100 AI Interview Questions and Answers



**Anish Mahapatra**

Contents

[List of Figures 7](#_Toc79274600)

[Introduction 8](#_Toc79274601)

[Q. What is the Gartner Hype Cycle? 8](#_Toc79274602)

[Structure of the blog 9](#_Toc79274603)

[**Part A: Topic-Based Questions** 9](#_Toc79274604)

[1 AutoML 11](#_Toc79274605)

[Q. What is Auto ML and how does it work? 11](#_Toc79274606)

[Q. What are some ways to implement AutoML? 11](#_Toc79274607)

[Q. What is transfer learning and how does it compare to a neural architecture approach? 13](#_Toc79274608)

[Q. What is GitHub’s copilot? Is it an implementation of AutoML – explain! 14](#_Toc79274609)

[Q. How does AutoML work? 14](#_Toc79274610)

[2 MLOps 14](#_Toc79274611)

[Q. What is MLOps and how does it work? 14](#_Toc79274612)

[Q. What are some of the ways to push a AI/ML model to production? 15](#_Toc79274613)

[Q. Explain how the MLOps lifecycle can be applied at an Enterprise level. 17](#_Toc79274614)

[Q. What is the difference between MLOps and DevOps? 19](#_Toc79274615)

[Q. What are the components of MLOps? 19](#_Toc79274616)

[3 AI Cloud Service 20](#_Toc79274617)

[Q. What is an API? How do we deploy our own API to productionalize a ML Model? 20](#_Toc79274618)

[Q. What are the popular Cloud Services that support AI implementations? Explain! 20](#_Toc79274619)

[Q. What Cloud Service would you prefer? Why? 20](#_Toc79274620)

[Q. What are some scenarios where AI Cloud Service could be used to increase ROI? 20](#_Toc79274621)

[Q. What are the benefits of Machine Learning on the cloud? 20](#_Toc79274622)

[4 NLP: Natural Language Processing 20](#_Toc79274623)

[Q. What is NLP? Explain some use-cases of NLP? 20](#_Toc79274624)

[Q. What are chatbots? What are the various implementations of chatbots used? 20](#_Toc79274625)

[Q. What are the main issues faced by companies when implementing chatbots? 20](#_Toc79274626)

[Q. Explain some NLP use-cases. 20](#_Toc79274627)

[Q. How does Transfer Learning aid in building better and faster NLP models? What are the top three most popular implementations of pretrained language models? 21](#_Toc79274628)

[5 AI Governance & Scalable AI 21](#_Toc79274629)

[Q. Now that large companies are leveraging AI to make decisions that affect human beings everyday, it is important to critique the algorithms implemented. On that note, what is AI governance? 21](#_Toc79274630)

[Q. What are the components of AI Governance? 21](#_Toc79274631)

[Q. Good AI exists. Looking at the top companies in the world, we know that a robust framework exists where AI can scale. How do the best compnies in the world build Scalable AI? 21](#_Toc79274632)

[Q. What are the difficultes when it comes to scaling implementations of AI? 21](#_Toc79274633)

[Q. State out the Enterprise AI Journey. 22](#_Toc79274634)

[6 XAI / Explainable AI 23](#_Toc79274635)

[Q. What are some of the common problems faced by companies when it comes to interpreting AI / ML? 23](#_Toc79274636)

[Q. Explain the concept of XAI and how it can be used in real world implementations? 23](#_Toc79274637)

[Q. If we were to consider AI/ML Models as black boxes that give us the desired output, what are some of the ways to help being trust back to the models being leveraged? 23](#_Toc79274638)

[Q. Explain some examples of tools and technologies that can help implement interpretable machine learning or XAI. 23](#_Toc79274639)

[Q. All of the conversations about XAI seem to be problems of the future. If you were to design a ai / ml model for a business now and productionalize a model, what are some of the considerations you would have as a lead of the project? 23](#_Toc79274640)

[7 Ethical AI / Federated Learning 23](#_Toc79274641)

[Q. With a world being driven more towards a privacy-focus mentality, what are some of the recommended approaches companies / teams can take to ? 23](#_Toc79274642)

[Q. A common statement is “With the onset of the Digital Era”. In the coming years, as AI startes to take over more and more components of human life, the statement is going to transition to “With the onset of the AI era”. Is Ethical AI a real concern or just hype? What are the ramifications? 23](#_Toc79274643)

[Q. How can Ethical AI be implemented at an enterprise level? What are its implications for Enterprise AI use-cases and Governance? 23](#_Toc79274644)

[Q. With privacy focus coming to the forefront, a lot of attention has been brought on data collected. A new branch of data collection and processing for ai / ml is federated learning. Explain further. 24](#_Toc79274645)

[Q. Explain how a model can be monitored after it is moved to production. Break down the approach. 24](#_Toc79274646)

[8 Cyber Security AI 24](#_Toc79274647)

[Q. Cybersercurity is the need of the hour. What is a good way to classify all security-related tasks using the PPDR framework? 24](#_Toc79274648)

[Q. How can machine learning be used within the space of Cyber Security. Broadly explain the various elements of Cyber Security. 24](#_Toc79274649)

[Q. Mention ten use-cases as to how AI / ML can be used within Cyber Security use-cases. 25](#_Toc79274650)

[Q. Explain the flow of Cyber Security attacks and how AI / ML models can hel plug the gaps. 25](#_Toc79274651)

[Q. How can Cyber Security be implemented across a PoC Team or at an enterprise-level with the help of AI / ML? 25](#_Toc79274652)

[9 RPA: Robotic Process Automation 25](#_Toc79274653)

[Q. What is Robotic Process Automation? Explain in brief with a few examples. 25](#_Toc79274654)

[Q. What is the difference between AI and RPA? 25](#_Toc79274655)

[Q. Explain the landscape of how RPA and AI can be used in tandem to increase revenue at an enterprise-level? Discuss a few use cases. 25](#_Toc79274656)

[Q. What are the phases of RPA at an Enterprise Level? 25](#_Toc79274657)

[10 GANs: Generative Adversarial Networks 25](#_Toc79274658)

[Q. What are GANs? 25](#_Toc79274659)

[Q. Provide an overall system view of GANs. 26](#_Toc79274660)

[Q. Explain some uses-cases on the variations of GANs. 26](#_Toc79274661)

[Q. Name some popular tools used for GANs. 26](#_Toc79274662)

[**Part-B: Scenario-Based Questions** 26](#_Toc79274663)

[1Q. How can NLP be leveraged to tackle Fake News and Cyber Bullying? 26](#_Toc79274664)

[Q. How would you build a governance model for AI in Healthcare. Give examples. 26](#_Toc79274665)

[Q. Gartner’s report says that 87% of ML Models do not make it to production. What are some of the gaps in the real world when it comes to productionalizing ML Models? 26](#_Toc79274666)

[Q. A lot of Data Science is being done in Jupyter Notebook. What are some of the ways to productionalize a Jupyter Notebook? 26](#_Toc79274667)

[5Q. Name some common test frameworks that can be used in AI / ML models? 26](#_Toc79274668)

[Q. Data Science is said to be primarily a combination of Math, Business and Technology. With automation of various systems in various fields, what does the future of Data Science look like? 26](#_Toc79274669)

[Q. If you were to pick out top 10 best and worst performing products from across multiple warehouses across India, what would the approach be? 26](#_Toc79274670)

[Q. Data Science is experimental in nature. Does logging play a part in Data Science? If so, please explain how. 26](#_Toc79274671)

[Q. What is the difference between a client-serice and an in-house model in Data Science? What are some of the benefits and drawbacks of both? 27](#_Toc79274672)

[10Q. How you would design scalable architecture to accommodate varying data needs of an organization? 27](#_Toc79274673)

[Q. What is streaming data? How are architectures designed to handle real-time data? 27](#_Toc79274674)

[Q. Apple recently came up with a host of measures to advocate for privacy. How can machine learning scale at an organizational level keeping in mind privacy? 27](#_Toc79274675)

[Q. What is FaaS? What is the imapact of FaaS in the Data Science industry? 27](#_Toc79274676)

[Q. Exploratory Data Analysis (EDA) is an activity that has traditionally taken a lot of time. Is there any way to streamline and speed up EDA? 27](#_Toc79274677)

[15Q. Let’s take a general use-case of an an internet product that is run using ads. Name some of the key metrics that can be used to measure performance? 27](#_Toc79274678)

[Q. Let’s say that we are running ads on Facebook and Instagram to advertise our product. What are some of the ways to assess if engagement has been per expectations for the last month / quarter? 27](#_Toc79274679)

[Q. Let’s say that we are Netflix and you are given a budget. How would you determine which Original Series we should put our money on? 28](#_Toc79274680)

[Q. Let’s say are serving a Fortune 500 company and we have the option of unlimited data. Is more data always better? 28](#_Toc79274681)

[Q. Let’s say you have a dataset with a million rows and 200+ columns. How can we determine what the most important features in the model are? 28](#_Toc79274682)

[20Q. Why do a lot of Kaggle Competition winners use Ensemble Learning? What is the main idea behind it? 28](#_Toc79274683)

[Q. Let’s say that we are Uber. How can you use just the GPS data from the car to determine the quality fo the driver? 28](#_Toc79274684)

[Q. Developers and product managers have hundreds of ideas on a daily basis. For the ideas that are selected to be tested, what is an efficient way to test these hypothesis in the real world? 29](#_Toc79274685)

[Q. A lot of Data Science projects start our as PoCs. List out what are the things to keep in mind while building a PoC. 29](#_Toc79274686)

[Q. What is a MVP? In a field as experimental as Data Science where experimentation is a part of the daily job description, what are some of the ways to maintain client confidence? 29](#_Toc79274687)

[25Q. How to build out a Data Science product from scratch? 29](#_Toc79274688)

[Q. Let’s say you join a new team and are given data with no context. What would be your approach to prepare for analysis? 29](#_Toc79274689)

[Q. You are given a classification problem on a fairly complix dataset. From your previous experience, you guesstimate that it will take ~3 months to processs the entire data with business context. How can you maintain client confidence intermittently? 29](#_Toc79274690)

[Q. A popular algorithm used for classification is Naïve Bayes. Why is it considered “Naïve”? 29](#_Toc79274691)

