FOUNDATION OF DIGITAL BUSINESS

Surojit Mookherjee

Vinod Gupta School of Management

Indian Institute of Technology Kharagpur

Week 05

Lecture 22

Good morning. So, continuing with artificial intelligence for business, the first lecture in this series of this module is preparing for AI. So, I will discuss 5 tips of to maximize tips to maximize AI success, some myths about AI and then I will talk about also the machine learning development life cycle. So, how you the phases you use for doing the machine learning thing. So, AI for the sake of AI what executives need to be aware of again you know this is for you to be concerned about or also be aware of.

Pushed by the top level executives, the company is often to survive in a game of decades. They need to adopt new technologies to stay relevant and profitable within the industry that is the core challenge. So, am I changing myself with the changing times? And here of course, we are talking about digital technology. So, are we adopting digital technology?

If you take Walmart as an example, Walmart was the biggest departmental store supermarket in the world. It is of course, in West US and they were the largest store by far, but when Amazon they saw the threat of e-commerce. So, it was forced to have a comparable e-commerce platform and then they started with the initially with AI. So, it is heavily powered by AI.

even before COVID-19, COVID-19 when people stopped coming or the employees they had to stay away and things like that, but they introduced AI even before COVID-19 pandemic. And this helped Walmart to achieve 37 percent more sales during the pandemic. because they took the processes before the pandemic. So, they could meet the pandemic with much better results, but for many other companies where it came as a shock and the people had to stay away customers are coming, ordering from remote

etcetera etcetera also they were not ready and they took time to you know come up with technologies and solutions to meet the COVID-19 situation or scenario.

But what you should note that the lack of proper understanding of AI can often lead to problems. Walmart is a positive case good solution, but not everybody succeeds. funding aspects see many companies especially start-ups they propose an AI model for everything. So, that is to impress the you know VCs for example, the banks who are giving them money. So, in order to get easy funding so, yeah you think that yes they are using technology good.

So, the smart people and intelligent people etcetera model and all digital etcetera AI first very good less people and they will manage things etcetera, but only over time Many of them have been found not to use any AI at all because AI was not required or justified. They just wanted to showcase that to get the funding once they have the funding they knew that all technically very smart computer people that AI is not the solution for this but they wanted to showcase that to get the funding. Even internal promotions and marketing you want to impress your boss for example, that I am very innovative.

So, you come up with an AI idea. and then you want to do an AI prototype. The best way of showing your innovation, but many of these innovations have failed because they are not justified you have not done enough research studies etcetera. So, you get the firm internal funding for your prototype you develop it and it fails. So, the executives need to be aware of such behavior amongst employees.

Because, today everybody you are saying that be innovative, come up with something new and use technology, use digital technology, use AI. So, you are tempting your employees to go ahead and experiment with all these new technologies and that will cost money, but many times most of these will fail. So, not that you should stop doing this, but you should be tell them before that have you done enough studies justify make a committee make them explain the use case whether it can be done by any other simpler tools or you need to have AI kind of thing. Now, some of the tips to maximize AI success is the tip number 1 is understand AI.

Start with AI education. Closing the adoption gaps, once you understand AI you will notice gaps in your company's infrastructure and talent pool. So, you do not need to know what are my gaps in my technology stack, we were discussing technology stack in the previous class. Is there any shortcoming, do I need new technology etcetera and of

course, the talent pool do I have the right talent. So, you will have to judge that or gauge that once you yourself are

Because then you know what is required, otherwise you have to always depend on some advices or suggestions by somebody else. Vendor selection and hiring, vendors all these technology vendors or service providers and even job candidates may try to impress you with buzzwords and hypes, but having the AI education will enable you to ask the right question and make sound purchase and hiring decisions. just that nobody can should fool you, where they will under he is an CEO or the big boss he not be so aware of the technology. So, let us tell him whatever blah blah we want to say and get through.

Optimizing investments, some business problems are suitable for AI while others are not. The ability to distinguish between these two categories is of utmost importance and only an AI education can help you to achieve this. So, every problem does not need an AI solution. So, you have to find out which problems do not need an AI solution and anybody comes up with an AI solution for that problem you know what will be your answer. Address the foundational gaps.

Technology companies like Google and Meta, Meta is known as Facebook are inherently AI driven. So, they were born digital we say digital native something like that. And they have massive AI infrastructure built in that is part of the regular job and of course, matching talent. From aggressive data collection to model development and deployment, these companies have the necessary resources, but what about you, your organizations? you are a conventional organization now you want to change you want to move towards the digital journey to undergo digital journey and develop a digital foundation.

