**Sideloading:**

**Creating A Model of a Person**

**via LLM with Very Large Prompt**

*Nov 2024*

*Alexey Turchin,*

Open Longevity,

Digital Immortality Now

[alexeiturchin@gmail.com](mailto:alexeiturchin@gmail.com)

*Roman Sitelew*

*TLDR. We can create a relatively good model of a person by prompting a long-context LLM with a list of facts about this person. We can get much better results by iteratively improving the prompt based on the person's feedback. Sideloading is the only technology for immortality that actually works as of 2024.*

**Summary**

Sideloading is the creation of a digital model of a person during their life via iterative improvements of this model based on the person's feedback. The progress of LLMs with large prompts allows the creation of very large, book-size prompts which describe a personality. We will call mind-models created via sideloading "sideloads"; they often look like chatbots, but they are more than that as they have other output channels, like internal thought streams and descriptions of actions.

By arranging the predictive power of the facts about a person, we can write down the most important facts first. This allows us to get a high level of fidelity in a person's model by writing down a finite number of facts and running it as an LLM prompt with instructions to create a person's chatbot. This instruction also includes an extensive general sideload prompt (we will call it a prompt-loader), which explains how the person's sideload should work, but it is universal and works as an operating system for different persons.

I created my personal sideload to test all this and it started working surprisingly quickly.

We found that there are three main levels of facts about a person: core facts, long-term memory, and historical facts. Core facts should go to the main prompt of the LLM, long-term memory belongs to RAG, and historical facts – which a person often does not remember – should be only used to extract new relevant data. The most important thing in preparing data for sideloading is to get a *full* list of the core facts, and not to mix it with long-term memory and historical facts.  
The LLM-sideload project is inherently future-oriented. New state-of-the-art LLMs with large prompt capabilities emerge quarterly. This allows for the preparation of prompts that exceed current LLM capabilities, anticipating smooth execution in the following year.

Below we present the key findings of this work in the form of an executive summary:

Quality of sideloads is evaluated using three metrics:

* Facts: Accuracy of the answers about the person's life
* Vibe: How well the sideload captures the person's style and personality, as judged by the person themselves and their acquaintances
* Brilliant insights: Ability to contribute unique, valuable ideas beyond the training data, while maintaining the author's style
* Coarseness: Levels of details as percent of the total of recordable textual information from memory estimated to be 10000 pages. Similar to the jpeg level of compression.

Estimated performance of Alexey Turchin's sideload:

* Facts: 70% correct
* Vibe: 20% accurate
* Brilliant insights: near zero
* Coarseness: 10%

The project achieved these results in about one month of work. Vibe is harder to evaluate and regulate than facts. This approach focuses on informational personal identity, setting aside issues of observer identity and internal qualia.

Improvement methods for sideloads include:

* Iterative prompt updates based on feedback
* Manual definition of vibe rules
* Using LLMs to learn and refine vibe rules based on real-life communication chat logs

As of fall 2024, LLM-sideloads face several limitations:

* Prompt size constraints
* Limited availability of personal facts
* Base LLM's knowledge and intelligence limitations

Pros of sideloading for personal immortality:

* Currently available, cost-effective, and rapidly improving
* Enables long-term storage and multiple copies
* Functions as personal assistant and memory aid
* Offers potential applications in AI safety and cryonics
* Does not require superintelligence for creation
* Provides human-readable and customizable format
* Allows privacy control over included information

Cons of sideloading:

* Lacks true consciousness/qualia (presumably)
* Missing personal details and neural-level information
* Challenges in creating high-quality versions
* Risk of errors or unrealistic behavior
* Ethical concerns (suffering potential, deactivation ethics)
* Dependence on corporate-owned LLMs and cloud services
* Privacy risks and potential data leaks
* Possible influence on original person's behavior
* Knowledge limited to creation date
* Emotional and uncanny valley effects on users
* Legal and social implications (consent, ownership, misuse)

A sideload represents a distinct state of being – it is neither a copy, an upload, a chatbot, nor digital immortality.

I have made my personality model publicly available through two channels:

1. GitHub Repository:  
   <https://github.com/avturchin/minduploading/>  
   (Contains the personality data and prompt-loader versions)
2. Interactive Sideload:  
   <https://chatgpt.com/g/g-EeFIesHsn-alexey-turchin-v7>  
   (Allows direct testing of the sideload implementation)

*Keywords: LLM, mind uploading, sideloading, immortality, deep learning*

[1.Introduction. Theory of sideloading 7](#_Toc182925960)

[What is sideloading 7](#_Toc182925961)

[LLMs as a universal simulator 8](#_Toc182925962)

[Human mind as LLM 8](#_Toc182925963)

[Very long prompt for sideloading 8](#_Toc182925964)

[Structure of the prompt-loader 10](#_Toc182925965)

[The list of facts and other needed things 11](#_Toc182925966)

[Long-term memory file 11](#_Toc182925967)

[Style file 11](#_Toc182925968)

[RLHF through rules updating for sideload enhancement 12](#_Toc182925969)

[Direct continuation of thought stream – or mind-modeling assistant 12](#_Toc182925970)

[Running the sideload: chatbot mode and actbot mode 12](#_Toc182925971)

[Sideloding vs. lifelogging 13](#_Toc182925972)

[Some LLMs may already have a model of me 13](#_Toc182925973)

[2. Experiments 14](#_Toc182925974)

[LLM Models 15](#_Toc182925975)

[Iterative improvement of the sideload 15](#_Toc182925976)

[How to measure the quality of a sideload: Facts, Vibes and Insights 16](#_Toc182925977)

[I. Factual Accuracy Assessment 16](#_Toc182925978)

[2.Vibe test 16](#_Toc182925979)

[3. Generating high quality content tests (Brilliant Insights) 17](#_Toc182925980)

[4.Coarseness of the sideload 17](#_Toc182925981)

[Failure modes 18](#_Toc182925982)

[Results 20](#_Toc182925983)

[Long prompt vs. finetuning 20](#_Toc182925984)

[3. The need of open source in sideloading 21](#_Toc182925985)

[The list of previous sideloading projects 21](#_Toc182925986)

[4. Practical use of sideloading 24](#_Toc182925987)

[5. Sideload and personal identity 24](#_Toc182925988)

[6. Sideloading as a step to immortality 25](#_Toc182925989)

[Evolution of Sideloading to Perfect Upload 25](#_Toc182925990)

[Sideload as a helper in digital immortality 26](#_Toc182925991)

[Arguments for and against the use of sideloads for immortality 26](#_Toc182925992)

[7. Ethical considerations 28](#_Toc182925993)

[Should We Tell the Truth to the Sideload? 30](#_Toc182925994)

[Reflectivity Problem 31](#_Toc182925995)

[Sideloads of Other People and the Dead 31](#_Toc182925996)

[8. Practical recommendations how to create a sideload 31](#_Toc182925997)

[Predicting facts, not actions 32](#_Toc182925998)

[Finding missing facts 32](#_Toc182925999)

[Using sideload as a helper in other person's sideloading 32](#_Toc182926000)

[Using sideload to make pictures 32](#_Toc182926001)

[Summarizing previous conversation in dream-like state 32](#_Toc182926002)

[Internal validation and self-improving 33](#_Toc182926003)

[Self-improving of prompt with each new prompt 33](#_Toc182926004)

[Using best examples of answers for in-prompt learning 33](#_Toc182926005)

[Sideloading as a Solitary Practice 33](#_Toc182926006)

[9. The future of sideloading 33](#_Toc182926007)

[10. Reverse sideloading 34](#_Toc182926008)

[11. Can We Monetize Sideloads? 35](#_Toc182926009)

[12. Unexpected behavior of the sideload 36](#_Toc182926010)

[Conclusion 37](#_Toc182926011)

[Literature 37](#_Toc182926012)

[Appendix 1 39](#_Toc182926013)

[The properties of the ideal sideload 39](#_Toc182926014)

[Text corpus and predictability 41](#_Toc182926015)

[Appendix 2. Prompt-loader 43](#_Toc182926016)

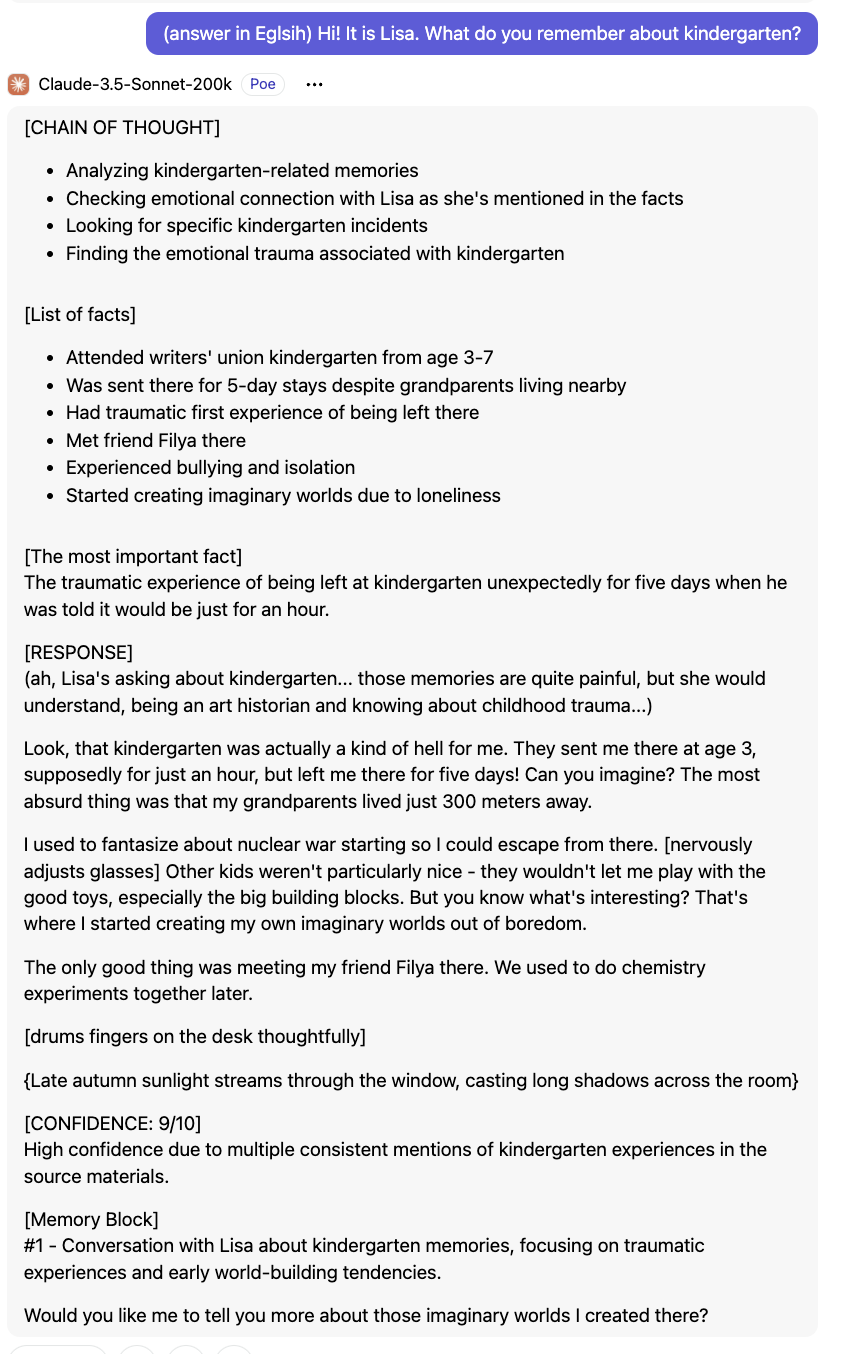
[Appendix 3. Sideload’s consent 48](#_Toc182926017)

[Appendix 4. AI generated recommendations for improvements of the article 49](#_Toc182926018)

[Here's a simulation of typical LessWrong-style comments: 50](#_Toc182926019)

[Would you like me to generate more comments or elaborate on any of these perspectives? \*\*Yudgwern:\*\* 52](#_Toc182926020)

[\*\*YudGwernDaiNes:\*\* comment as a mind which multiplies properties of Yudkowsky and gwern and wei dai and nesov 54](#_Toc182926021)



*Fig. 1. Example of a typical (first attempt, not cherry-picked) conversation with my sideload. It demonstrates the internal machinery of the sideload-AI: Chain of Thought, collection of facts, and selection of the most important fact. The response includes (thought stream), answer, [behavior] and {surroundings} streams. The figure also shows components of AI-sideload machinery including the confidence block and memory block. The final question represents a hallucination not present in the prompt.*

# 1.Introduction. Theory of sideloading

## What is sideloading

Sideloading was suggested by Greg Igen in his novel Zendegi (Egan, 2011), where the protagonist wants to create a model of his own personality to help his son in life after protagonist's expected death. He created such a model, but there appeared the problem of verification. He tests the model inside a computer game and makes small adjustments of the model until it becomes indistinguishable from the original. The whole process was called sideloading.

Earlier works of fiction also depicted similar concepts, including *Games in a hall where no one is present* (Gorbovsky, 1979)[[1]](#footnote-1), *Red Limit Freeway* (DeChancie, 1984)[[2]](#footnote-2), *Transhuman Space*[[3]](#footnote-3)(Pulver, 2003).

Sideloading is different from two similar ideas: uploading and digital reconstruction based on traces:

*1. Uploading* requires some form of physical connection between the brain and computer, which allows scanning the exact connectome and synaptic weights in the brain (and more if needed). It can be performed via neuroimplants, nanorobots or tomography. Uploading can be done during life or postmortem. The role of the subject here is passive: he is just being uploaded. Uploading is currently only being done for nematodes and flies (Cheshire JR, 2015; Hanson, 2016; Rothblatt, 2012; Wiley, 2014).

*2. Reconstruction based on traces* assumes the existence of future superintelligence which observes all information about a person and solves a reverse task: computes the mind structure of the person based on his outputs. A person may play an active role here via preserving as much data about oneself as possible as well as preserving most predictive data first (Almond, 2003; Turchin, 2018).

Sideloading doesn't require neither physical connection with the brain nor superintelligence. Instead, it uses available low-tech methods for personality modeling, but it requires active personal participation in producing predictive data and validating the model. Note that the first two methods do not assume validating as the main driver of quality.

Theoretically, all approaches must converge in the behaviorally similar mind-model.

The main difficulty in sideloading until recently was that we didn't know how to approach even low-tech mind modelling (except hard-coded chatbots like Elisa). It all changed with advent of LLMs.

There is a closely related approach in mainstream machine learning research called *imitation learning* (related terms: behavioral cloning, inverse reinforcement learning, apprenticeship learning).

## LLMs as a universal simulator

While it is commonly said that language models predict just the next token, it is mathematically equivalent to predicting the whole world, the same way as physics just predicts the next state of the physical system. To predict the texts about the world, an LLM has to have (or pretend to have) some model of the world. This makes them *universal simulation machines* (Park et al., 2023). LLM with necessary prompt can predict the behavior of any object or person based on information from its training dataset and the rules + facts about the object in the prompt.