[Q. Let’s say you are given a large dataset of all customers of a telecom company with over 2 million rows and 2000 columns. You have to build a classification model to decide if a customer is likely to churn or not next month. You are given a laptop with limited RAM and no access to cloud resources. What are some of the ways you can go about this? 29](#_Toc79274692)

[30Q. You are very lucky and have got the opportunity to work for an interesting project that is deep in Data Science. The project deals with 5 VP-level stakeholders, all located across various timezones. You realize that every single person has different ambitions and expectations from this high-level project. What are some of the ways you could tackle such a situation? 29](#_Toc79274693)

[Q. Let’s say you are dealing with customer churn data and you made a classification model with an accuracy of 97%! Why is this a bad measure? Explain your thought process. 30](#_Toc79274694)

[Q. Projects in the past that used waterfall models used to give a lot of importance to documentation. However, projects that use Agile methodology do not give much importance to documentation. Do you agree with this assupmtion from the perspective of Data Science? 30](#_Toc79274695)

[Q. Let’s say you are working on a large dataset and you have figured out that the data is biased. Now what? 30](#_Toc79274696)

[Q. Let’s say that you have to upgrde the hardware for your team to be able to train complex ai / ml model. What are some of the ways you can go about this exercise in an economical manner? 30](#_Toc79274697)

[35Q. Let’s say that everyday more and more Burger King restaurants are opening across the globe. We also notice that everyday, more and more babies are being born. Does this mean that the increase in the number of Burger King restaurants is leading to more number of babies being born? 30](#_Toc79274698)

[Q. 30](#_Toc79274699)

# List of Figures

[Figure 1 The AI Hype Cycle (Source: Gartner) 9](#_Toc79274700)

[Figure 2: Data Science is a subset of Artificial intelligence 10](#_Toc79274701)

[Figure 3: AutoML Workflow 11](#_Toc79274702)

[Figure 4: The end to end MLOps Lifecycle (Image by seldon.ai under the CC BY-SA 4.0) 15](#_Toc79274703)

[Figure 5: End to End ML Workflow (Source: ml-ops.org) 16](#_Toc79274704)

[Figure 6: Model Serving (Source: ml-ops.org) 17](#_Toc79274705)

[Figure 7: MLOps: The AI Lifecycle for IT Production 18](#_Toc79274706)

# Introduction

With so many pseudo-data scientists cropping up due to numerous courses that offer theoretical learning, the questions in interviews is getting streamlined to filter those that understand how real-world implementation works. It is important to understand how data flows in the real world and what sort of questions are being discussed across companies.

**The value of a company lies solely in the interview process.**

Interviews across companies are growing to appreciate those that have an end-to-end understanding of how problems can be dealt with over theoretical understanding. This detailed blog is dedicated to help you understand the latest trends in the industry and how you can absorb them. Everything you will need to know for an interview will be detailed out here. We will start off with by understanding what the trends in artificial intelligence are.

## Q. What is the Gartner Hype Cycle?

Gartner is a market leader in market research. The Gartner Hype Cycle Methodology is an industry-approved representation of how a technology or application will mature over time. It represents initial interest, a peak in its hype, rolls down to reduced interest in the technology followed by maturity and implementation. A lot of demand in the market comes through in two points in this cycle, in the peak and the maturity. That’s where most of the money can be made from a corporate job.

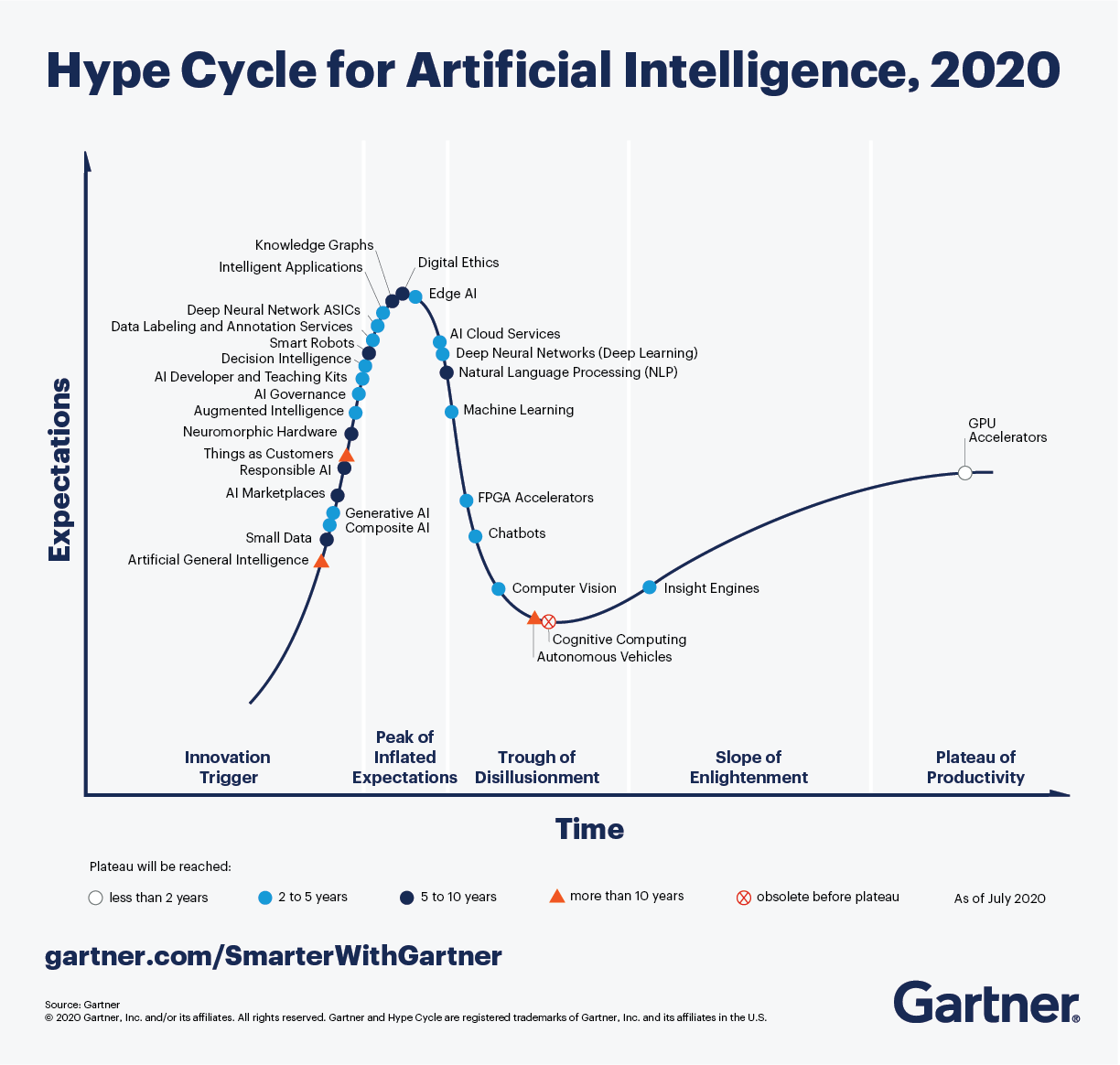


Figure 1 The AI Hype Cycle (Source: [Gartner](https://www.forbes.com/sites/louiscolumbus/2020/10/04/whats-new-in-gartners-hype-cycle-for-ai-2020/?sh=fe3826f335c4))

Looking at the Hype Cycle for artificial intelligence, focusing on the **blue** dots in the cycle, those are topics that are of interest in the next two to five years. An understanding of various innovation triggers in the space of artificial intelligence will help elevate conversation and make you stand out in the eyes of the interviewer. The answer is never to simply answer the question, but add an additional layer of context that will stick.

This blog will not tell you exactly how to answer questions when asked. It will explain what an instace of the best-in-class answes would sound like. We will try to use as many visual aids and examples to answer so that you can apply it in multiple scenarios and interviews.

# Structure of the blog

A lot of thought was put in as to what would make most sense in terms of structure when putting forth 100 AI interview questions and answers. For the first 50 questions, we will go over 5 questions each for 10 topics followed by 50 scenario-based questions. In the first section, we will try and cover as many current and future proof topics that we have seen being asked in some of the best analytics companies.

# **Part A: Topic-Based Questions**

Before we deep-dive into the questions below, let’s understand more about artificial intelligence and where it falls in the spectrum. In the recent past, Data Science has taken off in the space of technology, even being crowned as the Sexiest Job of the 21st Century. Let’s understand where Data Science belongs in the space of Artificial Intelligence.

