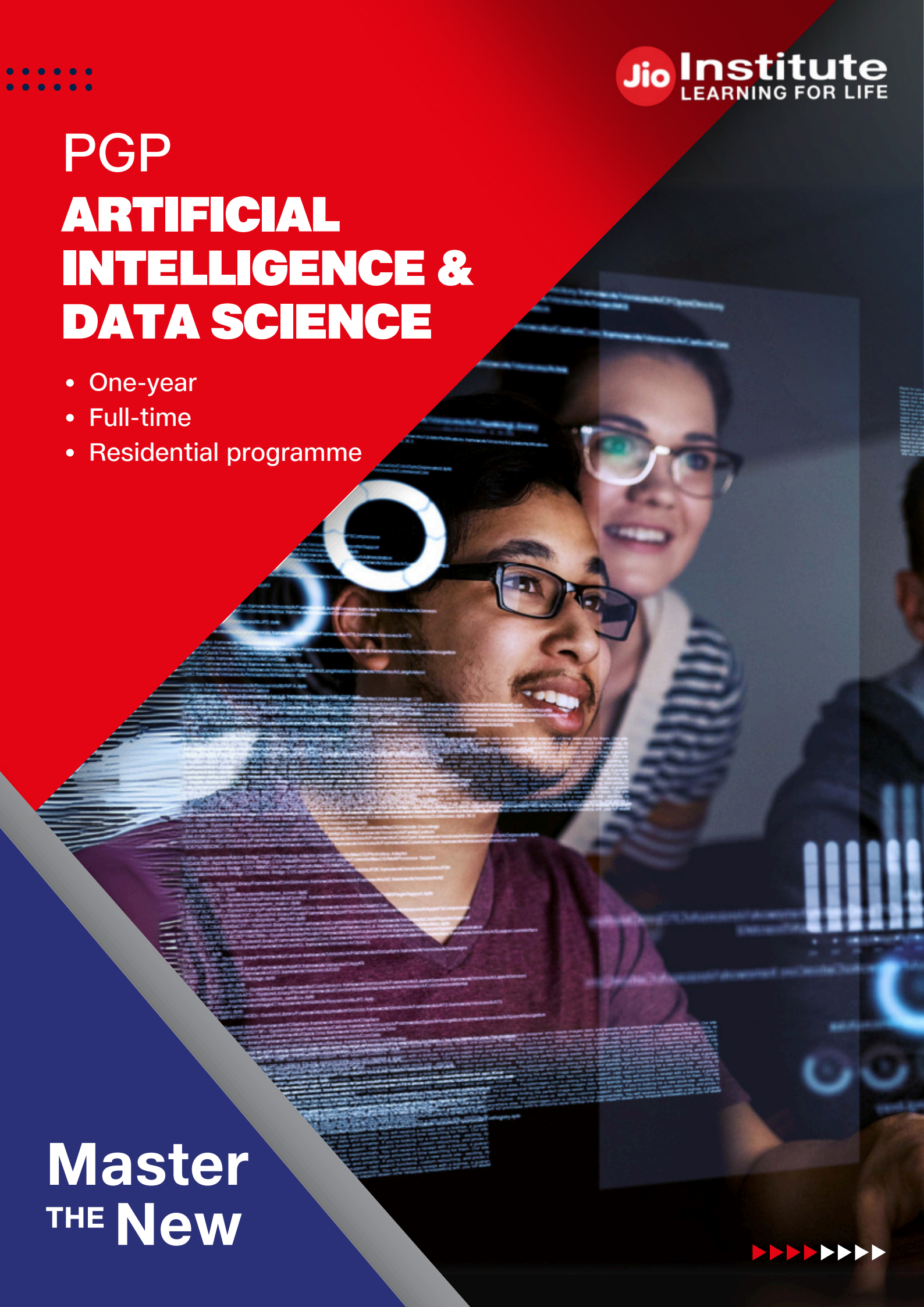
Jio Institute
LEARNING FOR LIFE

PGP
ARTIFICIAL INTELLIGENCE & DATA SCIENCE

- One-year
- Full-time
- Residential programme



Master THE New

Programme Details

In this programme, students will embark on a journey that explores the fascinating realm of cutting-edge AI technologies that are reshaping industries, revolutionizing decision-making, and transforming the way we interact with data/information. Our comprehensive and rigorous curriculum covers courses from foundation to advanced levels. Key concepts and technologies include Machine Learning, Deep Learning, Fundamentals of AI, Natural Language Processing (NLP), Optimization, Bigdata Engineering, Data Visualisation, Responsible AI, Reinforcement Learning and Time Series Analysis. The various courses are taught by renowned research-oriented faculty and industry practitioners from India and from around the world. The programme aims to instil strong theoretical foundations as well as the know-how to create practical solutions for enterprises and society. Students learn to convert business problems and workflows into AI&DS products and solutions across multiple verticals/industries. Enriched by exposure to real-life AI&DS applications through capstone projects and lectures from industry veterans, students are exposed to hands-on exercises, practical projects and quizzes to reinforce their learning.

