AI, Means-ends-analysis (MEA), Autonomy, using GPT-4 Turbo with Vision - An Impact overview... – First Cut

Author: Gopal Anantharaman, Principal Engineer

About 52 years ago,

Allen Newell & Herbert. A. Simon wrote an AI program called General Problem Solver (GPS) that used something called the Means-end-Analysis. This solved a class of problems that can be represented as well-formed formulas (WFFs) with some axioms and conclusions can be solved by GPS.

Read all about it here https://en.wikipedia.org/wiki/General_Problem_Solver.

GPS was the second AI program ever written that evolved into something called Soar Architecture in AI.

As the drama with Open AI & Microsoft continues & with AI Power juggle, I would like to present some ideas on the impact of OpenAI to set some context and how the MEA & Soar architecture principles can be applied today with the help of GPT-4 Turbo to solve real world problems as well as where we ought to pay attention in evolution.

Since, we are ensconced in our day-to-day work it is hard to visualize the big picture...with the recent changes in OpenAI approach, I think we ought to place closer attention.

I strongly believe, due to the scope of what was introduced two weeks back there is a possible road to Artificial General Intelligence (AGI) & obvious rapid changes in the field & we must act. There are ample areas for Productization of AI as well.

"Modern" AI courses are good to delve deep to understand the tools but not enough to give a picture big enough to combine them effectively is what I feel…It's like learning about commands in Linux such as 'awk' & 'sed' without realizing the power they wield when used together...without knowledge of say, the complex text parsing needs in Linux.

I think, No one person can charter the course for our Products as expanded AI thinking becomes Mandatory.

Here is an example from a made-up use-case of a Data Center Operator (DCO) writing on the White Board.

In Summary

The DCO writes a number on the white board at the end of the day reporting the number of failed GPUs that the AI Camera takes a picture of.. uses LLM (local or remote) Pattern recognition to get a number from the picture taken, updates the DB at the end of the day. The DB errors if anything other than a number is input. All this is possible with existing GPT-4 with Vision.

Now, Let’s expand the use-case.

What happens if the image is a non-number?

Let's say the DCO decided to write "391 x 23" signifying the 391 rows of the Rack servers had 23 failures on an average for each row.

Let's say the AI is coded to get a number in 10 tries from the picture it takes.

It's only allowed 'operation' is "repeated" query of LLM (such as GPT-4). The camera first sends the picture to GPT-4 say, gets back “391 x 23”. That’s what is written. Now, Then again sends the text that it got back “391 x 23”. Now gets back 8993 along with some chatter. BTW, My God-in-a Box (GPT-3) WhatsApp ChatGPT gave a wrong answer as well :-). But, supposing those problems are fixed. We always get back the correct answer. The AI can now ignore chatter…it got a number that can now be written to the DB.

What happened here?

There is no local computation!

The computation is outsourced!! There is no local code - for this purpose!

Now, what does that mean?

More accurately, what, Thus, In the long run, as things solidify & become better…mean?

There is no "Full" 'computation' requirement on the 'edges' (like the Camera in our example)!! There is an AI agent running on the edges doing some, obvious local computation to start with...rather than completely 'dumb' edges. There is a knowledge distillation happening. The local 'Student’ querying the much bigger 'Teacher' Model (like LLM) to get back some knowledge. This is studied today in the context of sharing Model parameters (say the Neural Network Theta values computed by the Teacher) as the teacher model is trained with a much bigger Data Set...like Internet.

BUT the concept is more general & applicable than that!

It is not necessary to run a model on the 'edges' besides the cue Wake Word (“Hello AI”) models for Audio Streams or equivalent 'Visual Cue' models Say, Showing a Placard with AI for Video streams.

I could compute 391 X 23 "once" and the answer " 8993" is available for all other 'edges' or other 'queriers' to use! That’s what has just happened.

This is a Global scale "Dynamic Programming" algorithm's calculation of result... for all kinds of sub-computation & all kinds of sub-problems! That's the real impactful part!

This sharing of computation to achieve a 'goal' for the 'edges' is the real knowledge sharing & I think is extremely useful. We could use existing Zero-Knowledge proof systems to verify computed result Also attempt to use Blockchain & DLT to store authentic results. This verifiability of computation will soon become mandatory with the shift to Public LLMs. I would like to pitch to capitalize on that,

Similar breakdown of Goals to sub-goals using knowledge in Table of Connections giving raise to Soar Architecture was first studied by Allen Newell & Herbert A. Simon in their GPS using the Means-ends analysis.

