14 NOVEMBER 2022 / PROJECTS

I record myself on audio 24x7 and use an AI to process the information. Is this the future?

What will happen when our phones record everything we say 24x7 using AI to process that information?, before

someone else does, I tried to do it myself, and this is the result.

First of all some clarifications.

- It's not really 24x7 because I didn't find it useful to leave it ON at night (maybe if I was a sleep talker it would be useful), but for now I leave it ON from when I wake up until I go to sleep, and I turn it off if I'm in a situation where I don't think it's appropriate.

- This is a "proof of concept" and not yet ready for production, everything described here works but probably "glued with tape", several of the processes are probably not automated or polished.

MOTIVATION.

I was doing an analysis a few years ago that show that the trend of increasing storage and lowering costs would result in smartphones with 1tb of space (to date there are already 38 models with that capacity), with this in mind, the question was: What would be the practical implications?, One of them was precisely that the smartphones will be able to record audio 24x7.

With the release of OpenAi's Whisper weeks ago, the stars are aligning to find the application of what could potentially be the future, and as the phrase "the best way to predict the future is to create it" says, this prototype is just an experiment to anticipate the future and have a window to what is possible

HOW DOES IT WORK.

I bought a couple of Chinese microphones, I wear them and turn them on all day recording everything I speak, at the end of the day the files are processed with OpenAi's Whisper and transformed into text files from which the information is extracted.

In that context I realized that you could create a rudimentary "Ok Google" style "Digital Assistant" to actively take advantage of the fact that it was recording everything anyway.

So I would have two types of information that I can generate and extract on a day-to-day basis:

- **Active:** I consciously tell the "Assistant" what to do

- **Passive:** is all the rest of the information that is extracted without requiring any action on my part.

Active functions:

They are the ones that, by using a keyword I am going to indicate that they will be processed asynchronously (at the end of the day). For example if I decide that the keyword is

```
"Robert"
```

Every time I speak that word I am indicating that everything I say after that will be taken as a command to my assistant until I say again

```
"End Robert"
```

Once the audio of the day is processed and transformed into text, a program takes care of searching the text for the command and extracts everything between `"Robert"` and `"end Robert"`

For example to register my weight for the day I simply say out loud

 `"Robert WEIGHT 60.1 end Robert"`

The format I define then is

 `"KEYWORD COMMAND data END KEYWORD"`

This way I generate commands for my assistant asynchronously.

The immediate question that arises is:

Why not just use "OK GOOGLE"?

Here are my main reasons

- **Limited Commands:** My biggest problem with "OK Google" is that I don't know by heart what it can do interactively and it wich commands will just return as a google search.

- **Privacy:** "OK Google" commands are saved in google with the full audio file, try entering google order history and you will be able to hear your voice, and many times the music and conversations in the background when you order commands to Google.

- **Synchronous:** When you send a command to Google, there is a delay, sometimes several seconds to receive the command, and another to see if it correctly understood what you said, this in my experience cut the flow and distracts you from the activity you are doing and makes it slower than just using the keyboard.

In this experiment all these points are solved:

- **Limited Commands:** Commands are defined by me so I know what to say.

- **Privacy:** I own the entire process, 100% privacy

- **Synchronous:** It's asynchronous, I don't have to wait for any response, I just say the command while I'm doing something else and trust that it will be executed at the end of the day.

DISADVANTAGES

This approach also has its drawbacks:

- The commands that I send will not be executed until the end of the day, so it will not serve to remind me of something that I have to do that same day.

- Being asynchronous, I will not find out if what I requested was successfully processed until the end of the day.

SHOW ME THE RESULTS NOW.

HOW IT WORKS IN ONE DAY.

'Robert WEIGHT 62.8 end Robert'

'Robert SLEEP 7 hours 14 minutes end Robert'

ANALOG API: Several electronic devices have connectivity only with my cell phone and have no way to export that information, now I use my assistant as an *"Analog API"* to communicate and centralize information between several devices.

'Robert LUNCH two toasts with a fried egg end Robert'

CALORIES AND MEALS COUNT: I indicate out loud all the meals of the day, at the end of the day I can have the list of meals consumed in the day and use an external API to calculate the calories consumed in each meal.

Or I can just make a routine for the assistant to register the calories of the food by reading it from the package.

'Robert NOTE the podcast talks about Morgan Housel's book the psychology of money end Robert'

NOTES: I take notes that are going to be centralized or I can summarize ideas without taking my hands off the wheel.

'Robert SPENT 250,000 on fuel end Robert'

EXPENSES: Every expense I make during the day I repeat it out loud to record it.

It feels a little strange to be "talking to myself" but it also frees me from the cell phone screen or from having to wait for any return and just talk trusting that what you are asking for will be executed later.