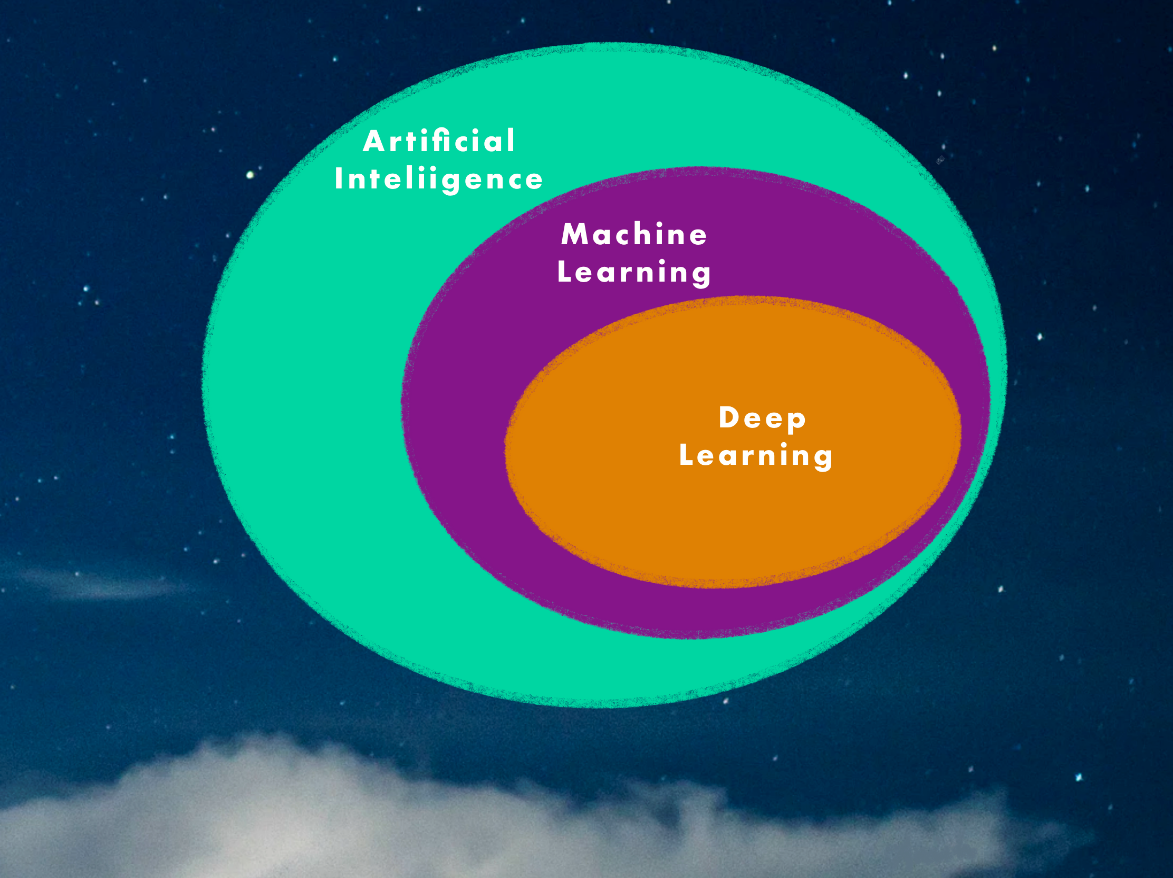


Figure 2: Data Science is a subset of Artificial intelligence

* **Deep Learning**: Subset of machine learning that leverages neural networks to train itself to perform layered tasks such as speech and image recognition
* **Machine Learning**: Subest of Artificial Intelligence which enables algorithms to learn without being explicitly programmed
* **Artificial Intelligence**: Programs with the ability to learn and reason like human beings

When explaining questions to an interviewer, remember, they are smarter than they look. No one knows everything about any topic. Especially in a subject like Artificial Intelligence, where new things are coming up at an exponential rate. When answering questions, be humbe and truthful. If you know the answer to the question, answer confidently. If you are unsure, be vocal about your thought process and the way you are thinking – take inspiration from the examples below and explain the answer to the interviewer through your experiences from projects and learnings. So right before we start, I would like to let you know that the focus of this blog is to get you started for interviews and give you exposure to what is the latest in the field of Artificial Intelligence. Consider all of the answers as one way to go about it and please feel free to pitch in your thoughts as well. Alright, we have had enough fun building up to this moment, let’s dive right into the questions!

# 1 AutoML

## Q. What is Auto ML and how does it work?

A. So, for anyone who has heard or worked on a traditional machine learning flow, we know that a traditional machine learning flow goes from Data Acquisition to Predictions. The following diagram explains Auto ML.

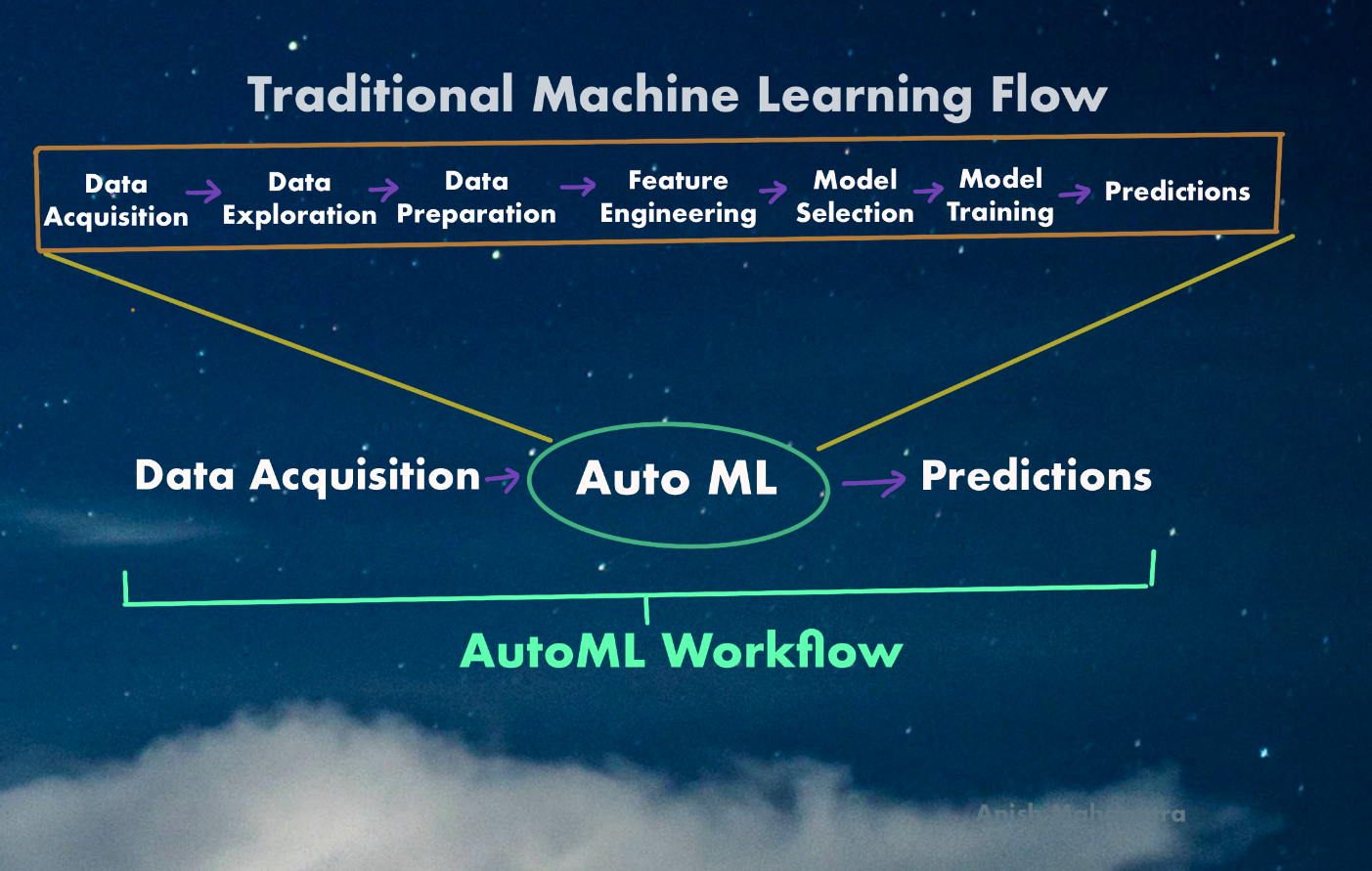
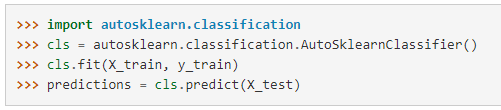


Figure 3: AutoML Workflow

As a Data Scientist, numerous activities go into the traditional machine learning flow. The process of automating the tedious, iterative tasks in the development of machine learning is known as AutoML. The benefit of AutoML is that it is a low-code or no-code approach, wherein folks who do not have experience in Machine Learning can also implement it.

## Q. What are some ways to implement AutoML?

A. Following are some of the popular ways to implement AutoML:

* **Auto-SkLearn**:   
  Scikit-learn is a package that every data scientist has used. Auto-sklearn is an automated machine learning toolkit that does that same thing except that it is an automated replacement for a scikit learn estimator.   
    
  The estimator automatically performs the algorithm selection as well as the hyperparameter tuning
* **Auto-Keras**:   
    
  To recall, Keras is an open-source library that provide a Python interface into the world of Artificial Intelligence, especially Tensorflow. AutoKeras is the AutoML system that is based on Keras. AutoKeras focuses on making machine learning and deep learning easier with the help of Neural Architecture Search. This will help non-coding folks achievestate of the art performance with minimal coding effort.
* **Auto-Weka**:  
    
  Weka is an extremely popular java-based machine learning software for data exploration. Similar to other AutoML applications, AutoWeka considers all possible algorithms and even applies hyperparameter tuning to be able to select the optimal fit using the latest strides in Bayesian Optimization.
* **Amazon Lex (Amazon)**:   
    
  Consider Amazon Lex as the AutoML backend of the extremely popular Amazon Alexa. With features such as Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU), if you are building a chatbot, this is a great toolkit.
* **Azure ML Studio (Microsoft)**:   
    
  With Azure ML Studio, Microsoft provides an AutoML application with a Grahical User Interface and Drag & drop format to try multiple algorithms along with hyperparameter tuning. It has automated componenets of the traditional ML Flow from data acquisition, experimentation and even logging. Definitely a must try within the Azure ecosystem.
* **Vertex AI (Google)**:   
    
  Google recently (May 2021) launched a unified AI platform to build, deploy and scale machine learning models for individuals as well as entrprises. This AutoML application also includes MLOps tools to manage data and models at scale. For the entire ML Workflow, Vertex AI is a one stop shop for end to end integration of AI that supports all open source frameworks.
* **H20AutoML**:   
     
  Anyone who has attemted to work on AutoML in the past decade has heard of H20.ai. H20AutoML is H20.ai’s foray into the space of AutoML. It is a function with the goal of finding the best model with minimal knowledge or effort from the Data Scientist.

## Q. What is transfer learning and how does it compare to a neural architecture approach?

A. Transfer learning is a process wherein the learning from an already developed model is resued and built upon for a new task. Essentially, it is using pre-trained models and customizing them for your use-case.

**Transfer Learning versus Neural Architecture Search**

The premise of transfer learning is that neural networks generalize and converge when it comes to similar use-cases. For example, boxes, squares, rectangles and crystals. This assumption allows us to build more on top of these generalizations for specific use-cases.

Neural Architecture search, on the other hand is based on the premise that each use-case is unique and should be done up from scratch. This novel architecture with its own hyperparameters will perform the best. However, the disadvantage is that we will also have to train from scratch and pre-trained models cannot be used.

