Rasa is an open-source conversational AI framework used to build context-aware chatbots. The core functionality of Rasa revolves around understanding user input, identifying their intent, and extracting relevant information, which is crucial for providing appropriate responses.

When a user inputs a query, Rasa’s NLU pipeline processes the message through several stages, including text preprocessing (tokenization, lowercasing, and punctuation removal). After preprocessing, Rasa uses a machine learning model to classify the intent behind the user’s message. This model is trained on labelled data where various intents, such as “cheap\_internet” or “fickle\_creek\_farm” are predefined.

Entity extraction identifies key information within the message, such as dates or locations. Rasa then assigns a confidence score to each predicted intent, ranking them based on the similarity to the user input. If the confidence score is below a threshold, Rasa triggers clarification questions to gather more information, either by asking the user directly or using models like SBERT to rank and select the most relevant clarification question.

Once the intent and entities are determined, Rasa performs the necessary actions, which may include triggering predefined responses or custom actions, such as API calls. The system then responds to the user with the appropriate answer, either static or dynamically generated.

Overall, Rasa’s intent recognition process combines machine learning and NLP to handle user input, enabling a seamless and interactive chatbot experience.