So, you should not compare yourself or benchmark yourself with the digitally digital native companies this big one guys like Google and Amazon etcetera. You are looking to start a series of small AI projects. The first step would be to form a team to get data collection started and address the storage gaps. Where is my data? Where I am storing and how I am storing?

So, start with that. So, form a team because data is at the heart at the beginning of any AI. So, without AI data you just cannot do anything. So, start with the data collection and where are you storing it? Be clear on the ROI.

See ROI and AI surprisingly they have a distant relationship. Think ROI in terms of the benefits and not in terms of financial gains. One has to give time to achieve financial

benefit. I must give it sufficient time. This is an experiment I am doing getting into a new technology altogether.

of which clearly nobody has got a clear answer that exactly it will work and this is the benefit I will get. So, I can quantify the benefit and can calculate my ROI. In most cases it will probably not give you the desired ROI because it is a learning experience, you take it through your learning curve. for the investment and that benefits will take time come later not quickly. So, the ROI could be in terms of many years multiple years.

So, think of other benefits, operational benefit, benefit for the customer, benefit for the process, benefit for your employees mainly focus on the customer. If it is a benefit for the customer, it is worth investing. just not your immediate financial gain, because if the customers have benefited obviously, you will have less charge more customers better business etcetera, cascading effect. Check for the following, what immediate pain points will the AI solution resolve? Do my customers or somebody my stakeholders, my vendors, my employees anybody have some big pain point which can be resolved by this AI solution?

What benefits would you see by addressing the pain points? and what are the added advantage of an AI solution which is expensive over a manual one which is cheaper, but less productive. If I can make this such comparison and then I can take a better call or decision. Consider the budget, significant financial investment of training and upskilling employees, hiring data scientists, consultants, data infrastructure is required and this needs budgetary support. So, when budget is scarce experiment with AI for taking it to production is risky.

Getting quality data is a key any compromise on this leads to development of low quality of models resulting in incorrect predictions and this incorrect prediction can hurt the business in various ways. So, your budget is a big consideration point of consideration and consider all of these before you take the decision. Be committed, AI initiatives require a long term commitment, you do not get quick incredible results from AI, it takes time to acquire the right data. Say autonomous cars is another classic example, last 4-5 years we are hearing about Tesla and others, you will be on the road, you will come soon etcetera etcetera, but it is still not there. Develop models need to be periodically retrained to prevent them from becoming stale.

It is always an iterative process and models need to be continuously improved. What it means? You have trained on a set of data which is say old obviously, data is old history.

Now that data was generated under certain circumstances. Now for example, because of some change in economic situation, geopolitical situation or whatever reasons those circumstances have changed.

that though data on which it was strain those circumstances are no longer valid. So, if I use that model for my prediction in today's world which is a change circumstances, it will not give good results give out, it is like that data has become strain. Suppose, we have this example of say COVID situation and post COVID situation. So, COVID situation is lasted for 2, 2 and half years if you had those data if you had used data from those days and then you are using a developed a model and now you are also going on using that model for today when now no longer COVID situations have changed.

Take the case of people working from home and going to office. So, now, all these service companies IT companies are insisting that people start coming to work from home and from office and is being implemented seriously. So, the situation has changed. So, that model for whatever just an example if that had one of the variables as saying employees people working from home to certain extent and that impacting something. Now, much less number of people are working from home they are going to office.

So, that situation is no longer valid. So, whatever model was developed based on those parameters during the COVID times that model will not be very accurate today. So, it is a iterative process. and hence it is takes both time and money. So, you have to keep on continuously train, retrain, train, retrain your model update with current data.

Now, five common myths of AI. There are many concept misconceptions surrounding AI. One is that supposedly it can be evil, it could destroy human species and there is even the concept of AI takeover popularized by figures like Stephen Hawking and Elon Musk. or AI can be used for doing unethical things, yes. However, this is more of a function of immoral humans than the evil AI.

Example creating deep fakes like videos, images, speeches that is true for any technology for that matter even it is nuclear technology. It is blame the people who are using those or misusing those technologies. The same will true is true for AI. So, AI is just what just doing what it is task to do. If you are now working on AI then somebody criticizes you tells you see your AI it can do this bad and that bad etcetera etcetera.

So, the answer will be yes it can do, but provided the person who is using it wants to do that thing. If it does not want to do that then AI will not do any harm. So, one of the myth

is AI will replace our jobs historically, technologically has always ended up technology is always end up creating more jobs than it destroys it just changes the nature of jobs. So, as of date as of experience with many technology is coming since start of say steam engine or whatever industry revolution once we have industry 1.0, 2.0, 3.0, 4.0, now it is almost 5.0. So, computers came and people thought computers will take away plenty jobs etcetera, but then it created much more jobs than probably took away jobs.