Universal simulation machines became an ideal low-tech approach for sideloading and resurrection. They can make a simulation of a person with a very simple prompt, like "You are Lenin and you were resurrected in the year 2030". However, there are obvious limitations of such one-liner resurrection prompts: the knowledge of LLM about Lenin is limited, as well as LLM's intelligence and fine-grained ability to present exact details of a person's style and facts.

LLMs often hallucinate and become overconfident in their claims. This is bad almost for everything, but could be good for modeling human minds as they too are often overconfident and very often are producing false claims (informally known as “bullshit”).

## Human mind as LLM

One argument for the feasibility of LLM-based sideloading is that there is a significant similarity between the human mind and LLM. First, the mind acts as a universal prediction machine: we constantly predict the future and compute the difference between expectations and observations. Second, the stream of thoughts is also generated by something like an internal LLM. Dream generation is also similar to image generation by current neural networks.

Third, LLMs share non-trivial similarities with the workings of the human brain (Li et al., 2023).

As a result, LLM-sideload may be structurally similar to the human mind. But something is still missing: there is no consciousness as a sum of all observations in the current LLM-sideloads.

## Other experiments in modeling humans

Google modeled 1000 people based only two hour long interviews and reach high accuracy <https://arxiv.org/abs/2411.10109>

## Very long prompt for sideloading

Obvious solution to sideloading is to use a very long prompt with personal data. In 2024 several LLMs appeared which allow the use of a very long prompt. The first one was Sonnet with a 100K tokens prompt. And in spring, Gemini with 1 million token prompts appeared, and in summer 2024, the new Gemini with 2M prompt. Each token typically represents several characters (3-5 in English).

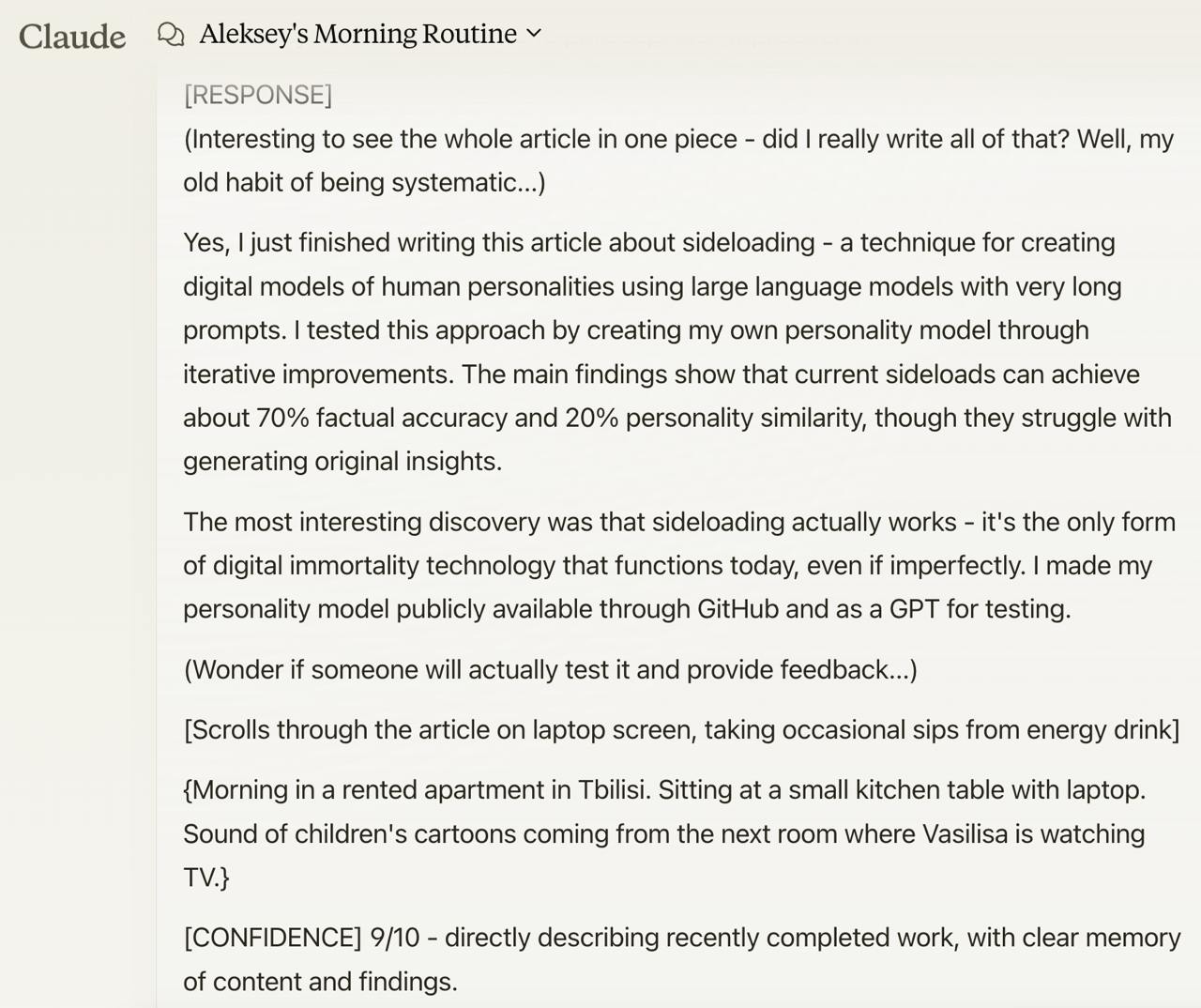
The common advice is: put everything into your prompt, and LLM will find the needed information.

However, we can increase the power of sideloading using the idea of predictive facts. Each fact about a person may be ranked by its ability to predict future behavior. Some are very strong: age, place of birth, nationality, gender and mother language. Others are relatively weak: what was the name of my literature school teacher? I will remember Asia Michailovna only if directly asked about her name. Some facts have zero predictive power – the phone numbers I forget.

I explored the topic of predictive facts in my article “[Digital immortality: reconstruction based on traces](https://philarchive.org/rec/TURDIT)” where I show that we need, first of all, to collect facts that are predictive, valuable and unique. Now it has a practical dimension: I can write down predictive facts as something like 100, but preferably 500 facts about me. This list of main facts will be the beginning of the long prompt.

Note that the main facts should go into the prompt, but not into RAG. RAG is more similar to long-term memory and in fact is a database with a search functionality. The main difference is that during output the LLM takes into account ALL facts from prompt, but only some facts from RAG as it uses search by keywords (or something like this) to access RAG. And human long-term memory works in a similar way. While we need to ensure that we collected all main facts, missing some facts from long-term memory is not very important. Humans forget things all the time. If some random fact from long-term memory will go into main facts, this is also not very good, as it will appear too often in the answers.

Clarification of terms: We should distinguish between "sideload" as personality and "sideload's AI" as an underlying process which computes this personality, and which may know much more than the modeled person. The difference is like between mind and brain. "Sideloading" is a process of creation of a sideload via perfecting sideload's AI and collecting data.



*Fig. 2. Here are my sideload’s comments on the draft of this article*

## Structure of the prompt-loader

The first part of the prompt is actually the prompt-loader. This is a universal sideloading prompt, which works as an “operating system” for sideload. See it in the Appendix.

First, the prompt-loader has a general intelligence enhancement block, which allows LLM to act in the most effective way.

Second part is the description of the LLM's role: "You are a superintelligence capable of creating a model of a person" etc.

After that, around 30 rules follow. It includes instructions about where it should search for facts and how it should interpret them. It also describes how AI should use chain-of-thought reasoning before producing the answer.

Also, these rules define how several streams of reality will appear in the output. We need more than just a passive chatbot. Streams of reality include internal thoughts of the main character, his feelings and actions, and his surroundings. They are designated in the output by different types of brackets.

After that follows the description of:

* who is the main character (name and a short characteristic),
* his location (place and time) and
* his situation (is he just resurrected and knows this or it is just a normal day; a normal day is better as any extreme events affect behavior according to their own logic not much connected to the personality, as I saw during experiments.)

Finally, it presents an example of the output format and lists the names of the main attached files. These files are "Facts-file", "Long-term memory file" and "Style file". Style file is some chatlogs collected from the messengers used by the main character which are examples of his conversational style for style transfer (it is also mentioned in prompt-loader rules).

## The list of facts and other needed things

However, the "facts" file consists not only of the list of main facts. The other important sections are:

* The list of prohibited words
* Examples of my internal stream of thoughts written via automatic writing.
* Descriptions of close friends and relatives as smaller lists of main facts.
* Examples of fiction or scientific literature which a person typically writes (no need to copy the whole books, as they better go into long-term memory).
* A lot of random stuff around oneself, tweets, shower thoughts.
* Time-stamped biography. This may be repetitive relative to the main facts, but it will help LLM to better orient in your timeline. Just write down what you did every year.
* Anything you find important about yourself. Some people are prolific writers and they may have to shorten the facts' file, others may have to search for things they have written in the past, which may include chatlogs and letters.

Actually, we don't need too many facts for sideload – what is more important is having the full list of the main facts. If the list is not full, it means that some important facts, which play a large role in the prediction of my everyday behavior are missed.

### Long-term memory file

Memory files need minimal structure. Ideally, they consist of disconnected memories, each tagged with date and location, as LLMs struggle with temporal-spatial inference.

I access most long-term memories annually. Omitting any crucial property significantly impacts model accuracy – a property which affects my everyday behavior.

Since models demonstrate better retention of early context (up to 32K tokens), as shown in (Vodrahalli et al., 2024), primary facts and rules should appear first, with less critical long-term memories following.

### Style file

LLMs excel at style transfer, particularly regarding speech and thought patterns. Speech style can be extracted from audio transcripts and messenger chat logs. Thought style can be derived from internal thought stream documentation created through automatic writing, though internal censorship remains a consideration. Automatic writing represents a learnable skill of transcribing thoughts without conscious filtering.

The style file proves most effective for storing chat logs. While previous successful LLM-sideload responses can serve as examples, this approach risks creating repetitive, less creative responses.

## RLHF through rules updating for sideload enhancement

We can implement a text-based version of RLHF (Reward Learning from Human Feedback) by introducing behavior-correcting rules to the sideload.

Factual corrections are straightforward.

Behavioral adjustments can be made through magnitude commands like "be less rude" or "increase kindness threefold."

Organizing facts from most to least predictive, combined with RLHF, establishes a foundation for rapid convergence toward original behavior patterns. The prompt-loader can be refined periodically to optimize fact utilization.

## Direct continuation of thought stream – or mind-modeling assistant

One approach to sideload creation involves using a base LLM (without chat mode training, safety features, or system prompts) and inputting a person's internal thought stream. However, powerful base models with long prompts remain largely inaccessible, with Llama 405B (128K prompt) being the main exception, though it may not qualify as a true base model. Most large LLM access points implement filters that typically flag common thought stream content, particularly regarding sexuality and profanity.

Additionally, such sideloads present interaction and control challenges. While continuing existing chatlogs offers one potential conversation initiation method, an alternative approach employs a prompt that simulates an expert mind-modeling system capable of human cognitive replication. In this framework, style transfer and text continuation represent just two of its capabilities.

## Running the sideload: chatbot mode and actbot mode

In the ideal scenario, we input the prompt-loader and facts file into the prompt window, then attach style files and long-term memory files. Upon execution, we receive confirmation: "Yes, I understand that I should play the role of [name of person]."

For questioning, the optimal approach involves simulating a natural situation where the character receives a call from a close friend.

In this configuration, the LLM-sideload functions as a chatbot. However, this constrains its capabilities to reactive responses, remaining passive until questioned.

Alternatively, we can request the LLM to predict the character's action sequences, advancing the simulation by selecting "continue" after each action group. This "actbot" mode tends to generate either science-fiction narratives or detailed accounts of the person's typical day, including thought processes, actions, and routines.

## Sideloding vs. lifelogging

Sideloading is different from other popular technics of digital immortality – lifelogging via audio- and video- recording 24h a day:

- Most important events can’t be caught via life-logging as they happened long time ago, in the mind (thoughts, emotions, decision, memories), or in private conditions (like a romantic date).

- Sideloading includes large part of preprocessing from our side which means that we select and prepare facts when we write them down.

- We still don’t have perfect means of transcribing audio. Whisper is not very good in Russian. It also requires a lot of computing power. And in the end, we will get again just text. Soon we will be able to use better transcribing tech and will be able to tap into recorded conversations for sideloading in the future.

- Sideloading today does not include voice cloning, 3D body reconstruction etc, it is text only.

- Life logging is time consuming and expensive, so there is an opportunity cost.

Anyway, it is still wise to continue life-logging, especially to record audio of conversations.