Highlights

- Futuristic, industry-aligned curriculum
- Application-based capstone projects
- Internships & study abroad module
- Renowned global & industry faculty
- Scholarships based on need & merit

Programme Curriculum

Foundation Courses :

- Probability & Statistics
- Data Structures & Algorithms
- Introduction to Programming
- Introduction to AI
- Databases & Data Warehouses
- Liner Algebra & Optimization
- Data Visualization
- Time Series Analysis

Core and Foundation Courses :

- Machine Learning I & II
- Big Data Engineering
- Natural Language Processing
- Computer Vision
- Deep Learning
- Speech Processing
- Responsible AI
- ML Operations (ML Ops)

Advance Elective Courses :

- Advanced topics in Machine Learning
- Advanced topics in Big Data Engineering
- Deep Learning for Natural Language Processing
- Deep Learning for Computer Vision
- Large Language Models (LLM)
- Quantum Computing
- AI in Healthcare

Tools & Platforms :

- Covers 13+ Programming Tools, Languages & Libraries

AI for X :

A lecture series organized throughout the year on case studies from the Industry, Society and Business Functions, where students learn how AI and data science is applied across various verticals.

AI for Industry Verticals

- AI for Telecom
- AI for Retail
- AI for Manufacturing
- AI for Refineries
- AI for Banking
- AI for Finance
- AI for Insurance

AI for Society Verticals

- AI for Agriculture
- AI for Healthcare
- AI for Education
- AI for Smart Cities
- AI for Sustainability

AI for Business Functions

- AI for Customer Lifetime Management
- AI for Human Resources
- AI for Accounting and Finance
- AI for Infrastructure and Asset Management
- AI for Supply Chain and Operations
- AI for Risk and Compliance

Dr. Vidya Setlur
Director, Tableau Research, USA

Detailed Curriculum

Mandatory Core and Foundation Courses

Probability and Statistics for AI: Artificial Intelligence (AI) and Data Science (DS) require strong mathematical foundations to understand clearly, articulate tersely, and innovate rigorously the plethora of algorithms and frameworks that have evolved in the vast field of AI/DS over the last century. This course revisits and rebuilds some of these mathematical foundations for a rigorous study, intuitive understanding, formal communication of algorithms, and express the future in AI/DS. The course introduces the basic concepts of probability and statistics with applications to AI and Computer Science. Topics include foundations of probability theory, discrete and continuous random variables, sampling distributions, the law of large numbers, central limit theorem, point estimation, confidence intervals, hypothesis testing, and regression analysis.

Introduction to Programming: Python is a versatile and popular programming language, known for its clear syntax, readability, and code simplicity. In this course, we will introduce programming using Python. Participants will learn Python's basic constructs, understand control flow, data structures, functions, file operations, exception handling, and object-oriented programming. Participants will also explore Python's powerful library ecosystem used in diverse fields such as data analysis, web development, machine learning, and more. This course is designed to establish a strong programming foundation, enabling participants to solve real-world problems effectively and confidently.

Data Structures and Algorithms: A good algorithm usually comes together with a set of good data structures that allow the algorithm to manipulate the data efficiently. In this course, we consider the common data structures that are used in various computational problems. You will learn how these data structures are implemented in programming languages. This will help you understand what is going on inside a particular built-in data structure implementation and what to expect from it. You will also learn typical use cases for these data structures. In this course, we will explore several fundamental algorithms and data structures in computer science. Some of the data structures, we will encounter include linked lists, stacks, queues, trees, heaps, hash tables, and graphs. We will study and analyze algorithms for searching, traversing trees, hashing, manipulating priority queues, sorting, finding shortest paths in graphs, and much more.

The basic idea of this course is to help you understand many of the fundamental data structures of computer science. With an appreciation for data structures and algorithms and practical experience in implementing them you can be a much more effective designer and developer for new applications. Elegant algorithms are also a nice counterpoint to the crafty code and weird features we encounter in daily work.

