**EK Meeting 11-10-23 – Settup AI Model**

Dataset labels:

* Will need the full dataset of 5258 labels
* A selection will be made from this with relevant (active) items for EK-Home (Reiner will do this)

What is the AI predicting: 1 or 4 categories?

* First Product: AI will predict only the level 4 categories (brick code)
* First product will infer category 1-3 purely from lookup table after prediction of category 4
* Further optimialisation can lead into AI predicting earlier level categories if confidence score is low

How often is the AI predicting?

* AI will predict initial dataset of 50k-200k images. Afterwards, 50-100 images/day need to be processed by the algorithm.
* Think about overall structure of AI model. Make it work in cloud (azure) / based on a daily schedule
* Document this architecture, potentially showcase what resources will be needed.

What is more important: accuracy/volume?

* Generally they are happy with ~80% accuracy. Moreso, the lowest percentage predictions, will go into a ‘low confident’ database that can be handled / fixed manually. (this is not done by me, but I can flag these datapoints with the scores and EK-Home will fix that)

Datasets images (nMedia/columbus)

* The nmedia dataset consists of ~200k-400k images, these do not have labels
* Columbus dataset consists of ~50k images, these do have labels and can be used to check performance of AI (ground truth should be somewhere, maybe ask yama)