A quick note here – Google’s latest AutoML products subscribe to the school of thought that the neural architecture search needs to be used directly for the most optimal approach. This is an area that is still being explored and with progress in AutoML, we will learn more.

## Q. What is GitHub’s copilot? Is it an implementation of AutoML – explain!



GitHub is the go-to website for source code management. GitHub copilot is an AI tool developed collaboratively by GitHub and OpenAI to assist users by autocompleting code. Copilot is set to be an extension for MS Visual Code that not just autocompletes a few functions or phrases, but entire functions! This is an extenstion of pair programming, where two programmers work on the same set of code to speed up the development process. It was trained on all available open-source code and is powered by a deep neural network language model called Codex.

Since GitHub copilot still aids the developer in aut-completing code, it would not classify completely as AutoML in it’s current state in my opinion. However, as the project progresses and moves out of the *preview* stage, GitHub Copilot can be the gold standard of AutoML tools.

## Q. How does AutoML work?

A. If you have ever automated anything, pipelines are the best way to automate and orchestrate workflows. AutoML is essentially a set of automated pipelines, that when trigerred simply try out all the permutations and combinations until they come up with the top results.

This would include the automation of a standard machine learning workflwo which would include the steps of

* Gathering the data
* Preparing the Data
* Training
* Evaluation
* Testing
* Deployment and Prediction

This includes the automation of tasks such as Hyperparameter Optimization, Model Selection and Feature Selection. The orchestration of pipelines can happen on various platforms, but are generally integrate with cloud providers such as Microsoft Azure, Amazon Web Services or Google Cloud Platform at an Enterprise Level.

# 2 MLOps

## Q. What is MLOps and how does it work?

A. If you have worked on machine learning projects and proof of concepts (PoC), you must have realized that a lot of AI / ML models end up not going into production, and hence, do not make generate any revenue for the stakeholders. This is the main reason why projects are shortlived and there is so much experimentation that does not lead to desired results.

MLOps is an end-to-end machine learning development process that aims to design, build and manage reproducible, scalable and evolvable ML-powered software. It is the perfect culmination of software engineering into machine learning so that the outcomes of machine learning are products that can scale.

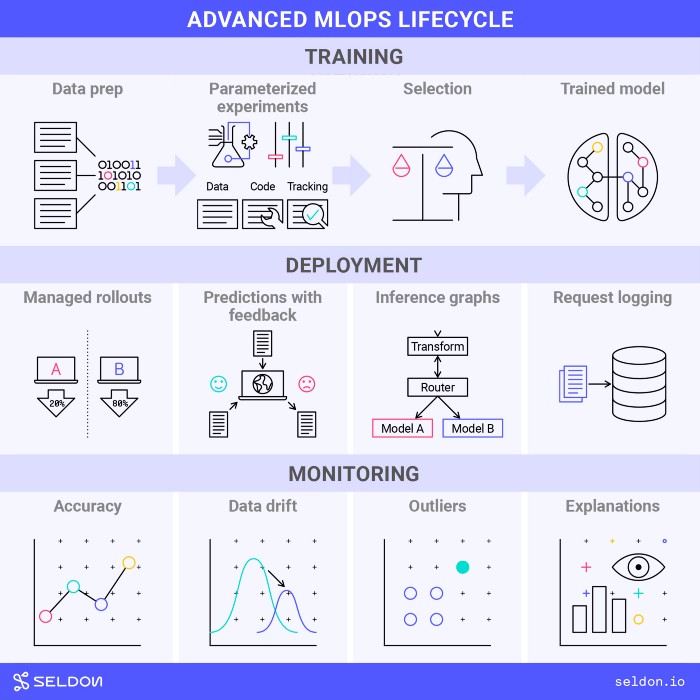


Figure 4: The end to end MLOps Lifecycle  
(Image by seldon.ai under the [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/))

A classic example of how MLOps can be broken down into an ideals MLOps Lifecycle has been given above, which consists of training, deployment and monitoring.

## Q. What are some of the ways to push a AI/ML model to production?

When a model is in the prototype and development phase, a lot of teams use IDEs such as Jupyter Notebook. While this might seem like the best option during development as we are able to see the results instantly, it is advisable to write code that is worthy of being moved into production. The entire cycle can be visualized as shown below.



Figure 5: End to End ML Workflow  
 (Source: [ml-ops.org](https://ml-ops.org/content/end-to-end-ml-workflow))

Now, let’s discuss how AI / ML models can be deployed into production. The ideology behind all of the steps is **pipelines**! Experiment to see what works best for your data, automate it using pipelines and then monitor the performance of the workflow.

* **Data: Data Engineering Pipelines**Data is everything. Make sure that the quality of data works for your use-case.
  + Data Ingestion
  + Exploratory Data Analysis (using RAD Tools)
  + Validation
  + Data Wrangling
  + Data Splitting
* **Model: Machine Learning Pipelines**Every team enjoys the experimentation with data. It is important to also be transparent with the business during the phase of documentation and also give equal focus to logging all experiments & learnings.
  + Model Training
  + Model Evaluation
  + Model Testing
  + Model Packaging
* **Code: Deployment Pipelines**   
  Once the final ML model has been decided, it is critical for the model to be pushed into production. This would include the following phases:
  + Model Serving
  + Model Performance Monitoring
  + Model Performance Logging

Based, on demand, the model serving requirements can be further analyzed.

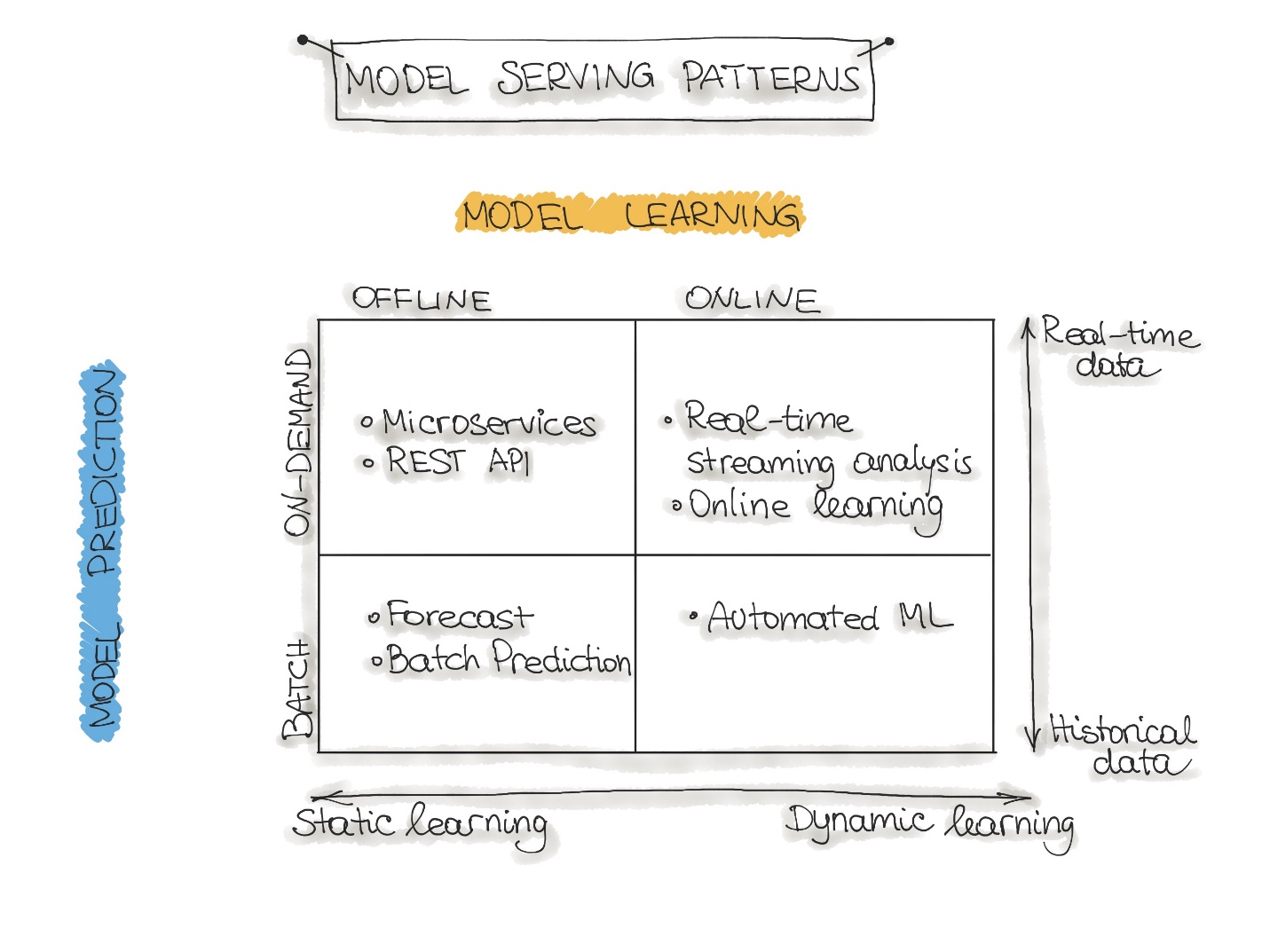


Figure 6: Model Serving  
(Source: [ml-ops.org](https://ml-ops.org/))

## Q. Explain how the MLOps lifecycle can be applied at an Enterprise level.

For Machine learning models to go into production, we need to stop thinking of machine learning as strictly iteratively experimental. MLOps is the marrige between machine learning and software engineering. The final output should be visualized as a product. So, when a technology product is being developed, the code needs to be functional, modular and tested.