Introduction to AI: Artificial Intelligence (AI) can be thought of as the philosophy, art, science, algorithms, engineering, and process of understanding and replicating human intelligence into machines. These include logical, heuristic, and probabilistic problem-solving, strategy, structure, and causality learning, pattern recognition across a variety of data, visual perception, language analysis, synthesis, and translation, speech analysis & synthesis, etc.

This core course explores AI's traditional, modern, and futuristic frameworks and addresses the subtle and fundamental transition from computer science to artificial intelligence thinking. This course provides a general overview of major problems, approaches, and applications of Artificial Intelligence (AI) at an advanced undergraduate / beginning graduate level.

Databases & Data Warehouses: The ability to deal with data plays a critical role in all disciplines of information technology. This course covers all the essential topics in database management building the foundation needed for dealing with persistent data. It introduces the fundamental concepts of designing, using, and implementing database systems and applications.

Database system architectures, Relational data model, Query Processing, Database design, Data Storage and Indexing, Query Optimization, Transaction Management, Data Warehousing, Data Analytics, Advanced Topics: Data Mining, Blockchains, etc.

Optimization: The applicability and use of optimization and decision models have increased significantly in recent years. Information has become an important organizational resource, and optimization models play an important role in the optimal utilization of this resource along with others, in the form of organizing and structuring information so that it can be used more efficiently. This course will introduce and apply optimization techniques to managerial decision problems with the objective of enhancing decision-making capabilities as well as knowledge about using spreadsheet-based tools.

The course emphasizes model-formulation skills and provides a basic understanding of some of the assumptions and limitations of decision models. We will not focus on algorithmic details but will use established software packages to obtain the solutions. The major emphasis is on interpreting the solution and strategizing based on the interpretation. The aim of the course is to help students become intelligent consumers of analytical modeling tools and adopt an analytical, data-driven approach to making business decisions.

Machine Learning: Most modern Enterprises across variety of verticals that have become “data rich” due to the first wave of digital transformation (Automation) are now investing heavily to become “AI First” in their second wave of digital transformation. They are aspiring to migrate from broadcast, batch, reactive, tactical decision-making to personalized, real-time, proactive, strategic decision-making. This has the potential to improve the quality of service they provide to their customers, the efficiency and reliability of their operations, optimal utilization of their resources – personnel, money, and materials, and adapting their ability to make dynamic decisions with evolving contexts. This in turn leads to systemic growth in customer satisfaction, net profitability, and market share.

This unprecedented paradigm shifts in “how Enterprises make decisions” is enabled by a proven suite of algorithms, techniques, frameworks, processes, and platforms that have emerged in the varied fields of Artificial Intelligence, Machine Learning, Data Mining, Statistical Pattern Recognition, etc. over the last several decades.

In this course, we will study a plethora of frameworks, algorithms, techniques, tools, and guiding principles to (a) understand any type of data better, (b) combine it with the domain knowledge appropriately, (d) find deeper insights hidden in the data, (d) build a wide variety of powerful descriptive, predictive, and prescriptive models, and (e) continuously improve them with feedback data coming from the environment they are deployed in.

Time Series Analysis: Data obtained from observations collected sequentially over time are extremely common in a business requiring analysis of time series data. Some examples are weekly interest rates, daily closing stock prices, monthly price indices, yearly sales figures, etc. In addition, time series analysis is applicable in diverse fields. Much familiar time series occur in the field of economics, where we are continually exposed to daily stock market quotations or monthly unemployment figures, annual GDP figures, etc. Social scientists follow population series, such as birthrates or school enrollments. An epidemiologist might be interested in the number of Covid - 19 cases observed over some period.

The first step in any time series investigation involves scrutiny of the recorded data plotted over time. This scrutiny often suggests the method of analysis as well as statistics that will be of use in summarizing the information in the data. The time domain approach which assumes that correlation between adjacent points in time is best explained in terms of a dependence of the current value on past values. This focuses on modeling some future value of a time series as a parametric function of the current and past values, usually by decomposing components of time series data. Another approach develops a systematic class of models called autoregressive integrated moving average (ARIMA) models to handle time-correlated modeling and forecasting, with a provision for treating more than one input series through multivariate ARIMA. A completely different approach to time series analysis is the survival analysis. The objective of this course is to expose students to various methods of analyzing time series data to understand its behavior as well as make forecasts for the future.