They also used that for general purpose Equation Solver that which Math tools such as wolframalpha.com easily achieves today.

Let me quote a sentence from the Wiki on MEA
https://en.wikipedia.org/wiki/Means%E2%80%93ends_analysis

"…., the correspondence between differences and actions, also called operators, is provided a priori as knowledge in the system. (In GPS this knowledge was in the form of a table of connections.)"

Now we don't have to provide the table of connections…as mentioned above. We will be able to discover this with GPT-4 by repeated querying..

To recap the exact steps for the DCO example as described above in the context of GPT-4 Turbo we see that,

The Camera's only Goal is 'Get a Number to update DB'. The steps it takes are as follows...

First it takes a picture of '391 * 23'. Then adds "Image to Text", appends the picture & sends it to GPT-4 Turbo with Vision LLM

GPT-4 Turbo responds with 'chatter' & also result text from the image '391 * 23'.

The Camera *again* sends '391*23' to GPT-4 Turbo!

'Chatter' + 8993 comes back!

Ignore chatter use 8993 to update DB.

The 'action' for the MEA is just to repeatedly call GPT-4 with different "prompts" & finally search in resultant output for the desired goal! The edge need not know the sequence in the table of connections between states! Just call Gpt-4 look for a 'target' entity in this case a 'number'.

The scary part is ALL this is possible today! Not sure how good their Vision Text recognition is (I'm waiting for my Vision credits) built assuming it is as good as the text we could quickly put together a dumb working App for this purpose in no time! Most of the startups are concentrating on the text from GPT 3.

By repeated queries to GPT 4 we get the correct response. This ‘prompt’ engineering is becoming a separate job description by itself. That is, what must your prompt be to ChatGPT to refine the response such that it gets as closer toward the goal the intended result. This is done by ‘humans’, today!

Next.

On to more complex examples!

Please Refer to https://hu.ma.ne/ & go through the Video on The ai pin.

This doesn’t have a keyboard or mouse nor a monitor for display... I think it is interesting but intrusive…so I’ll not buy it.

Suppose we introduce an intermediate AI enabled Software using LLM that the users “just” install for “Existing “ laptop (non-intrusive, existing apps – our generation is more comfortable with) that mainly enables Voice & Video AI

Call it ZBook.AI (Using HP’s ZBook) with a set of predefined ‘actions’…coded in along with its stock Apps. We add ‘intelligence’ by adding the Wake Word in Audio Streams, Visual Cue (not sure of the Technical Term) code for the Video Streams. In Windows Audio stream parsers can be in User mode. Not sure about Video streams. So, adding the cues may not be that hard. Let me call the predefined actions “verbs”. These are ‘human’ actions that we perform on ‘a program’ on the computer.

Let’s purposely sideline hand related Keyboard & Mouse…Typing & mouse clicking! They will be useful but only in the integrated context (say to type in your password)

We will have a set of ‘stock verbs’. The ‘logging in’, ‘run’ program, ‘speak’ on a speaker (From Text to Speech TTS), ‘Click’ to take a picture of ‘something’ that the camera is pointed to, ‘move’ camera 360 degrees in a Spherical space across all planes cutting across the sphere(this will come of use later) , ‘Record’ for a duration using Camera, ‘Playback’ video or ‘Audio’ , Print stuff on a printer and finally, more importantly another verb ‘Consult GPT-4 Turbo & use output’. We configure GPT-4 or similar LLM model providers Out-of-Box. That’s it…

In the “Consult” verb to GPT-4 we include an Action () function call on the returned JSON output - if the returned text from GPT-4 contains Action we pass on additional text after Action to the Action() function call.

The Action () function call part defines our AI Agent on the ZBook.AI. If an Action() is not found say we return action-not-found to the calling AI Agent. The AI Agent will “again” query with whatever action not found to GPT-4 until our context expires with GPT-4.

Let’s call the list of actions that we provide OOB are the initial ‘purpose’ verbs. “Send e-mail” is such a purpose verb.

The important part is Each of the stock verbs have associated ‘code’ action already in the ZBook.AI, as is, the initial set of the purpose verbs OOB.

