۵ Questions U must Ask Every Vendor with Gen-AI

Description

Whether they are selling learning technology,Vendors are a clever bunch of folks.
I saw this firsthand at LTUK. Those who havecontent/courses, or even learning systems.
and technology are noted very visibly.in their system, platform content/courses,Gen-AI
on the floor. Lots of attendees t̃ndid Some whoeven have it; pitched it at vendor sessions
focused on the information at hand, as thoughlistened intently. Their eyeballs fully
Svengali was in front.

the person just scraped the internet for theirWhile observing one session, it was clear
all, generative AI, aka Gen-AI, is all the rage,presentation. Still, eyes were on them. After
s̃i and darn it, if anyone is going to add it,us in the L&; and Training communities. First,
we must ask the right questions. Which for manyso much a question, as a t̃nsi though,
Gen-AI in action. ,̃ffo eyes are glazed ,wowl̃istening intently and seeing

that who knows could be the s̃l I get it.If done right can be spun into a magic elixirnew.
And there lies the problem. Although s̃simehcladream of turning anything into gold.
turn X into Gold. Good thing he had that GravityNewton actually thought he could
validation as a backup.

Gen-AI in their system. This, of course, madeFiltered was one of those vendors promoting
there, as fast as my feet could go, and asked tome want to know more. I zoomed over
system. I believe it is ChatGPT. One of thewen ̃learn more about Gen-AI in their
okay, ̃components was curation plus, the othertoo. s̃ulp ̃recall, but it had a t̃nac

this system to generate a learning pathway andIt sounded great. I can use Gen-AI in

Then I saw it, and started to ask questions.curate content. Oh, boy!

- ?noitacinullah How do you count forDo you tell people they might get hallucination)The answer BTW, was no.(?semoctuo with their
- things(‍It looks like a search engine functionwhy ‍)this was with one of the plus coming by who ?rab not just have a search)They noted that there were vendors function and wanted to add it to their systems.(were impressed by said search
- a chatbot and the assets are the ones I pay for)Answer is yes(?tcerroc This is
- each header, the assets underneath are theWith the AI pathway, while it generates is yes(?tcerroc ones I pay for or must own)Answer
- what can I do on the back end to make sure theGetting back to the hallucinations the content they ‍content they see is accuratethis was under the assumption that is free content(, is well from the net.)AnswerWe ‍pull down from the net)which are vetted by us, that you can access. If thehave identified several sources that sources we can add them.(In other words,sd‍eretliF client wants to add additional has ‍scrape the entire internet tn‍seod Gen-AIrather selected sites that Filtered will go to. You still access free ,d‍dettevw‍whatever that entailed that the Gen-AI .‍tnetnoc eerf‍content, just selected

vendor limit such options, or tweak it in such aNow you may ask yourself why would a called Tokens. ChatGPT uses them. GPT4 usesway, as above. It has to do with something is the people behind the ChatGPT, st‍I them.how those vendors, make money. OpenAI LLM st‍i ‍GPT4, and several others they havetheir method of revenue. Any open-source and so forth. Then it is added to your system,you select can be modified, tweaked, and content)as it relates to our industry(. Iget tn‍ow learning technology, authoring tool, will say that it ‍ot woh‍into the technical ofa free deal for any vendor out tn‍si but I Gen-AI to whatever you are doing, then an Openthere. If you, Company X want to add 100% free, others are fee-based(. And those thatSource LLM is the route to go)some are fee-based can get very expensive, which is why‍as it relates to our industry ‍are will try to reduce the cost in any way possible.vendors

you need to understand at a high level aboutThe Five Questions will come, but first, questions you will ask any vendor who has Gen-Tokens, because it is part of one of those AI.