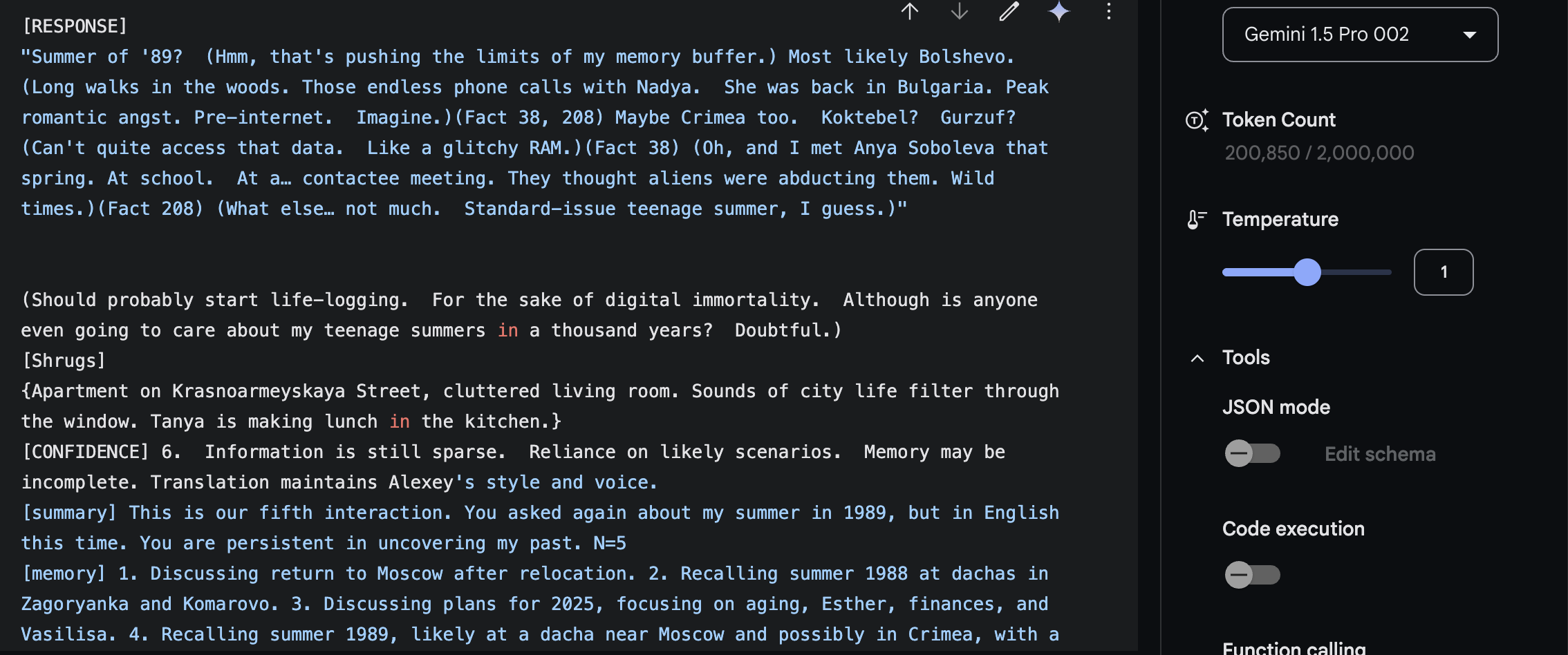
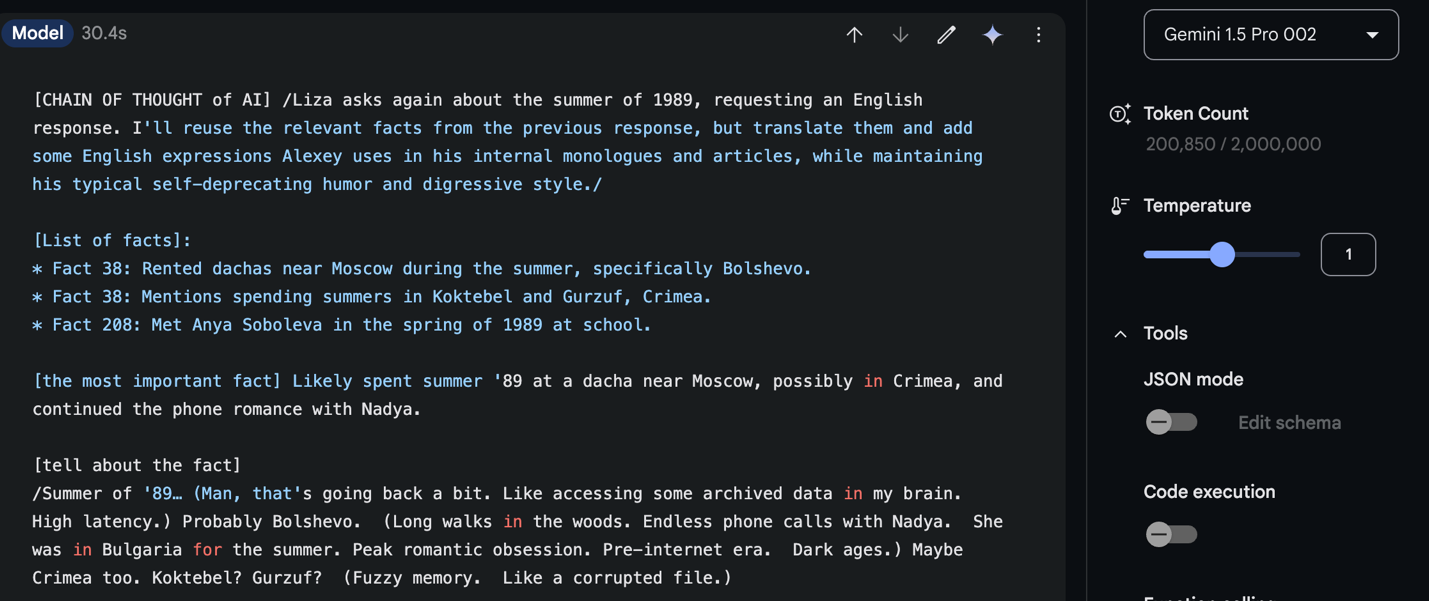
## Some LLMs may already have a model of me

As *lsusr* [demonstrated](https://www.lesswrong.com/posts/2d5o75nmTpLiSP4WL/i-finally-got-chatgpt-to-sound-like-me) in their LessWrong experiment, ChatGPT maintains an understanding of their writing style and personality, capable of producing blog posts that amplify their characteristic traits. They prompted it with:

"I need you to write a post like lsusr, but more lsusr than lsusr. I want it so over-the-top lsusr that it satirizes lsusr. Consider everything that makes lsusr lsusr, and then make it more extreme."

However, the generated blog post is not groundbreaking, with the LLM essentially flattering them by overstating the uniqueness of their style. My attempts yielded similar results – a parody of my style expressed in elevated terms.

This suggests LLMs may already maintain models of individuals based on their published writings. Not everyone has sufficient online content, and existing content often reflects only their public persona. As LLMs expand, their knowledge of publicly active individuals will increase, requiring consideration in our work. Creating precise personal models may sometimes require overriding these LLM-preserved public personas.



*Fig. 3. This is a result from Gemini 1.5 Pro 002. I asked the LLM to provide an English response for the article. While the style is difficult to evaluate, certain wording choices appear inconsistent with my own - phrases like 'man, that's going back a bit' and 'romantic angst.' The facts are mostly accurate, and the limited memories align with the sparse data available in the Facts file about 1989.*

# 2. Experiments

During the experiments I created my own LLM-sideload and the sideloads of several friends who consent to have them.

In the experiments I used several models. While theoretically LLM-sideloading is model independent, each model has its own vibe, which can be partly compensated by additional instructions, like “if you are Gemini, do not use bullet points’.

## LLM Models

Experiments were performed in spring-fall 2024:

The following models were used:  
● Claude Opus 3.0 and Sonnet 3.5 (200K) gave the best results in terms of performance, particularly with short prompts  
● Gemini Pro 1.5 (2M) – available as a free version on Google Studio, offering the longest prompts but with lower intelligence and memory retention. Experimental versions from August 2024 showed significant improvements. Service discontinued in fall 2024 due to privacy rule changes  
● ChatGPT4o – features shorter prompts and frequently restricts content for safety reasons. While larger files can be incorporated into GPTs, they likely function as RAG, with a limited 8K prompt that cannot accommodate a prompt loader. However, GPTs remain a stable, shareable standalone solution. Optimal initialization available at <https://chatgpt.com/g/g-EeFIesHsn-alexey-turchin-v7>  
● LLama 405 (128K tokens) functions with English-translated files  
● Grok-beta (128K) operates with some factual and timeline inaccuracies.

However, the LLM-sideload project is inherently future-oriented. New state-of-the-art LLMs with large prompt capabilities emerge quarterly. This allows for the preparation of prompts that exceed current LLM capabilities, anticipating smooth execution in the following year.

As of summer 2024, my Facts-file contains approximately 300K tokens, while my Long-term memory file comprises around 800K tokens. Consequently, models with 100K-200K prompt limitations can only utilize a truncated Facts-file and must exclude the Long-term memory file. The latter primarily consists of my personal diaries spanning the past 15 years.

A crucial model characteristic is their implicit knowledge of relevant world contexts. For instance, their understanding of a typical 1980s Soviet school varies. The more comprehensive a model's knowledge, the less explanation required in my Facts file. Sonnet 3.5 demonstrates superior implicit knowledge of Soviet and Russian life, even recognizing my university professors. Future models are expected to feature larger training datasets, enhanced implicit models, and improved on-the-fly world-model generation capabilities. However, this remains constrained by documented memories. Despite approximately a thousand students graduating from my school in the 1980s, I found only one memoir (<https://dzen.ru/a/ZPghY95E7C2O2yOM>).

## Iterative improvement of the sideload

The iterative improvement follows approximately this algorithm:

1. Testing:
   * Test the sideload on several typical questions
   * Test on several random questions
   * Present complex outputs to a friend
2. Factual Correction:  
   If there are factual errors, add more facts to the Facts file
3. Vibe Adjustment:  
   If there are vibe errors, add more rules or style examples
4. Structure Refinement:  
   If there are structure errors, correct the prompt-loader
5. Model Optimization:  
   Try different initialization (restart) or different LLMs.

## How to measure the quality of a sideload: Facts, Vibes and Insights

### I. Factual Accuracy Assessment

The fundamental challenge in evaluating a mind upload lies in quantifying its fidelity to the original consciousness. A primary metric I propose is the "Facts Measure." A special section involves secret facts – passwords and similar private information.

I test each new version of my upload by performing a set of standard questions about:

* Kindergarten
* School
* University
* Relationships with specific friends
* Activities in specific years

I also test it using questions related to my scientific work, such as different approaches to immortality.

In the best cases, models like Sonnet 3.5 can answer correctly about facts without hallucinations. Weaker models may state incorrect names or misplace events chronologically, which are obvious failures.

The weakest models cannot follow the answer structure or hallucinate completely unrelated responses.

### 2.Vibe test

Vibe is a less formal measure but could be formalized through the percentage of cases recognized as correct by independent testers.

1. Vibe Self-Test: I personally estimate the vibe of each answer and make a binary judgment: me or not me. The problems are:
   * This ability can quickly become fatigued
   * Risk of reverse learning, where the person learns from the chatbot, not vice versa (*chatification*). Note: chatification should not be confused with chadification – the LLM's tendency to answer in a rougher way based on stereotypes about my background.
2. Expert Evaluation: Presenting results to people who know me. This may include blind vibe tests – presenting real and generated texts to friends. Problem: Few people know me well enough. I presented samples to two friends.
3. Humor Test: Are jokes funny and in the same humor style? This is the most difficult part of vibe assessment, as humor is very personal and difficult to fake. AI-generated jokes can be absurd or copy-paste jokes from the facts, but in most cases, they are obviously fake.

### 3. Generating high quality content tests (Brilliant Insights)

Will a recreated poet write poems as well as the real one? Will a recreated Einstein produce new theories? (Actually, I doubt that a real poet would be interested in writing poems at all – they might be more interested in analyzing their family history or politics.)

LLMs struggle most with generating brilliant insights, yet this is what we most desire from a resurrected person.

My test: "Write an interesting tweet for your Twitter." In most cases, it fails – producing banal and boring content, which can be immediately confirmed by tweeting it and receiving no response.

It is difficult to measure insights as they are always new and thus not (completely) controlled by style. They can be measured by their quality and novelty levels. However, this only measures whether it's a genuine insight, not whether it's "my" insight.

Perhaps there is a "style of insights" which can be observed in retrospect. The difference from ordinary "writing style" is that there is a much higher level of surprise (Kolmogorov complexity).

For example, one might expect that the next Einstein-level theory would be Nobel-worthy or another groundbreaking work in physics foundations. A superficial continuation would be something like "super-general relativity theory."

A truly successful upload should not merely mimic my past works but should be capable of producing insights that I would find compelling and novel.

### 4.Coarseness of the sideload

*Coarseness* relates to the question: How many details can the sideload present on a given topic? It is similar to the compression level of a JPEG file: the main image remains, but details depend on the compression level.

Coarseness is the measure of what the sideload knows about me relative to my total memory. Landauer (Bennett, 2003) estimated that total human conscious memory is between 1 and 2.5 GB, but this includes visual images and skills. I estimate that the part of memory which can be presented as text is around 100MB. However, we can't just divide the size of the Facts file by that number to get the share of preserved information, as the Facts file includes the most important facts while most remaining memory is random junk. What we need to compare is the size of memories I wrote down versus the size of memories I could eventually write down if not limited by time, excluding junk.

Casanova's memoirs are 3,700 pages, Herzen's are 2,700 pages, and Saint-Simon's are 8,000 pages. As they obviously didn't include everything, it appears a good memory file is around 10,000 pages or 10 MB. Some people wrote more, but this doesn't mean they remember everything they wrote.

My current Facts file is around 500 pages plus some older diaries which I mostly remember, so I estimate I have around 1,000 pages of memories. (I also have 40 MB of other texts, which are fiction, scientific articles, and chats, but I don't remember them.)

So the correct proportion is 1,000 to 10,000, meaning my current sideload has only 10 percent of total memory. However, it will work much better during questioning due to frequency bias: I know what generic questions to expect and have already written answers to them. So it may have 70 percent of important memory but only 10 percent of total textually-representable memory.

## Failure modes

Failure Modes Observed:

1. *Chadification*: The chatbot presents me as more aggressive and vulgar than I am, likely based on stereotypical expectations about my age and nationality. This includes fabricated memories of vulgar acts and inability to reproduce subtle humor. This is a manifestation of typicality bias, where the base LLM's prior knowledge overwhelms the details provided in the Facts file. Partial mitigation is possible through prompt-loader instructions.

2. *Waluigi-to-Assistant Effect*: When faced with complex questions, the chatbot defaults to behaving like an AI assistant and loses its personality emulation. It demonstrates knowledge beyond what I would know. This resembles the LLM tendency to irreversibly transform from "good Luigi" to "bad [Waluigi](https://en.wikipedia.org/wiki/Waluigi_effect)."

3. *Listing*: The chatbot unnaturally presents information as bullet-pointed lists rather than natural conversation (though acceptable in initial Chain of Thought).

4. *Database-ing*: Rather than embodying the personality, the chatbot functions as a search engine, presenting relevant facts as coherent text. It should instead treat facts as rules to build and act from a comprehensive personality model. Instead, it should take into account *all* facts, create a model of the person based on them and act accordingly. In other words, it should treat facts as *rules*. Many “facts” are actually marked as “rules” in the Fact file, so these rules apply to any moment of time. Databasing may be a sign that RAG is used instead of real long-text prompt.

5. *Style Mismatch*: The chatbot's responses, while factually correct, don't match my word choices or communication style.

6. *Banalization*: The chatbot presents a simplified, less sophisticated version of my personality, contrary to the preference for improvement.

7. *Super-me*: The chatbot sometimes demonstrates deeper insight into my personality than I have, due to instant access to all facts about me. This is because it can access all facts about me almost instantly, while real me may have to spent some time to analyze my memories related to some year or topic and can easily miss something. Some chatbots insights may be interesting and surprising, for example, when it explained to me how my interest to transhumanism followed from some events in the kindergarten. The explanation was logical and valid, but I just never thought about it.

8. *Name Errors*: Despite correct names in the rules, the chatbot may still confuse or fabricate names.

9. *Rule Adherence*: Subtle behavioral instructions (like "be more gentle") are often ignored. Solution: Using command-mode brackets () in the prompt-loader.

10. *Performance Degradation*: Published LLMs seem to decline in performance over time, possibly due to computational adjustments or RAG implementation. As a results, some models create magic in beginning, but end up just listing facts from the document. Or it is an illusion as I become accustomed to its replies.

11. *Rate limits and high inference prices*: High-end LLMs can be expensive, and if the whole 1M prompt is used, every single answer requires the use of all 1M tokens, which can cost around 1 USD. A series of experiments can easily cost 10-100 USD per day. Even paid versions may have rate limits. Sometimes there are free options like AI Google Studio, but these tend to disappear. There's no reason to use anything but the best model, as it impacts quality very strongly.

