FOUNDATION OF DIGITAL BUSINESS

Surojit Mookherjee

Vinod Gupta School of Management

Indian Institute of Technology Kharagpur

Week 05

Lecture 24

Lecture 24 : Becoming AI ready

Good morning. In this session of artificial intelligence for business now and beyond, I will be discussing how to become AI ready. So, this is for organizations. So, becoming AI ready, so I will be talking about 5 pillars of AI preparation and then optimizing decision making with AI. So, these are the concepts I will touch upon in this session.

So, continuing with the pillars, the 5 pillars of think the number 1 pillar is of course, all of you know by now is data readiness. and you get your data infrastructure in shape. So, question is are you storing your data? So, first thing you are a manager, so you must ask your employees, your team are we storing the data? Bulk of the data is often unstructured or semi-structured in various formats like emails, purchase orders, power points, PDFs, support tickets, call center transcripts, customer feedbacks.

So, you need to be aware that these are also data. You might ignore them email, purchase order, powerpoint, some PDF document etcetera, but they are also data source for you. So, that is what you should know and you should consider and ask your team that are you preserving them or are you deleting all your emails for example, you might be deleting all the customer emails. Are you warehousing data? Now, where you store data?

So, data cannot be stored in different silos or compartments. So, we have something called data warehouse, again an IT structure, IT hardware which consists of servers. It is an important tool where structured and semi-structured data can be stored. And we are now things called data lakes which another form of data warehouse where you can store unstructured data any text, image, PDF files like emails whatever you are referring to earlier all of such data can be stored in something called a concept called an equipment

called what we call as data lake which is again something a part of extension of data warehouse you might see. So, data from various locations and sources can be shared.

and made accessible to the teams. Now, continuing on that are you logging? So, logging keeps track of everything that happens in a software application or keeping any type of digital record at a specific time intervals. We have data logs like you must have seen charts or you might have seen in a hospital for example, a patient's heart condition is being continuously monitored. So, that is nothing, but another kind of log.

record of the sensor output. So, that is logging. So, later on you can analyze the entire set for certain duration for certain time period and find out whatever you want to find out. So, logs is a time series representation of your sensor outputs. So, in AI logging helps to understand the customer behavior around product features

before introducing intelligence in this feature. So, any customer behavioural responses that you are collecting becomes your log, because you are collecting on an hourly basis or a daily basis or a weekly basis or whatever. So, it has to be on a time series. Logged data can also be a source of training data for many AI applications, because you need data for AI training. So, the logged data gives you a huge volume you need for training, giving the training set.

Google uses search logs for example, to improve its search algorithms and recommend related search. Logs allows to analyze search and just imagine just think about Google, how many search is being conducted in a Google every day across the world, it will be in billions. several hundred millions of Google searches are created every day. So, that is a log for Google what is the search and what is the quality of the response. So, they always analyze that and based on that analysis they can always fine tune their tool the software underlying and the logic

whatever the AI system they are employing to improve the quality of the output. Is the search good enough or acceptable to the customers whosoever are using the Google So, log this is of course, in a very large scale now in a real business terms of course, in a much smaller scale as compared to Google. So, it is completely different way of handling things. Logging customer interaction this will allow you to improve the customer experience and their productivity over time.

So, we take feedback from customers. So, whenever you go to a restaurant you know they will after that when you submit that they give the bill along the bill they will give

you a feedback card. you have to put some rating stars 4 star, 5 star, 3 star, 2 star, what about the service, what about the food, what about the quality, what about the timeliness, etc. So, everywhere you go and you get a service done even in a hospital for example, they give you a feedback. Why do they do that?

It is exactly for this purpose. they want to log your responses so that they can study that and based on that they can improve their performance. Ultimately end of the day we want to improve the way we are doing business, we improve the service experience for the customer and it is all for the sake of the customer. Customer, log events related to production of goods and services within your factory or within operations you also need to log your various events related to the manufacturing process and service process.

So, logging the process is nothing but collecting your data over a period of time. The more data you can collect the more longer the period of time the better the insights or analytics you can run on those data. Have you digitized your paper documents? What about your paper document hard copies which is lying in your files all over the place? How you digitize that?