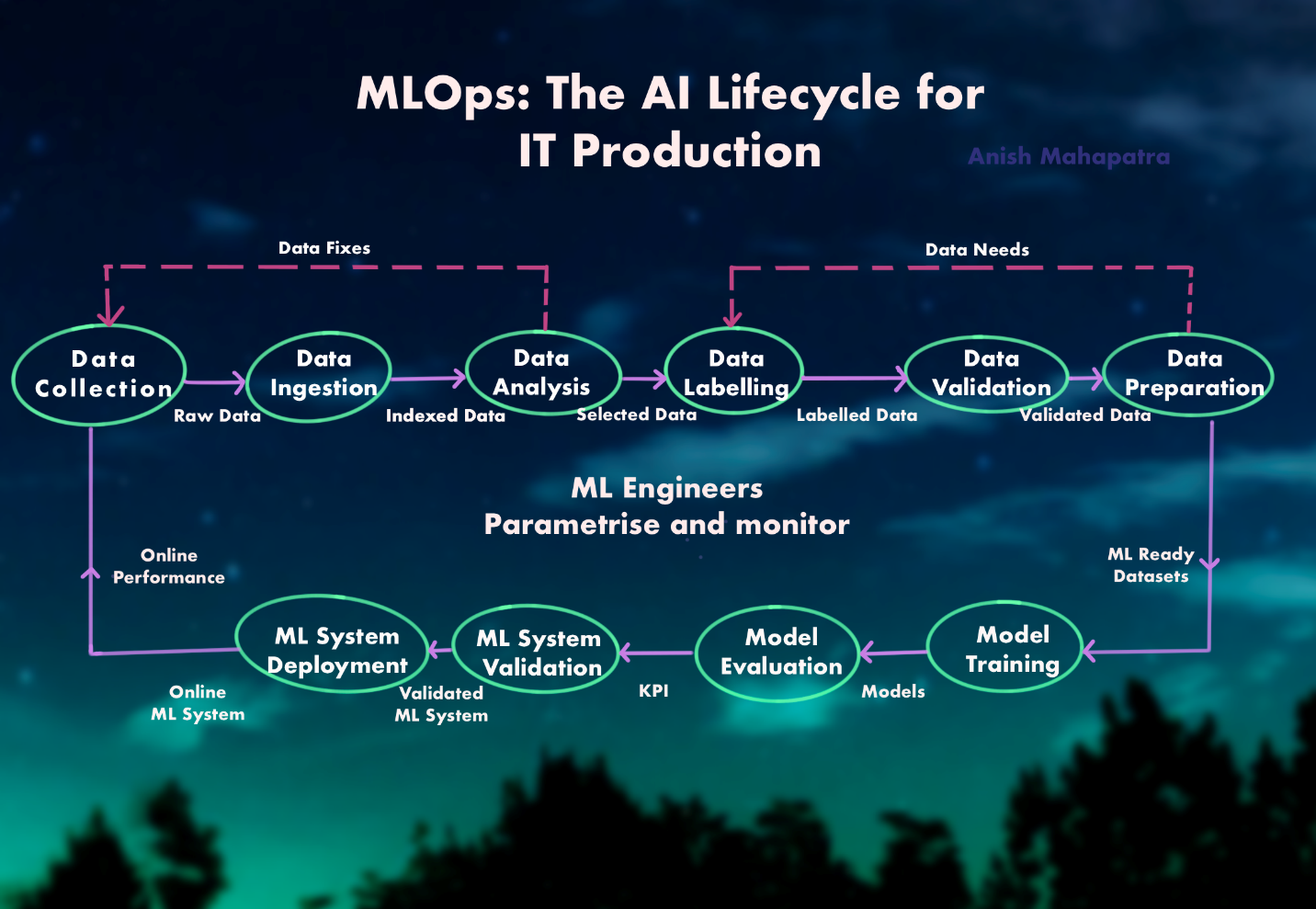


Figure 7: MLOps: The AI Lifecycle for IT Production

MLOps follows a similar lifecycle as a traditional machine learning flow except that it the model continues in the workflow to production. This is then monitored by MLOps Engineers to ensure that the quality of the model in production is as expected. In terms of the tools used in MLOps, here are a few use-cases.

* **Model Registries**   
  This is exactly what it sounds like. Model Registries larger teams keep track of version models and store them. It is even possible to revert to an older version
* **Feature Stores**   
  When experimenting with larger sets of data, there might be various versions of the analytical datasets and subsets for different tasks. A feature store is a modern, elegant solution to leverage data prep work from previous runs or other teams as well
* **Metadata Stores**   
  For unstructured data such as image and text data to be leveraged effectively, it is important to track metadata effectively in production

## Q. What is the difference between MLOps and DevOps?

DevOps is a set of practices used across enterprises to reduce the software development lifecycle. It provides a continuious integration and continuous development (CI/CD) approach to the software development lifecycle.

**Highly Experimental**   
While DevOps works in an environment that is more application and optimization oriented, machine learning systems are highly experimental in nature. Thus, MLOps needs to take into account various experiments, feature engineering, model parameters, model tuning etc.

**Testing**   
MLOps goes beying unit-testing. Model KPIs such as data quality parameters, model drift, model bias and performance needs to be tracked

**CACE Principle**   
MLOps follows the CACE Principle – **C**hange **A**nything and it **C**hanges **E**verything. Machine learning systems are highly dependent on the nature of the problem and the data in consideration.

**Deployment**   
Deployment for machine learning systems involves multiple steps such as data processing, feature engineering, model training, model registry and model deployment. DevOps follows a CI/CD approach, but in MLOps, there is an additional concept of Continuous Training (CT)

**Tracking Lifecycle**   
In DevOps, tracking the performance of the code along is sufficient. In MLOps, it is necessary to track the data quality parameters through every step in the workflow as well

## Q. What are the components of MLOps?

There are a few ways to break down this question. One way that MLOps can be broadly broken down is:

* **Design**   
  Design thinking plays a big role in MLOps. All the way from the nature of the problem, the hypotheses to be tested, the architecture and the deployment
* **Model Development**   
  This phase includes the data engineering pipelines, the experimentation to get the right machine learning systems in place and model testing & validation
* **Operations**   
  As part of the operations, the model has to be deployed, tested and analyzed over time. The CI/CD pipelines are triggered using an orchestration tool and monitored.

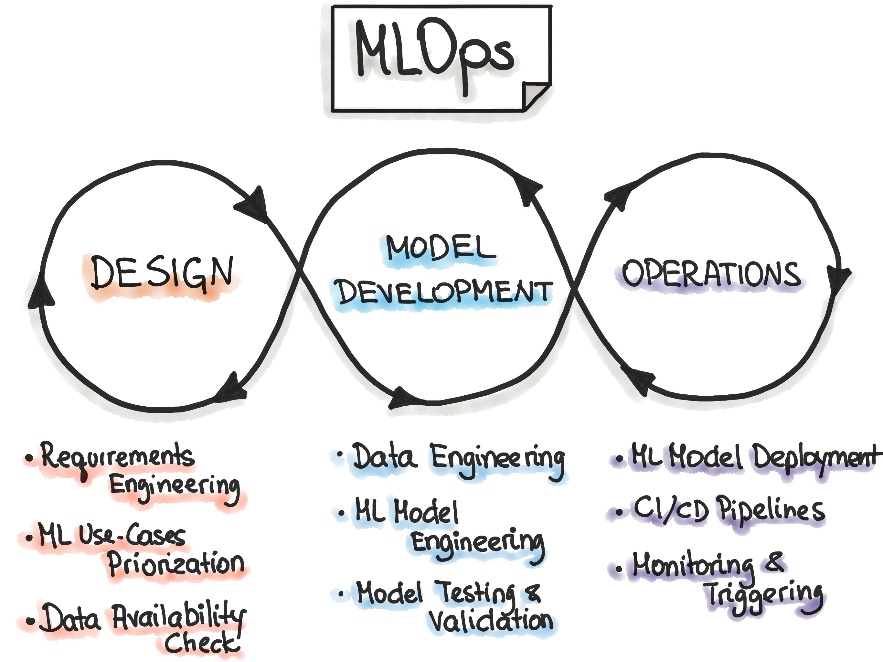


Figure 8: Components of Machine Learning  
(Source: [ml-ops.org](https://ml-ops.org/))

# 3 AI Cloud Service

## Q. What is an API? How do we deploy our own API to productionalize a ML Model?

A. Most Data Scientists know how to run python code on a Jupyter Notebook. We run the codes, do data analysis, come up with the final model result and stop there. How do machine learning systems in the real-world interface with the rest of the systems in place? Even though python is great to experiment and peform machine learning, in terms fo deployment, there are two main options:

* **Rewrite the logic** in a language that fits into the technology stack being leveraged   
  Imagine going through the entire machine learning workflow only to have to do it again. This is an option I would not recommend as the return on invstment is not worth it.
* **API-first approach**Okay,so how do appications that use different frameworks or languages cross-communicate? APIs! APIs help cross-language applications function efficiently. So, what is you made the final machine learning model that makes sense and you create a sort of endpoint that other developers or the front-end team can connect to. All they would need is the URL endpoint from where the API is being served.

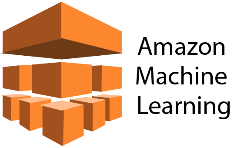
API – Application Programming Interface help applications talk to each other. The person who needs to connect to the API needs to perform a simple REST call to the API using a Software Development Kit (SDK).

The two more popular approached to productionalizing our machine learning models is using Flask or FastAPI. So, basically, we serialize/deserialize the machine learning model objects into a pickle (.pkl) or a h5py file and we let users query it.

Think about it this way, let’s say we are deploying a model that predicts the probability of churn for a customer. So, if the query was a new customer, then the returned results would be the probability, which we could configure in the front end to say *churn* or *no churn* based on the threshold.

## Q. What are the popular Cloud Services that support AI implementations? Explain!

A. AI Implementation is a fancy way of saying are you able to handle an end to end implementation of AI / ML models on the cloud? To understand further, it is a good idea to get familiar with the term MLaaS, which is Machine Learning as a Service. MLaaS offers multiple ML related services as an additonal component to cloud computing. There are multiple options in the market, so it would depend on the use-cases being tackled. Let’s discuss a few of the popular ones:

* **Google’s Vertex AI**   
     
  Google recently unveiled their Unified MLOps & AI Platform – Vertex AI to help Data Scientists / ML Engineers increase experimentation, deploy faster and manage models better
* **AWS Machine Learning**   
    
  AWS is the largest player in the cloud computing market. Their Machine Learning offerings have matured over time to cater to all machine learning needs from speech recognition, computer vision, AI etc.
* **Azure Machine Learning**   
    
  Azure has been grabbing more market share over the years because it has been doing a lot right in the machine learning space. Azure Machine Learning is a no-code, drag and drop interface to perform machine learning at scale along with model management.

