

## Deploying AI Models for 5W1H Extraction

The extraction for “Who?”, “What?”, “When?” and “Where?” are performed on-demand, using standard NLP tools. Therefore, these extractions are handled directly by the backend service (having a dedicated “summary” endpoint). However, since extracting “Why?” and “How?” requires a deeper understanding of the context (underlying intents and causes are not explicitly mentioned in conversations), we use custom-trained, dataset-driven models. Consequently, data-intensive tasks need to be offloaded to cloud-based ML platforms. For that, we are using Google Cloud Platform (AI platform for training and deploying, Google Cloud Functions for serverless execution & Natural Language API for NLP tasks).