### **ML Production Series — Establish Model Baseline 3**

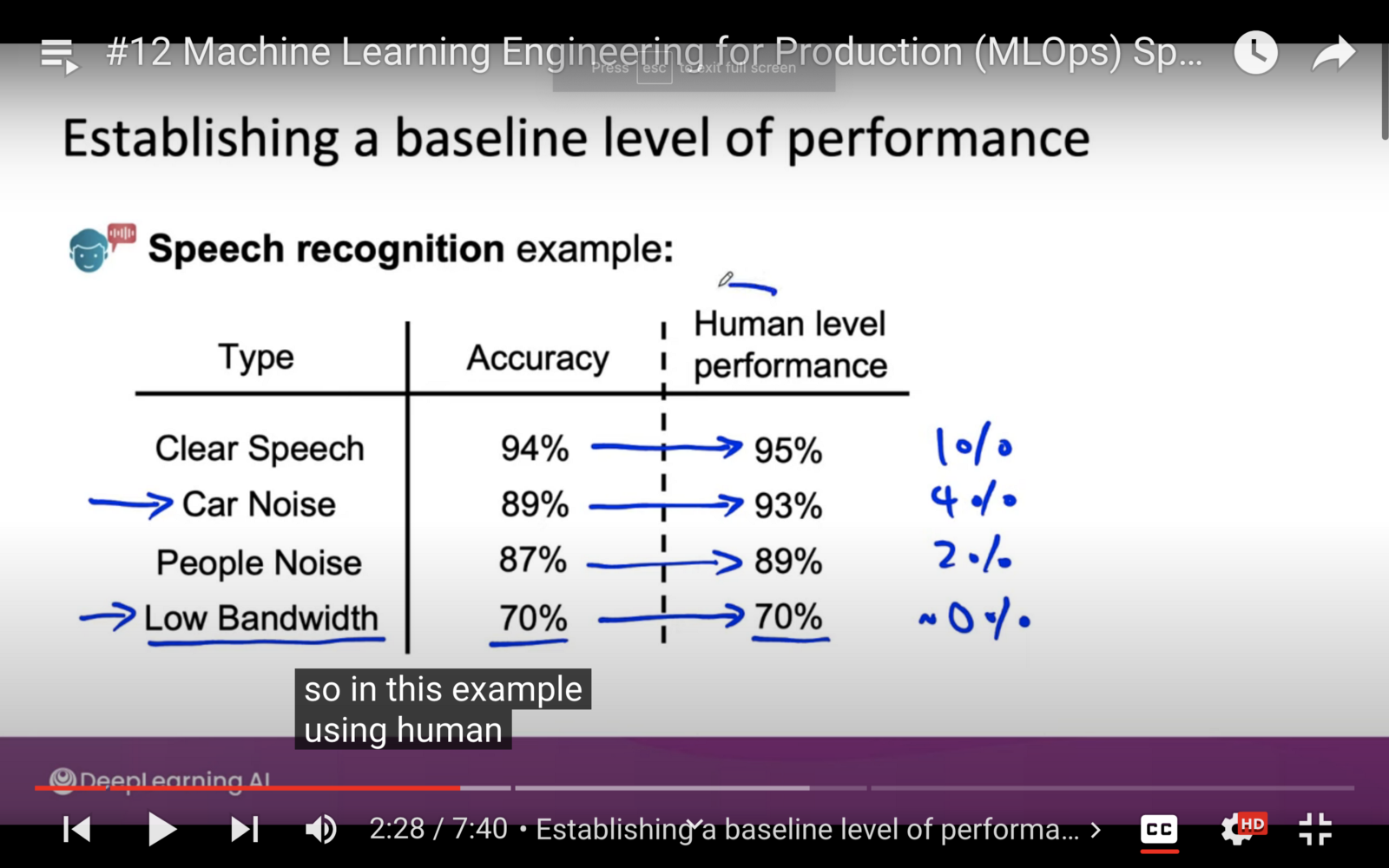
In this tutorial, we are gonna discuss the best practices for establishing a model baseline. It is important as ML is an iterative process and continuous improvement of the baseline helps to make a high-quality application

### **Example**

Let us understand this concept with a speech recognition system as shown in the **pic (fig-1)** below, considering we have four scenarios of different types and the respective accuracy of the model in a particular category.

