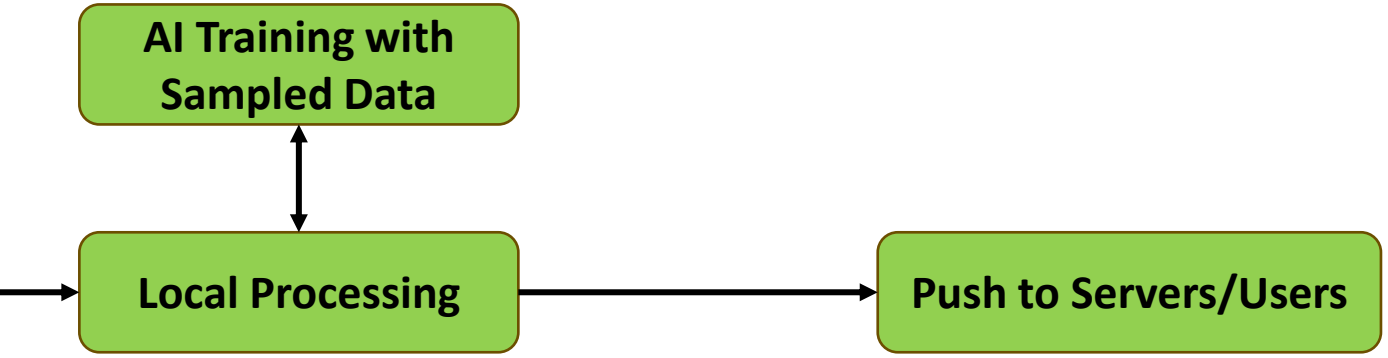
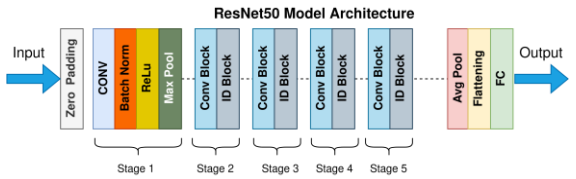
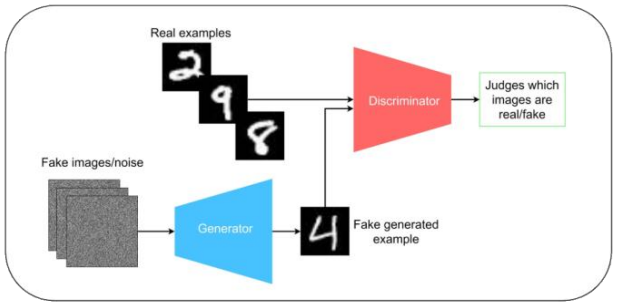
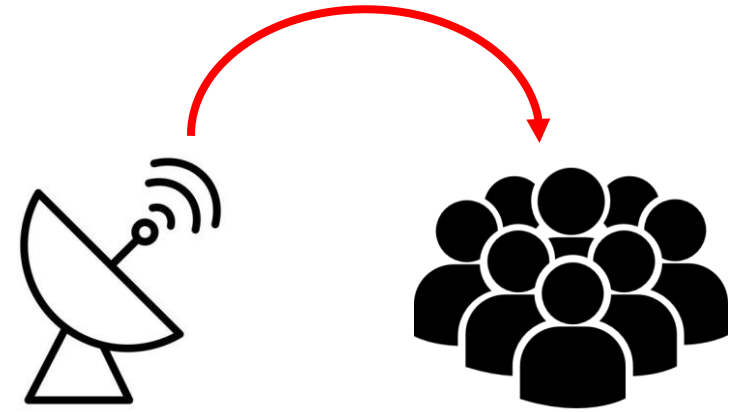
How do Different Data Collection Strategies Affect the Quality of AI Models for Environmental Sensing?