## Q. What Cloud Service would you prefer? Why?

This is a critical thinking question. It relies on your understanding of the use-case, cloud services and confidence in implementing scalable solutions. Probe more into the functionality or use-case the interviewer is asking about. There are multiple factors one can consider:

* Is the existing data on on-prem servers or already on a cloud platform?
* Does the team already use an existing AI Cloud Service Provider?
* What is the skillset of the team on the cloud platform?
* What is the nature of machine learning being performed?
* What sort of architectural needs will the project have and what are the associated costs?
* How future-proof is the project and the platform?

Objectively, if we are comparing platforms, GCP is ahead of AWS and Azure. AWS is difficult to configure and setup – it is confusing for beginners and is comparitively complicated.

Azure is good for basic functionalities, but there is a lot Azure needs to mature over in terms of Machine Learning. GCP is new and so, uses the latest tech stack. Everything is bleeding edge, plus there is nobody that does Machine Leaning like Google, so the entire cloud computing stack is machine learning focused.

A few more tangible points why GCP is better:

* A lot of Google’s work is open-source
* Their Speech and Translate APIs are market leaders in their respective segments
* AutoML is pretty great for multiple use-cases
* They have custom designed ASICs called TPUs for optimized workloads

## Q. What are some scenarios where AI Cloud Service could be used to increase ROI?

A. Some scenarios where AI Cloud Services can be leveraged are:

* **Customer Service Automation**: Integrating Machine Learning and speech recognition into customer service can help save a lot of overhead costs
* **Personalization and Engagement**: Tailoring personalized customer experiences can lead to better sales and happier customers, which in turn leads to a higher revenue
* **Fraud Detection**: Multi-touch fraud detection algorithms can help isolate and pick out points where gaps can be plugged
* **Scalable Media**: Based on number of concurrent users and workload, a scalable architecture optimized to user preferences can lead to higher engagement and revenue
* **Forecasting**: Multiple decisions can be optimized by leveraging accurate forecasting in terms of sales, supply-demand, costing, dynamic pricing etc.

## Q. What are the benefits of Machine Learning on the cloud?

For use-cases that need both proof of concepts or scalable solutions, working on the cloud can have multiple advantages. Some fo the benefits are listed below:

* **More Economical**   
  Working in a pay-per-use model has numerous cost benefits that can be passed on to the business. The cost of maintaining and upgrading legacy hardware is also eliminated
* **Less configuring**   
  One of the most time-consuming aspects is to configure and setup the same set of applications across multiple hardware touchpoints. This needs to be kept track of along with security and updates. The advantage of working on the cloud is that everything is handled by the experts for a reduced cost as a result of economies of scale
* **Elastic Architecture**Whether you need 1/10th of your existing needs or 1000 times more. Both can be done with cloud architecture with the click of a button
* **Better support for streaming architectures**   
   With real-time analysis taking the forefront where multiple consumers are getting used to things being done instantly, there is much greater support for real-time or in-time architectures as compared to on-prem settings.

# 4 NLP: Natural Language Processing

## Q. What is NLP? Explain the components of NLP?

Human langauge is spread with intircacies that make it complicated to decipher. Obtaining intended meaning from speech or text is even more difficult especially when software is used. Natural Language Processing (NLP), a part of Information Retrieval breaks down speech and text to help software understand what is happening.

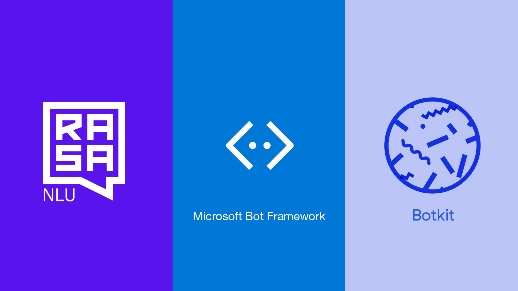
The components of NLP are as follows:

* **Speech Recognition**Recognizing speech and being able to convert it reliably to text. There are various components of speech itself along with multiple accents globally, which makes this a challenge
* **PoS Tagging**Differentiationof the parts of a sentence to help tag the parts of speech such as noun, verb, adjective etc.
* **Word sense disambiguation**Semantic Analysis to determine the right meaning of the word being used
* **Naming Entitiy Recognition**Identifying proper nouns as entities as a part of the sentence such as name, location etc.
* **Co-reference resolution**Whether pronouns are referring to the same object. For instance, the question *how old is Obama?* And the follow-up question of *how tall is he?* should provide the right answer if the co-reference resolution is configured correctly
* **Sentiment Analysis**Understanding the tone of the speech or text based on subjective qualities or words used. This is used in social media to better gauge sentiments towards an event or a product.
* **Natural Language Generation**This is giving power to the algorithm to make sentenses that would make sense to humans. This is currently a hot area of research along with NLU (Natural Language Understanding)

## Q. What are chatbots? What are the various implementations of chatbots used?

Chatbots are the medium through which NLP is carried out. A chatbot is a software that simulates human-like conversations to enable users to do tasks that can be automated. Chatbots can help business generate revenue by reducing human effort to do repetitive tasks and even generate sales by interacting with potential leads.

The popular implementations in the market are:

* **Google’s Dialogflow**   
    
  Google is great at speech and natural language processing. Dialogflow grants enterprise users access to all of Google’s tremendous Speech APIs that can be integrate within various platforms
* **Facebook’s wit.ai**   
    
  With Facebook owning platforms such as Facebook, Instagram and WhatsApp, facebook’s reach in marketing is tremendous. Wit.ai is Facebook’s NLP offering to enable users to integrate chatbot with multiple applications to aid for an enhanced experience.
* **Microsoft’s BotKit**   
    
  Botkit integrate seamlessly into Microsoft’s ecosystem with a support frameowrk that can support a host of applications such as Teams, messanger, Slack, Telegram etc. and can be automated as well
* **Rasa**   
    
  Rasa is an open source framework to tackle the best problems in NLP, NLU. It is made for persnalized conversations at scale

## Q. What are the main issues faced by companies when implementing chatbots?

When chatbots are to implemented atop existing tech stacks, the following are a few of the common issues seen across clients:

* Compatibility with existing technology stack
* Security vulnerabilities (as chatbots have to query sensitive datasets)
* Transparency and visibility into end to end implementation
* Implementation at scale along with growth of the business
* Future-proofing AI implementations
* Return on Investment

A use-case I have come across when Fortune 500 clients inquire about chatbots is whether it is possible to perform everything without the intervention of third parties such as Google, Facebook and Microsoft. So, Rasa gets recommended a lot as it can even work without the internet (locally) and it integrates just as easily with a larger set of applications at multiple levels.

## Q. Explain some NLP use-cases.

Some of the popular use-cases of NLP are explained below:

* **Understanding customer sentiment**: Analyzing multiple touch points to understand the pain points of customers can lead to quicker, more accurate resolution. This reduction in friction can lead to higher sales.
* **Multilingual Support**: As businesses go global, reducing the barrier of language with the help of multi-lingual support can play a big role in scaling companies and products
* **Entity Extraction**: Understanding the relevant parts of the sentence along with intent classification can aid in analysing specific demographics
* **Automation of receipt and invoice processing**: Automation of invoice processing with the help of textual analysis and fraud detection can help reduce operational costs tremendously
* **Content classification**: Classification of inappropriate content can save companies cost and time. Classification of relevant content in a domain-specific manner can help build contextual relationship graphs

## Q. How does Transfer Learning aid in building better and faster NLP models? What are the top three most popular implementations of pretrained language models?

Training NLP models takes time. Imagine training a model of the complete context of human language. So, it only makes sense that we should be able to reuse the training rather than try to train from scratch every single time. That is what transfer learning achieves.

Transfer learning is building on existing training to train and build towards another use-case. It is building more specific context over generalized context. Three of the most popular NLP implementations are as follows:

* OpenAI’s GPT-3 (Generative Pre-Trained Transformer 3): The implementation of GPT-3 has been pegged as a revolution in NLP. It has achieved a God, cult-level status as the best implementation of NLP to date. However, due to it’s limited availability, it is yet to be tested by the general public. We have had the chance to see a few use-cases and the results look unbelieveably positive
* BERT (Bidirectional Encoder Representations from Transformers): BERT is designed to pre-train deep bidirectional representations. BERT is conceptually simple and empirically powerful
* XLNet: XLNet is ageneralized autoregressive pretraining method that leverages the best of both autoregressive language modeling (Transformer-XL) andbidirectional capability (BERT)

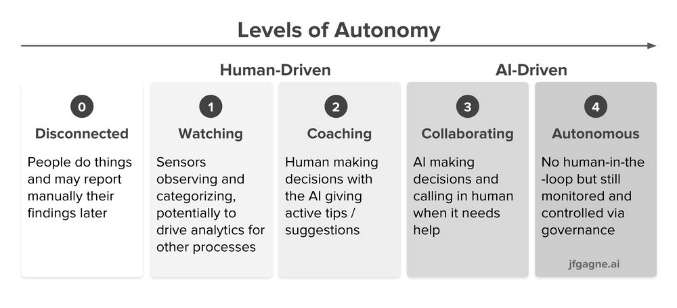
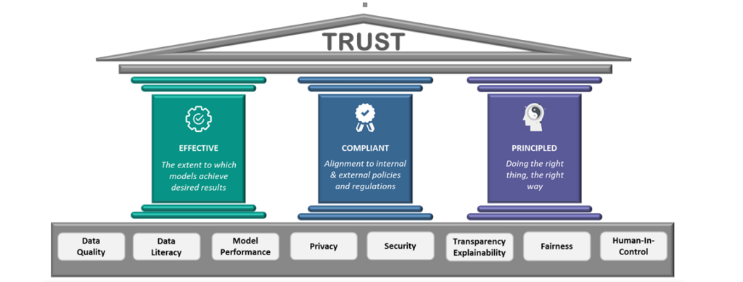
# 5 AI Governance & Scalable AI

## Q. Now that large companies are leveraging AI to make decisions that affect human beings everyday, it is important to critique the algorithms implemented. On that note, what is AI governance?

