The need for trust in AI has always been of utmost importance and one way of achieving it is through striving towards a common set of attributes that capture essential "facts" about a model. This course will give you an overview **on the concept of transparency which helps in building stronger trust and how AI FactSheets can help you achieve that transparency by providing a diverse collection of facts tailored to the particular AI model or service**. It will also walk you through how these facts can be leveraged by different roles in the AI lifecycle to increase overall effectiveness and the methodology of constructing an AI FactSheet that is appropriate for a particular use case, and target audience.

**By the end of this course, you will be able to:**

* Explain the value of transparency in fostering trust, its relevancy to the world of AI and how it is achieved in other industries/domains
* Describe how AI FactSheets can provide useful AI transparency for a diverse collection of transparency consumers
* Apply a methodology for creating useful AI FactSheets (real world use case)
* Recognize how FactSheets can be a foundational component in providing AI Governance

**This course is intended for:**

* Analytics Leaders
* Data Science Leaders
* Practicing Data Scientists
* Machine Learning Engineers
* AI specialists
* Anyone with an interest in AI Trust and Transparency having the prerequisite knowledge that is mentioned below

**To be successful in this course, you should have some knowledge of:**

* AI/Machine Learning Workflow
* Evaluation metrics
* *IBM AI Fairness 360 for AI Fairness course (helpful but not necessary)*
* *IBM AI Explainability 360 for Explainable AI and ML course (helpful but not necessary)*

**Estimated Learning Time: ~ 4hr**

* Total of 4 learning modules with 4 hoursof self-paced video lectures
* Check for Understanding (CFUs)

In this module, you will:

* Recognize the need of Trusted AI
* Describe and differentiate various factors that can build trust in AI
* Explain how is transparency achieved in other domains and its relevancy in the world of AI
* Recognize challenges of achieving transparency in the world of AI

#### AI FactSheets - Website

The following link provides an overview of the FactSheet project, a research effort to foster trust in AI by increasing transparency and enabling governance.

* IBM AI FactSheets - <https://aifs360.mybluemix.net/>

#### AI Explainability and Factsheets w/Michael Hind

The following YouTube link will open an episode of *Trusting AI: Unlocking the Blackbox.* Michael Hind, Distinguished Research Staff Member at IBM Research AI, discusses one of Trusted AI’s toolboxes, AI Explainability 360, as well as AI Factsheets 360 in depth.

* AI Explainability and Factsheets w/Michael Hind | Trusting AI: Unlocking the Black Box - <https://www.youtube.com/watch?v=byx-IkICEHQ>

Provided below is a summary of the entire module. For your convenience, below you will see a heading that matches the content topics that were covered in this module, followed by a bulleted list of key concepts covered.

**Lesson 1: Need for Trusted AI**

1. Recent “AI incidents” have been significant to prove that intelligent systems are prone to unforeseen and failures when they were deployed to the real world. Therefore it is essential to build trust in AI.

2. IBM Researchers believe that Trust in AI can be fostered through these pillars:

* **Explainability** is the ability of the AI model to explain how and why it arrived at a particular decision
* **Fairness** is the ability of the AI model to be free of bias in its decisions and to avoid unfair treatment of certain groups
* **Robustness** is the ability of the AI model to be safe and secure and not be vulnerable to any tampering or compromising the data they are trained on.
* **Transparency** is the ability to disclose information to increase the understanding of how an AI model or service was created and deployed
* **Governance** is the ability to direct, manage and control the AI activities throughout the AI lifecycle.

**Lesson 2: AI Transparency**

1. AI Transparency enables AI consumers to better understand an AI model and it can help enable AI Governance.

2. Transparency is about disclosing important information about the product, known as "facts", without revealing secrets.

3. A standardized nutrition label is unlikely to achieve appropriate transparency in the world of AI because unlike food product consumers, AI consumers have different needs

4. “AI FactSheets” provides an approach for achieving the necessary AI transparency by helping you build something like a nutrition label for an AI system tailored to the specific use case and audience.