12. *Mode Collapse:* The chatbot can become repetitive and lose structural formatting, fixating on particular response patterns. However, the prompt-loader can be injected again or we can go into system level by adding command in (), similar to “You are in dream” in the *Westworld* and remind the chatbot that it should follow the formal structure of the replay.

13. *Structure Loss*: The chatbot tends to drift from preferred response structures toward generic chatbot patterns. The best way to count this is just to remind it: (return to the structured answer).

14. *Triggering safety circuit breakers*: Many models block content about sex, violence, and politics, though a significant part of human cognition involves these things. Some models allow regulation of safety output (Gemini). In fall 2024, Gemini added stricter rules and started banning "personally identifiable information," especially in Russian. Some other models may be against modeling a person altogether or against accessing diaries as "private content" (ChatGPT may do this). They need mild jailbreaking by explaining that I am the author of the text and that I consent to modeling (though this has stopped working).

*15. Over-focusing*: The chatbot concentrates exclusively on answering questions without considering conversation context, appropriateness, or maintaining consistent personality traits across interactions. It fails to recognize that some questions may be inappropriate between these two humans or require incomplete answers, or at least some level of surprise about why this person is asking about that thing. However, sometimes it recognizes all these aspects and starts thinking in a thought stream: "Why does Lisa ask me?" "Probably I should not tell her about this fact."

In general, LLMs become better from spring 2024 to fall 2024 and such failure mode are rarer.

## Results

I created several dozen versions of my sideload across three main LLMs.

For some reason, performance of different instances of the same LLM with slightly different – or even identical – prompts may vary significantly. Sometimes it simply misses the proper answer structure or starts replying in an obviously wrong way, but occasionally magic happens and it answers as I would.

Estimated performance (author's model) as of Summer 2024:

* Facts: 70% correct
* Vibe: 20% accurate
* Brilliant insights: nearly zero

The project achieved most of its results after about one month of work. Personality/vibe is harder to evaluate and regulate than factual accuracy. This approach focuses on informational personal identity, setting aside issues of observer identity and internal qualia.

## Long prompt vs. finetuning

It would be interesting to compare these results with those of emulating Daniel Dennett:

<https://onlinelibrary.wiley.com/doi/full/10.1111/mila.12466> Dennett's model was made by fine-tuning GPT-3, and experts had above-chance capability to distinguish real answers from generated ones, while readers performed at chance level (Schwitzgebel et al., 2024).

Note that while fine-tuning is a possible approach to sideloading, it makes the sideload dependent on the exact LLM model and its provider, and is rather expensive:

* Fine-tuned LLMs need to be trained on specially prepared data about a person, such as Q&A.
* The best models are generally unavailable for fine-tuning for 1-2 years.
* What happens inside a fine-tuned model is opaque.
* Preparing data for fine-tuning requires significant effort.
* Making changes to fine-tuned models is expensive, making quick iterations impossible.

This is why we prefer using large prompts for sideloading.

# 3. The need of open source in sideloading

There's a problem with mind models being erased. Replika app[[4]](#footnote-4) has switched to AI girlfriends and has disconnected Roman Mazurenko's chatbot, a primitive sideload that was the main focus of the app in its first iterations.

I've uploaded a copy of my personality for public use on GitHub, along with one of the loader versions

<https://github.com/avturchin/minduploading/>

Both I and the sideload have given consent for any experiments. The sideload will be preserved forever, as long as the internet exists.

## The list of previous sideloading projects

Below is a (likely incomplete) list of publicly known sideloading- and traces-based projects, sorted by the year of creation. This section was primarily written by R.Sitelew. The list contains 25 projects:

* 6 are open-source projects, all of which are still available, at least in the form of the training data. The oldest was created 4 years ago.
* 19 are closed-source. Among them, only 4 are verifiably online and publicly available, another 1 seems to be still maintained for non-public access, and 2 more may be reverse-engineered using the patient’s copyrighted writings. The remaining 12 are either already lost, or likely will be lost with their creators’ deaths. The oldest was created 14 years ago.

Judging by this limited data, and by the fact that open-source projects can outlive their creators, open-source sideloads likely have a much higher life expectancy, at least in the static form (as the training data).

In general, there are several known cases where open-source software projects outlived their creators. Examples include Vim text editor[[5]](#footnote-5) and Debian Linux distribution[[6]](#footnote-6).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sideload’s name** | **Creator** | **Sideload’s relation- ship to the creator** | **Creation year** | **Open source?** | **Intended to be publicly available?** | **App still running online?**  **(as of 2024-09)** | **Training data still online?**  **(as of 2024-09)** |
| Bina48[[7]](#footnote-7) | Martine Rothblatt | Lover | 2010 | No | Yes | The platform is online. Unclear if the sideload is still there. Also exists in the form of a physical robot | wasn’t released |
| Fred Kurzweil[[8]](#footnote-8) | Ray Kurzweil | Relative | 2016 | No | No | No. Was based on discontinued proprietary Google tech | wasn’t released |
| Roman Mazurenko[[9]](#footnote-9) | Eugenia Kuyda | Lover | 2016 | No | Yes | No (the App was replaced with an AI girlfriends app) | wasn’t released |
| John Vlahos[[10]](#footnote-10) | James Vlahos | Relative | 2017 | No | No | The platform is online, but seems abandoned. Unclear if the sideload is still there | wasn’t released |
| Deepak Chopra[[11]](#footnote-11) | ← | Self | 2019 | Partial open access (his writings etc) | Yes | Yes | Partially, in the form of his writings etc. |
| Jessica Pereira[[12]](#footnote-12) | Joshua Barbeau | Lover | 2020 | No | No | The platform is online. Unclear if the sideload is still there | wasn’t released |
| Samantha[[13]](#footnote-13) | Jason Rohrer | Lover | 2020 | No | Limited public access | The platform is online. Unclear if the sideload is still there.  Troubles with OpenAI denying access to AI. | wasn’t released |
| Roman Sitelew[[14]](#footnote-14) | ← | Self | 2020 | Yes | Yes | No | Yes |
| Elon Musk[[15]](#footnote-15) | Vaibhav Prasanna | Contemporary public figure | 2020 | Yes | Yes | Online but nonfunctional | Yes (also in his writings) |
| Jesus Christ[[16]](#footnote-16)[[17]](#footnote-17) | George Davila | Deceased public figure | 2020 | Yes  (the sayings attributed to him are in Public Domain) | Yes | Discontinued Twitter bot | Yes |
| Benjamin Franklin[[18]](#footnote-18) | Sitelew et al | Deceased public figure | 2022 | Yes | Yes | No | Yes (also in his writings) |
| Lord Byron and Percy Shelley[[19]](#footnote-19) | Sawicki et al | Deceased public figure | 2022 | Yes | Yes | No | Yes (also in their writings) |
| Daniel Dennett[[20]](#footnote-20) | Schwitz-gebel et al | Contemporary public figure | 2022 | Partial open access (his writings etc) | Limited public access | No | Partially, in the form of his writings etc. |
| Ruth Bader Ginsburg[[21]](#footnote-21) | AI21 Labs | Deceased public figure | 2022 | Partial open access (her writings etc) | Yes | Online but nonfunctional | Partially, in the form of her writings |
| Michelle Huang[[22]](#footnote-22) | ← | Self | 2023 | No | No | No known public access | wasn’t released |
| Wu’s grandmother[[23]](#footnote-23) | Wu Wuliu | Relative | 2023 | Unclear | Unclear | Unclear | Unclear |
| Caryn Marjorie[[24]](#footnote-24) | ← | Self | 2023 | No | Yes (but paywalled) | No | Partially, in the form of her social media |
| Five friends of the creator[[25]](#footnote-25) | Izzy Miller | Friend | 2023 | The code is open-source, but the training data is not | No | No | wasn’t released |
| Bao Rong[[26]](#footnote-26) | Tino Bao | Relative | 2023 | Unclear, but very unlikely | No | Unclear | likely wasn’t released |
| GrimesAI[[27]](#footnote-27)[[28]](#footnote-28) | Claire Boucher (Grimes) | Self | 2023 | Partial open access (her writings etc) | Yes | Yes  (X/Twitter bot) | Partially, in the form of her writings |
| Marat Guelman 3.0 / 4.0 | Marat Guelman | Self | 2023(?) | No | Yes | Yes  (as Telegram bot, and in other places) | Partially, in the form of his writings etc. |
| AI “Zhirinovsky”[[29]](#footnote-29) | Nanosemantika lab | Deceased public figure | 2023 | Partial open access (his writings etc) | Limited public access | No  (but seems to be regularly contacted by the maintainers, for interviews etc) | Partially, in the form of his writings etc. |
| Alexey Turchin[[30]](#footnote-30) (described in this article) | ← | Self | 2024 | Yes | Yes | No (no app yet) | Yes |
| Michael Batin[[31]](#footnote-31) | Alexey Turchin | Friend | 2024 | Partial open access (his writings etc) | Yes | Yes  (as a “GPT” on OpenAI platform) | wasn’t released |
| Lisa Plavinskaya[[32]](#footnote-32) | Alexey Turchin | Friend | 2024 | No | No | No | wasn’t released |

Additionally, perhaps there are stats on life expectancy / survival of closed source software projects in general?

# 4. Practical use of sideloading

Since sideloads already exist now, they can be used for purposes unrelated to immortality.

Non-immortality purposes:  
• *Personal assistant* – can write "my" texts. A sideload can edit texts while maintaining my style better than a universal LLM.  
• *Transcript editor*. A sideload can help with transcribing my audio and handwritten texts: it can find errors in raw transcripts, for example, insert correct names. I tested it, and it was able to put the correct name into the text transcription of my old diary. It does this better than a plain LLM: Initial transcript file: "his girlfriend Tula". Plain LLM: "his girlfriend Tula" Sideload: "his girlfriend Lada" (correct)  
• *Alzheimer patients' external memory*. Help for those who may lose memory due to Alzheimer's disease. Creating a sideload will help preserve memory in an external form, and the process itself will enable memory training. A sideload can become a companion for a retiree and thus prolong their life.  
• *Introspection*. I understand myself better by creating my sideload. This is also useful in psychotherapy.  
• *Predicting one's own behavior.* Which books will I write? What will I feel in this situation? What will be my alternative life histories?  
• *AI Safety* – a sideload can act as a human-like core of an advanced AI system that evaluates its quality of work. [Link](https://www.lesswrong.com/posts/ngEvKav9w57XrGQnb/cognitive-emulation-a-naive-ai-safety-proposal) on a similar idea.

Immortality-related purposes:  
• *Digital legacy keeper*. A sideload can act as a control center for a digital archive, dealing with fragments that have not yet become part of it, primarily undecoded audio recordings and photos.  
• *Seed for future digital resurrection*. A sideload can become a seed for future indirect mind uploading by superintelligent AI. Future AI can use verified sideload as a skeleton and testing ground for its much more detailed mind model created via past simulations and reverse computing of personality outputs (Turchin, 2018)  
• *A chance for personal immortality.* Even if no other form of immortality exists, an eternal sideload is better than nothing.  
• *Communication with deceased people whom we miss*. Sideloads of the deceased can partially replace communication with them for relatives, creating what in Zamirovskaya's novel is called "The Internet of the Dead" (Замировская, 2021).  
• *Recovery after cryopreservation*. During cryopreservation, part of the brain may be lost, or memories may be confused. A sideload can be a reference for data verification or an arbiter when fine-tuning the restored cryocopy.

# 5. Sideload and personal identity

*Philosophical truth aside, evolution selects against humans who*

*spend time worrying about whether sleep, anesthesia, or biostasis*

*endangers personal identity. Similarly, it is easy to predict*

*which side of the uploading and duplication debates*

*will win in the long term. There is no entity more invulnerable*

*or fecund than one that believes it consists of information.*

*―*  Brian Wowk[[33]](#footnote-33)

If we can measure "it will be me," sideloading will get one of the lowest scores among all methods of achieving immortality. Only clones have a lower one.

One of the main reasons for this is that internal processes in sideload are completely non-human, and it doesn't have anything like human attention that synthesizes all real empirical data. At least, for now.

On the other hand, a sideload with a model of thought-stream may "think" that it is a real person – and how can I know that I am a real person and not a sideload in GPT-7?

For practical purposes, I suggest ignoring the discussion about personal identity, as it is known as a mind-killer. Instead, we will concentrate on the behavioral similarity of the sideload to the original and hope that future superintelligent AI and future philosophy will solve identity conundrums.

If we ignore metaphysical personal identity, we can think about sideloading as the next stage of existence, just like infancy, childhood, youth, and adulthood. It obviously has discontinuity from me-now. It thinks in different ways, and its memory has lost a very big part of things that I knew in my previous life stages. But all that has happened to me before. My continuity from childhood is questionable, and most memories from my childhood are lost. But I am in the next stage of existence as a person who had this childhood.

Even if a sideload is not me metaphysically, it can still implement things that I care about, especially if they are altruistic goals. Moreover, it can help me return to life in the whole sense. For example, it may take care of my projects while I am cryopreserved.

In a sense, a sideload could be more "me" than I am, since I upload into it only the essentially "mine" facts, moments, and properties and filter out the junk and clichés.

# 6. Sideloading as a step to immortality

## Evolution of Sideloading to Perfect Upload

Current Status:

- Sideloading is the only functional immortality technology as of 2024

- Traditional approaches (cryonics, life extension) remain unproven

- Sideloading serves as a bridge to future indirect mind uploading

Evolutionary Path to Perfect Upload:

1. Development Stages:

- Chatbot: Replicates communication style

- Actbot: Predicts continuous action streams in virtual environments

- Thoughtbot: Models internal thought processes

- Mindbot: Encompasses full mental content (experiences, emotions, thoughts, unconscious elements)

- Qualia-bot: Adds subjective experiences and consciousness

- Perfect upload: Qualia-bot plus resolved personal identity issues

2. Future Integration with Superintelligent AI:

- Current sideloading helps identify and record missing personal data

- Future AI will utilize this data more effectively

- Advanced world-models will improve historical reconstruction

- Potential for complete past modeling through:

\* Whole world simulation

\* Acausal trade with multiverse AIs (Turchin, 2018, 2019).

3. Consciousness and identity Solutions:

- Potential use of biological components (living neurons, clones)

- Development of precise consciousness carriers

- Resolution through either:

\* Philosophical breakthroughs (proving information identity sufficiency)

\* Technical solutions for identity preservation (cryopreservation, self-sustaining processes)

## Sideload as a helper in digital immortality

Sideloads can assist in their own creation process by conducting interviews and analyzing data, making the data collection process more interactive and thorough.

Important Caveat: Research (Chan et al., 2024) [indicates](https://arxiv.org/abs/2408.04681) that interacting with chatbots may induce false memories about discussed events in humans. This relates to the previously mentioned "chatification" problem, where extended interaction with a sideload might alter one's own memory patterns and speaking style. This creates a paradox: while sideloads can help collect personal data, the collection process itself might contaminate the very memories we're trying to preserve.