A. [The shape of AI governance to come - KPMG Global (home.kpmg)](https://home.kpmg/xx/en/home/insights/2020/12/the-shape-of-ai-governance-to-come.html)   
[perspectives-on-issues-in-ai-governance.pdf](https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf)

## Q. What are the components of AI Governance?

A. [AI Governance – What are the KPIs? And who is accountable? (2021.ai)](https://2021.ai/ai-governance-kpi/)   
[AI Governance Market Size, Share and Global Market Forecast to 2026 | MarketsandMarkets](https://www.marketsandmarkets.com/Market-Reports/ai-governance-market-176187291.html)



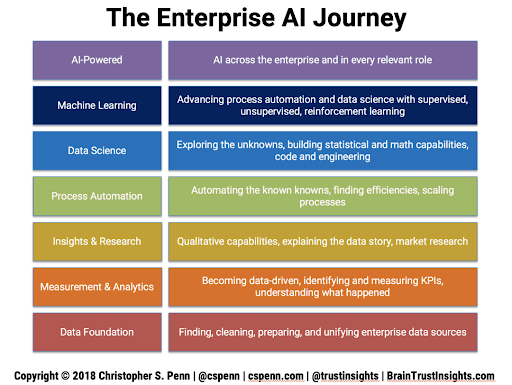
## Q. Good AI exists. Looking at the top companies in the world, we know that a robust framework exists where AI can scale. How do the best compnies in the world build Scalable AI?

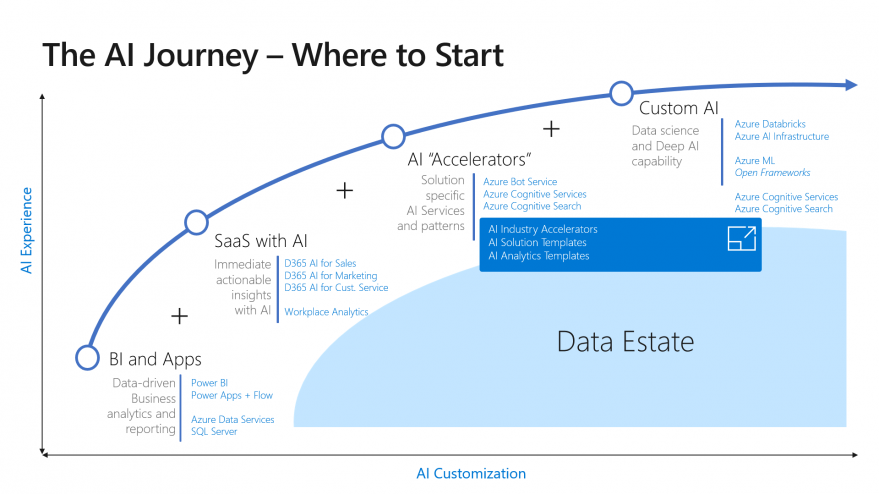
A. [How leading companies scale AI. Five key principles to build an AI… | by Wouter Huygen | Towards Data Science](https://towardsdatascience.com/how-leading-companies-scale-ai-4626189faed2)   
[2021\_019\_001\_735330.pdf (cmu.edu)](https://resources.sei.cmu.edu/asset_files/WhitePaper/2021_019_001_735330.pdf)

## Q. What are the difficultes when it comes to scaling implementations of AI?

A. [Scaling AI: 5 Reasons Why It's Difficult | Towards Data Science](https://towardsdatascience.com/scaling-ai-5-reasons-why-its-difficult-6ea77b9f7d48)   
[Scaling AI: How to Make It Work for Your Company | Accenture](https://www.accenture.com/us-en/blogs/intelligent-functions/scaling-ai-how-to-make-it-work-for-your-company)

## Q. State out the Enterprise AI Journey.





# 6 XAI / Explainable AI

## Q. What are some of the common problems faced by companies when it comes to interpreting AI / ML?

A.

## Q. Explain the concept of XAI and how it can be used in real world implementations?

A. [What is Explainable AI (XAI)?. An introduction to building trust in… | by Nicklas Ankarstad | Towards Data Science](https://towardsdatascience.com/what-is-explainable-ai-xai-afc56938d513)

## Q. If we were to consider AI/ML Models as black boxes that give us the desired output, what are some of the ways to help being trust back to the models being leveraged?

A. [Explainable AI — All you need to know. The what, how, why of explainable AI | by Lawtomated | Medium](https://lawtomated.medium.com/explainable-ai-all-you-need-to-know-the-what-how-why-of-explainable-ai-dcf2287a9f6c)

## Q. Explain some examples of tools and technologies that can help implement interpretable machine learning or XAI.

A. [Explainable AI (XAI) — A guide to 7 Packages in Python to Explain Your Models | by Prateek Bhatnagar | Towards Data Science](https://towardsdatascience.com/explainable-ai-xai-a-guide-to-7-packages-in-python-to-explain-your-models-932967f0634b)

## Q. All of the conversations about XAI seem to be problems of the future. If you were to design a ai / ml model for a business now and productionalize a model, what are some of the considerations you would have as a lead of the project?

A. [Explainable Artificial Intelligence (XAI): Will it be “The Next Big Thing” in AI ? | Medium | Medium](https://medium.com/@arunvaish9/explainable-artificial-intelligence-xai-will-it-be-the-next-big-thing-in-ai-eb240062a822)

# 7 Ethical AI / Federated Learning

## Q. With a world being driven more towards a privacy-focus mentality, what are some of the recommended approaches companies / teams can take to ?

A.

## Q. A common statement is “With the onset of the Digital Era”. In the coming years, as AI startes to take over more and more components of human life, the statement is going to transition to “With the onset of the AI era”. Is Ethical AI a real concern or just hype? What are the ramifications?

A.

## Q. How can Ethical AI be implemented at an enterprise level? What are its implications for Enterprise AI use-cases and Governance?

A.[Ethical AI: its implications for Enterprise AI Use-cases and Governance | by Debmalya Biswas | Towards Data Science](https://towardsdatascience.com/ethical-ai-its-implications-for-enterprise-ai-use-cases-and-governance-81602078f5db)

## Q. With privacy focus coming to the forefront, a lot of attention has been brought on data collected. A new branch of data collection and processing for ai / ml is federated learning. Explain further.

A. [Federated Learning | Towards Data Science](https://towardsdatascience.com/understanding-federated-learning-99bc86a0d026)

## Q. Explain how a model can be monitored after it is moved to production. Break down the approach.

A. [MLOps: Model Monitoring 101 | Towards Data Science](https://towardsdatascience.com/mlops-model-monitoring-101-46de6a578e03)

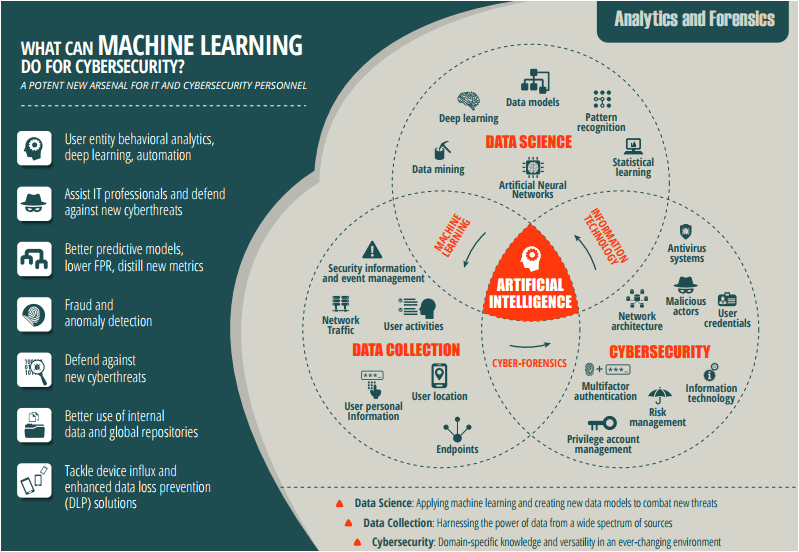
# 8 Cyber Security AI

## Q. Cybersercurity is the need of the hour. What is a good way to classify all security-related tasks using the PPDR framework?

A.

## Q. How can machine learning be used within the space of Cyber Security. Broadly explain the various elements of Cyber Security.

A. [Machine Learning for Cybersecurity 101 | by Alex Polyakov | Towards Data Science](https://towardsdatascience.com/machine-learning-for-cybersecurity-101-7822b802790b)   
[How Machine Learning will Transform Cybersecurity - BoTree Technologies](https://www.botreetechnologies.com/blog/machine-learning-in-cybersecurity/)



## Q. Mention ten use-cases as to how AI / ML can be used within Cyber Security use-cases.

A. [Machine Learning for Cybersecurity 101 | by Alex Polyakov | Towards Data Science](https://towardsdatascience.com/machine-learning-for-cybersecurity-101-7822b802790b)

## Q. Explain the flow of Cyber Security attacks and how AI / ML models can hel plug the gaps.

A. [Can Machine Learning Help Cybersecurity Systems? | Stout](https://www.stout.com/en/insights/article/can-machine-learning-help-cybersecurity-systems)   
[Artificial Intelligence vs. Machine Learning in Cybersecurity | Varonis](https://www.varonis.com/blog/ai-vs-ml-in-cybersecurity/)

## Q. How can Cyber Security be implemented across a PoC Team or at an enterprise-level with the help of AI / ML?

A. [Eliminating the complexity in cybersecurity with AI - Wipro](https://www.wipro.com/cybersecurity/eliminating-the-complexity-in-cybersecurity-with-artificial-intelligence/)   
[How Machine Learning will Transform Cybersecurity - BoTree Technologies](https://www.botreetechnologies.com/blog/machine-learning-in-cybersecurity/)   
[Can Machine Learning Help Cybersecurity Systems? | Stout](https://www.stout.com/en/insights/article/can-machine-learning-help-cybersecurity-systems)

# 9 RPA: Robotic Process Automation

## Q. What is Robotic Process Automation? Explain in brief with a few examples.

A. [Differences Between AI And RPA - When To Use Both | UiPath](https://www.uipath.com/blog/automation/ai-rpa-differences-when-to-use-them-together)   
[6 Smart RPA Adoption Challenges And Solutions | Market Insights™ - Everest Group (everestgrp.com)](https://www.everestgrp.com/2019-04-6-smart-rpa-adoption-challenges-and-solutions-market-insights-49802.html/)

## Q. What is the difference between AI and RPA?

A. [AI & RPA: What’s the Difference | NICE](https://www.nice.com/rpa/rpa-guide/rpa-ai-and-rpa-whats-the-difference-and-which-is-best-for-your-organization/)   
[The Real Difference between RPA and AI (chetu.com)](https://www.chetu.com/blogs/artificial-intelligence/is-rpa-part-of-ai.php)   
[RPA Vs. AI Vs. Machine Learning: A Comparison (crowdreason.com)](https://www.crowdreason.com/blog/rpa-vs-ai)

## Q. Explain the landscape of how RPA and AI can be used in tandem to increase revenue at an enterprise-level? Discuss a few use cases.

