

### **Attribute Extraction**

The metadata contained within each product listing is critical for customer to make a purchase decision. At our scale, it is not uncommon for product attributes like brand or color to be missing from product listings, which adversely impacts product discovery and lowers customer trust. Addressing this problem of missing attributes at the scale of billions of products requires multiple ML capabilities. Some examples are - entity extraction from semi-structured textual content (product titles, descriptions), precise understanding of multimedia content like images and video, mining customer-feedback to identify defects thus reducing dependency on manually curated defect datasets. We adapt and extend state of the art techniques like Deep Learning and Sequential Models to build these capabilities.

### **Inaccurate Listings**

Incorrect data and information contained in our catalog leads to erosion of customer trust and poor customer experience. We have to ensure that the information contained in the product metadata catalog is factually correct, consistent and adheres to preset guidelines. The Amazon catalog is vast and diverse which poses challenges to solving this problem at scale. To weed out inaccurate listings, we leverage a wide range of ML techniques to detect mismatches between product titles, descriptions and images, precisely classify billions of products into product taxonomy and identify images that do not conform to prescribed guidelines. We have built ML solutions to generate embeddings of the different product attributes such as titles, descriptions and images into a relatively low-dimensional feature space using fine-tuned Deep Learning models. These embeddings are used as features in models trained to perform diverse set of tasks within defect identification and attribute generation.