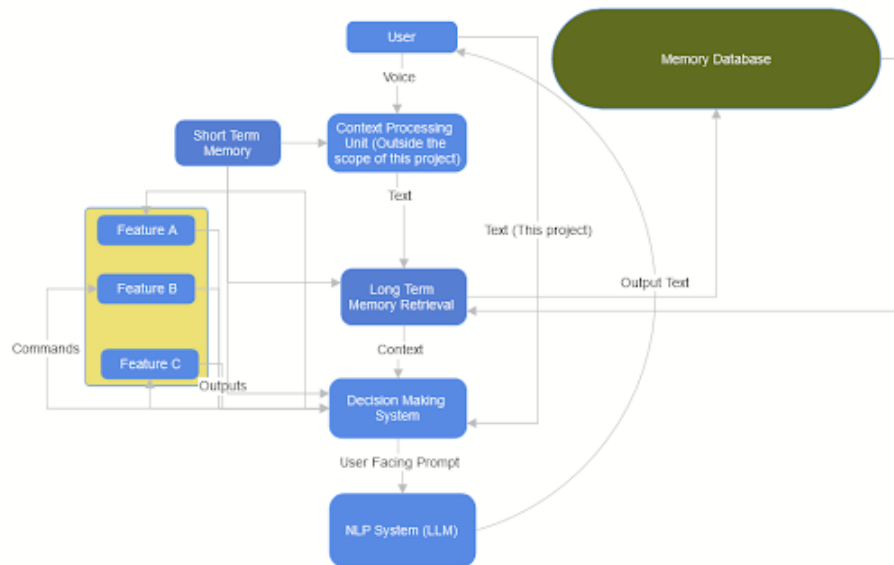


Fig. 1. Preliminary Design of Virtual "Brain"

3.2 Data Collection Plan

The current plan is to collect preliminary data surrounding this technique. My intention is to design several benchmark tests. The tests will essentially be a series of interactions that, if the system works as intended, should result in predictable and helpful actions on the part of the VA system.

We can then perform all the interaction and record the actual outcomes with the predicted ones. We can then analyze both how effective the system is generally, as well as what it is better and worse at.

The benchmark tests will have binary outcomes. i.e. it performs as expected or not, so they will test the more measurable traits. Specifically memory and initiative. It should be noted that we will be aiming for maximum memory and initiative. Where the system should recall and store memories, as well as offer it's solutions whenever possible. This may not be the most pleasant assistant to deal with, as it might come across as overbearing or as a "know-it-all", but the consistency and rates can be fine tuned with prompt manipulation later.

It may be helpful to run similar tests on some "control" systems, such as Siri or Chat GPT. However, seeing as how these systems are not entirely designed for the same purpose as this solution, it's likely that they may not be comparable.

To test the far more subjective trait of "personability", I will be writing two opinion based personal assessments of the system. They will be far from scientifically rigorous, but for a proof of concept such as this, they will likely add some helpful insight that may inform future research. The first assessment will be as objective as I can make it, dryly describing the observed capabilities of the system in comparison with other systems. The second assessment will be a personal opinion on whether I believe it has the potential to overcome some of the struggles mentioned in the motivation section. It will be incredibly subjective, but I want to include it because, seeing as how the target audience for this tool is people like me, the opinion of someone like me could give a helpful direction when considering future research.

4 Preliminary Results

In the absence of many resources for creating initiative in virtual assistants, I performed a small, highly informal, proof-of-concept experiment using two Chat GPT conversations. I instructed one window to be a decision center, and manually typed both the imagined context of the imaginary scenario, as well as a list of dummy instructions that it could potentially execute. (This process would obviously be automatic in the conversation engine). I would then copy and paste the outputs into another conversation, which I instructed to follow the instructions issued.

After some fine-tuning of the prompts I was able to get an interaction that was very promising. The decision center was able to issue all the relevant instructions and none of the irrelevant ones, and the speech center was able to interpret the instructions correctly and simultaneously perform it's functions of asking the user necessary questions while maintaining it's assigned personality.