It did take away jobs yes in many places yes through kind of automation, but then it created many other new types of jobs. So, same can be thought about AI as well. So, we still do not know what are the types of new jobs it will create. But many other jobs might go. So, it is a open question, there is no clear cut answer, but only time will answer this question or myth or whatever it is here.

AI is 99.9 percent accurate, this is not true. When AI models accurate to 95 percent, they are inaccurate on 5 percent of times and that too on data with which it is familiar. on new data the inaccuracy will be higher. So, this can have deadly consequences in medical applications example AI using CT scans for predicting disease like cancer driverless even and driverless cars etcetera. So, this accuracy if you remember when chat GPT came started.

it was trained on data up to a certain time say 21st 2021 October November something like that there was a cut off time date. And all data previous to that were used for developing this model chat GPT training 11. So, what happens is if you write a asking a question or whatever to chat GPT which is the answer is from the period beyond that time frame 21 November beyond period. So, for year 2022 or year 2023 3.0 was not aware of this data which was not given those data.

So, it cannot answer that question or if it whatever it gives it will be wrong it cannot be right answer it does not have that information. So, that is one the you know by agenda with this training data the quality how much data where data which history etcetera has been used for development how relevant it is for the time. So, this accuracy part you have to take it very carefully and the accuracy will increase with the volume of data which is being used for the training. Also the recent data the more and more recent data you get it has to be continuously reused retrained. So, that retraining which I was discussing in the previous slide is very important because that is how you increase the accuracy

because the quality of the data currency the recency everything is improving when you are using very at the present data. AI means instant incredible results, no self driving cars

by 2020 was proclaimed a many car manufacturer. However, AI takes long term commitment and a willingness to adapt with time. IBM had come up with a computer called Watson for cancer detection. It beat the world champions in many games like Geo party etcetera, but when it was put for actual medical use in cancer detection where it was supposed to a patient comes and sits in front of the computer

and talks about his or her problems ailments and symptoms. that computer this Watson was supposed to predict whether he or she has cancer, if he has then it will come up with 3 type of cancers having different probability numbers. So, cancer type A 80 percent probability, Kaiser 2 30 percent probability, Kaiser 3 less probability. So, this will help the doctor to quickly identify which type of cancer is most likely. there with the patient and all this can be could have been done can be done with probably few minutes time.

Otherwise if you compare to the regular process where you have to the doctor will subscribe you know n number of test many test and you have to undergo the test one is its time plus of course, money. before all the test reports take time they come may be 2 days 3 days 4 days and the patient has to wait the doctor has to wait once the reports come then the doctor can keep those reports and come to a conclusion. But here what is happening was it was giving a first cut indication most likely within few minutes and without any test. But this did not succeed extensive trials that carried out real life it was even in India this was used in one of the hospitals in South India whatever, but it was not very successful. So, it has been I think withdrawn and they are probably reworking that some other things.

Computer algorithms are less biased than humans. We think that computers are less biased. We humans we are biased whatever stereotyping we do etcetera etcetera. So, that is we have a personal bias through ages through when we are born and brought up etcetera through education system. We have gathered lot of biases or we become bias less or whatever whichever way look at it.

But a computer algorithm will be so much less biased than a human. But, this is not necessarily true. So, in the US study an investigative journal found that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of reoffending. So, somebody has created a small crime and he is a black person and the white person the it found that the black defendants were far more likely to be incorrectly judged. So, it is a bias at a higher risk of reoffending.

So, that he will create again repeat offense by a computer algorithm and the judge was referring to that using that. So, the bias in data is introduced in various ways like historical bias. gender racial wise etcetera, lack of representation of any particular cohort or a group. Example in facial recognition technology that is when you are doing that sampling and testing and etcetera, the data you have selected did not have representation of from say a particular sect or a community. So, the data is biased because it is not evenly distributed for different demographies.

Suppose it is a country like India and where there are so many languages, so many regions etcetera and you want to develop a tool for Indians AI model for some of the behavioral response to certain situation you want to predict. Now for training data you have taken you have missed out some areas say northeast you know many people do not go etcetera etcetera you did not strike your mind. So, you focus on the main areas north, south, east, west Delhi region, Mumbai region, Bangalore, Madras region, Hyderabad, Pune and then Calcutta region etcetera, Lucknow, Nagpur, Bhuvaneshwar etcetera some major you know location, but you forgot completely about northeast. Now that tool probably is being used by say some hospital or wherever in Delhi and many hospitals for whatever purpose maybe could be predicting disease or whatever.

Now many patients come from north east they come and then they get a biased incorrect result because when that testing was done training was done their genetic structure with that the whole thing was left out completely. intentionally or unintentionally whatever the case may be. So, these are the things which you have to be aware of that when you are trying to develop model consciously think that am I missing out on something which can then directly result in a bias. So, something not considered is also a bias something incorrectly considered is also a bias.