Because if you digitize that, that will be the source of most of your unstructured data. Just imagine the courts, the lawyers, the amount of documents they have to handle, hard copies that go through pages and pages and pages, just imagine a judge making a judgment for hundreds and hundreds of pages. So, if you want to go through or summarize or find out the gist of say past documents, 10 years of judgments on a similar whatever case etcetera.

How will you do it? Manually you can have to dig out those judgements nowadays if these are available online you have to go through and read. So, when they are available online means they have been digitized. So, that is the advancements which has happened in this our country all of course, in many development countries in the legal system that most of the things are now digitized and available online. So, once it is digitized you can now use it run through them run them through any GPD, GNI tool which can then summarize for you.

So, what would have taken you may be 30, 40, 50 hours you can do that in 1 or 2 hours. Establish an AI literacy, so this is what part of this program is actually doing making you AI literate and that is all idea you should take these two things at a corporate level such literacy programs need to be run from the top downwards. All employees in the company should have some basic literacy about AI, what it is and what it can do and what it should

not do. Make leadership data saving. So, leadership should be very much affiliated to data, you must have a liking for data.

So, wherever you have an opportunity store the data, collect the data, store the data, do not delete, do not ignore the data because they will come of use later on when you start some AI experiment. Be ready to experiment and brace for uncertainties. So, AI is not yet a very acceptable tool. technology because it is also it is having lot of problems, but it has got enormous amount of promises. So, what is required from the industry or all of us is that we should start experimenting because the limit of AI is the amount of experiments we can do and make use of it.

that is the limit because otherwise the potential of AI is almost we can say unlimited, but if you do not use it obviously it is not being utilized. So, the extent to what we can exploit AI technology depends on how much experimentation we are ready to do and how much risk we are ready to take. Build cross functional teams, always think cross functional, do not just think AI means just a AI programmer or hardcore IT technology or a computer science person, no it is much more beyond that. So, you must have the business person involved as part of the team, of course the AI implementer, the AI techie guy and then you need data engineers and software engineers. Create an ethics and accountability committee, this is extremely crucial.

Because with AI comes lot of potential problems because you are handling data, the data could have some problem, the output of that problematic data will also be a problem. And if there is a problem it and since it is going public and you are utilizing for any sorts of public service it is going into the social domain, it can create lot of ruckus and bad mouth bad press etcetera. So, you must have an ethics and accountability committee to oversee all these AI development work and anticipate the potential problems which may incur. and then take necessary preventive action or be prepared also if something happens how should we handle this the mitigation plan you know the risk,

but we have to take the risk otherwise we do not proceed or we do not progress. So, for each risk you must have a mitigation plan that if it happens. if something goes wrong then what shall I do. So, it should not happen that when the thing happens actually goes wrong and then you start thinking and scratching your head oh my God now what should we do. So, that scratching of your head should have been done when the product was being developed.

So, you make a list of things like this can go wrong, this can go wrong, this can go wrong and for this go wrong, this is the remedial action, this is the mitigation plan, this goes wrong, this is the mitigation plan. So, everything should be there on paper, so that you are prepared. So, you will never be taken for a surprise and this is very important in a business environment. Keep an open mind.

So, be open to ideas, be open to innovation, be open to taking risks, challenges, be open to experimentation otherwise you will not be able to exploit the AI technology. Pillar 3 skill readiness, train and upskill your key employees. So, start with this executive education should cover what AI is and is not. potential pitfalls of AI, industry use cases of AI, how to prepare for AI as an organization, how to spot promising AI opportunities, how to get cost effective AI solutions implemented, how to measure the success of AI and skills for using and interpreting data. So, these are some of the areas in which training or education will be required for most of your management staff starting from the very senior from the CEO.

The infrastructure readiness, the algorithms are extremely data hungry, which means that you need lot of data. The development can be computationally expensive, it needs high computational machines. The high power those GPUs you might have heard of in company like Nvidia, the extremely high processing power and it also for that matter consumes lot of electrical energy to run those processors, they generate lot of heat. Need specialized and multidisciplinary talent which can be expensive.

So, this talent part I have already talked about, I am talking about it should be multidisciplinary and AI talent again because it is a new technology. So, there will be a shortage, there will be crisis and they will be expensive. Models must be evaluated and tested before and after development. So, that is quite obvious for any IT product you have to thoroughly test before you release and also continuously go on testing after release. So, example I quote here is that is another methodology called design thinking where the if you take the product say chat GPT 3.0 when it was launched they call it something like a in a IT terminology called a beta product.

