**Course description:**

So, you have all ML algorithms under your belt and your coding skills are awesome, but… The data. The blood and bone of every ML/AI model. Alpha and omega of any modeling. Do you have it right? Is it the feet of clay that will collapse with the first blow or is it rock solid diamond that will shine through in your AI implementation?

Are you really-really sure? Well, we are here to put this to test in this course on Responsible Data acquisition in ML.

**Why should I take this course**: because if your data is not right, your model is wrong straight away. Learn about responsible data acquisition: what it involves and how to achieve this throughout all steps of the ML workflow.

**Who is this for**: this is for every ML/AI enthusiast from a novice citizen data-oriented person to a savvy AI engineer. It is for a ML/AI developer and a ML/AI consumer, because it is not only about developing a solid ML/AI model, but also understanding data use in that model to strengthen the “trust” in the model, understand the “risk” and appreciate the “benefit” of ML/AI at all levels!

**What will I learn**: we will cover the foundations of responsible data acquisition, including main regulations and policies, third party data vendors,

**Get your data but do it responsibly!**

It all starts with data – it’s a cliché but. It still true! Long gone those days when data was just “there”. It is still “there” but the joy of using it comes with duties and responsibilities. Ethics and data laws are not only buzz words, but it is also the reality. We must make ML fair and responsible data acquisition is the heart of it. We will start talking about Fair ML and why ethical data acquisition is important. We will cover principles of responsible data acquisition, ethics, laws and duties and responsibilities we have acquiring our data. From understanding the sources of data compliance to integrating ethical data practices throughout the ML workflow, we will get you well-versed in key regulations like GDPR and CCPA. You'll learn to strategize data compliance and navigate the traps of third-party data licensing, ensuring legal and ethical integrity in your ML projects. While navigating the maze of data use laws is tricky - remember understanding of principles is always better than referencing rules! And ethics has been around since Ancient Greece. So, In the fast-changing world of data regulations – principles are more likely to remain the same! So, get your ethics right from the start, but consider modern day realities of dialing with noisy and biased data.

**Plans are useless but planning is indispensable (Eisenhower)**

In this section we will look into the “how” component of responsible data acquisition – designing a roadmap to ethical data collection. Remember – practice beats theory every single time, so we will see how to transform theory to practice and get our hands dirty with developing data acquisition strategy, from setting objectives to navigating ethical and regulatory requirements. We will showcase possible difficulties that may emerge and remedies to be applied to ensuring you're equipped to handle all the bumps on the road of data collection in the ML/AI landscape and enjoy the ride!

**The start is only the beginning!**

Getting your data right is only the beginning of responsible practices of data use to ensure Fair ML practices. Just because you “have” data, does not mean its “usable” straight away (or in the worst-case scenario – usable at all!). Your data responsibilities continue with safeguarding ethical data use by examining the data to validate it and use ethically and responsibly through data cleaning, model development and evaluation and implementation.We will look into validation of data, especially data that comes from different sources, issues that may arise with data aggregation, working with different data types and strategies to use navigating the quest for diversity, representation, transparency and fairness in data. We will talk about validating data to achieve fairness, non-discrimination, accuracy, and quality.

Navigating data responsibly throughout the ML journey will bring us to critical concerns in data – BIAS.

**It most definitely has a bias…**

ML works with real data. Dirty data. Lopsided data. No data is perfect, and we have no choice but work with this.

Using biased data is the same as using a map to get to the venue in a new city just to realize that you were using the wrong map! You can surely get to Paris, but this Paris can be in Texas, US or in France. Which one do you actually need?

The biggest concern in our responsible data ML is watch for bias at all stages of ML workflow. To irradicate it from the start but be careful as not to introduce bias ourselves throughout feature selection, feature engineering, model development, evaluation, and deployment. Bias never sleeps, so don’t sleep on it either!

We will talk about different type of bias and ways to identify it. We will talk about data audits, bias identification, and mitigation strategies. We will equip you with best practices and techniques to ensure data validity. We will focus on fair ML/AI to refine your ML workflow and develop accurate and unbiased insights.

**Get your hands dirty (and ethically “clean”)!**

This is a conceptual course, but it is NOT just about learning concepts. As we will embark on our journey, we will take plenty of practical examples and exercises to see those concepts in action and take them out with you on your ML journey.

Fair AI starts with … you. So, empower yourself and make a difference in the world of AI and ML

Whether you are a data professional, a student, or an AI enthusiast, this course is your gateway to becoming an ethical leader in the data-driven future. Join us on this journey and unlock the full potential of responsible data acquisition and fair ML/AI.