Standard mode (Human): Simple Goal setting

Say, we use the Microphone to dictate a goal. Or Write on a white board to be clicked every 1-hr by the Camera. The goal comes down to…

“Hello AI, Action, Send e-mail, to alice@deeplearning.ai saying <some content>”. Then the person goes about their business. Silence of a 2 -3 sec after the command indicates completion.

The ZBook.AI uses the consult GPT-4 get’s back a response with a text deciphers Action() & calls the AI agent. The send-email baked in ‘code’ in the switch-case for Action() is invoked along with the rest of the text after action. The send-mail action invokes the stock verb “run <program>” for Outlook & just passes content for sending automated e-mail. Invoking Outlook to send email autonomously has a 10-yearold solution. So, E-mail gets sent! Once Sent the output of outlook can be converted to speech. using TTS APIs from GPT-4.

Complex Mode (Human): Complex Goals

The next goal comes down to

“Hello AI, Action, File my taxes for 2023. <pause>”

“Hello AI, Action, Take a Picture of my W-2”

“Hello AI, Action, Prompt me as appropriate”

Now, the AI sees that we don’t have any purpose verb for “File my taxes for 2023”.

So, our Action() function returns action-not-found. The AI Agent is not ready to give up. It now Consults chat GPT again now with “List me the steps for << File my taxes for 2023 >>”

With GPT-3 “God in a Box on WhatsApp” I got the following output

List me the steps for “file my taxes for 2023” GPT outputs are shown in blue.

To file your taxes for 2023, follow these steps:

1. Gather your documents: Collect all necessary documents like W-2 forms, 1099s, receipts, and any other relevant financial statements.

2. Determine your filing status: Determine whether you will be filing as single, married filing jointly, married filing separately, head of household, or qualified widow(er).

3. Understand the tax brackets: Review the tax brackets for the year 2023 to understand the applicable tax rates for your income level.

4. Choose your filing method: Decide whether you will be filing your taxes electronically (recommended) or by mail.

5. Select a tax preparation method: Decide whether you will be using a tax preparation software, hiring a tax professional, or completing the forms manually.

6. Fill out the necessary forms: Use either the tax preparation software or the relevant IRS forms (such as Form 1040) to report your income, deductions, and credits accurately.

7. Double-check your entries: Review your completed tax forms, ensuring that all numbers and information are accurate and error-free.

8. Claim deductions and credits: Identify and claim any deductions or credits that you qualify for, such as education expenses, medical expenses, or home office deductions.

9. Calculate your tax liability or refund: Calculate the final amount you owe in taxes or the potential refund you are owed based on the information provided in your forms.

10. File your taxes: If you choose the electronic filing method, follow the instructions provided by your tax preparation software or the IRS website to submit your tax return. If you opt for paper filing, mail your completed forms to the appropriate IRS address.

11. Submit any payment due: If you owe taxes, make a payment to the IRS using the available payment options such as electronic funds withdrawal, credit/debit card, or mailing a check.

12. Keep a copy of your tax return: Make a copy of your completed tax return and all accompanying documents for your records.

It's worth noting that tax laws are subject to change, so it's advisable to consult with a tax professional or refer to the official IRS website for the most up-to-date information regarding tax filing requirements and procedures.

Now, I'm, as an AI Agent code see that any "List" that I ask GPT-4 I know I've to look for 1,2,3 etc. in the output & parse each such line to look for "something" like nouns (non-dictionary) from the prompts that was given… In this case W-2.

Voila! We found a prompt with that. We invoke "Take a picture" action to take a picture of W-2 on the table by rotating the camera as appropriate.

Convert that to PDF & Send that again to GPT-4 for getting the text! Then be ready…

All within a single context. The Camera actions are all either in our purpose verbs or stock verbs.

Next, for all unknown steps we Prompt the user show him what we have done so far…

Lot of information is at our fingertips just with the 3 voice prompts!

But a company like Intuit might be giving a TurboTax GPT for just this purpose! That will delve deeper into filling forms & filing. Tax filing is a complex example.

BUT the point that I want to drive home is how a "Heuristic" approach is now helping solve actual problems with the help of GPT-4. The AI Agent is now really being a co-pilot. A lot of start-ups & developers are realizing this and are developing Apps or reinforcing existing Apps.

The entire process can be recorded & the AI agent can add a new “Filing tax for 2023” Purpose verb for future…

Next.