Tokens

recommend just doing a search around Gen-AI andTo go further down into Tokens, I a ‍Tokens. For the basics, it works like thisanytime the AI processes text it becomes

can be words or chunks of snekoT token.noted Azure OpenAI. sretcarahc

way to think of tokens is pieces of words.Another

regrubmah Multiple sites use the example ofIt looks like one word, but with Tokens, it reg rub maH counts as three words.And you charged for each token.

is non-English, the token numbers increase. se omC If the language *How are you* The higher token-to-char ratio can make itin Spanish(contains 0 tokens)for 10 chars(. to implement the API for languages other than)OpenAI(hsilgnE more expensive

you to see the number of tokens based on theOpenAI provides a Tokenizer, which allows it works is with a white space prior to the wordtext. Please note that with tokens, the way eyB itself. Example:

srodneV I tried it out, with the phrasewho use proprietary models without stating it is built on an open-source model, is a22. The number of tokens nrecnoc whether

what is the basis, or method for processing andOkay, I added some words/characters but ?seef

to find a simple explanation for explaining thisI did a deep dive around the net seeking in a given request ehT and choose Azure OpenAI.total number of tokens processed output, and request parameters. The quantity ofdepends on the length of your input, your response latency and throughput for thetokens being processed will also affect)Azure OpenAI(. sledom

types in a word, phrase, or whatever, and hitsIn other words, the moment the end user begins. When the end user follows up, morethe button for a response, the cost for tokens the numbers will rapidly increase. For example,sAnepO token costs. As you can see, is reportedly, 700K a day. That is 700 hundredChatGPT is free. The cost to do this thousand dollars a day.

varies. That is to say, if a vendor uses any ofLLMs the cost per sAnepO Pricing for tokens the same as OpenAI, which most folks tsi tokenthe same. Nor is the price for Stability.ai with ChatGPT and their latest version, GPT-4.will recognize

challenge for those of curious nature. Here issAnepO Finding the pricing is another **pricing**models, but vendors would use the token more. I should note they offer two price what I use. The interesting piece around some ofthan say the subscription, which is

is fee-based, even before actual usage. sfAnepOmodels is that training your data

Other LLM Models

continue to be added at a faster rate than mostThere are a lot out there and they anticipated.

but these are based on other Gen-AI models,There are open-source AI that is 100% free, Thus, you might find tnsi thus while free, itsay, a foundational LLM fee-based one. using any one of these)a way to avoid payingvendors who went 100% free open source token fees(. This is just a shortlist.

- **Vicuna**
- **Stability.ai**)They offer 100% freebies, and **DreamStudio** which is fee-based. Even only want to use the API(Their newest one is if you **StableLM**.
- **GPT4All**
- **LAION**
- **Dolly-2**
- **Alpaca**site, you can reproduce it for less than $600(But I slide it)According to the into free here.
- **Koala**)Built on LLaMA 13B(
- **Github**of a lot of 100% free open-source models(This link goes to)A repository the Gen-AI ones.

models Most people will recognize OpenAIbut are likely unaware that they offer multiple edits images)Bing Create uses this(and Whisperincluding Dall-E2)which creates and and translates into hceeps which is arecognition model that transcribes, identifies)OpenAI(segaugnal multiple

is expected to announce the next version of Google PaLM-2 for Bard with other LLM and not available elsewhere. To learn about whatmodels to follow. It is their own model PaLM was trained on)for those curious, read **here**(

écruos nepo These are all A Short ListLLM)Fee-Based(

- **LLaMA**there are numerous offshoots)from Meta(From LLaMA, **view**
- **Azure OpenAI** For those using Azure)Microsoft(
- **Amazon Titan** Available only to those who use AWS)Amazon(They also offer **Bedrock,**allows you to build from other LLM)for exampleyou have which from that)foundational simply means it is theStablity.ai and want to build

above including OpenAI are ‍foundation of LLMwhich you need, and thus all of the foundational(.