THE CONTROL PANEL.

The result of all the information that I explicitly indicate to my assistant can be centralized in a daily control panel:

❤️ Health Info

Weight

Sleep

Calories

832 Cal.

☕ Morning

- Fried egg
- Two toasted loaves

🍴 Noon

- Lamb sandwich
- Orange juice
- A portion of Almonds

⚙️ Afternoon

- One cereal bar
- One avocado

🌙 Night

- Pumpkin soup
- Potato salad

💵 EXPENSES TODAY

- 250.000 - Gas
- 325.000 - Supermarket
- 15.000 - Bread

590.000 - Total today

📓 TODAY'S NOTES

Podcast note, Morgan Housel launch a new book

Remember to try Whisper from OpenAi

The control panel is a basic example of what we can do with the information that we naturally feed our assistant with.

Do our days all pass the same? We don't know where the hours went? We can also create *"My Journal"* so that it generates a summary of what we did that day.

Mi Diario
20/11/2022

🕐 06:12 - 07:22

- Despertar
- Desayuno: Huevo Frito, Dos panes tostados
- Salgo al trabajo

🕐 07:56 - 12:39

- Llegando al trabajo
- Reunion de proyecto Whisper
- Investigacion segmentacion de voces en python
- Almuerzo: Sandwich de cordero, Jugo de naranja, Porcion de almendras

🕐 13:22 - 16:12

- Trabajando en primer MVP de segmentacion de voces
- Saliendo de la oficina

🕐 16:37 - 17:53

- Cargando combustible: 250.000
- Compras en el supermercado: 325.000

Initially I am implicitly saying out loud what I am doing, but ideally it would be possible to do it automatically according to the context of

what was spoken and the clock time, at the end of the day it generates a *"Journal"* of everything what i did in the day

All **Active** features require conscious action on my part that will translate in results I expect from our assistant.

But the interesting results will be in the **Passive** information that we get for the simple fact of recording our daily conversation, some of the ideas are:

PASSIVE INFORMATION - IN PROGRESS!

RELATIONSHIP THERMOMETER

According to studies on couple relationships, it is possible to predict with an accuracy of up to 90% if the couple is going to divorce by studying the interactions, specifically the relationship between positive and negative interactions between the couple:

https://www.gottman.com/blog/the-magic-relationship-ratio-according-science/

The magic ratio is 5 to 1 according to this study, for every negative interaction during the conflict a stable and happy couple has 5 positive interactions.

SENTIMENT ANALYSIS.

I can try to do a sentiment analysis by hour and situation to be able to identify under which situations the greatest stress, relaxation, happiness and sadness are generated during the day.

Do I use positive or negative adjectives when referring to a topic or person?

TOTAL RECALL.

Implement a search engine to show, for example, all the instances where I talk about X topic, or what my opinion was at some point about X topic.

PRELIMINARY CONCLUSIONS.

- **I'm a little less paranoid if everything I say is recorded:**
  In my tests, using the cell phone and letting it record audio at the distance where it would normally be, the accuracy with which it is transcribed is quite mediocre - according to the distance - I would say that it is around 50% or less accurate, that would still allow me to eventually capture if I am talking about a brand to advertise, but not enough to do something precise or complex, I do not leave aside that those who can have access to the audio of the phones already have AI models that improve quality but my tests give me a bit more peace of mind in that regard, it's not as easy as it seemed to record 24x7 and get something out accurately enough to do something important.

- **In addition to the audio, the context is needed:**

  Reading the result of the transcripts, without the proper context (time, location, who I am talking to, the previous context of what was said) is very limiting

- **The potential to get this right is immense, both positive and negative:**

  POSITIVE

  Perfect Memory : Something magical is basically being able to perfectly "relive" everything I did that day, from trivial conversations and knowing their context, something that is lost in the ether today can remain available now.

  Personal Psychologist/Coach : Recording everything that came out of our mouths, every interaction we had, and analyzing that can give us a vision perhaps impossible until now to have, a psychologist/coach in real time who accompanies us to everywhere.

  A virtual clone : Imagine training a GPT3 with everything you said in the last year?, and making it, under certain parameters, answer your messages, or manage your agenda.

  NEGATIVES

  Exactly the same points, but in the hands of someone other than us.

The difference between utopia or dystopia is who has access to that information

— RoberDam.com —

Projects

Xaddress - An Algorithm that can give an instant solution to 7 billion people

See all 1 posts →

PROJECTS

Xaddress - An Algorithm that can give an instant solution to 7 billion people

Decentralized open source physical addresses, 1.2K stars on Github

1 MIN READ

Latest Posts   Facebook   Twitter   Ghost based Theme