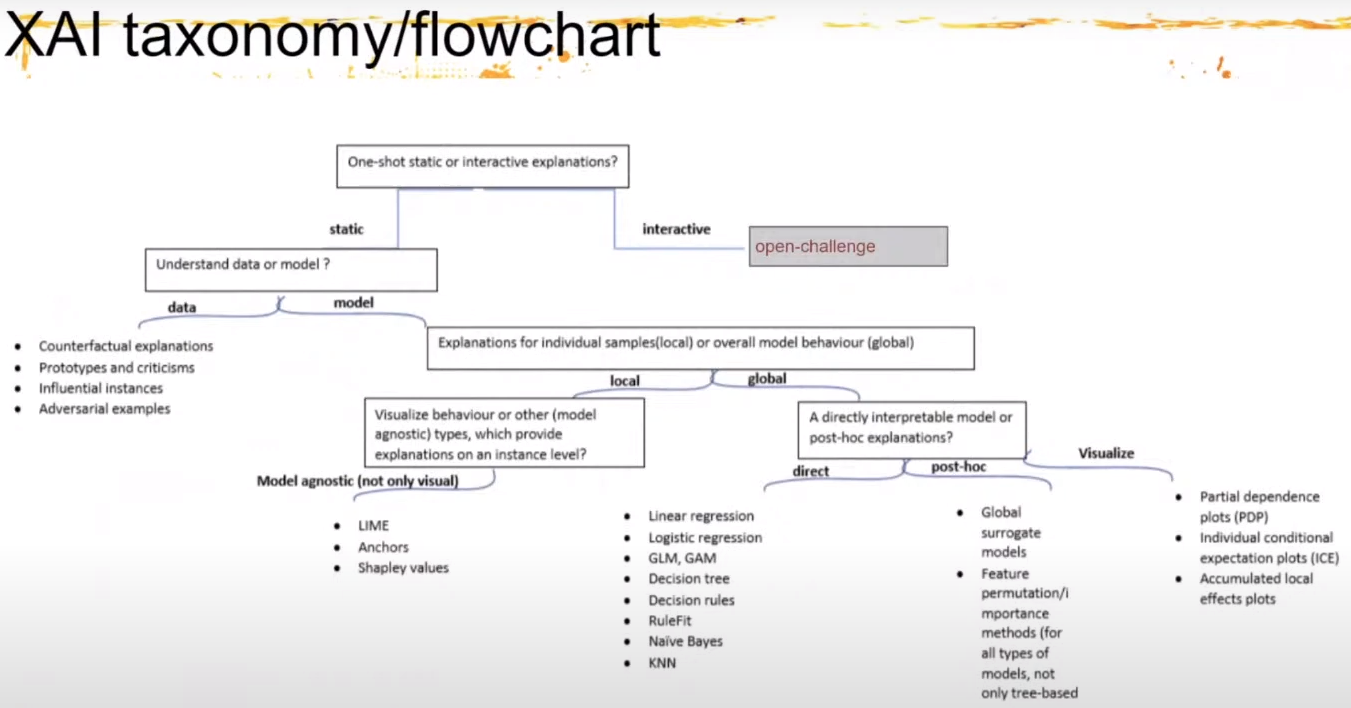
XAI Data Scientist User Journey

<https://www.youtube.com/watch?v=8FvCuNJO5a8>

* What is a good explanation?
  + Contrastive
  + Selective
  + Consistent with prior beliefs
  + Social
  + Focus on the abnormal
  + General and probable
* How do we know an explanation is useful?
  + Goodness of explanations
  + Are users satisfied with the explanations
  + How well the users understand the AI system
  + How curiosity motivates the search of explanations
  + User´s trust and reliance in the system
* It is difficult to validate the explanations



* Methods for XAI:
  + Local explanation
    - LIME
      * Surrogate models that explain individual predictions
    - Anchors
      * Finds a decision rule that ‘anchors’ the prediction sufficiently (changes in other variables do not change the prediction)
    - Shapley values
      * Marginal contribution of a feature on an instance level, compared to the average prediction
  + Visual explanations
    - Partial dependence and Individual Conditional expectation plots (PDP and ICE)
      * PDPs: that marginal effect of a feature on the predicated outcome (average across the ICE plots)
      * OCE plots: display one line per instance showing how the prediction changes when the instance changes
    - Accumulated local effects (ALE) plots
      * ALE plots: describe how a feature of interest influences the prediction on average
  + Example-based explanations
    - Influantial instances
      * An instance is ‘influantial’ if deleting it from the data substantially changes the parameters or predictions of the model
    - Counterfactual explanations
      * A counterfactual explanation of a prediction describes the smallest change to the feature values that results in a change in the prediction
    - Prototypes and criticisms
      * A prototype is a data instance that is representative of the data; a criticism is one which is not well represented by the set of prototypes
* No one-size-fits-all approach, it is a process rather than single product !!!