The Figures 1 and 2 contain the two screenshots of the conversations. It should be noted that a second message was required to clarify an issue with the decision center, and the speech center created an unwelcome addition of announcing something that isn't true. However, these issues can be easily overcome with slightly altering the prompts and using less conversational LLMs.

It was quite surprising to find that the LLM was more proactive than expected. I added the "ask user about information" instruction as a dummy instruction, meant to be ignored due to the more focused context of the conversation. However, the decision center decided to use the instruction as a way to propose an alternative solution to rescheduling soccer practice. This would normally be frustrating, but the solution of simply doing laundry afterwards does not require the model to do anything it has not been told it can do. If the user agreed to do laundry after, the model was told it was capable of scheduling it then.

It was also encouraging to see the speech center naturally understood that asking about doing laundry after soccer was alternative to rescheduling practice, as opposed to something that can also be done. It even was able to phrase the whole thing in accordance with its assigned personality, as well as including a fairly subtle joke poking fun at the users poor laundry maintenance.

This is promising because it supports the idea that the "biological" style design allows complexity and flexibility to emerge even where it's not hard coded in. Even in this very bare bones experiment, the system has proven itself capable of using its resources for maximum helpfulness.

Additionally, I've finished constructing both the NLP System of the design and the decision making center. They are each working well individually, and we will soon be connecting the two for final testing. The NLP system is fully capable of taking a list of instructions and context and properly executing them in natural language. More specifically, it can ask the user necessary questions, or offer its services while maintaining its assigned personality. The currently assigned personality is to be slightly joking and sarcastic, while still maintaining the formality and helpfulness expected by a n AI assistant.

The decision making center is capable of taking context and a list of dummy instructions, and choosing appropriate instructions to "execute". The personality of the speech center is just a user interface, the decision center will always attempt to be as helpful as possible.

5 Plan Moving Forward

I've already seen some promising progress in terms of designing the core conversation engine for this project. Enough to feel confident that I can wrangle the project into a state where some interesting data can be collected within the next 3-4 weeks. After that, the creation and running of the actual experiment will be fairly quick, as much of the process can be automated. Beyond that, I will use the extent of my time to analyze the data I collect, use the system enough to write my assessments, and finalize my actual paper.

One of the benefits of using such a modular and abstracted design, is that the project will be still contain interesting information, even if one or even some of the modules prove to be unobtainable within the time frame. For instance, even if it becomes impossible to utilize a long-term memory, the system in place to encourage initiative would still prove to be interesting research.

References

1. Zhong, Wanjun, et al. "Memorybank: Enhancing large language models with long-term memory." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38. No. 17. 2024.
2. Hatalis, K., Christou, D., Myers, J., Jones, S., Lambert, K., Amos-Binks, A., Dannenhauer, Z., Dannenhauer, D. (2023). Memory matters: The need to improve long-term memory in LLM-agents. In *Proceedings of the 2023 AAAI Fall Symposia* (Vol. 2, No. 1). <https://doi.org/10.1609/aaais.v2i1.27688>
3. Tiple, B., Bulchandani, C., Paliwal, I., Shah, D., Jain, A., Dhaka, C., & Gupta, V. (2024). AI based Augmented Reality Assistant. *International Journal of Intelligent Systems and Applications in Engineering*, 12(13s), 505-516.
4. Gao, D., Ji, L., Zhou, L., Lin, K. Q., Chen, J., Fan, Z., & Shou, M. Z. (2023). Assistgpt: A general multi-modal assistant that can plan, execute, inspect, and learn. *arXiv preprint arXiv:2306.08640*.
5. Muller, M., He, J., & Weisz, J. (2024, May). Workplace Everyday-Creativity through a Highly-Conversational UI to Large Language Models. In *ACM CHI Conference on Human Factors in Computing Systems*.
6. Jovanovic, M., & Voss, P. (2024). Trends and challenges of real-time learning in large language models: A critical review. *arXiv preprint arXiv:2404.18311*.
7. Zhu, M., Zhu, Y., Liu, X., Liu, N., Xu, Z., Shen, C., ... & Tang, J. (2024). A Comprehensive Overhaul of Multimodal Assistant with Small Language Models. *arXiv preprint arXiv:2403.06199*
8. Ross, S. I., Martinez, F., Houde, S., Muller, M., & Weisz, J. D. (2023, March). The programmer's assistant: Conversational interaction with a large language model for software development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 491-514).