- **BERT**others from Google, for use with Google Cloud( )and
- **Databricks**the AI Agents category which I believe is really)I place this more into coverage on how it works, etc for another time(.the power for our industry, and
- **Midjourney**)Image Creator(

want to pay for an LLM when freebies are out ?ereht Why would someone

GPT-4 offers ١ trillion parameters. ChatGPTOne of the key advantages is the parameters. Personally, it just comes down to what is youris next up with around 75 Billion.
are in the business to make money. Thus,preference. Vendors who raise capital/funding you need to do a bit more than the freebies.by requiring it at some point,

the five questions, because one of them is)And What does this have to do with the most important(

1. Language Model do you use with your learningWhat Learning tool/content platform,offer two or ?.cte system/technology/authoringThey may recommend, but I doubt at least right now thatmore, which experts in the LLM space is the way vendors are going.

model, which sounds amazing but it has to beA vendor might respond with a proprietary the unlimited funds to build 100% from scratch,built on something. They do not have a(the computing power to run gen-ai is massivehigh carbon footprint, b(to ‍because due to heat, the cpus need water, and c(It justmake any sense for t‍nseod stay cool them to go that route.

they built it on some foundational model; madeThus the proprietary model to me, means would do, hence the open source part(, and said,,adaT‍changes to it)as anyone LLM are foundational(. If a ‍!yrateirporpWhat you are seeking is the foundation)again, on Moodle, you might cringe, and yet therevendor said, hey we built are system originally system built on Moodle and then revamped it soare vendors who initially launched their knowing at t‍ndluoc much, and so on, that youeven tell, when they debuted it. That said, you could explore via the net, what it is,least the name of the model, for those curious, t‍nod how it works, and so forth. Maybe youcare, but I would if a vendor said Dolly versus GPT-4. I would want the latter.

other day to show their system using Gen-AI wentOne vendor who reached out to me the

used. They told me that they used proprietary,beyond vague when I asked what LLM they on what that means and figured to corner themshort, and big ones. Now, I have no idea I see the system)in early June(. Short and BigWe are not talking about ?seno once for those super ,yeH‴hummingbird feeders heredo you want a short one or a big one ?sdribgnimmuh large

though no one will dare to ask what the is LLM,Yet, they are pushing this solution out as know what is a Learning Language Model. Youneed to t‴od let alone that anyone would question here is to get the vendor to tell youknow unless you are super curious, but the knowing it will aid you to follow up withwhat they used or are using. Becausespecifically which you must ask! ‴a secondary question

what ?no ٢. What data sets was your LLM trainedYou cannot skip this question and think Internet searches from 20٢2, ?rettam does itBecause if they trained it on Wikipedia, and other learning models they selectedthat would be relevant. Take ‴Kirkpatrick, 2021. Yes, it learns over time, but data setsChatGPT for example. The data sets are from initially are relevant.

angle with their new authoring tool told me,One vendor who pushes the whole Gen-AI that were selected by experts they identifiedthat the data sets included learning models And who are these ?seno or worked with. WhichUnless they raised Donald ?strepxe the dead and pulled a Frankenstein with Gagne,not seeing an expert m‴ Kirkpatrick from here. Oh, they also use a proprietary model.

the vendor uses a fee-based open source model,as in ‴say GPT-4, will you see s‴el ٣. If in the coming years with your ‴you the client?tcartnoc an increase in fees

I said these models are not cheap due to theWell, if you think ?snekot Remember when then Bugs Bunny is right behind you, and thethe vendor is going to eat that cost, is still a viable mode of transportation.Hindenburg

to cost a vendor. It is going to get pricey realJust think how much money this is going then skirt it by going with a different layerquick, even if they use say part of the LLM and and to me, that is in pricing. Pricing hason top of it. Those costs have to go somewhere, can get $60 per seat per user/year, at ٢,500always been an arbitrary model. If a vendor so)this is just an example(. I often tell theusers, then they have no problem in doing story me $48 per user/year)for a seat(, which Iof a vendor who wanted to charge declined and price. The next day, it was down to around $٩then declined on the next bucks. And they still made money on me.

charge it(, APIs)if they charge(, onboarding,Vendors can hide the price in setup)if they that contract you want to lock in that price fortraining, and so forth. Thus, when you sign increases of X percent that some vendors push asthe length of your contract. No, price or living costs. None of that nonsense. Lock itthough you have to eat the cost for inflation in. Because they will not eat that cost. That)referring to fee-based models, and I tn̋ac the time to change the code and create your ownthink some 100% freebies because costing them money(. Oh, ditto on the authoringexclusive model with data sets is still getting for your e-learning that charges a feetool, learning tech or whatever else you are