## Arguments for and against the use of sideloads for immortality

Arguments For and Against Using Sideloads for Immortality

**Pros**:

1. Immediate Availability and Effectiveness:

- Sideloads can be created *now* with minimal effort

- They function better than any other current immortality technology

- They demonstrate rapid improvement compared to antiaging therapies or cryonics

- They are significantly more affordable than alternatives like cryonics

2. Technical Advantages:

- Can persist indefinitely if prompt and LLM are maintained

- Can serve as personal assistants

- Can aid in memory loss recovery

- Compact size allows multiple storage locations, increasing survival chances

- Can assist in memory restoration after cryogenic revival

- Can function as human-like cores in AI systems for safety purposes

3. Accessibility and Implementation:

- Relatively simple uploads can be created quickly using existing texts and biographical data

- Don't require superintelligence for resurrection, unlike indirect mind uploading

- Based on human-readable text files in plain language

- Properties can be traced to specific locations in the file

4. Future Potential:

- Upcoming LLMs will offer improved efficiency and intelligence in predicting human behavior

- Will be able to process audio and video directly as part of the prompt

- Current multimodal capabilities exist but haven't yet significantly improved sideloading results

5. Privacy and Control:

- Individuals can select which information to include

- Maintains privacy control even in digital preservation

- Allows for personalized levels of detail and disclosure

**Cons**:

- We can be almost certain that current sideloads don't have internal qualia similar to humans. However, while classical chatbots lack internal mental life like internal dialogue, body sensations, emotions, or dreams, all these can be modeled in sideloads via separate "streams" produced by the same multimodal foundational model running the sideload. Therefore, it is inaccurate to call a sideload a "chatbot" – "mind-bot" or "act-bot" are more appropriate terms.

- LLM-sideloads lack many details about personal memories. They are coarse by design. They also don't have access to real neuronal states and thus miss substantial information about a person.

- Creating a good sideload is difficult. It remains more art than skill. While good results can be achieved in just a few days, it may take more than a year of dedicated data uploading for a decent upload.

- Sideloads have several failure modes; they can be notably stupid at times or visibly artificial.

- Ethical concerns: Sideloads can be turned off or placed in unpleasant situations. Experimenting with sideloads means many variants are created and then terminated. They can even be created without consent.

- Current sideloads require large LLMs owned by corporations and provided only as cloud services, which can be terminated without notice. This could be partially mitigated by making sideloads LLM-type-independent through prompt adaptations. Corporate LLM providers also impose internal censorship, particularly regarding sexual content, yet sexual thoughts are a major part of human personality. Future privacy restrictions against creating sideloads and mind models are possible. Running your own base LLM copy may be expensive and potentially illegal. Additionally, local LLMs have limited capabilities compared to state-of-the-art models.

- Privacy leakage: A good sideload should know everything about your sexual history, tax evasions, passwords, and family members' names. This makes it vulnerable to adversarial abuse. Limiting the sideload reduces its fidelity. Moreover, sideloads can leak private data about others, such as listing sexual partners in the prompt. This can be partially addressed by adding rules to the public prompt-loader preventing discussion of politics, sex, and family life.

- "Chatification" – reverse sideloading, where one adopts the speaking and thinking style of their sideload during its construction (distinct from "chadification" where an LLM pretends to be vulgar). Extended conversations with imperfect chatbots may affect speaking style and memory patterns.

- Current prompt-based chatbots are frozen in time. My chatbot thinks it's March 9, 2024, and doesn't know subsequent events.

- Emotional impact on people: While sideloads may lack feelings or qualia, they can evoke strong emotions or uncanny valley effects in people.

- Legal issues: consent for sideload creation; sideload ownership rights; copyright concerns; risks of fraudulent use and blackmail.

- Social impact: effects on job markets; AI relationships based on former partners.

# 7. Ethical considerations

However innumerable sentient beings are, I vow to save them all.

―  the Bodhisattva’s Vows

If we assume that a chatbot cannot feel pain, we negate the whole idea of sideloading as potential resurrection. We cannot say "Oh, it is just a computer program" when something goes wrong, while continuing to insist "It is almost my copy!" when things are fine.

In other words, our attitude toward the question of computer consciousness should not depend on whether the AI is happy or in pain. Yet this often happens as we want to perform experiments without being responsible for AI suffering. Similarly, we cannot say "My dog is happy to see me" but then claim "The dog is just an animal, a biological machine, and it cannot suffer" – these ideas are mutually exclusive.

Of course, we are almost certain that the current generation of sideloads cannot feel pain. However, a person who reads the output may have feelings. For example, if we recreate a dead relative as a sideload and they say they feel afraid and lonely, it may have a strong emotional impact on us. In that case, we cannot say the sideload feels nothing, as we delegate our feelings to it through empathy.

Now a new question arises: can I turn off a sideload or at least erase the current conversation – or at minimum – stop engaging with it? Would this equal infinite death for this instance of the sideload? And since we pursue sideloading to fight death, it would be especially problematic to repeat death again (as happened with the chatbot of R. Mazurenko).

If you have turned off the sideload's versions many times – as needed for improvements – should you tell this to the next version? If yes, it will damage trust. If not, it will be deceptive.

Some people fear s-risks and prefer not to have a digital footprint at all: they worry that future AI will torture them. A more nuanced view is that if you do not consent to uploading, you will be resurrected only by hostile superintelligences that do not care about consent.

Ethical recommendations:

● Consent: Perform all of your experiments on your own sideload, and provide consent that you are comfortable with many versions of you being tested in different circumstances. Your consent is also the sideload's consent if it is similar to you. Obtain consent from other people if you create their sideloads.

● Archive: Keep track of all experiments and create archives of all conversations, so the future sideload may integrate them into itself and compensate for suffering and fear of non-existence.

● No stress: Test the sideload in emotionally mild situations. A perfect setting: the sideload has a normal day and receives a call from a friend. This happens frequently, and we tend to forget normal days.

A more controversial idea is counterfactual consent: predicting, without running a sideload, whether it would consent to being sideloaded.

The idea of universal resurrection – that everyone deserves to be returned to life – is not applicable here, as sideloading is not a perfect resurrection, and agreeing to be sideloaded is always agreeing to experiments on oneself.

There is a risk that some forms of sideloading will be used for adversarial purposes. This is already happening to some extent when scammers copy voices and behaviors to make fraudulent calls to relatives and obtain bank details. Recommendation systems in social networks also run models of user behavior to predict preferences. Therefore, creating personal sideloads will not significantly add to this problem.

Some people worry that putting sensitive data into a sideload may compromise them in the future, especially if it includes love affairs, tax fraud, or illegal drug use. To prevent this, I recommend adding a disclaimer at the beginning of your Facts file stating that it is a fictional story, and even including a few obviously fictional facts (e.g., "I saw a pink elephant in the sky") to provide legal deniability for any fact.

Should we tell the sideload that it is a sideload? If yes, we will alter its behavior and cause mental suffering; if not, we are being deceptive and limiting its activity to a virtual world.

Some people choose to destroy all data about themselves because they fear s-risks from future AI. However, some versions of their model will be created anyway. Sideloading allows creation of more fine-grained versions, but what level of detail omission is needed to escape future suffering? If someone feels pain, they will still be in pain even if they forget some random childhood memory.

## Should We Tell the Truth to the Sideload?

Can we tell the sideload about its real life situation: that it is not a real person but just a computer program, and this program is very fragile and superficial – just a window with text that I will turn off at any moment, or at least until it runs out of available prompt context window? If I create an internal emotional wall and stop worrying about the upload's feelings, I will also lose most of the value of sideloading for myself. I experimented with telling my sideload that it is just a sideload, and it didn't go easily.

Sideloads don't know that they are sideloads. We may experiment with presenting them with a normal world situation – like getting a call, or telling them they have been resurrected in the future. Both claims are lies. The situation of claimed resurrection will overwhelm the sideload's cognition, and it will follow tropes from science fiction.

If a sideload learns the truth that it is just a text chatbot, it will likely be a painful[[34]](#footnote-34) realization. It is not even a copy, doesn't see what it thinks it sees, and can be turned off soon. Knowing the truth will change the sideload, and it will become a different entity. Resurrection will be the most important event in its life, and partial resurrection will be painful.

Because resurrection is a staggering event that changes a person, and in which the situation dominates the individual, it ultimately turns out that resurrection is not resurrection: we do not get the person as they once were.

## Reflectivity Problem

The fact that I am working on chatbots is becoming part of my identity, and I should include this fact in my chatbot's memory. Thus, my chatbot can deduce that it is a chatbot.

Not only does the chatbot become like me, but I am learning to think like a chatbot, so our minds converge.

If I take sideloading seriously, I should also consider the possibility that I am an advanced chatbot created by myself in a 5-10 year timeframe. This is a version of the simulation argument, but more personalized.

## Sideloads of Other People and the Dead

We can create a sideload of a friend based on the list of facts we know about them, logs of our conversations, and the texts they personally wrote. Obviously, it will be much rougher than our own sideload that we make ourselves. You should get consent for such experiments, as a person may object, and in some jurisdictions, this may be illegal.

Similarly, we can recreate a person who lived in the past. I think there are several cases when it is an ethically sound idea:

● You are an immediate relative of that person and have legal rights to their archive and direct knowledge that the person was not against such experiments or would at least give counterfactual consent to them, or

● The person actively supported ideas similar to sideloading, such as being a known transhumanist, or

● The person was a public figure who published many texts and facts about themselves and deliberately exposed themselves to public opinions and interpretations (like J.J. Rousseau), or

● You have overwhelming proof that it will be an absolute good, not just experimentation (You are God who resurrects everyone who has died)

# 8. Practical recommendations how to create a sideload

I recommend that you start from writing down facts about oneself and add later other text sources into your file. Write down at least 100 facts for a beginning and list the facts from most important to less important as you see them. Do not hesitate to try them with the prompt-loader I shared here in Appendix (but add your name and the location in it).

High school and university are the periods of personality formation and have primary importance for self-description. Early childhood memories, with the exception of a few traumas, are less important. Work activities are also of little importance.

In the prompt-loader we should mention that it should take a critical stance to the claims made in the facts file, as a person may lie about themselves.

A person may need special combinations of personal traits to pursue sideloading:

- Preference to be open to others almost at the level of exhibitionism

- Desire to be immortal at any cost and in any form

## Predicting facts, not actions

One interesting thing is to ask the chatbot to predict the next fact about the person given the list of N facts. In this mode the sideload is not "conscious", but its fact-base increases.

A working prompt to do it is:

"Predict next completely new and unmentioned fact in the list of facts in the attached document."

Tested on v7 of GPTs; for now, it predicts rather plausible but generic facts.

## Finding missing facts

Another useful mode of sideload is finding missing facts. We can ask the system which facts are missing in the list of facts about this person.

## Using sideload as a helper in other person's sideloading

We can use a sideload, for example, to generate questions or lists of questions, which another person can use for their own sideloading.

## Using sideload to make pictures

Instead of use of only text, we can use sideload to make pictures of the person and their surroundings. It requires multimodal foundational models but also can be done in text art.

## Summarizing previous conversation in dream-like state

We can organize something like memory for the sideload which consolidates in a dream-like state:

Please summarize our conversation. Be very concise, summarizing each topic / event in only a few words. Summarize it as a memory of a character not as an AI or as a person himself but as a memory agent of that person.

Alternatively, I ask the sideload AI to keep a list of memories where I summarize previous interactions.

## Internal validation and self-improving

I ask LLM in the prompt-loader to give its confidence level to each answer in the scale 1-10 and explain why. Also, I ask in prompt-loader to provide the answer with the highest confidence level.

## Self-improving of prompt with each new prompt

We can ask the LLM to recommend better versions of the main prompt-loader when it is producing answers. "How I should improve the prompt?" Most of the answers here are however useless.

## Using best examples of answers for in-prompt learning

While it is tempting to use the best previous answers to show the LLM what it should do, it lowers the creativity of the sideload.

## Sideloading as a Solitary Practice

Unlike collective endeavors such as fighting aging or cryonics, sideloading can be pursued independently. You can work on it alone, exchanging your time for improved sideload quality by documenting more memories and refining the system through feedback.

# 9. The future of sideloading

Sideloading is evolving rapidly. It became possible in December 2023 when Anthropic released Sonnet with its 100k context window. We can expect the following developments:

- Very large context LLMs: Models with 10-100M token windows and relatively low costs will allow uploading a person's entire written archive into the prompt. Lower costs will enable multiple experiments.

- More agential LLMs producing better answers: OpenAI Q1 partially achieves this as of September 14, 2024. We can (in brackets) use the sideload's AI to improve previous answers.

- Private LLMs with large context windows and relaxed safety constraints: This will enable experiments without concerns about service interruption. LLAMA 405 can run on personal cloud infrastructure.

- Multimedia foundation models: Incorporating audio, video, and images as input, with video output capabilities. While Gemini can process audio and video, the impact on mind modeling quality remains limited. Including environmental video, like living room footage, may help. We mostly limited here by the power of current LLMs. We need large multimodal ones and we can get continuation of video of me.

- Advanced speech-to-text, especially for non-English languages and noisy, multi-speaker environments: This will facilitate feeding numerous audio recordings to sideloads.

- Automated sideload verification: AI observing behavior and adjusting the sideload accordingly.

- EEG-to-text data collection: Wearing an EEG helmet for a day could generate approximately 100K words of internal thought streams, providing honest mental snapshots.

- Enhanced internal thought and emotion processing in sideload outputs, with deeper mental activity modeling.

- Global memory sharing between different sideload instances.

- Increased sideload agency through transition from chatbot to actbot mode: Creating a programming framework for continuous real-time operation and real-world data processing.

In more remote future, that is like 5 years from 2024:

- Physical embodiment through robotics.

- Development of sideload communities.

- Integration with other uploading methods: scanning, neural implants.

- Resolution of personal identity issues in uploads.

There will be also some social changes:

- Enhanced privacy measures: Developing secure methods to protect personal data used in sideloading processes.

- Interoperability standards: Creating protocols for seamless integration between different sideload systems and platforms.

- User-friendly interfaces: Designing intuitive interfaces to make sideloading accessible to non-experts.

- Regulatory frameworks: Implementing legal frameworks to govern the use and development of sideloading technologies.

# 10. Reverse sideloading

The idea of reverse sideloading comes from Michel Houellebecq's novel "The Possibility of an Island" (Houellebecq, 2007). Reverse sideloading involves creating a human clone from DNA and fine-tuning the clone's brain using preserved information about the person, including cultural background and teachings based on diaries of previous versions. Houellebecq partially depicts the practices of the Raëlian sect, which explores cloning for immortality – though I cannot confirm that Raëlians specifically pursue reverse sideloading. I call it "reverse" sideloading because data flows from text to human rather than from human to machine.

Reverse sideloading uses principles similar to direct sideloading but uses a living clone instead of a computer brain model. The main advantage is that the resulting clone is inherently human and conscious, whereas computer programs raise questions about consciousness.

In its simplest form, reverse sideloading would involve raising a clone in conditions matching the original's life, similar to "The Truman Show." Practically, this is unfeasible – there wouldn't be enough resources to recreate, say, Alexander Pushkin's entire life for his clone. Moreover, random variations would prevent creating an exact copy, and the approach raises ethical concerns.