#### **Analysis for — comparison with human-level performance**

consider we **first think**: of establishing a baseline where we want to improve people's noise and low bandwidth audio accuracy

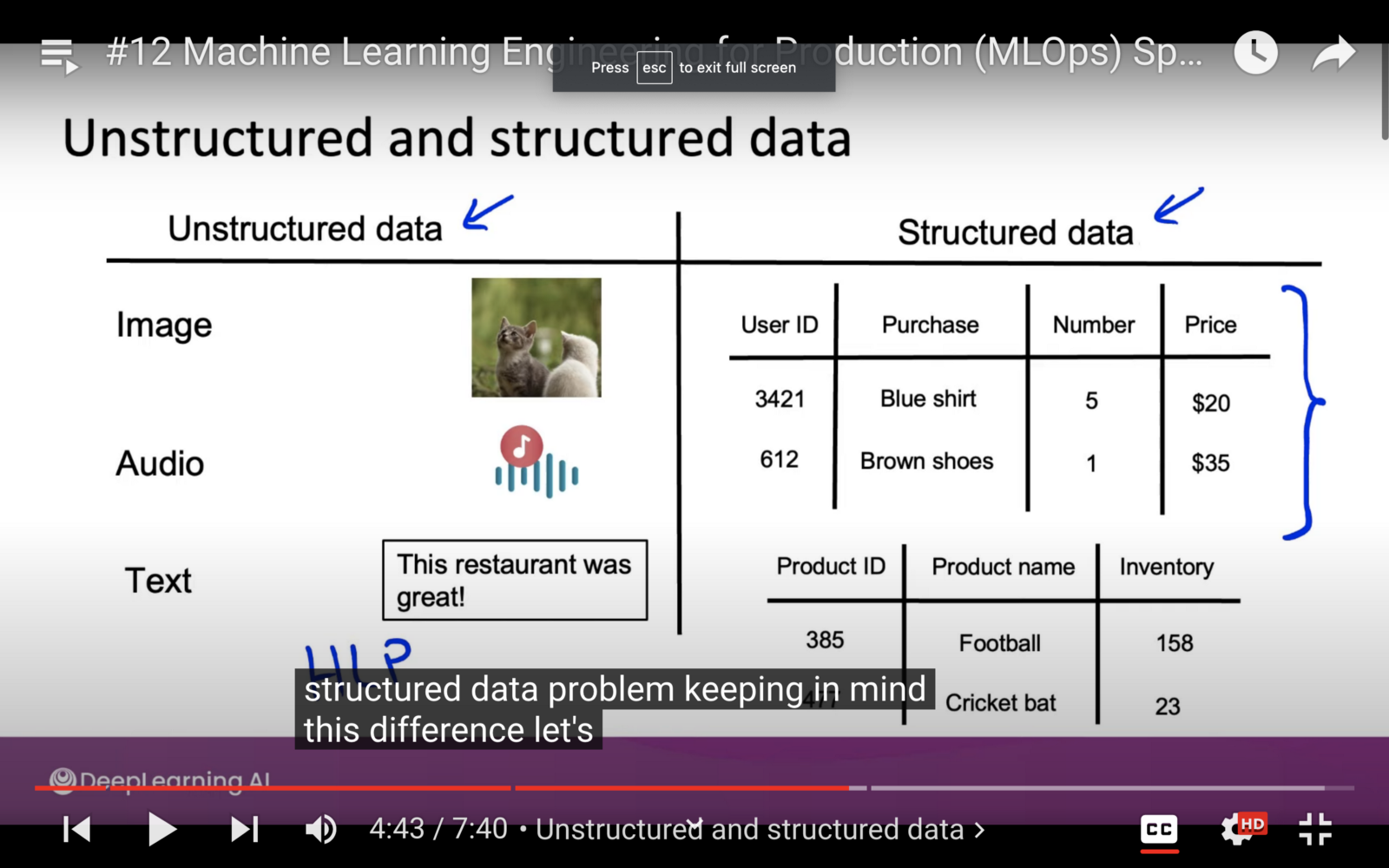
Fig-1

**second thought:** it is required to compare our algorithm performance with human-level performance to make a decision for establishing a baseline as In the above picture you can see let's say you hired humans to do the transcription and you got the above-mentioned accuracy of human-level performance.

**Conclusion —**we came to make us decide to improve people's noise and low bandwidth level performance as humans also cannot get much accuracy, instead focus our attention to improve **Car Noise** level performance

### **Scenario : Unstructured Data and Structured Data**

As described above the establishing basline- there can be a scenario in which data for modeling can be structured or unstructured — as seen in **fig-2** below unstructured data include **image, audio, text —**in which human level performance comparison is best practice to establish a baseline- **whereas structured data** might include **Inverntory system database or other application based databases** which also includes excel sheets or spreadsheets in different format contains lot of columns in which humans can have difficulty to make a decisions

fig-2

### **Now the Question is How to establish baseline for unstructured data?**

As **fig-3 describes** the ways to establish the baseline performance — for unstructured data Literature comparison is one of the best way to establish a basline in which you compare the performance of your model with state of the art.