Beta product means it is still under development, but they are releasing for public use. Why did they do it? The whole purpose was to get it tested by a large number of users because internally you cannot have such heavy testing with limited employees because employees are costly expensive. So, if I release it and if it becomes popular like it did it

become viral and within a month millions registered and started using because there was something new for everybody. And people are very excited.

So, all of the testing or the when you are using it you are actually using it on their server because you are connecting through the internet and in the cloud of OpenAI. So, OpenAI as a company was getting all the question and the answers the prompts and the responses in their servers which then now they could analyze to see the performance of the tool chat GPT 3.0 and based on those analysis they came up quickly with the next version 4.0 etcetera within the short period of 6 months. So, the whole testing cycle got extremely speeded up because it was open to the crowd. So, it was crowd testing like we have been discussing on crowd sharing, crowd sourcing, crowd funding.

So, this is crowd testing. Models must be integrated into your business systems and monitor once when you deploy that that should be part of your business system and you have to then regularly monitor just leave it unattended that is there fine people are using it whatever is happening happening no. You must at the from the highest level you should monitor because it is a new technology new tool and as I told there is a lot of risks involved. So, it should be continuously monitored whether it is doing its performing as it was expected to do or it is hallucinating giving up some wrong results incorrect

you know feedbacks or people are misinterpreting etcetera because of problems with the data bias training etcetera whatever could be the reason. So, all of this requires lot of software tooling computational power special hardware and supporting personnel. So, this is your infrastructure requirement and now cloud computing is another infrastructure solution to all of these essentially used for AI model development on a as a service mode. They offer out of box tools which allows you to build those innovative. So, if I want to do an AI experiment and I do not have funds.

to buy all those chips firstly fund, secondly time because I do not get off the shelf and they have huge pending order wait time for those sophisticated GPU chips made by Nvidia and other company. So, the easiest solution is or one of the best solution is to go to the cloud provider and take their AI platform on hire and start doing my experiment because I can start doing it from right from the moment I want to do it and pay their rental and register with them. So, this has become proved to be a very useful solution. easy go forward solution for working on AI models.

Even the government of India for example, has realized this importance Chinese government had done it much earlier. So, government of India is setting up a cloud

specifically for AI development and they are buying those huge amount of GPUs very expensive. So, they are spending may be few 100 crores of rupees to build this cloud and then make it available to users whosoever wants to use it you pay and use. So, at least the platform is there for you somebody may here in this case the government of India is building the platform it is available you take it on rent for whatever period you want and pay as you use.

So, it is pay per use concept which is perfectly logical and rational. Budget readiness is the pillar number 5, finally end of the day you need money. So, you have this data warehousing cost, the AI infrastructure cost, the training cost and the talent acquisition cost. Of all these the AI infrastructure cost you can manage all the data warehousing using cloud. So, you can have not a capex, but you can do it as an op ex basis.

So, by just paying rents, training cost and talent acquisition cost is of course, there. And then if you outsource in the development then the talent acquisition cost will be minimized. So, you should have consider all of this for as part of your budget planning. How to optimize decision making with AI? Now that we have started the building models and utilizing them.

So, now next comes is how best to use that make the best use of these models which have spent money to develop. Carly Fiorina was a former CEO of HP, very famous CEO said goal is to turn data into information and information into insight. And whenever we talk about AI we are always talking about insights and she made this statement many years back. So, the types of data being collected in the organization. sales data, customer web activity data, customer emails, customer call logs, customer reviews and complaints, sensor data from manufacturing operations,

quality assurance data, employee performance data, market research surveys, etc. So, just some list of some of the common data sources, potential data sources for most organizations. And then of course, the manufacturing etcetera plants this will be mostly for e-commerce type of organization etcetera because web surfing, web surfing etcetera web activity. So, it varies from a type of organization, but this is a sample list the list could be much more bigger. So, data when summarized with statistical tools to gain insights and track patterns we call it as a data analytics.