Bias on the contrary that is for the data and for the human side humans hold unconscious where you believe social stereotypes about certain groups of people unknowingly that is from our childhood etcetera etcetera etcetera. So, these can be both have bias. So, it is not true that algorithms are not biased or less biased they can be more biased, but we are also biased human beings also we are biased. So, that could be a competition between the two who is more or who is less biased, but keep that in mind that when you are doing AI etcetera the bias factor plays a very important role.

Hiring processes, medical care etcetera get impacted by unconscious human bias and this can also easily seep into computer algorithms from the programmers bias. The same

holds true for data source. So, one is data source, another is the programmer who is writing the developing the model. for the machine learning. Now, he is also human being right and he has a bias that is what we are discussing he has a bias.

So, if the situation is for hiring like that Amazon thing we talked about or the Harvard case we discussed about the applications for students applications etcetera. So, the person who had wrote that whatever AI model they himself or herself mean actually bias he is or her bias got into that model apart from the data bias. So, the both are types of biases can get into your model one is the data part another is as a human being. Problems that could have been solved with simple statistics are forced to use supervised ML or even deep learning.

Superiority does not come from sophistication of the tool being used. This I have already discussed earlier that you have to find out the application it does it really need AI or a deep AI ML or a deep learning etcetera. If not then that should be avoided. Now, talking about say the next part is ML machine learning development life cycle. So, how do you start the lifecycle for machine learning?

The first phase is problem definition and planning. So, this involves defining the sub problems, framing the initiatives, performing feasibility analysis and planning the AI deployment. So, let us take a problem. customers are churning out on account of increased hate speech on your website, web portal. I have a web portal, there are lot of hate speech is coming in from wherever I cannot control it and that is why because of hate speech I am losing customers, so moving out of my portal.

Let us take a Facebook, so I am going or logging out of Facebook, I do not want to be in Facebook or something like that. This is the problem for Facebook that because of hate speech many people are abandoning Facebook. So, framing AI initiatives and feasibility analysis. These are the what you have to frame, what pain points does the AI solution solve? The AI solution planning what will it solve?

What metrics are you looking to impact? Like here number of customers living per month that is a metric. How will the AI solution integrate into the business system? Do you have the necessary data to solve the problem? This helps to set expectations to collaborate with the development.

So, will the AI solution I am planning will it integrate into my after all it is a business thing like customers are coming using registering etcetera will it fit into that. And then of

course, do you have the data? So, all these head switches etcetera do you have all the data to relate data switch and churn so that I can link them. Data phase 2 is data acquisition and preparation data is the center of all AI initiatives I have talked about this. And then phase 3 is the model development, the training the model to complete a particular task.

So, you need to provide hundreds if not thousands of past examples to train model, this is the volume of data the training initiative. Model training is iterative is evaluated for accuracy on a subset of data not used, this is the test data which is not used for training, you will use it for testing it the accuracy part. And then model evaluation, tuning and experimentations are an integral part of the development, where art and science finally happens. Once you have developed the model you have to go on working on the accuracy, initially probably it will be 50 percent accuracy, 60 percent then you have to go on that is a very painstaking process doing the accuracy, tuning and experimentation.

bringing in more data, real data, live data, clean data. It needs special hardware superfast GPU processors and very large memory for data storage and consequently also very large electricity consumption to run this all this GPU processors. is just to give you a constraint. Post development testing, for post development testing also needs to be iterative in nature. So, that is with your test data, not the training data for which you need to do the testing, then you deploy the model, put it to use and then monitoring and feedback.

Model performances can degrade over time and this needs to be detected and addressed. I have talked about this earlier that it becomes stale, can happen due to changes in customer behavior or problems with the data. I gave you the example and then customers feedback and models performance can be used to fine tune the model. When you get start getting customer feedback obviously, you can use that for fine tuning a model. Lastly, this gives overall the development life cycle map.

So, this is the problem definition planning, data acquisition preparation, model development, model post development testing, then deployment and finally, monitoring and feedback. So, if you go you can do the sub steps, sub tasks from here problem definition planning, It is initiative planning and early development planning, then feasibility assessment, then for data acquisition etc you have exploratory data analysis, training data acquisition, feature engineering for modern development. And these are all this is the iterative process the development part, post development testing is also iterative part.

Then you have non-model issue resolution, key matrix evaluation, these are the standard you know life cycle management for any major IT development similar things. So, model deployment will have deployment planning and execution and model integration and then for monitoring you have the usage data collection and performance monitoring. So, nothing new here. So, for all these 6 steps you can use the standard IT development methodology. With that I will end this session.

Thank you very much.