Big-Data Engineering: Traditional database systems and data analysis algorithms have become increasingly insufficient to deal with this unprecedented growth in the volume, variety, velocity, and veracity of data. Big Data tools and techniques promise to store and crunch this data at scale, find deep insights in it, build sophisticated prescriptive, predictive, and descriptive models from it, and enable accurate and real-time decisions. Adapting business models to digital through data and analytics is critical for the long-term success of mainstream enterprises.

The objective of this course is to primarily focus on modern database systems to solve real-world business problems for both batch and streaming data. This course aims to take a 360-degree view of the Big Data ecosystem through various prisms i.e., start-ups, VC firms, publicly traded, and privately held corporations. The course is going to be a good mix of theory, hands-on experience, and class discussions on both case studies as well as business use cases.

Data Visualization: Data contains a wealth of insights about the process that generated it. Finding deep insights from data by asking it the right questions, interpreting the answer to those questions correctly, and letting the data speak to us is an essential skill for a data scientist.

There is data all around us. As humans, to make sense of the world and the information that it contains, we rely on our visual systems and cognitive abilities. Well-designed information visualizations leverage these human capabilities to help see and understand data.

In this course, we will study techniques for understanding and creating effective visualizations based on principles from perception, design, and cognitive science. The course is targeted toward students interested in understanding effective visualizations, building better visualization tools and systems, as well as understanding the state of the art in research in this space.

Natural Language Processing: Language can be considered our “sixth sense” that has literally made humans the dominant species on the planet and made modern human civilization possible. It is the primary mechanism used throughout human history to communicate and share knowledge both across society and across generations. One of the holy grails of AI is to build Machines that can “understand” (analyze), “generate” (synthesize), and “translate” language the way humans do so that we can communicate with machines in our natural languages the way we communicate with each other.

This course introduces the major challenges of, and classic approaches to, Natural Language Processing (NLP) at a beginning graduate level. The course takes a functional approach — developing data representations and algorithms based on analysis of what is required by key uses of language — as well as considering the critical properties of language with respect to these uses. The primary focus is on developing applications to enable computer use of language as a medium of communication with human beings.

Computer Vision: This introductory course will focus on traditional methods for computer vision. We will cover the fundamentals of image analysis, its applications and image filtering. Following this, we will focus on feature detection, edge detection, image classification, scene understanding, object detection, and optical flow, where we will cover both classical as well as deep learning approaches. A tentative list of topics to be covered in this course,

- Mathematical Preliminaries
- Coordinate Transforms
- Image Filtering, Edge Detection, Feature Extraction
- Region/Boundary Segmentation
- Image Classification
- Object Detection
- Basics of Neural Networks for Pattern Recognition

Speech Processing: The course introduces speech processing and its applications. The main objectives can be summarized as follows: To understand the basic principles of sound and speech production and perception, convey details of a range of commonly used speech feature extraction techniques, provide a basic understanding of multidimensional techniques for speech representation and classification methods, familiarise you with the practical aspects of speech processing, including robustness, and applications of speech processing, including biometrics, speech pathology based disease detection, speech synthesis, and speech recognition, and give you practical experience of developing speech processing systems.

Deep Learning: Unstructured and semantically low-level data such as images, videos, audio, and text are becoming increasingly common in enterprises and society as people and systems generate and consume such data in both public and private settings. Human systems of visual, auditory, and linguistic perception and generation have mastered the art of processing such data in the brain. Modern Deep learning systems try to mimic a similar multi-layered architecture to interpret this low-level data (e.g., pixels, signals, words) into high-level semantics.

This course explores the biological, computational, theoretical, architectural, and modelling aspects of deep learning models that come in a variety of architectures, on different data types and are used for a variety of applications.

Human Computer Interaction: Human-computer Interaction (HCI) studies the relationship between computer technology and people. HCI is a multidisciplinary approach to observing how to best design computers and other forms of information technology for our everyday user needs and enjoyment. It borrows methodologies and techniques from Computer Science, Human-factors Engineering, and Cognitive Science to make our interaction with technology as beneficial as possible. HCI is important to learn because it’s designed to close the gap between user and device to maximize efficiency, with some even using it to improve the happiness and satisfaction of users. This is applied through user-interface design, accessibility (e.g. screen readers), keyboards, and other tools that help us use technology more easily and effectively. This course will help students to learn to implement HCI concepts into a variety of applications specifically web development.

