PART I: QUESTIONS

- What is your favorite definition for an anomaly and why? (Other panelists might argue against this in later discussion!)
  What is your favorite anomaly mining application? Why is it challenging and (why) do most definitions fail on this application?

- Hypothesis driven definitions (based on the context) derive anomalous objects. Data driven definitions (without any context information) derive unexpected/rare objects. How do you rate these? Is there a third option from your point of view?
  In slightly different terms: Should we pre-define what an anomaly is and simply search for what we have defined (known unknown)? Or should we instead be looking for the unknown unknown? Which one do you think is more compliant with the scientific process (observation—hypothesis/theory formulation—prediction)?

- In anomaly mining, the problem statement is often open-ended and solution-agnostic (e.g., Is there political unrest in Alabama?). One may need to scope and formalize the problem as well as decide what data sources may be relevant to address it, before the detection process. In that way, is the anomaly-mining task similar to the data science process? Are there key differences?

- In the last past few years, a large number of different models (i.e. definitions/formulations) for anomaly mining have been published. Would it be possible to have one common modeling for this field just like in classification so as to focus our work power on a unified problem definition? Or rather, should we have various definitions to meet the needs of different applications? Can we compile a short list of fundamental models? How many do you think are there?

- What is your best way to transform an industrial/scientific anomaly detection problem into a formal/algorithmic solution? Please provide your best-practice workflow.

PART II: CASE STUDY

Goal: Could the global warming be cast as an anomaly mining task?

Problem statement: Are the weather conditions around the globe critically abnormal in the last several years?

How would you solve this problem with your means of anomaly definitions and best-practice workflow? How would you approach transforming the problem statement into formal anomaly definition(s)? How would you decide what data sources to leverage? How would the algorithmic solution look like?