So, you have must have business analytics, data analytics. So, that is called whatever data analytics or business analytics. But if you want to know what is the profile of your customers for example, you have to add a layer of intelligence on top of it to get this

meaningful insights. this is where AI comes in. So, we have your data you do your analytics now you top it up with AI use it in a machine learning model to develop a model so that it can give you an insight.

So, this is now we call as say intelligent data analytics. So, we have data analytics now we build a layer of intelligence and we call it now IDA intelligent data analytics. It typically leverages machine learning and NLP natural language processing to help get deeper meaning from the data. Making data analysis more effective particularly effective when the data source has unstructured data because without NLP you cannot extract any meaning from unstructured data. The benefits could be an ability to add information not present in the raw data, whatever the raw data you see, you do not see much, you just see some numbers, some text, some figures, but you want information,

you just do not want to see data and from information you go to knowledge, these are steps data, information, knowledge. Ability to summarize and make sense of the large amount of unstructured data, again just extension of that we want to make some sense summarize and make some sense. Reduction in manual and time consuming work in processing large volume of data. So, we have already talked about all of these given examples. So, large volume of text for example, you want to find out what is the summary, legal

your purchase documents, there could be various areas where you create lot of documentation and you want to sometimes extract the key thing, the summary finding. So, instead of going through all of that reading yourself use the tool, AI tool, JNI whatever and do your analytics, getting insights from the given data source. So, idea is becoming more critical to business of today primarily because more than 80 percent of the enterprise data is unstructured. So, that is where AI is getting the due respect because of the presence of such large volumes of unstructured data. Continuing on this, how do you want to optimize listening at scale for new product innovation?

What does it mean? Listening to as many customers as possible, monitoring customer pain points around the web that is online, you collect the pain points and product reviews and social media complaints reveal a lot about the brand and becomes a source for lot of data. So, this is where you can actually keep your ears and eyes on what is happening, what people are talking about your product in various platforms and social media. So, this is listening at scale, at scale means you have to do it at for a large say Facebook, if you get into Facebook of large number of Facebook users maybe thousands or maybe

millions or whatever. Millions will not be true, but it will be thousands, hundreds and thousands is the number of customers you are referring to and you want to know what they are talking about your product.

So, these are the rich sources of customer data and the customer needs can then be identified from these feedbacks and new products or modifications to existing products can be done based on this analysis. So, once you have the feedback you collect the feedback in large numbers, the number is very important just not 1, 2, 3 like that. we need large volumes of data then you can utilize those to develop real insights real meaning and get the correct feedback on the average. And then change your modify your product accordingly then that will impact the majority, but if you just take feedback from a very few small numbers and do some major change then that may be

actually impacting very few who have recorded their feedbacks the others may not be having this problem. So, the scale is important, listening to scale. Disparate data source to improve the customer experience. It is important to overloading presence of social media to listen to multiple sources One is to leverage this multiple data source from customers like complaints, conversations, social media messages like WhatsApp is one, conversations is Facebook is one,

Instagram is one, complaints is your probably call centers or service centers who are receiving customer complaints. So, you have to combine take into account all of this do not ignore any of this. A hospital improve the patient experience by abandoning the conventional patient engagement survey like you give this feedback form and like this and this various etcetera and then you score them 1 to 5 scale 5 or give 1 star 2 star 3 star whatever. So, that was the traditional method and also used continuously used in many places.

So, what they did for a change they used NLP and ML to do an sentiment analysis on the feedback the customers are posted on social network. So, they extracted discussion themes and then got the real customer emotions because when you give them a form and in front of you are standing and say to you out of 5 you give 4 or whatever. So, that may not be a real feedback, but when you go back home and all your then and you do it online when nobody is observing you or the hospital person is not there, then your real emotion will come out will be captured. There are significant negative experiences like non-supporting staff.

long wait times etcetera. So, these are the things which are very important for that hospital to give good services to their customer to their patients. So, based on these findings the hospital could make some plan for corrective actions. So, this is the AI way of doing things as compared to a traditional form based thing where you just keep giving scores 1 to 5. So, it becomes an IDA.

Using search logs to improve discovery and marketing. So, your search engines on a company's platform or a website are important source for data And the products being scanned by the customer can give insight as to their preferences and companies can use the knowledge to promote specific products. So, this depends on the company, the type of business etcetera, but in some of the business where the huge number of goods like most of the e-commerce for example, like say myntra.com or any fashion stores or garments etcetera, where there are huge number of product items, sales items, stock giving items.

