ABSTRACT

Dr. Barclay Brown

Engineering Fellow
Raytheon Technologies

Data Requirements and the Green School Bus Problem

As AI and machine learning-based subsystems and components become more prevalent in complex systems, the importance of training data is driving a new specialty within the systems engineering discipline of requirements engineering. Traditionally, system requirements included functional and non-functional requirements, including the "-ilities" such as reliability, dependability, durability, sustainability and others. For systems that include machine learning capabilities we propose the additional discipline of data requirements.

Successful data requirements engineering is necessary to prevent headline-making AI failures, which appear mysterious and frightening to both engineers and the public, but which can often be traced to relatively simple problems in the engineering of the data used to train the system. The best algorithm cannot overcome poor training data.

The systems engineer must be concerned with a new kind of system requirements--data requirements. Data requirements specify how much and what kind of data must be made available to the AI subsystem for training and testing. Systems engineers must learn enough about the environment in which the system will operate, and about the situations it could encounter to fully specify the data required to successfully train it. More than the machine learning developer or even the data scientist, it is the systems engineer who will be aware of the system of which the AI subsystem is a part, and also the larger context of the entire system in the environment where it will operate. In this session, we'll take a closer look at the role of data in AI systems and how systems engineers will need to learn to deal with it effectively. We'll use the framework of effective requirements engineering to show how data requirements can be elicited, analyzed, validated and implemented in the data design.