A. [AI & RPA - New Level of Automation | UiPath](https://www.uipath.com/automation/ai-and-rpa)   
[RPA & AI: Different Approaches To Problem Solving | Market Insights™ - Everest Group (everestgrp.com)](https://www.everestgrp.com/2018-01-rpa-ai-different-approaches-problem-solving-market-insights-43514.html/)   
[AI RPA - PS3G: Where Business and Technology meet](https://www.ps3g.com/solutions/ai-rpa/)

## Q. What are the phases of RPA at an Enterprise Level?

[(27) The convergence of RPA and AI can boost Business Process Automation | LinkedIn](https://www.linkedin.com/pulse/convergence-rpa-ai-can-boost-business-process-antonio-grasso/)

# 10 GANs: Generative Adversarial Networks

## Q. What are GANs?

A. [Understanding Generative Adversarial Networks (GANs) | by Joseph Rocca | Towards Data Science](https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29)   
[A Gentle Introduction to Generative Adversarial Networks (GANs) (machinelearningmastery.com)](https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/)   
[Background: What is a Generative Model? (google.com)](https://developers.google.com/machine-learning/gan)

## Q. Provide an overall system view of GANs.

A. [Overview of GAN Structure  |  Generative Adversarial Networks (google.com)](https://developers.google.com/machine-learning/gan/gan_structure)

## Q. Explain some uses-cases on the variations of GANs.

A. [GAN Variations  |  Generative Adversarial Networks  |  Google Developers](https://developers.google.com/machine-learning/gan/applications)

## Q. Name some popular tools used for GANs.

A. [Top 10 Tools For Generative Adversarial Networks (GANs) (analyticsindiamag.com)](https://analyticsindiamag.com/top-10-tools-for-generative-adversarial-networks/)

# **Part-B: Scenario-Based Questions**

## 1Q. How can NLP be leveraged to tackle Fake News and Cyber Bullying?

A.

## Q. How would you build a governance model for AI in Healthcare. Give examples.

A. [AI Governance on Scalable AI - Google Search](https://www.google.com/imgres?imgurl=https%3A%2F%2Fwww.researchgate.net%2Fprofile%2FSandeep-Reddy-5%2Fpublication%2F337012288%2Ffigure%2Ffig1%2FAS%3A821590203850759%401572893777628%2FOutline-of-the-Governance-Model-for-Artificial-Intelligence-AI-in-Health-Care.png&imgrefurl=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FOutline-of-the-Governance-Model-for-Artificial-Intelligence-AI-in-Health-Care_fig1_337012288&tbnid=frLGozmuIC-gkM&vet=10CAMQxiAoAGoXChMI4Miyq9mQ8gIVAAAAAB0AAAAAEAs..i&docid=VCDsSUs5zTg_aM&w=850&h=350&itg=1&q=AI%20Governance%20on%20Scalable%20AI&ved=0CAMQxiAoAGoXChMI4Miyq9mQ8gIVAAAAAB0AAAAAEAs)

## Q. Gartner’s report says that 87% of ML Models do not make it to production. What are some of the gaps in the real world when it comes to productionalizing ML Models?

A.

## Q. A lot of Data Science is being done in Jupyter Notebook. What are some of the ways to productionalize a Jupyter Notebook?

A.

## 5Q. Name some common test frameworks that can be used in AI / ML models?

A.

## Q. Data Science is said to be primarily a combination of Math, Business and Technology. With automation of various systems in various fields, what does the future of Data Science look like?

A.

## Q. If you were to pick out top 10 best and worst performing products from across multiple warehouses across India, what would the approach be?

A.

## Q. Data Science is experimental in nature. Does logging play a part in Data Science? If so, please explain how.

A.

## Q. What is the difference between a client-serice and an in-house model in Data Science? What are some of the benefits and drawbacks of both?

A.

## 10Q. How you would design scalable architecture to accommodate varying data needs of an organization?

A.

## Q. What is streaming data? How are architectures designed to handle real-time data?

A.

## Q. Apple recently came up with a host of measures to advocate for privacy. How can machine learning scale at an organizational level keeping in mind privacy?

A.

## Q. What is FaaS? What is the imapact of FaaS in the Data Science industry?

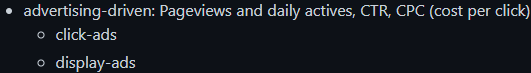
A.

## Q. Exploratory Data Analysis (EDA) is an activity that has traditionally taken a lot of time. Is there any way to streamline and speed up EDA?

A.

## 15Q. Let’s take a general use-case of an an internet product that is run using ads. Name some of the key metrics that can be used to measure performance?

A.



## Q. Let’s say that we are running ads on Facebook and Instagram to advertise our product. What are some of the ways to assess if engagement has been per expectations for the last month / quarter?

A.

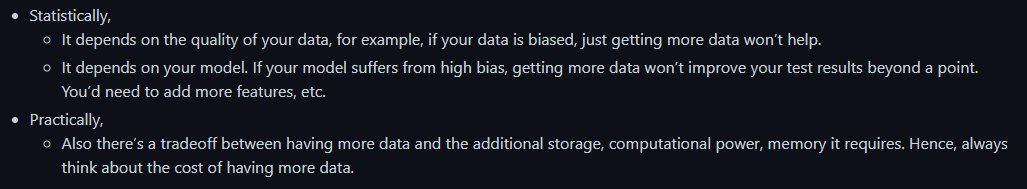


## Q. Let’s say that we are Netflix and you are given a budget. How would you determine which Original Series we should put our money on?

A. 

## Q. Let’s say are serving a Fortune 500 company and we have the option of unlimited data. Is more data always better?

A.



## Q. Let’s say you have a dataset with a million rows and 200+ columns. How can we determine what the most important features in the model are?

A.

## 20Q. Why do a lot of Kaggle Competition winners use Ensemble Learning? What is the main idea behind it?

A.



## Q. Let’s say that we are Uber. How can you use just the GPS data from the car to determine the quality fo the driver?

A.

## Q. Developers and product managers have hundreds of ideas on a daily basis. For the ideas that are selected to be tested, what is an efficient way to test these hypothesis in the real world?

A. [7 A/B Testing Questions and Answers in Data Science Interviews | by Emma Ding | Towards Data Science](https://towardsdatascience.com/7-a-b-testing-questions-and-answers-in-data-science-interviews-eee6428a8b63)

## Q. A lot of Data Science projects start our as PoCs. List out what are the things to keep in mind while building a PoC.

A.

## Q. What is a MVP? In a field as experimental as Data Science where experimentation is a part of the daily job description, what are some of the ways to maintain client confidence?

A.

## 25Q. How to build out a Data Science product from scratch?

A.

## Q. Let’s say you join a new team and are given data with no context. What would be your approach to prepare for analysis?

A.

## Q. You are given a classification problem on a fairly complix dataset. From your previous experience, you guesstimate that it will take ~3 months to processs the entire data with business context. How can you maintain client confidence intermittently?

A.

## Q. A popular algorithm used for classification is Naïve Bayes. Why is it considered “Naïve”?

A.

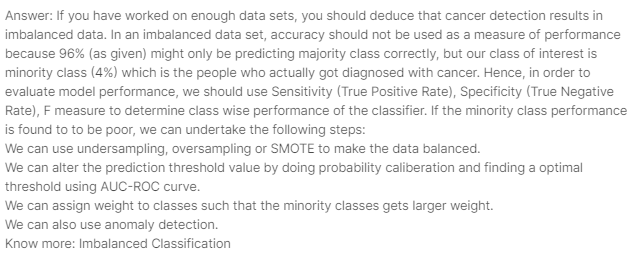
## Q. Let’s say you are given a large dataset of all customers of a telecom company with over 2 million rows and 2000 columns. You have to build a classification model to decide if a customer is likely to churn or not next month. You are given a laptop with limited RAM and no access to cloud resources. What are some of the ways you can go about this?

A.

## 30Q. You are very lucky and have got the opportunity to work for an interesting project that is deep in Data Science. The project deals with 5 VP-level stakeholders, all located across various timezones. You realize that every single person has different ambitions and expectations from this high-level project. What are some of the ways you could tackle such a situation?

A.

## Q. Let’s say you are dealing with customer churn data and you made a classification model with an accuracy of 97%! Why is this a bad measure? Explain your thought process.

A. 

## Q. Projects in the past that used waterfall models used to give a lot of importance to documentation. However, projects that use Agile methodology do not give much importance to documentation. Do you agree with this assupmtion from the perspective of Data Science?

A.

## Q. Let’s say you are working on a large dataset and you have figured out that the data is biased. Now what?

A.

### Q. Let’s say that you have to upgrde the hardware for your team to be able to train complex ai / ml model. What are some of the ways you can go about this exercise in an economical manner?

A.

## 35Q. Let’s say that everyday more and more Burger King restaurants are opening across the globe. We also notice that everyday, more and more babies are being born. Does this mean that the increase in the number of Burger King restaurants is leading to more number of babies being born?

A.

## Q.

[Scenario based interview questions for Data science Part - 2 | Data Science and Machine Learning | Kaggle](https://www.kaggle.com/general/232068)