Ivy.ai

Steps in the IvyQuantum **Answering Algorithm**

Breaking it down ...

The algorithm underpinning IvyQuantum's generative chatbot's response entails a series of steps to facilitate seamless and secure interactions. Initially, when a user submits an utterance, it undergoes a meticulous preprocessing stage. This step involves not only sanitizing the input to remove any potentially sensitive or personally identifiable information（ PII） but also anonymizing it to further protect user privacy. This sanitization process typically involves a combination of techniques like pattern recognition, named entity recognition, and regular expressions to detect and redact or mask PII elements such as names, addresses, phone numbers, and more.

Following data sanitization, the chatbot initiates a document search within a vector database. This database is constructed using deep learning-based embeddings, which convert textual data into high-dimensional vectors. Each document in the database is represented as a vector, enabling efficient similarity and relevance calculations. The chatbot then retrieves documents that are deemed contextually relevant to the user's query based on vector similarity. This search process ensures that the chatbot has access to a rich knowledge base to draw upon when composing responses.

The next step involves document ranking, a critical phase in which the chatbot assesses the relevance and importance of each retrieved document. A neural net is used to assign relevance scores to these documents. The chatbot considers factors such as the document's recency, source credibility, and how closely its content aligns with the user's query to rank the documents accordingly.

Finally, the chatbot processes the ranked documents using the state-of-the-art language model, GPT 4. This advanced model leverages deep learning architectures, such as transformer networks, to generate coherent and contextually accurate responses. It does so by considering the content of the top-ranked documents, the user's original query, and contextual information from the conversation history. The generated response is carefully crafted to provide a natural and informative answer while avoiding sensitive information and adhering to privacy guidelines.

In summary, this sophisticated algorithm ensures that the generative chatbot can deliver technically sound, privacy-conscious, and contextually relevant responses to user queries, offering a high-quality conversational experience while safeguarding user data.