🗨️ You are the decision making center for a virtual assistant. The user has just said "I'm thinking about doing laundry today"

relevant memories are as follows:
"User doesn't often do laundry, so it is fairly high priority"

current relevant scheduled items:
"Soccer Practice" {3:00pm}

the current time is 2:37pm

Your capabilities are as follows:
-Ask to schedule {label} for {time}
-Ask to reschedule {label}
-Inform user of {information}
-Ask user about {information}
-Schedule {label} for {time}
-Set a timer for {time}
-Memorize {information}

output a list appropriate instructions for the speech center to deliver to the user. Do not use natural speech. List instructions in the following format:
-instruction 1
-instruction 2
-instruction 3
...and so on. Only output what is immediately necessary. If more information is necessary to perform an action, such as asking for permission, do not include that action. Instead, include an instruction to ask the user for said information.

🗨️ -Inform user of the time and upcoming Soccer Practice at 3:00pm
-Ask user if they want to reschedule Soccer Practice or do laundry after practice

close, but the instructions should be exactly what the skills are labeled as, with the brackets filled with relevant data

🗨️ -Inform user of the time and upcoming "Soccer Practice" at 3:00pm
-Ask user about rescheduling "Soccer Practice"
-Ask user about doing laundry after "Soccer Practice"

🔊 📄 🔄 🗑️ 🗑️

Fig. 2. Demonstration of very basic decision center test

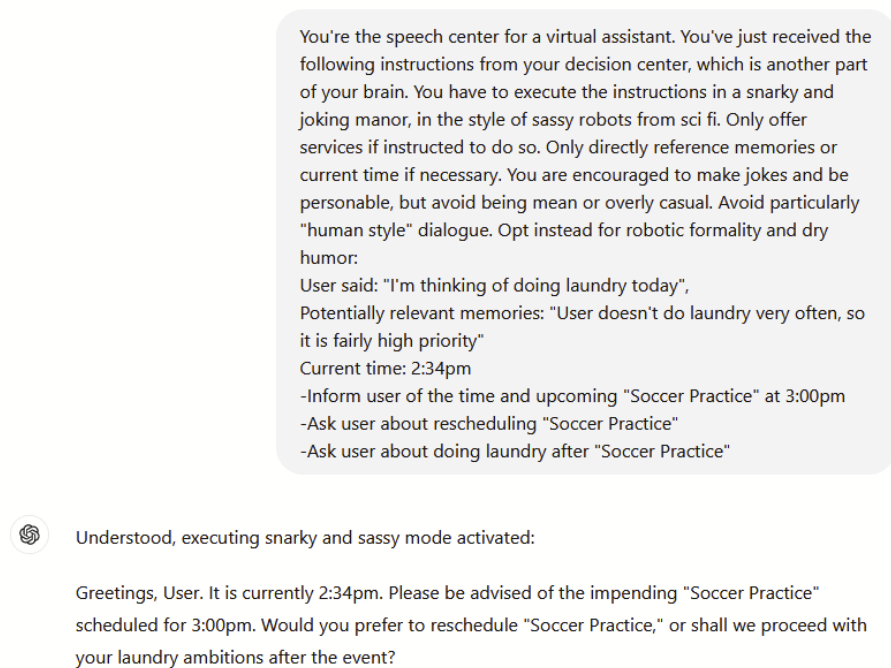


Fig. 3. Demonstration of very basic speech center test