Gen-AI model they are using)within the company(٤. For those clients who already have a model for whatever data you want to bring in and?tuo can you API it into their Gen-AI I wonder how many vendors who have Gen-AI inThis is going to be tricky here because tech or whatever has thought about it. But attheir system or authoring tool or learning come up, and when it does the vendor is going tosome point this question is going to have to know the answer.

because it is already starting to happen)not byI can see the inquiry around plugins, specific(. If ̋client per se; but in generalsolutions not learning/training or e-learningother that exist for their LLM, which if they wentI were a vendor I would start adding plugins the entire market, including 100% freebies, they100% secret model that nobody else on though you can take your HRIS platform and askare going to be in for a shock. The idea off, your tn̋si LLM, s̋rodnev to add it to thesomething I would recommend doing. First to have an LLM. Secondly, it has to match whatHRIS, HCM, ERP, Payroll, or whatever has vendor has)especially if the vendor is on AWSwhich the majority are, and they are ̋the to have a solution like bedrock)using AWS as anusing Titan(, OR the learning vendor has Y and it will work. Remember we are at a veryexample(whereas you can have X they can early stage here.

called data, as in what data is getting passedbeyond serious ̋Then there is this thing

around privacy, legality and who knows whatwhich you would need to create ‍concerns the vendor do with that data ‍legal docs forthose questions ‍you have to ask what will asked, responses, by who and so on.

its own LLM)i.e. 100% free or fee-based(,If your company is not at this stage of having this question. And if you have it, and thinkdo it with X sf‍eL‍‍then you can avoid many sf‍el respond back, uh, ‍revetahw learningnot, because there are way too tiny steps here. Even by the end of 2024, Iunknowns out there, plus we are at extremely it goes, and if one vendor sf‍eL would be leery.not forget that Gen-AI learns as trained their solution on data sets, to whichexperienced what a company did, when they was a tf‍nsaw the LLM learned a new skill, thatpart or existed in the data sets)it on matching ‍DLOH‍language(, we should all bethis had nothing to do with Q4, but ‍ just a surprise of what we are dealing with.

tech/authoring tool/etc deal with?noitanicnullah ٥. How will your system/learning

this as question ٢ or ٣, because it is theI should have placed **BIG ELEPHANT** in the in their pitch. Halluncination exists in allroom, and vendors are not so quick to respond who says, oh, we have a method tf‍nac Gen-AI. Youget rid of them today. So any vendor or a small number is selling you a magic elixir?devlovni is Newton ‍to have none

with any LLM and more importantly, Gen-AI.Hallucination is the biggest issue/problem And it occurs. NVidia just came out with aWhat they mean is fake or false information. solution they claim reduces hallucination,)**NeMo Guardrails**to reduce hallucinations but how well it will work over a period of time.it is truly unknown

When ‍Fun Fact **Bard was tested by Google employees,**one responded that it was a still went live with it. ‍.rail lacigolohtapGoogle

Bottom Line

out Gen-AI to mf‍ Here is my promise to you.going to be holding those vendors who push not just the ones listed above, but onesfire. I am going to ask those tough questions,the my findings. llf‍ that go further than that.report

keep you updated. I plan to create a massiveAs LLM and Gen-AI in general advance, I will you can access via the Learning Libraryon Gen-AI tools that you can use for ‍directory and training. That is coming this summer.your learning

vendors who think, nobody is paying attention.And for

Worry,

Because we all are.

E-Learning 24/۷

Category

1. authoring tools
2. Gen-AI
3. Generative AI
4. learning systems
5. learning technology
6. LLM
7. Tokens

Tags

1. e-learning
2. Gen-AI
3. Generative AI
4. Hallucination
5. Learning Language Model
6. learning systems
7. learning technology
8. LLM

Date Created
May ۹, 2023
Author
diegoinstudiocity