Consider a simple schematic for a sensor network…





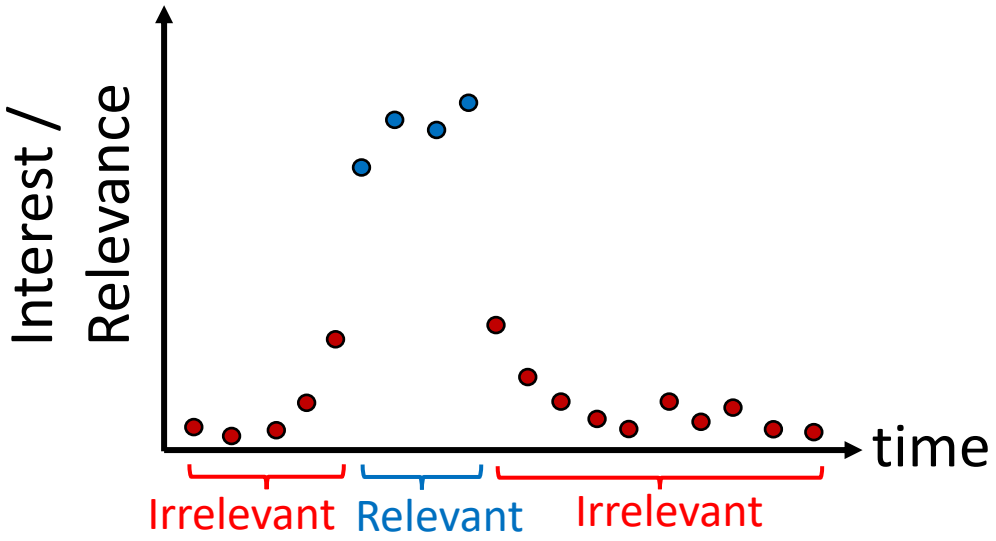




In many cases, **most of the data collected is unnecessary.** Hence, the data collection is *inefficient*, and wastes both memory to store it and energy to process and transmit it.

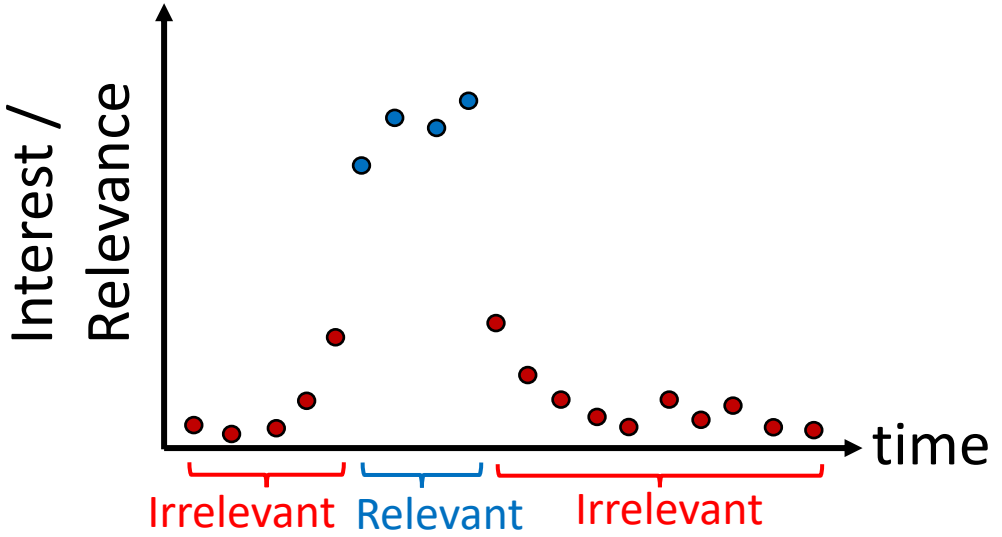
How do Different Data Collection Strategies Affect the Quality of AI Models for Environmental Sensing?

To start this project, we will consider a binary **Relevant**/**Irrelevant** flagging scheme. In many cases, the relevant data (blue) are a minority of the data collected, which is mostly irrelevant (red).