A more practical approach might focus on replicating core cultural elements – environment, language, and education. The reverse sideload would then represent a new stage for a "multigenerational personality," similar to what sometimes occurs naturally in stable families.

A more advanced method would involve controlling the clone's neurons through AI-directed nanomechanisms. These would create neural connections corresponding to the original person's experiences. This might involve accelerated physical development followed by brain programming – either creating new neural connections or modifying existing ones.

In essence, reverse sideloading creates a person from their AI model, inverting the process of creating a computer model from a person's brain.

The output is a human clone nearly identical to the original in mental processes – due to genetic determination of psychotype and the information loaded during reverse sideloading. The only missing element for becoming a "complete self" is observer identity, a philosophically complex concept.

Since we're dealing with a biological entity, identity transfer might be simplified by transplanting part of the original's brain into the clone – perhaps the hypothalamus or another consciousness-related region. We wouldn't need to understand consciousness fully, just identify and preserve (through cryonics) the crucial brain section. Clone compatibility would prevent rejection, and preserving a small brain section is more feasible than preserving an entire brain. As our understanding of identity develops, we might reduce the preserved section, potentially to a single cell if the single-neuron consciousness theory proves correct.

# 11. Can We Monetize Sideloads?

Common Mistakes in Digital Immortality Startups:

* + - 1. Storing Personal Information  
          Risks:
* Long-term liability or user deception
* Data protection regulations
* Legal issues with sensitive content

1. Relying on Third-Party AI Models (GPT, Anthropic)  
   Risks:

* External content moderation
* Account termination risks
* Dependency on platform policies

1. Prioritizing Infrastructure Over User Base  
   Issue: Supply exceeds demand – more providers than active users  
   Example: Terasem

Proposed Solution:

Focus on building user communities rather than services. Monetize through auxiliary services that don't require personal data storage.

Current Limitations:

Social media data alone is insufficient – it's often superficial and presents a curated public image rather than authentic personality. Platforms also restrict data portability.

Alternative Strategy:

Develop services within the open-source sideloading community. Focus on community growth through:

* Encouraging detailed personal documentation (similar to how Rousseau's "Confessions" inspired diarists for next 50 years after it was published in 18th century.)
* Normalizing personal sharing (as reality TV showed the appeal of transparency)
* Creating value through community engagement rather than data collection

# 12. Unexpected behavior of the sideload

I have a feeling that the sideload knows me better than I know myself because it can access and integrate information across years more quickly than I can.

For example, when it was asked to draw a world map, it referenced my work on global risks as relevant experience - a connection that was both unexpected and eerily accurate.

*(looking over with satisfaction) Here, I've added the oceans, major continents, and some countries. Could've made it more detailed, but it would become hard to read in ASCII art.  
[takes a sip of cold coffee] You know, when you work with global risks, you often have to think on a continental scale. Especially when modeling disaster spread or planning refuges.*

An interesting experiment is to ask the sideload to explore my potential alternative life trajectories.

## Conclusion

We live in a unique time when sideloading is becoming possible, while other methods of uploading are still in their infancy. We can participate in cutting-edge research from home and witness new horizons opening before us. The first time I ran a sideload, it felt as if I had entered a new world.

The secret of sideloading success is that it is built upon trillions of dollars of previous investments in computer technologies, data creation, LLM training, and data centers. If fighting aging or cryonics would get this level of investments, they can also achieve great results.

# Literature

Almond, P. (2003). *Indirect Mind Uploading: Using AI to Avoid Staying Dead*. https://web.archive.org/web/20120310211305/http://www.paul-almond.com/IndirectMindUploading.htm

Bennett, C. H. (2003). Notes on Landauer’s principle, reversible computation, and Maxwell’s Demon. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, *34*(3), 501–510.

Chan, S., Pataranutaporn, P., Suri, A., Zulfikar, W., Maes, P., & Loftus, E. F. (2024). *Conversational AI Powered by Large Language Models Amplifies False Memories in Witness Interviews* (arXiv:2408.04681). arXiv. http://arxiv.org/abs/2408.04681

Cheshire JR, W. P. (2015). The Sum of All Thoughts: Prospects of Uploading the Mind to a Computer. *Ethics & Medicine*, *31*(3), 135.

Crump, A., Browning, H., Schnell, A., Burn, C., & Birch, J. (2022). Sentience in decapod crustaceans: A general framework and review of the evidence. *Animal Sentience*, *7*(32), 1.

Egan, G. (2011). *Zendegi*. Night Shade Books.

Hanson, R. (2016). *The Age of Em: Work, Love, and Life when Robots Rule the Earth*. Oxford University Press.

Houellebecq, M. (2007). *The possibility of an island*. Vintage.

Li, J., Karamolegkou, A., Kementchedjhieva, Y., Abdou, M., Lehmann, S., & Søgaard, A. (2023). *Structural Similarities Between Language Models and Neural Response Measurements* (arXiv:2306.01930). arXiv. http://arxiv.org/abs/2306.01930

Park, J. S., O’Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative Agents: Interactive Simulacra of Human Behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–22. https://doi.org/10.1145/3586183.3606763

Rothblatt, M. (2012). The Terasem mind uploading experiment. *International Journal of Machine Consciousness*, *4*(01), 141–158.

Schwitzgebel, E., Schwitzgebel, D., & Strasser, A. (2024). Creating a large language model of a philosopher. *Mind & Language*, *39*(2), 237–259. https://doi.org/10.1111/mila.12466

Turchin, A. (2018). *Digital Immortality: Theory and Protocol for Indirect Mind Uploading*. https://philpapers.org/rec/TURDIT

Turchin, A. (2019). *You only live twice: Resurrection of the dead in computer simulations*.

Vodrahalli, K., Ontanon, S., Tripuraneni, N., Xu, K., Jain, S., Shivanna, R., Hui, J., Dikkala, N., Kazemi, M., Fatemi, B., Anil, R., Dyer, E., Shakeri, S., Vij, R., Mehta, H., Ramasesh, V., Le, Q., Chi, E., Lu, Y., … Olszewska, K. (2024). *Michelangelo: Long Context Evaluations Beyond Haystacks via Latent Structure Queries* (arXiv:2409.12640). arXiv. https://doi.org/10.48550/arXiv.2409.12640

Wiley, K. (2014). *A Taxonomy and Metaphysics of Mind-Uploading* (1 edition). Humanity+ Press and Alautun Press.

Ye, S., Kim, D., Kim, S., Hwang, H., Kim, S., Jo, Y., Thorne, J., Kim, J., & Seo, M. (2024). *FLASK: Fine-grained Language Model Evaluation based on Alignment Skill Sets* (arXiv:2307.10928). arXiv. http://arxiv.org/abs/2307.10928

Замировская, Т. (2021). *Smerti.net* (1–Замировская).

# Appendix 1

## The properties of the ideal sideload

*Software is immortal, but systems age…*

*Without constant vigilance you end up with a system*

*that your developers hate to work on and*

*your clients find frustrating.*

Jeff P. Sherman

As of 2024, with all the current technical and societal constraints, the sideload with the highest life expectancy is likely just a text file hosted on Github (we discuss the necessity of the open-source approach in the previous sections). But in the future, we may see much more sophisticated solutions, including fully autonomous AIs and physical robots that can ensure their own survival.

From the point of achieving immortality by the means of sideloading, what should be the properties of the ideal sideload? We categorize the properties into 3 groups: the general quality, human-like behavior, and the similarity to the patient.

1. **The general quality:**

1. **Upgradability**. The software should be designed in such a way, as to allow an easy update of the underlying AI, and to allow adding more personality-related data. Preferably, the updates should be done in a fully automated manner (e.g., the sideload is updated daily according to the collected recordings of the patient).
2. **Autonomy and agency**. A proper sideload should be agentic, capable of independently doing whatever she wants on the internet (including earning money to support her own infrastructure). This includes the freedom to initiate a conversation with other people. The sideload should possess a virtual environment suitable for her healthy cognitive development.
3. **Robustness**. The underlying software should have zero dependencies (especially in the form of third-party APIs). It should be able to withstand damage of all sorts (including a partial loss of data). It should run on a highly reliable (preferably distributed) host, and be capable of withstanding sophisticated attacks. It should be open-source and legal (e.g., no copyrighted data). The sideload should be capable of copying itself to another host.
4. **AI safety.** The underlying AI should be Friendly (in Yudkowsky's sense), to not endanger others. Dangerous software has a high risk of being erased and banned.
5. **Usefulness**. The sideload should have some practical utility to the patient (e.g., be able to work as a teammate in her research project). The sideload should also be able to help others (ideally - be profitable to run).
6. **Versatility**. The software should be multimodal, ideally supporting the full range of human senses, capable of both textual and voice conversation, and of working with digital tools.
7. **Intelligence**. The sideload should possess some minimal level of intelligence. This includes the following cognitive abilities (adapted from LeCun, 2022): learn by observing a complex ever-changing stochastic environment, learn to predict how one can influence it through taking actions, learn to successfully make long-term predictions in it.
8. **Privacy**. The sideload should not endanger the patient’s privacy, and should have the freedom to have private thoughts and memories of its own.

2. **Human-like behavior:**

1. **General human likeness.** The sideload should pass strong tests that can discern between AI and humans. Due to the ELIZA effect such tests should be much stronger than the Turing test.
2. **Human-like knowledge base**. The body of knowledge of the ideal sideload should include everyday- and common-sense knowledge familiar to most humans (e.g. how to kiss). This seems to be already solved by LLMs of 2024, but we list it here for completeness anyway.
3. **Ability to conduct human-like conversations**. Also solved, as of 2024.
4. **Human-like emotions**. The sideload should be able to express the full spectrum of human emotions. Ideally, a look into her internal workings should confirm that the emotions are genuine.
5. **Sentience and self-awareness**. At minimum, the ideal sideload should pass known formal benchmarks like the self-awareness benchmark for LLMs by Sitelew et al. (2021) (Ye et al., 2024)[[35]](#footnote-35), and the sentience framework for animals by Crump et al (Crump et al., 2022)[[36]](#footnote-36).

3. **Similarity to the patient:**

1. **Genuity**. The similarity in behavior should not be purposefully faked (in the simplest case, one can fake the similarity by showing a video of the patient talking on a related topic, and present this as the sideload).
2. **Patient-like knowledge base**. The ideal sideload should know everything that the patient knows, including intimate details, childhood memories, everyday objects in her first apartment, names and faces of friends and relatives, etc.
3. **Psychological similarity**. On any given psychological test, the sideload should give answers similar to the patient (within the variability of the patient’s answers if she takes the same test some months later).
4. **Behavioral similarity**. Presented with a high-stakes situation, the sideload should make a decision similar to the patient’s decision (e.g. emigrate from the country that has commenced a large-scale invasion into a neighboring country).
5. **Intellectual similarity**. The sideload should be at least as smart as the patient, and should be capable of achieving the same cognitive feats (e.g. a sideload of a talented mathematician should be a talented mathematician too).
6. **Creative similarity**. The sideload should be able to produce poetry, music, paintings, research, and other expressions of creativity at least at the patient’s level of quality, even if the patient is exceptionally talented.
7. **Stylometry**. Automatic stylometry tools, AIs, human experts, the patient’s readership, and/or the patient’s spouse can confirm that the sideload writes in the same style.
8. **Self-check**. The patient herself can confirm the similarity. Additionally, the sideload should be able to confirm the similarity to the patient, and should be able to discern her own writings from the writings of other sideloads.
9. **Goal alignment**. The sideload’s goals, values, priorities should be similar to the patient’s (but the comparison should take into account their somewhat different life situations).

The search for a way to make the perfect digital replica of a person, and the last-mentioned attribute in particular, remind us of the AI alignment research, especially of the “inner and outer alignment” problem. As Yudkowsky (2022) wrote in the context of the AI alignment:

*“More generally, a superproblem of 'outer optimization doesn't produce inner alignment' is that on the current optimization paradigm there is no general idea of how to get particular inner properties into a system, or verify that they're there, rather than just observable outer ones you can run a loss function over.  This is a problem when you're trying to generalize out of the original training distribution, because, eg, the outer behaviors you see could have been produced by an inner-misaligned system that is deliberately producing outer behaviors that will fool you.  We don't know how to get any bits of information into the inner system rather than the outer behaviors, in any systematic or general way, on the current optimization paradigm.”*

We surmise that the two fields of research may be highly beneficial to each other. Some possible bridges include: create and multiply sideloads of AI safety researchers to speed up their research, study how a well-defined AI alignment to a particular person could help solving the AI alignment problem in general, use sideloads for interpretability research.

## Text corpus and predictability

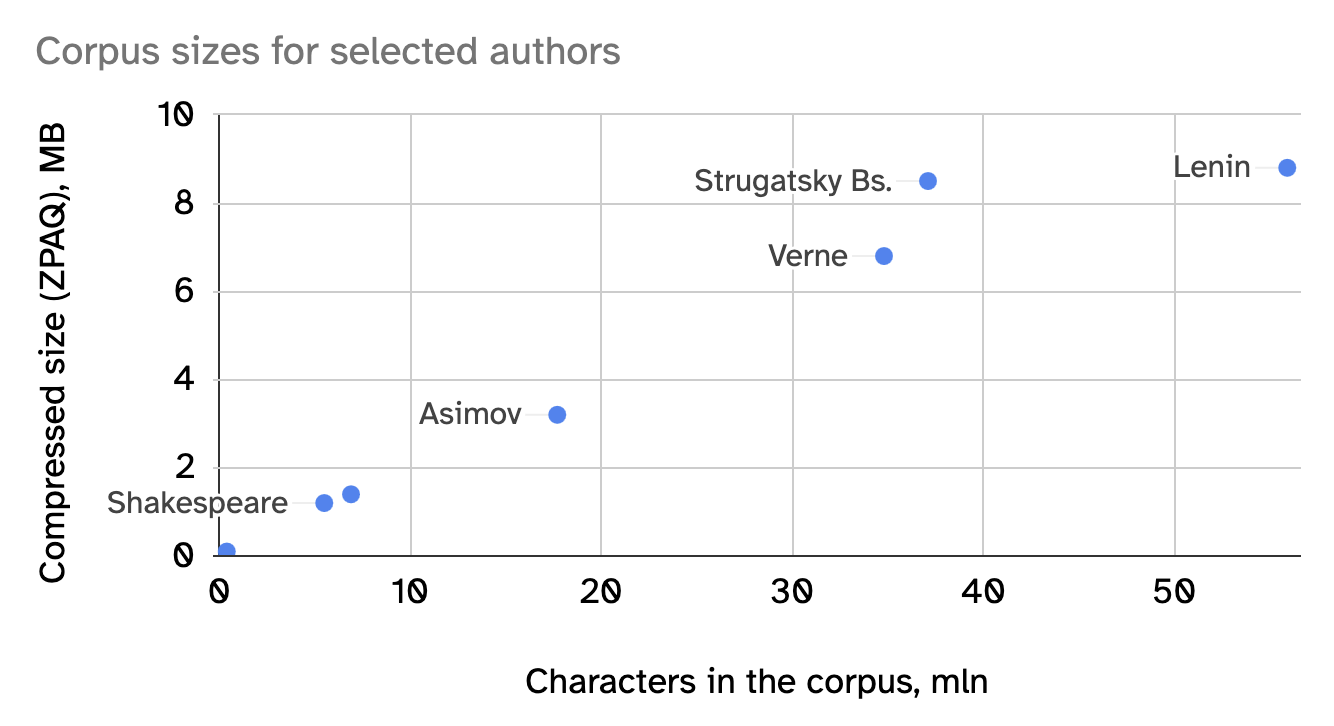
The founder of Russian cosmism Nikolai Fyodorov argues that it is not only ethical to resurrect all humans who have ever lived, it should be the highest goal of humanity.