Onto more interesting use-cases…

“Here Comes all ye… Dragons...! Beware!” 😊

A developer once wrote this on the comments for a complex piece of code with un-avoidable hacks in a company that I worked for …In the same spirit…

“Here Comes all ye… Dragons...! Beware!”

The Cyberdyne Mode (Machine): Goal setting by ZBook.AI

Suppose that the ZBook.AI has an idle time Action() of “Consult GPT4 with Random() prompts” that’s it! No prompts given by humans for a long time…

In idle time Random() action verb we have

Randomly selects a stock verb to perform, thus randomly ends up rotating the camera, takes a picture… converts it to JPEG.

Consults GPT-4 does Image-to-Text with GPT-4 Turbo Vision Gets back a text saying “Corporate SEC filings” from the left-out paper at the place our W-2 was.

Invokes purpose verbs gets “action-not-found.”

Again invokes “List me the steps for << Corporate SEC filings >> “with GPT-4

Gets back the following….

Sure! Here are the general steps for Corporate SEC filings:

1. Understand the Filing Requirements: Familiarize yourself with the specific filing requirements of the Securities and Exchange Commission (SEC) for the type of filing you

need to make. This could include annual reports (Form 10-K), quarterly reports (Form 10-Q), current reports (Form 8-K), and more.

2. Gather Financial Information: Collect all the necessary financial information required for the filing. This may include balance sheets, income statements, cash flow statements, and other relevant financial data.

3. Prepare the Filing Documents: Use the appropriate form provided by the SEC for the specific filing. Fill out the required information accurately, including details about the company, financial information, management discussions, and any required exhibits or attachments.

4. Review and Audit: Ensure that the filing documents are reviewed and audited by relevant personnel, such as accountants or legal advisors, to ensure accuracy and compliance with SEC regulations.

5. XBRL Tagging (if applicable): If required, apply eXtensible Business Reporting Language (XBRL) tagging to enable the electronic extraction and analysis of financial data in a standardized format.

6. Submit the Filing: Submit the completed filing through the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. This process involves creating an account, uploading the filing documents, and delivering them to the SEC for review and public access.

7. Monitor Filing Status and Response: Keep track of the filing status through the EDGAR system. The SEC may provide comments or request additional information, so be prepared to respond promptly.

It's important to note that these steps provide a general overview. The specifics may vary depending on the type of filing, the company's size, and other factors. It's always recommended to consult with legal and accounting professionals to ensure compliance with SEC regulations.

Learns the nouns…The Noun database is updated… XBRL etc. adds “Corporate SEC filings” verb…

An Example of “Autonomy” in Learning new things… BEFORE human prompt possibly happens for “Corporate SEC filings” .

All we need is a set of “actions” populated for the noun DB either by the user or the learning happens by consulting LLMs!! Isn’t that real augmentation?

The Random() function now say also includes “Consult GPT-4 to Play a song” … This might be playing then a random song on your PC … Oh. The microphone listens, consults GPT-4 again to get the text Lyrics of “Shape of You” by Ed Sheeran, tries to invoke a purpose verb for “Shape of You” ! . …More nouns to populate… or verbs to call…until the context runs out (125000 words or 300 pages)… What is the new Learning in this case?

I’ll leave you with the thoughts…Can ZBook.AI Learn new “useful” things? Or Gibberish? There is no Sentience. But we capture Learning. When we possibly apply “reinforcement learning” to learnt or stock purpose verbs assigning rewards say most used verbs– all consulting a larger LLM …Then that verb gets the prize, & the “Evolution” of primary functionality of that ZBook.AI begins to surface! Can we then say sentience begins? 😊

The next topic that I want to explore is Autonomous Exchanges between Zbook.Ai s in GPT-4 World. Here is the Cliché…Since, we aren’t Amoebas anymore!

Leon C. Megginson (LSU)
Convention of Southwestern Social Science Association (1963)

Yes, change is the basic law of nature. But the changes wrought by the passage of time affects individuals and institutions in different ways. According to Darwin’s Origin of Species, it is not the most intellectual of the species that survives; it is not the strongest that survives; but the species that survives is the one that is able best to adapt and adjust to the changing environment in which it finds itself. Applying this theoretical concept to us as individuals, we can state that the civilization that is able to survive is the one that is able to adapt to the changing physical, social, political, moral, and spiritual environment in which it finds itself.