Another way — is to make a first version of your model and do the continuous improvements in which you will compare the older implementation of the system

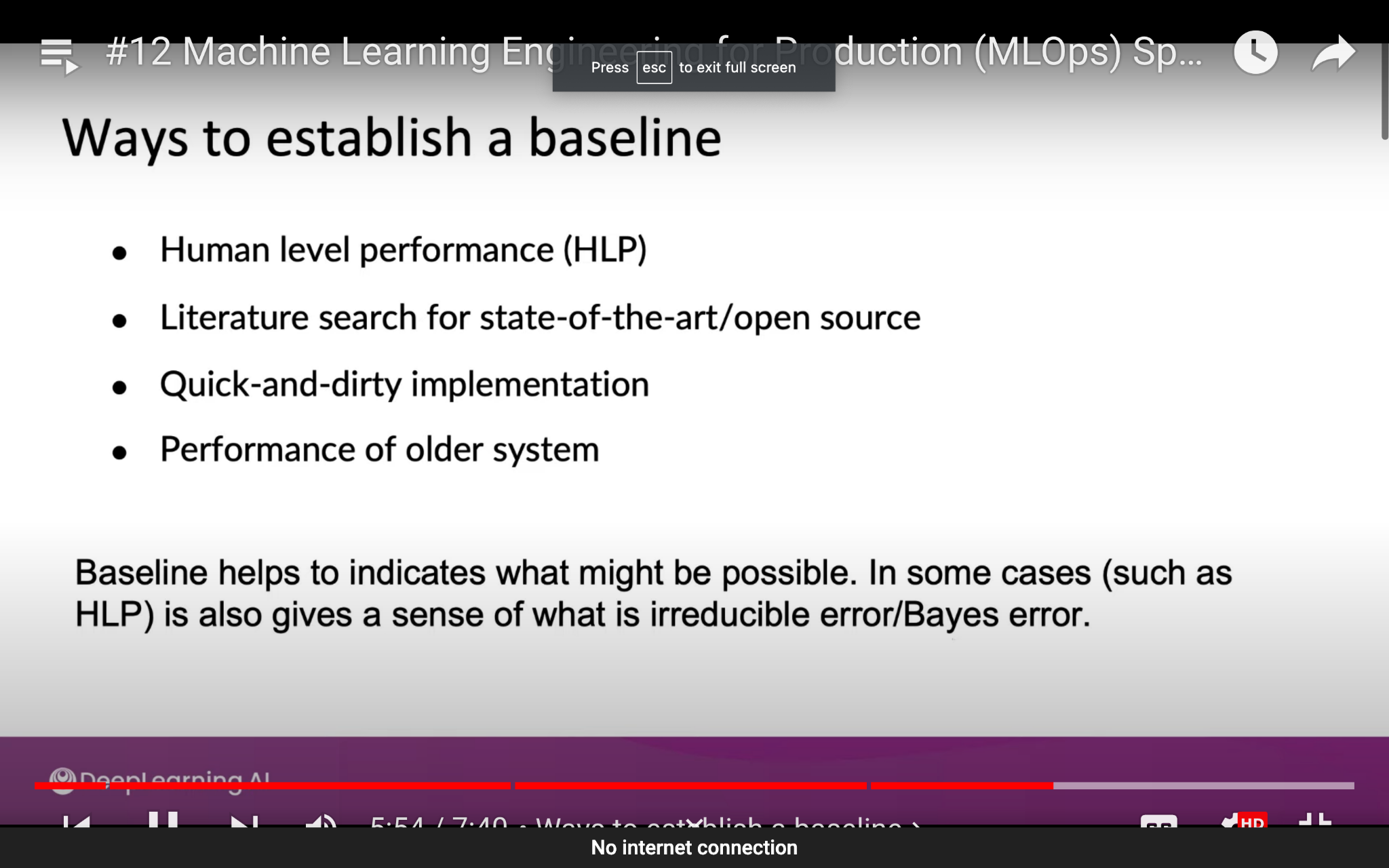


fig-3

**Benefit of establishing baseline —**is to find out the irriducable error.

### **Real World Problem : Companies and AI team conflict**

**Andrew Ng said —**he have seen some scenarios in which companies with AIteams have a discussion to reach a particular accuracy. for example client ask you to reach accuracy of 80 percent at least. In such cases AI team should ask them to make a baseline before deciding about error tolerance or system performance commitments.