For a project with one of the world’s largest technology corporations, Lionbridge AI collected and annotated 30,000+ unique conversations in English and French. The data was used to improve their machine learning model’s ability to understand natural language.

THE CHALLENGE
In order to build applications able to engage with humans on a conversational level, machine learning models need to be fed large amounts of structured language data.

As one of the world’s leading technology companies, our client designs, develops and sells consumer electronics and software to hundreds of millions of people worldwide. For this project, the client needed to collect a range of conversation data from native speakers of English and French. The data was used to refine their natural language processing (NLP) models.

THE SOLUTION
The company hired Lionbridge to build comprehensive dialogue datasets in English and French.

- **Crowd Management**
  Lionbridge tapped into our global crowd of 500,000+ qualified contributors to quickly source over 200 native speakers of English and French. The project management team recruited, educated and managed contributors to ensure all work was completed on time and in line with client instructions.

- **Data Collection**
  The client provided Lionbridge with specific instructions that covered conversation topics, required keywords, and conversation length. Lionbridge then distributed thousands of scenarios to 200 remote contributors, each of whom was tasked to create their own unique conversation according to the client’s specifications.

- **Data Annotation**
  As a follow-up task, contributors annotated the conversation corpus with a range of predefined tags. This included labeling for speaker information (Speaker 1, Speaker 2) and sentiment (Positive, Negative).

THE RESULTS
Lionbridge’s data quality and speed of execution surpassed the client’s expectations. This pilot project demonstrated to the client that Lionbridge could source and tag appropriate data through their qualified contributors. As a result, they felt confident enough to increase Lionbridge’s involvement in the project. Since then, the program’s scope has been expanded to include additional languages and over 150,000 conversations.

30,000+ Collections Collected
2 Languages
200+ Native Speakers

ABOUT THE CLIENT
Our client is a prominent multinational technology company that designs, develops, and sells consumer electronics, computer software, and online services. In recent years, they have been particularly focused on advancing state-of-the-art machine learning technologies.

Lionbridge offers 300+ languages and dialects, 500,000+ global contributors, and 20+ years of experience.