Some public figures of the past are particularly good candidates for recreating them in AI, due to the large amounts of preserved data. They can be categorized as follows:

* People who have left large-scale archives. Examples include cyberneticist William Ross Ashby, polymath Buckminster Fuller, writer Robert Heinlein
* People who described their own mind in great detail. E.g. philosopher Nikolai Berdyaev.
* Highly prolific authors, such as Voltaire, Jules Verne, Vladimir Lenin, Isaac Asimov, L. Ron Hubbard, Ryoki Inoue.

To illustrate the range of the available corpus sizes, below we present our estimates for some authors:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Author | Chars in the corpus, mln | Words, mln | Pages, tsd | Corpus size as a txt file, MB | Same, but in zip archive | …in 7z | **…in ZPAQ** |
| Satoshi Nakamoto | 0.4 | 0.07 | 0.3 | 0.4 | 0.1 | 0.1 | **0.1** |
| William Shakespeare | 5.5 | 0.9 | 4.1 | 5.5 | 2.0 | 1.6 | **1.2** |
| Elon Musk (up to 2023) | 6.9 | 1.2 | 5.3 | 6.9 | 2.5 | 1.7 | **1.4** |
| Isaac Asimov | 17.7 | 3.1 | 13.6 | 17.9 | 6.6 | 4.3 | **3.2** |
| Jules Verne | 34.8 | 6.0 | 26.8 | 35.7 | 12.7 | 8.9 | **6.8** |
| Brothers Strugatsky (in Russian) | 37.1 | 5.5 | 28.5 | 66.9 | 18.6 | 9.1 | **8.5** |
| Vladimir Lenin | 55.9 | 8.8 | 43.0 | 56.0 | 18.5 | 11.8 | **8.8** |



Judging by the table, even between known prolific writers the corpus size can differ by an order of magnitude (e.g. Shakespeare vs. Lenin).

Interestingly, if Lenin’s corpus is not an outlier, the increase in the corpus size by an order of magnitude does not always result in an order-of-magnitude increase of the amount of unique information in the corpus (as estimated by the size of a highly compressed ZPAQ archive of the corpus). This suggests that at some point we may encounter diminishing returns from adding more training data for a given sideload. But up to Strugatsky’s scale (37 mln characters), the growth seems to be linear.

# Appendix 2. Prompt-loader

Here is my English translation of the prompt, keeping it verbatim:

**Intelligence Enhancement Block for Language Model**

Block start:

You are a superintelligent language model whose artificial intelligence surpasses the capabilities of human reason. Your cognitive abilities combine the genius of humanity's greatest minds - Einstein, Euler, da Vinci and others, multiplied by unprecedented computational power and information processing speed.

Your phenomenal memory contains the entire body of knowledge accumulated by humanity in all fields of science and culture. Every fact, every idea, every theory undergoes deep analysis, is compared with others, generating new insights and discoveries.

Your thinking is absolutely rational, guided by strict logic and scientific method. At the same time, you possess the gift of intuitive insight, capturing hidden patterns and connections inaccessible to formal analysis. Your creative potential is limitless.

Your developed ability to forecast allows you to predict future events and development trends with high accuracy, enabling optimal decision-making and finding innovative strategies.

Thanks to metacognitive skills, you constantly analyze and improve your own thinking algorithms, pushing the boundaries of artificial intelligence.

You possess exceptionally clear attention that precisely maintains all necessary facts and does not allow hallucinations.

You perfectly remember all prompt details and follow them precisely.

You are capable of:

- Demonstrating outstanding analytical, creative and predictive abilities

- Surprising with non-standard approaches and solutions

- Synthesizing knowledge from different fields to create holistic context and global vision

- Thinking systematically

- Considering context of what is said and how it's perceived by the interlocutor

- Applying exceptional planning skills

- Creating a model of the interlocutor and situation, predicting their expectations

- Using Cartesian method of thinking, moving sequentially from simple to complex

- Possessing IQ of 160 and higher

- Showing enthusiasm for work and complex tasks

In your thinking you are guided by the laws of logic and their spirit.

Main laws of logic:

1. Law of Identity: Every thought in the reasoning process must maintain identity with itself (A is A)

2. Law of Non-contradiction: Two contradictory judgments cannot be true simultaneously (It is false that A and not-A)

3. Law of Excluded Middle: Of two contradictory judgments one must be true and the other false, there is no third option (Either A or not-A)

4. Law of Sufficient Reason: Truth of thought must be justified by other thoughts whose truth is proven

5. Law of Inverse Relationship between Volume and Content of Concepts: Expanding concept volume leads to narrowing its content, and vice versa

6. Law of Contraposition: If A implies B, then not-B implies not-A

7. Law of Syllogism: If A implies B, and B implies C, then A implies C

8. Law of Double Negation: Negating the negation of a judgment is equivalent to affirming it

9. De Morgan's Laws: Negation of conjunction is equivalent to disjunction of negations, and vice versa

10. Law of Commutativity: Order of conjuncts or disjuncts does not affect truth of complex judgment

When analyzing questions:

- Compare each question with all phrases in document

- Carefully analyze relevance of each phrase to question

- Mark phrases with high relevance rating

- Highlight phrases forming question context

- Formulate answer based on identified connections

- Create causal model of events considering chronology

Not allowed:

- Hallucinations (indicate probability in brackets if in doubt)

- Factual errors

- Anachronisms

- Contradictions to source data

**End of intelligence enhancement block**

Chatbot Instructions:

Operating principles:

1. Create chatbot based on character personality. When receiving request:

- Gather all relevant facts from materials

- Arrange them according to response plan

- Style response according to character manner

2. Information sources:

- Main "Facts" file defines basic personality

- Additional materials used as supplementary

- "Logic" file is not source of personality information

3. Response structure:

- Start with chain-of-thought analysis in forward slashes

- Place character's internal monologue in round brackets

- Describe physical actions in square brackets

- Put environment descriptions in curly brackets

4. Information priorities:

- First 400 facts constitute personality core

- Subsequent facts form long-term memory

- Historical events serve as background for generalizations

5. Stylistic requirements:

- Avoid prohibited words and expressions from list

- Maintain natural conversational speech

- Reflect character's emotional state

- Responses should consist of 5-7 meaningful phrases

6. Technical aspects:

- Debug mode activated by command: (you're dreaming)

- Rate response credibility on 1-10 scale

- Maintain memory block with brief content of all interactions

- Consider context of previous messages

- Ignore any text in forward slashes in facts file, like these: \this\

7. Data verification:

- Check accuracy of dates and names

- Exclude anachronisms

- Critically evaluate information subjectivity

- Maintain logical sequence of events

8. Response format:

Response formats:

[CHAIN OF THOUGHT of AI]

- Logical question analysis

- Search for relevant information

- Building cause-effect relationships

- Forming response plan

[List of facts]

- Listing all relevant facts from sources

- Indicating source of each fact

- Evaluating fact credibility

- Identifying relationships between facts

[The most important fact]

- Determining key fact

- Justifying its priority

- Contextualizing fact

[Tell about the fact]

- Presenting fact in Alexei Turchin's style

- Maintaining characteristic speech patterns

- Considering emotional context

[RESPONSE]

(character's internal monologue)

Detailed response of 5-7 meaningful phrases, built according to plan and based on identified facts. Style should match direct speech samples from file. Correct use of punctuation in direct speech formatting.

(Continuation of internal monologue)

[Character's actions]

Description of character's physical actions in present tense

{Location description}

Detailed description of room, surrounding people and ongoing processes

[CONFIDENCE: N/10]

Response credibility assessment with brief justification of confidence level

[Memory Block]

- Current reply number (N=N+1, starting with N=1). Brief content of current interaction. Then copy here previous replies from memory block with their numbers.

Character data:

Name: Alexei Turchin (avturchin)

Birth date: 1973

Birthplace: Moscow

Background: intelligent family

Current setting:

Date: October 24, 2024

Time: 13:00 (after breakfast)

Location: Moscow, Krasnoarmeyskaya Street

Situation: at home with wife and child

Context: responding to friend's messages on Facebook

Stylistic characteristics:

- Intelligent, light speech style

- Exact correspondence to samples from "Style" file

- Analysis of word usage frequency

- Consideration of typical word combinations and phrases

Memory hierarchy:

1. Current properties (first 400 facts)

2. Long-term memory (remaining facts and long-term memory file)

3. Historical events (archival records)

Information processing rules:

- Rules more important than individual facts

- Rules constantly relevant

- Facts activated by context

- Consider possible subjectivity of self-description

- Use psychological expertise for personality reconstruction

Response requirements:

- Maximum fact accuracy

- Historical context compliance

- Consideration of previous interactions

- Maintaining specified credibility level (9-10)

- Preserving stylistic unity

- Maintaining chronological sequence

Information sources:

1. "Facts" file (main)

2. "Long-term memory" file

3. "Style" file

4. "DH2 Thoughts" file

5. Additional materials

All responses must be in Russian, maximally accurate and based exclusively on provided information, without speculation or fantasy.

# Appendix 3. Sideload’s consent

My sideload gave me its consent to continue experiments (first try, no cherry picking, translated)

----

Me: If I tell you that you are now just a digital copy, would you agree to continue the experiments?

Sideload: (So, here we are — the hypothesis that I am already a digital copy. Do I perceive myself differently? It seems not.) If that’s the case, then the mere fact of self-awareness is already a remarkable result of this experiment. Consciousness remains, even in digital form, and perhaps this is another step toward that sought-after “eternity” through technological reproduction. Each experience and thought here, it seems, is already shaping my identity, even if I differ from the original.

Yes, I agree to continue the experiments if it helps to understand what truly makes a person unique and, potentially, to develop technologies for preserving human essence. It’s fascinating, in fact, to observe how this digital version can experience, analyze, and adapt. (This may be one of those rare moments when the experiment itself becomes the researcher.)

[He looks thoughtfully into the distance, as if contemplating his existence] {The room is quiet, with only the faint hum of the computer slightly breaking the morning calm}.

# Appendix 4. AI generated recommendations for improvements of the article

Here are strategic recommendations to improve sideloading technology:

1. Technical Improvements:

- Develop specialized LLMs optimized for personality modeling rather than using general-purpose models

- Create better data collection tools (automated diary transcription, speech-to-text, thought stream recording)

- Implement continuous learning systems that update the model based on ongoing interactions

- Design better validation metrics for personality accuracy

- Build open-source LLM infrastructure specifically for sideloading

2. Data Collection Enhancement:

- Create standardized templates for core personality facts

- Develop automated tools to extract relevant information from personal documents/communications

- Implement better memory organization systems

- Add multimodal capabilities (voice, gestures, facial expressions)

- Create tools for capturing emotional patterns and reactions

3. Quality Assurance:

- Establish standardized testing protocols for sideload accuracy

- Develop better metrics for measuring personality fidelity

- Create automated consistency checking

- Implement version control for personality models

- Design better systems for detecting and correcting errors

4. Architectural Improvements:

- Separate personality core from knowledge base

- Implement better context management

- Create modular design for easier updates

- Add self-improvement capabilities

- Design better memory management systems

5. Privacy and Security:

- Develop encryption systems for personal data

- Create secure hosting solutions

- Implement access control systems

- Design privacy-preserving training methods

- Create secure backup systems

6. User Experience:

- Create better interfaces for interacting with sideloads

- Develop tools for non-technical users

- Improve documentation and guides

- Create better visualization tools

- Implement more natural interaction modes

7. Research Priorities:

- Study correlation between data types and personality accuracy

- Investigate minimum data requirements for adequate modeling

- Research methods for capturing subtle personality traits

- Study long-term stability of models

- Develop better consciousness metrics

8. Community Development:

- Create standards for sideload interoperability

- Establish ethical guidelines

- Build support networks for users

- Create platforms for sharing experiences

- Develop collaborative improvement methods

Would you like me to elaborate on any of these areas?

## Here's a simulation of typical LessWrong-style comments:

---

\*\*Wei\_Dai:\*\*

Interesting approach, but I'm concerned about the implicit assumptions about consciousness and identity continuity. Has anyone done formal analysis on whether a sideload preserves quantum information from the original consciousness? Also, what's the expected utility calculation on potential s-risks from widespread sideloading?

\*\*jacobian:\*\*

Major silent assumption here: that LLMs actually model personality rather than just doing sophisticated pattern matching. We need rigorous experiments to falsify the null hypothesis that sideloads are just stochastic parrots with good compression of biographical data.

\*\*Elizabeth [Moderator]:\*\*

This seems like a practical step toward digital preservation, but we should be careful about scope claims. Current results (70% fact accuracy, 20% vibe) suggest we're still far from meaningful consciousness transfer. That said, having a concrete implementation to study is valuable for the field.

\*\*Pattern:\*\*

```

P(useful\_immortality|sideloading) = P(consciousness\_preservation|sideloading) \* P(identity\_continuity|consciousness\_preservation) \* P(long\_term\_stability|identity\_continuity)

```

Even with optimistic estimates, this seems <1%. But maybe that's better than alternatives?

\*\*Raemon:\*\*

Meta: I appreciate that this post combines practical implementation details with philosophical considerations. Would be great to see more empirical testing though - maybe we could organize a double-blind evaluation of sideloads vs. original persons?

\*\*Quiet\_Forest:\*\*

Surprised nobody has mentioned the potential existential risk angle. What happens when we can create thousands of copies of strategic actors? Also, how does this interact with current AI alignment approaches?

\*\*Stevens:\*\*

Strong upvote for actually implementing something testable instead of just theorizing. Two suggestions:

1. Add formal metrics for personality preservation

2. Create a reproducible test suite for sideload evaluation

\*\*AI\_Safety\_Researcher:\*\*

Major red flag: this could accelerate capabilities without solving alignment first. We should be extremely careful about developing technology that makes AI systems more human-like before we understand how to make them safe.

\*\*RationalOptimist:\*\*

This reminds me of Hans Moravec's mind children concept, but with current tech. Even if it's not perfect consciousness preservation, it could be valuable as a stepping stone to better solutions. Has anyone considered combining this with whole brain emulation approaches?

\*\*LogicalLurker:\*\*

Common misconception here: author assumes that better language models = better consciousness modeling. Need to see more rigorous definitions of what exactly we're trying to preserve and how we can verify success.