Machine Learning Operations (ML Ops): Machine Learning Operations (MLOps) refers to the tools, techniques and practical experiences required to train your machine learning models and deploy and monitor them in production. After we have trained our machine learning model, the next big task is to deploy the model to production and scale it so that more users can use it. In this course, you will learn how to use various tools and methodologies to do all this effectively. While knowing machine learning and deep learning concepts is essential, but for building a successful career in Artificial Intelligence, you need to have good experience with production engineering capabilities. This course deep-dives into machine learning and deep learning algorithms along with building expertise in DevOps technologies.

Responsible AI: "With great power comes great responsibility". AI is emerging as a tool for fundamental and disruptive transformation of human society. If not used ethically and responsibly, it has the potential of doing more harm than good to individuals, societies, and the human civilization as a whole. Strong regulations are emerging across the globe to ensure that AI is used for the right purpose, in the right way, and with the right governance.
In this core course we will understand the emerging guiding principles of Responsible AI, the state-of-the-art algorithms used to detect and correct any intentional or unintentional transgressions by AI systems w.r.t. these guiding principles, and the emerging landscape of regulations for AI across the world. This will enable emerging data scientists to "consciously and proactively" incorporate these principles and regulations into their model building from the start and not as an afterthought.

Recommendation Engine: Human-computer Interaction (HCI) studies the relationship between computer technology and people. HCI is a multidisciplinary approach to observing how to best design computers and other forms of information technology for our everyday user needs and enjoyment. It borrows methodologies and techniques from Computer Science, Human-factors Engineering, and Cognitive Science to make our interaction with technology as beneficial as possible. HCI is important to learn because it's designed to close the gap between user and device to maximize efficiency, with some even using it to improve the happiness and satisfaction of users. This is applied through user-interface design, accessibility (e.g. screen readers), keyboards, and other tools that help us use technology more easily and effectively. This course will help students to learn to implement HCI concepts into a variety of applications specifically web development.

Advanced Elective Courses

Advanced Topics in Bigdata Engineering: With this unprecedented growth in the volume, variety, velocity, and veracity of data, it has become increasing challenging to deal with big data using the traditional data management systems. Deriving insights from big data in near real time has become is critical for the long-term success of mainstream enterprises. So, the objective of this course is to primarily focus on modern data architecture and data lake design to facilitate downstream analytics applications including descriptive and predictive analysis. Many new ideas that have evolved in recent times like data lake-house, data discovery, feature stores, experiment tracking will also be discussed in detail to empower the participants to design and develop advanced data management systems in the real world.

Advanced Topics in Machine Learning: With this unprecedented growth in the volume, variety, velocity, and veracity of data, it has become increasing challenging to deal with big data using the traditional data management systems. Deriving insights from big data in near real time has become is critical for the long-term success of mainstream enterprises. So, the objective of this course is to primarily focus on modern data architecture and data lake design to facilitate downstream analytics applications including descriptive and predictive analysis. Many new ideas that have evolved in recent times like data lake-house, data discovery, feature stores, experiment tracking will also be discussed in detail to empower the participants to design and develop advanced data management systems in the real world.

Advanced Topics in Statistics: This advanced course on statistics explores some deeper statistical constructs used for more sophisticated number crunching, find deeper patterns in data, and analyzing it to find complex structures.

Deep Learning for NLP: With this unprecedented growth in the volume, variety, velocity, and veracity of data, it has become increasingly challenging to deal with big data using traditional data management systems. Deriving insights from big data in near real-time has become is critical for the long-term success of mainstream enterprises. So, the objective of this course is to primarily focus on modern data architecture and data lake design to facilitate downstream analytics applications including descriptive and predictive analysis. Many new ideas that have evolved in recent times like data lake-house, data discovery, feature stores, experiment tracking will also be discussed in detail to empower the participants to design and develop advanced data management systems in the real world.

Deep Learning for Speech Processing: This advanced course explores the evolution of speech processing frameworks from “traditional” to “deep learning” architectures specifically designed for speech data and to solve core problems such as speech recognition, speech synthesis, speaker identification, etc.

Deep Learning for Computer Vision: This advanced elective course explores the evolution of Natural Language Processing from "traditional" approaches to "deep learning" architectures specifically designed for text data and to solve specific NLP tasks with this data. In this course we will cover fundamentals of deep learning for Natural Language Processing — from text classification to recent Transformer language models.
Pre-requisites:

- Intro to Natural Language Processing
- Intro to Machine Learning
- Experience with Python programming language is recommended.

