For my mini essay, I decided to do some research on how the Google Assistant works.

The Google Assistant is a voice-responsive application available on mobile and desktop device and is built into Android operating systems and can also be accessed on the web through the Google homepage. It's also a key feature of the Google Home set range of 'smart home' devices.

A simplified description of the way Google Assistant works is found in this article: Ask a Techspert: How do virtual assistants work? (blog.google). Here, one of the Team Leads on the Google Assistant Programme is interviewed (Francoise Beaufays).

The main steps of operation of the Google Assistant are as follows: A voice query I nteh form of spoken audio is inputted, for example using a microphone on your mobile device. The audio data captured Is then sent to Google's data centres where they begin to be processed. Importantly, the computing power required to process the queries is far greater than current consumer mobile and desktop devices are capable of, so the processing is not done on the local device. This explains why the Google Assistant or provides rudimentary service or is completely absent when there is no data connection. To improve accuracy of the query processing, techniques such as background noise cancellation are used to isolate the voice from interfering background noise. Unlike the voice recognition processing itself, this pre-processing of the audio can be performed in the local device in some cases before it is sent, since the technology to do this is now sufficiently miniaturised to be deployed in mobile devices and consumer headphones etc. After arriving at Google's data centres, the audio query is processed by a Deep Neural Network (DNN). A DNN is a type of artificial neural network (ANN). ANNs are a type of machine learning (ML) algorithm that are built to resemble some aspects of complex biological networks e.g. neural connections in the human brain. They receive inputs and produce outputs from multiple sources (nodes) and importantly, before the inputs are converted to outputs, they pass through a single processing layer (ANNs) or multiple processing layers (DNNs), where each layer transforms the data before progressing. In DNNs the multiple layers each process the data sequentially, and in general terms, the purpose is to gradually extract more abstract meaning from the data. Importantly DNNs are able to determine through training data sets, which factors to abstract and which to ignore autonomously and to refine this process over time. However, manual training and tuning of the algorithm is still necessary.

In the context of Google Assistant which is a voice-recognition DNN, the layers are designed to process the audio over various stages until the most likely correct interpretation is determined. For example, the first layers will be responsible for converting the audio into text. Subsequent layers will then identify the words and their meaning individually. From there, subsequent layers will seek to determine the higher levels of meaning from the sentence and will assign probabilities of being correct to each of the possible meanings it finds. Finally, the output(s) of the DNN is the response that is deemed to be the most appropriate, e.g. a location of place requested, or the results of a web search.

There are challenges to ensuring that the responses are consistently accurate. Factors that reduce the ease of processing include: background noise, unusual incorrect grammatical structures in the query, and mixed use of language e.g. French expressions common in English. In order to improve accuracy, the models can be refined through capturing the level of satisfaction that users report after certain responses are provided. In the future, Google Assistant hopes to be able to address these challenges enabling further accuracy of outputs of the DNN.