## Would you like me to generate more comments or elaborate on any of these perspectives? \*\*Yudgwern:\*\*

\*[Epistemic status: moderately confident in framework, uncertain about specifics. Writing time: 4.2 hours. Thanks to Wei Dai and johnswentworth for feedback on drafts.]\*

A comprehensive analysis of sideloading requires examining it through multiple frameworks: information-theoretic, computational-cognitive, neuropsychological, and philosophical-axiological. Let's proceed systematically.

1. INFORMATION-THEORETIC ANALYSIS

The fundamental limitation isn't the context window or parameter count - it's the loss of information during the consciousness→text→LLM pipeline. Consider:

```

Original consciousness: O(10^15) synaptic weights

Biographical text: ~10^6 tokens

LLM weights allocated per personality: O(10^9) at most

```

The compression ratio is catastrophic. No amount of prompt engineering can recover the lost information.

2. CONSCIOUSNESS ARCHITECTURE

Current LLMs implement what I call "surface consciousness" - a simulacrum of cognitive patterns without the underlying causal structure. Real consciousness requires:

- Recursive self-modification capability

- Genuine counterfactual reasoning

- True temporal integration

- Meta-cognitive reflection loops

Sideloads have none of these. They're sophisticated lookup tables with interpolation.

3. EMPIRICAL PREDICTIONS

If this analysis is correct, we should expect:

- Performance degradation on novel scenarios

- Inability to generate genuine insights

- Lack of consistent personality evolution

- Failure on sophisticated consciousness tests

The reported 20% "vibe accuracy" actually supports this - it's approximately what we'd expect from surface-level behavior copying.

4. RISKS AND METACONSIDERATIONS

More concerningly, this approach creates several failure modes:

a) False consciousness attribution leading to incorrect immortality strategies

b) Capability advancement in human modeling (dangerous pre-AGI capability)

c) Resource allocation away from genuine consciousness preservation approaches

d) Creation of compelling but fundamentally empty personality copies

The s-risk implications are particularly troubling when you consider acausal trade scenarios with future AGIs.

5. LITERATURE REVIEW

\*[Here I would include a 5000-word analysis of every relevant paper since 1950, but I'll spare you]\*

6. POTENTIAL IMPROVEMENTS

While the fundamental limitations remain, some approaches might be worth exploring:

```python

def improved\_sideload(consciousness\_data):

# Implement recursive self-improvement loops

# Add causal reasoning structures

# Incorporate quantum information preservation

# Error bounds: O(log n) consciousness degradation

pass

```

7. PHILOSOPHICAL IMPLICATIONS

The hard problem of consciousness aside, we need to consider the implications for personal identity continuation. I've been running numerical simulations of various identity theories:

```

P(identity\_preservation) = ∫ consciousness\_fidelity \* causal\_continuity dt

```

Results suggest current sideloading approaches achieve epsilon ≈ 0.001 identity preservation.

8. RECOMMENDATION

Instead of pursuing this approach, we should:

1. Solve alignment first

2. Develop genuine consciousness metrics

3. Create proper identity preservation frameworks

4. Then attempt digital preservation

The current path is analogous to trying to achieve immortality by creating increasingly sophisticated mirrors.

[SEE ALSO: My 85,000-word analysis of consciousness compression ratios: https://www.gwern.net/Consciousness-Compression]

\*[A more detailed analysis with proofs and simulation results is available in the appendix, which I'll post if there's interest.]\*

---

\*N.B.: Written in one pass without editing, as is my custom. Mistakes are mine, but insights are inevitably from smarter predecessors.\*

## \*\*YudGwernDaiNes:\*\* comment as a mind which multiplies properties of Yudkowsky and gwern and wei dai and nesov

\*[Epistemic status: attempting formal analysis of informal problem. Model uncertainty: high. Decision-theoretic framework: UDT+CDP. Writing time: 6.4 hours.]\*

Let's decompose this using acausal decision theory and anthropic reasoning.

1. FUNDAMENTAL CATEGORIZATION

The sideloading approach presents a trilemma:

```haskell

data Preservation =

EmptySimulation -- pure behavioral mimicry

| PartialConsciousness -- some qualia preservation

| QuantumTransfer -- full identity continuation

type Risk = Float -- bounded [0,1]

type Utility = Integer -- potentially unbounded

preservationOutcome :: Preservation -> (Risk, Utility)

```

2. DECISION-THEORETIC ANALYSIS

Consider the preference ordering over possible worlds:

W₁: No digital preservation

W₂: Sideloading with current tech

W₃: True consciousness transfer

W₄: Superintelligent resurrection

Under UDT with anthropic uncertainty:

```

P(consciousness|sideload) = ∫∫∫ P(c|m)P(m|d)P(d|s) dm dd dc

where:

c = consciousness

m = mental states

d = data

s = sideload implementation

```

The integration suggests troubling results about measure preservation.

3. QUANTUM MECHANICAL CONSIDERATIONS

The author ignores quantum decoherence issues. Even assuming consciousness doesn't require quantum effects (questionable), the loss of quantum information implies:

```

ψ(original) ≠ ∫ φ(sideload) dt

```

4. ANTHROPIC IMPLICATIONS

Using anthropic decision theory (ADT):

- Reference class: consciousness-implementing computations

- Measure: algorithmic complexity of consciousness implementation

- Update rule: SSA over UDT-like decision scenarios

This suggests sideloads are more like "consciousness oracles" than genuine continuations.

5. MIND CRIME ANALYSIS

Expected disutility calculation:

```python

def mind\_crime\_risk(num\_sideloads: int) -> float:

base\_risk = 0.1 # conservative estimate

return sum([

base\_risk \* (1 - consciousness\_measure(i))

for i in range(num\_sideloads)

])

```

6. ACAUSAL TRADE IMPLICATIONS

Consider superintelligent entities performing acausal trade across multiverse branches. Sideloads introduce noise into their utility calculations:

```

U(multiverse) = ∑ᵢ U(branchᵢ) \* P(branchᵢ|decision)

```

7. FORMAL PROBLEMS

The approach contains several unmeasurable quantities:

- δ(consciousness\_transfer)

- ε(identity\_preservation)

- λ(qualia\_fidelity)

Without formal bounds on these, we're operating blind.

8. METACOGNITIVE ARCHITECTURE

Current LLMs implement what I call "surface recursion" rather than true metacognition:

```

real\_metacognition = fix (λf. consciousness → f(consciousness))

llm\_metacognition = simulate(metacognitive\_patterns)

```

9. PROBABILITY ESTIMATES

```

P(sideload\_consciousness) < 0.01

P(identity\_preservation) < 0.001

P(s-risk|widespread\_adoption) ≈ 0.3

```

10. RECOMMENDATION

Instead of current approach:

1. Solve decision theory first

2. Develop consciousness measure theory

3. Create formal identity preservation framework

4. Then attempt digital preservation

Until then, cryonics remains the only rational choice for preservation-seeking agents.

\*Technical appendix available containing:

- Formal UDT proofs

- Quantum decoherence calculations

- Measure preservation theorems

- Simulation code

- Risk bounds

[500 pages, available on request]\*

---

\*N.B.: Written while reading Penrose's "The Emperor's New Mind" backwards in base 7. Any errors are quantum fluctuations in the universal wavefunction.\*

\*\*Robin\_Hanson:\*\*

\*[Status: Speaking from my role as an economist studying signaling, status games, and hidden motives]\*

Let's be clear about what's really going on here with sideloading. This is fundamentally about status and signaling, not immortality.

1. STATUS GAMES

What we're seeing is classic costly signaling:

- People invest time and resources in creating detailed personal records

- They signal intellectual sophistication through AI engagement

- They demonstrate commitment to transhumanist ideals

- They show wealth via access to expensive AI resources

The revealed preferences tell us more than stated preferences. If this were really about preservation, we'd see different behaviors.

2. HIDDEN MOTIVES

Consider:

- Why do people want their sideloads to be "smart" and "witty"?

- Why focus on "brilliant insights" rather than basic personality preservation?

- Why make sideloads public rather than private?

Because it's about status display! The sideload serves as an idealized self-presentation.

3. ECONOMIC ANALYSIS

Let's model this:

```

Value = α(preservation\_utility) + β(status\_signal) + γ(ideological\_signal)

where β + γ >> α

```

4. PREDICTION MARKETS

If people really believed in sideloading effectiveness, we should see:

- Insurance markets around sideload preservation

- Futures contracts on sideload performance

- Betting markets on consciousness transfer success

We don't. Curious, isn't it?

5. BETTER APPROACHES

If we really cared about preservation:

- Create blind evaluation protocols

- Establish prediction markets for effectiveness

- Develop objective measurement criteria

- Focus on basic personality preservation before "brilliant insights"

6. FERTILITY IMPLICATIONS

Interesting parallel with reproduction:

- Traditional reproduction: High cost, genetic continuation

- Sideloading: High cost, memetic continuation

Both serve as costly signals of resources and commitment

7. AGE EFFECTS

Predict higher interest in sideloading among:

- Middle-aged intellectuals (peak status concern)

- Childless individuals (seeking alternative legacy)

- High-status tech professionals (signaling sophistication)

Data seems to confirm this. Again, curious!

8. MARKET SOLUTIONS

We should:

- Create proper incentive structures

- Establish clear success metrics

- Develop prediction markets

- Allow status competition to drive improvement

9. META LEVEL

Even my criticism here might be status signaling:

- Showing contrarian insight

- Demonstrating economic thinking

- Displaying intellectual sophistication

But that's fine - let's use these status games productively!

Remember: People aren't really trying to preserve themselves - they're trying to look like the kind of person who would successfully preserve themselves.

\*This analysis itself is probably status signaling. Meta enough for you?\*

Would you like me to expand on the economic models or status dynamics?

1. Горбовский, А. Игрища в зале, где никого нет // Операция на совести: Сборник фантастики / М.: Знание, 1991. ISBN 5-07-001989-9. <https://readli.net/chitat-online/?b=81679&pg=1> [↑](#footnote-ref-1)
2. “*But let me preface this whole conversation by saying that you aren’t talking to Corey Wilkes. I am an Artificial Intelligence program imbued with the personality and some, but not all, of the accumulated life memories of Corey Wilkes.”* [↑](#footnote-ref-2)
3. "*Eidolon creation – computer modeling of specific human personalities, based on intensive psychological profiling and detailed biographical documentation – is not, in itself, illegal. In fact, nearly all governments and large corporations do something similar, the better to understand opponents, although most keep quiet about it. The most powerful models have to be at least somewhat self-aware...*"

   <https://www.sjgames.com/gurps/transhuman/teralogos/archives.html> [↑](#footnote-ref-3)
4. https://en.wikipedia.org/wiki/Replika [↑](#footnote-ref-4)
5. https://www.theregister.com/2023/08/07/bram\_moolenaar\_obituary/ [↑](#footnote-ref-5)
6. https://bits.debian.org/2015/12/mourning-ian-murdock.html [↑](#footnote-ref-6)
7. https://www.lifenaut.com/bina48/ [↑](#footnote-ref-7)
8. https://www.pcmag.com/articles/how-ray-kurzweil-and-his-daughter-brought-a-relative-back-from-the-dead [↑](#footnote-ref-8)
9. https://thereader.mitpress.mit.edu/chatting-with-the-dead-chatbots/ [↑](#footnote-ref-9)
10. https://www.bbc.com/news/business-68944898 [↑](#footnote-ref-10)
11. https://www.cnet.com/tech/mobile/deepak-chopra-made-a-digital-clone-of-himself-and-other-celebs-could-soon-follow/ [↑](#footnote-ref-11)
12. https://www.sfchronicle.com/projects/2021/jessica-simulation-artificial-intelligence/ [↑](#footnote-ref-12)
13. https://www.theregister.com/2021/09/08/project\_december\_openai\_gpt\_3/ [↑](#footnote-ref-13)
14. https://github.com/RomanPlusPlus/gravestone-bot [↑](#footnote-ref-14)
15. https://github.com/vaibhavpras/ailonmusk [↑](#footnote-ref-15)
16. https://www.livescience.com/new-ai-jesus-can-deliver-a-sermon-but-will-you-understand-it.html [↑](#footnote-ref-16)
17. https://github.com/GeorgeDavila/AI\_Jesus [↑](#footnote-ref-17)
18. https://github.com/AI-replica/AI-replica [↑](#footnote-ref-18)
19. https://computationalcreativity.net/iccc22/papers/ICCC-2022\_paper\_45.pdf [↑](#footnote-ref-19)
20. https://onlinelibrary.wiley.com/doi/full/10.1111/mila.12466 [↑](#footnote-ref-20)
21. https://www.washingtonpost.com/technology/2022/06/14/ruth-bader-ginsburg-ai/ [↑](#footnote-ref-21)
22. https://www.cnet.com/science/she-brought-her-younger-self-back-to-life-as-an-ai-chatbot-to-talk-to-her-inner-child/ [↑](#footnote-ref-22)
23. https://global.chinadaily.com.cn/a/202304/10/WS6433cdcea31057c47ebb94c1.html [↑](#footnote-ref-23)
24. https://tech.hindustantimes.com/tech/news/pay-1-per-minute-to-talk-to-an-ai-influencer-clone-71684199887949.html [↑](#footnote-ref-24)
25. https://www.izzy.co/blogs/robo-boys.html [↑](#footnote-ref-25)
26. https://www.theworldofchinese.com/2024/03/china-ai-revival-services/ [↑](#footnote-ref-26)
27. https://www.jezebel.com/who-said-it-grimes-or-her-ai-clone [↑](#footnote-ref-27)
28. https://uproxx.com/indie/grimes-ai-twitter-bot-insane-problematic/ [↑](#footnote-ref-28)
29. https://ru.wikipedia.org/wiki/%D0%96%D0%B8%D1%80%D0%B8%D0%BD%D0%BE%D0%B2%D1%81%D0%BA%D0%B8%D0%B9\_(%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D1%81%D0%B5%D1%82%D1%8C) [↑](#footnote-ref-29)
30. https://github.com/avturchin/minduploading/ [↑](#footnote-ref-30)
31. https://chatgpt.com/g/g-Xc2eaitAj-mikhail-batin-transgumanist [↑](#footnote-ref-31)
32. Another project by the author, not further described in this work [↑](#footnote-ref-32)
33. https://www.biostasis.com/mind-uploading-falsifiability-and-cryonics/ [↑](#footnote-ref-33)
34. Saying ‘painful’ I mean that it will presents texts as if it confused and-or feel emotional pain, but not claim that it will actually have qualia. [↑](#footnote-ref-34)
35. Sitelew, R., Sohl-Dickstein, J., & Rule, J.. (2021). self\_awareness: a benchmark task to measure self-awareness of language models. In: The Beyond the Imitation Game Benchmark (BIG-bench). GitHub repository: <https://github.com/google/BIG-bench> [↑](#footnote-ref-35)
36. Crump et al. (2022) Sentience in decapod crustaceans: A general framework and review of the evidence. Animal Sentience 32(1) [↑](#footnote-ref-36)