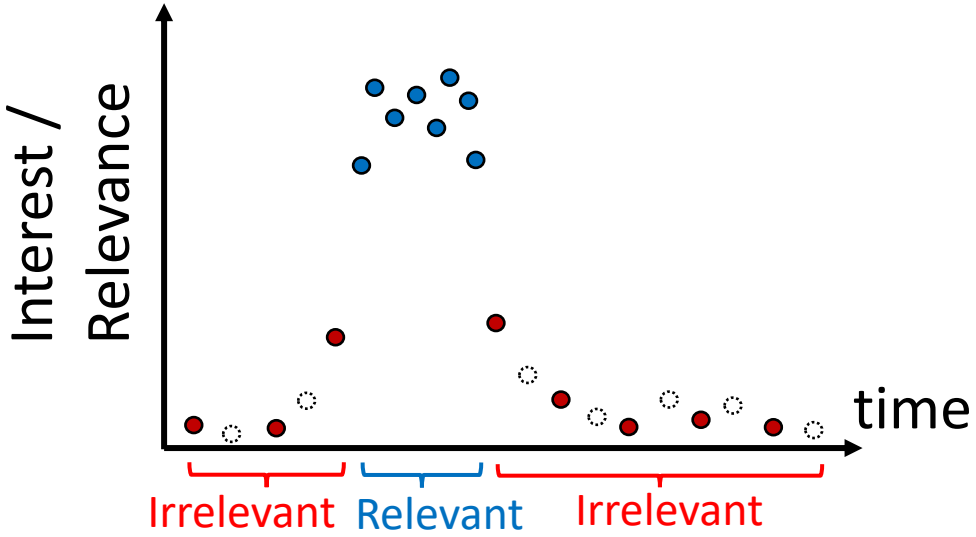
How do Different Data Collection Strategies Affect the Quality of AI Models for Environmental Sensing?

To start this project, we will consider a binary **Relevant**/**Irrelevant** flagging scheme. In many cases, the relevant data (blue) are a minority of the data collected, which is mostly irrelevant (red).



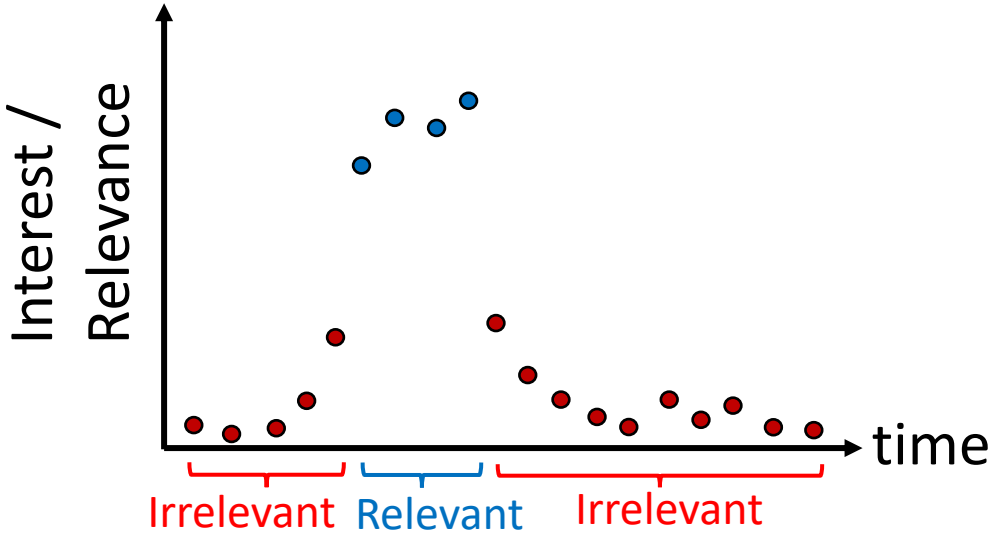
Instead, we can gather data in a smarter way by:

Changing the Sampling Rate



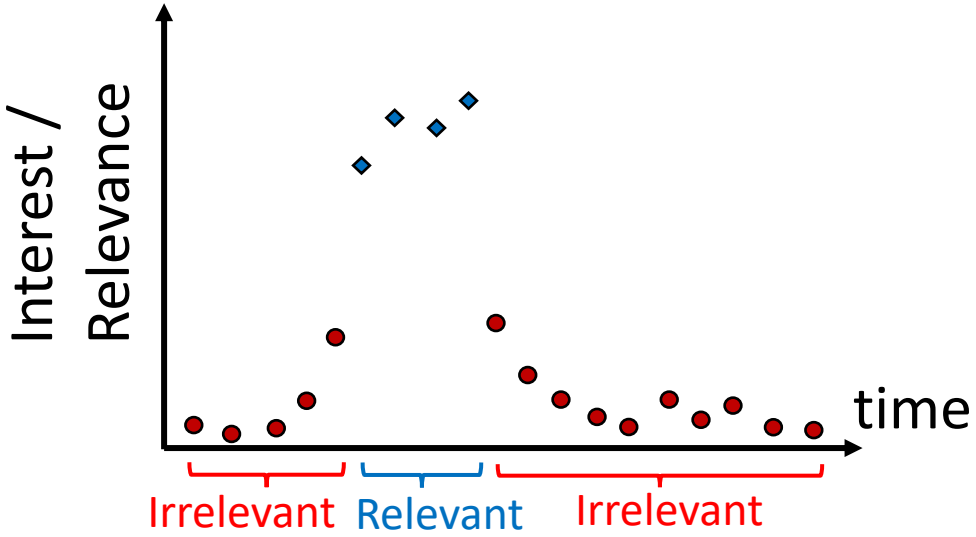
How do Different Data Collection Strategies Affect the Quality of AI Models for Environmental Sensing?

To start this project, we will consider a binary **Relevant**/**Irrelevant** flagging scheme. In many cases, the relevant data (blue) are a minority of the data collected, which is mostly irrelevant (red).



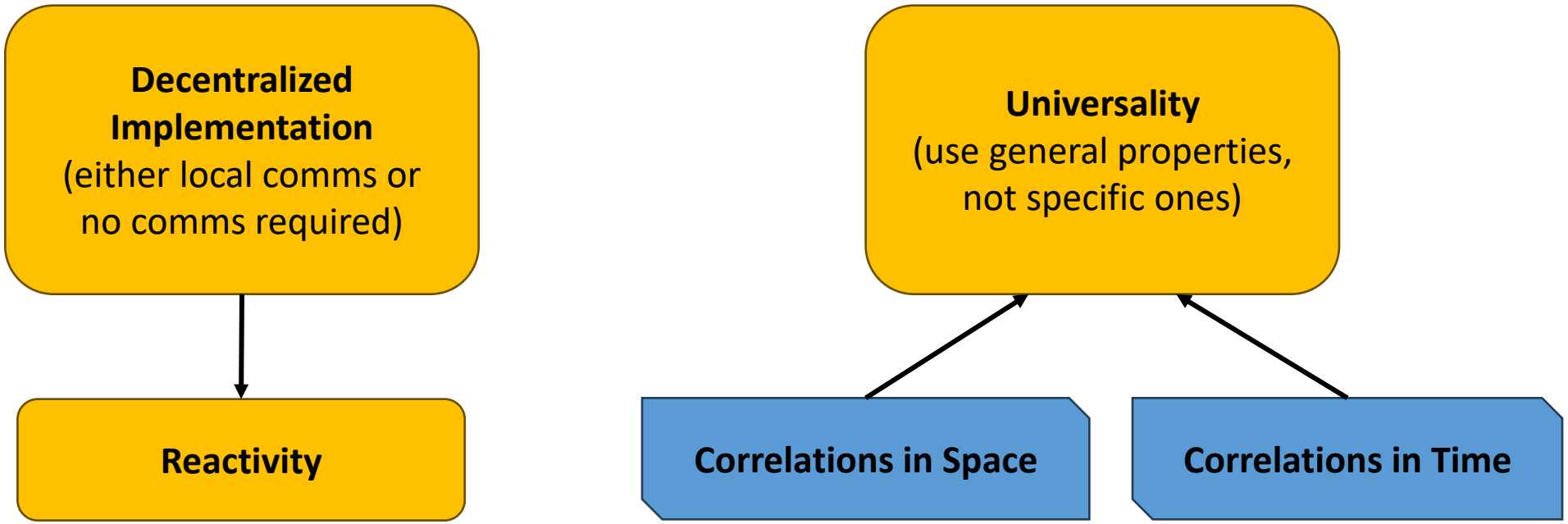
Instead, we can gather data in a smarter way by:

Changing the Sampling Quality



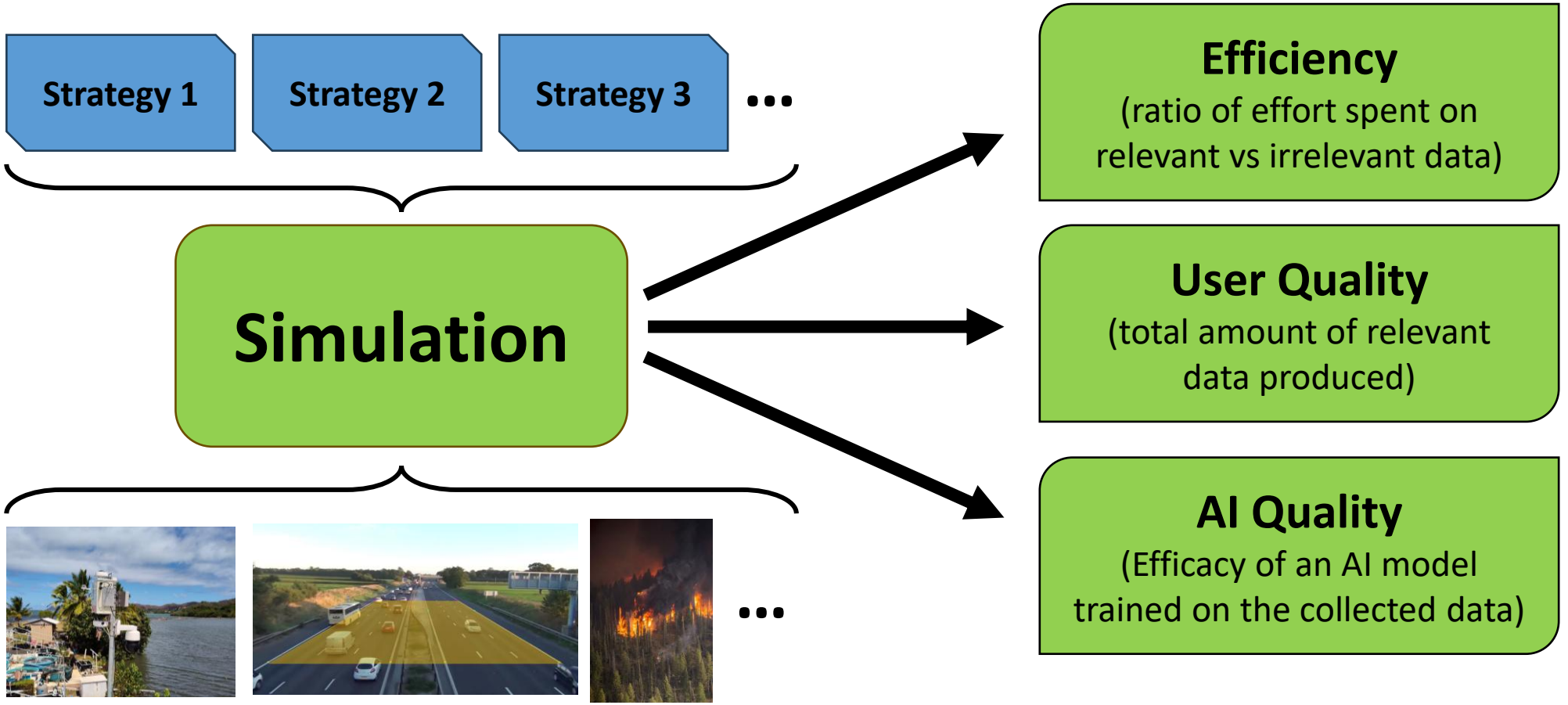
How do Different Data Collection Strategies Affect the Quality of AI Models for Environmental Sensing?

Our goal is to design a few **data collection strategies** to try and gather data with greater *quality* and *efficiency* than a uniform strategy. The strategies we create will all have two main properties:



How do Different Data Collection Strategies Affect the Quality of AI Models for Environmental Sensing?

To investigate our strategies' performance, we will use a **simulation** motivated by actual SAGE applications to analyze the *efficiency* and *quality* of the collected data.



How do Different Data Collection Strategies Affect the Quality of AI Models for Environmental Sensing?

Questions to the users (scientists):

- **Is a binary/discrete flagging scheme preferable to a continuous one?**
  - Discrete schemes may be useful if there are discrete levels of relevance or data quality standards
  - Continuous may be useful if relevance evolves gradually or if the data is often at a “medium” level of interest

- **Are there other types of relevance-correlations in your data? Other types of behaviors that you would want the sensors to have available?**