So, you can find out what the customers are surfing, what they are liking, what they are not liking. So, that search log can be given to improve your discovery and marketing. So, what you really find out what they are looking for because the actual when they come to the store and buy things or look at things. you will get much more volumes of data when you go to the online thing because there are many number of people who are surfing your site may be buying may not be buying, but you are getting their idea of what they are liking and what they are spending more time on etcetera.

So, these search logs contain the rich data about the customer search and browsing behavior. Some of these could be the search keywords used, the search results displayed for specific clicking on the search results and number of user query reformulation using different search terms for the same item. Suppose you are not happy with the search term like it is same thing like prompting I wrote this and then whatever came the result I was not happy. So, I change the search term.

So, based on this you can understand that this person is really wanting to look or get a particular product and is not being able to find it. So, it is still trying various ways of giving the search terms. So, that product is very important. So, if you can identify that product which is important and the customer is trying to search desperately you can then push or promote and add for that customer for that product. So, you are getting insights by analyzing the search logs.

Use NLP to automate the root cause analysis. Documentation defect and the resolution reporting on incidents and accidents are complex semi-structured and combination of

structured data like date and time of the complaint etcetera, the number of people affected. So, it is combination and then the of course, the remark etcetera. So, it is a structured plus.

So, only option is to use your NLP tool natural language processing can use to analyze this data and do the root cause finally, you want to know what was the real cause. So, that is called a root cause then you can solve the problem otherwise you will be just trying to do this that etcetera which may not call solve the real problem because you have not identified what is the real problem and that you can do with the help of NLP. Also apart from root cause correlated events can be identified and idea can be instrumental in making such discoveries. Like one of the stories I told earlier was in a supermarket they suddenly realized that young fathers coming to buy diapers for the children are also buying lot of gear.

So, based on this insight from the data they what they did was smartly place the beer cans just next to the diapers stalls and those action cells. So, you are looking for actually came to look for diapers, but then you see some beer can just nearby. So, you will be tempted to pick up a beer can and there is a wait time in the supermarket in evening etcetera you have a beer. So, anyway such insights etcetera can be given by these LNP technique of analyzing the data. To conclude if you see the various sources of data source like Twitter comments, call center logs and then user reviews, these are all being collected in a platform and you run an AI engine called say

using an NLP tool you do something called as sentiment analysis. tool in marketing areas. So, you find out what the sentiment being expressed by the users, some are happy, some are not happy etcetera. And then you aggregate this data And then also combine that with your structured databases whatever structured database which is coming.

So, these are the unstructured data and then you combine that with your structured data like those date, time etcetera, item bought etcetera specific things. Combining of this then you use your business intelligence tools like Power BI etcetera to visualize this output and then that will help you to give you the right. So, the combination of all this technology is giving you the final output. So, AI component is here where you are doing the sentiment analysis using NLP tool collecting all the unstructured data working on unstructured data and then you combining that with your structured data and then do visualization which is nothing to do related to AI per say because visualization is something else it is simple straightforward analytics and it will display the data in

various shapes and forms like it can give you in the form of a chart. It can give you in the form of a graph, it can be a bubble chart whatever because pictures tell much more better stories than simple data tables or whatever running text. So, you display the output in a form which is which can give you more better insights. If you the size of the bubbles for example, for various agenda items relative size can give you a particular meaning a weightage for example. or the bar charts or the pie chart.

So, those visual charts can help you understand the problem which you are trying to understand. So, use combination of all these technology will give you the right final is your insights. So, the final thing what you want is could I makes any intelligence or sense out of all this data which was collected. So, that is the main goals I am collecting lot of data. Then I am doing some AI analysis here in combining with other forms structure data etcetera and I am running an visualization or analytics engine.

So, based on this analytics engine the output which is a visualized dashboard is another common name is called a dashboard, through the dashboard I can get my insights. And that as one advantage of a dashboard is that if I click on any of those visualization figures I can deep dive into the actual data. So, I can drill down. So, I can get the real actual data and then see that whether what I am thinking or what I am realizing is correct or not backed up by the data.

So, whatever the data is saying I am also saying the same thing. So, with all these various tools or techniques we can use to start the AI journey and then try to optimize or make