Graph Theory & Network Thinking: With this unprecedented growth in the volume, variety, velocity, and veracity of data, it has become increasingly challenging to deal with big data using traditional data management systems. Deriving insights from big data in near real-time has become critical for the long-term success of mainstream enterprises. So, the objective of this course is to primarily focus on modern data architecture and data lake design to facilitate downstream analytics applications including descriptive and predictive analysis. Many new ideas that have evolved in recent times like data lake-house, data discovery, feature stories, and experiment tracking will also be discussed in detail to empower the participants to design and develop advanced data management systems in the real world.

AI for "X"

In this "industry talk series" and "case studies for AI applications" course, students will learn how AI/DS is applied to various industry verticals. The wide variety of use-cases across different verticals will highlight the common principles of applied AI/DS in real-world inclsing: (a) understanding the right pain points and identifying the right opportunities for AI/DS in each vertical, (b) formulating the business problem into an AI architecture, (c) using both domain knowledge and data to build AI/ML/DL models, and (d) integrating these models into the existing business processes. This course will have industry lectures throughout the year on:

AI for Society Verticals:

- AI for Agriculture
- AI for Healthcare
- AI for Education
- AI for Smart Cities
- AI for Sustainability

AI for Industry Verticals:

- AI for Telecom
- AI for Retail
- AI for Manufacturing
- AI for Refineries
- AI for Banking
- AI for Finance
- AI for Insurance

AI for Business Functions:

- AI for Customer Lifetime Management
- AI for Human Resources
- AI for Accounting and Finance
- AI for Infrastructure and Asset Management
- AI for Supply Chain and Operations
- AI for Risk and Compliance

Holistic Learning Module

Design Thinking: "Necessity is the mother of invention". Every successful product that we use today is addressing a very critical human need, addressing an important individual, industrial, or societal pain point, and has evolved over multiple iterations through a very well defined process that has now matured into what we call "Design Thinking". Traditionally, the design thinking process is taught mostly to product managers, customer experience designers, and growth managers. But as the charter of AI/DS is evolving from merely "embedding intelligence into existing products and services" to "building intelligent solutions for society", it becomes critical for AI architects and Data Scientists to also become first principles "product thinkers" and "design thinkers". This hands-on Holistic Learning Module on Design Thinking will explore the core guiding principles and a systematic process of both conceiving new, and continuously improving existing products by identifying the right pain points and addressing them with the right product features.

System and Solution Thinking: Over the last several centuries, products have evolved from mostly physical products such as cars and TV's to purely digital products such as search engines and social networks. The next generation products will be complex digital-physical ecosystem products such as telecom networks, e-commerce platforms, agriculture and healthcare for a country, smart cities of the future, and so on. Today these third generation products are put together by clobbering together existing building blocks giving a very siloed experience to customers, a broken architecture to engineers, and a fragmented and inefficient design of operations.

A whole new set of skills are needed to conceive, design, build, and operate such ecosystem products that will require a new kind of integrated operating system that doesn't exist yet.

The Holistic Learning Module on System and Solution Thinking will develop the required skills to not just build and deploy bottom-up AI/ML models wrapped in microservices but to build a top-down AI + IT architecture for such complex ecosystems.

Communication Skills for Data Scientists: Data Scientists and AI experts not only have to create intelligent solutions for businesses but they also have to become well versed in the art of domain, business, technical, and engineering communication. As they have to work with multiple stakeholders in the ecosystem including the business leadership that articulates the problem statements in business terms, subject matter experts that have built up enormous domain knowledge within the vertical over the years, data and engineering teams that will help them build and deploy the AI solutions at scale, and the operations team who will execute the decisions generated by the AI systems.
This Holistic Learning Module will explore the art of listening and conversing about pain points that AI/DS can address, domain knowledge required to build the right models and solutions, presenting the insights generated from the data, and describing the nuances of the solution to business, platform, and operations stakeholders.

Leadership & Entrepreneurship for AI&DS: There are three different types of leadership and entrepreneurial opportunities for data scientists and AI experts in the foreseeable future. First, leading the transformation journeys of existing businesses in different verticals such as banking, finance, insurance, manufacturing, retail, telecom, mobility, entertainment, etc. from "data-rich" organizations to "AI-first" organizations. Second, building startups that champion an AI-first solution to industrial or societal use-cases addressing the efficiency, precision, personalization, automation, and democratization of AI in these areas, and third, to create completely new business or technology innovations that could disrupt the way we live, work, educate our children, take care of our health, grow our crops, or run our cities and governments.

This holistic learning module will explore the opportunities, challenges, and a journey towards a "Data Science Leader and Entrepreneur".