The how and why behind using A.I.

Civinc connects people with different perspectives in chat conversations, and distills the prevailing sentiments. We use Artificial Intelligence (AI) to extract insights on general ideas, contention and sentiment. Here's how it works:

Context & goal

During a Civinc session participants anonymously engage in chat conversations with peers who have a different view. These conversations are always about a specific statement on which the partners had a different opinion. If AI is enabled for a session, all the messages that are sent per statement are aggregated and analysed with A.I. before being permanently deleted. The sole goal of the use of AI is to provide the organizeer and participants a general understanding into the different existing perspectives and lay the foundation for deeper (joint) reflections post-session.

AI in practice & principles

The analysis of the aggregated chat messages is based on 'prompt-engineering'. We instruct the Large Language Model (LLM) to run different types of analyses based on specific prompts that are formulated by us. We can for example instruct it to summarize all conversations on a specific topic, in order to distill the general sentiment. In addition, we can instruct it to produce specific output such as the most common ideas, arguments or reservations expressed in the context of the conversations. There are a few principles that apply no matter what:

- AI analysis is always optional. If a customer or admin, for whatever reason, decides that they do not want to use AI, it can easily be disabled per session.
- The AI analysis only runs on a statement when a minimum number of messages has been sent in multiple different conversations, to ensure a critical mass.
- The output is always generic in nature, never contains direct quotes and never includes personal information and can never be traced back to an individual user.
- The chats themselves are only visible to the participants and are deleted permanently after the session / AI analysis.

LLM processing & privacy safeguards

Civinc uses the Mistral AI API as the foremost frontier Europe-native solution for AI analysis, with specific privacy and data protection safeguards in place:

- The data is solely processed on servers in the EU
- The data is not available to other customers, is not used to improve AI models, and is not used to improve any 3rd party products or services
- The data will only be processed for the purposes as defined by Civinc and will be permanently deleted from Mistral servers within 30 days
- As Civinc does in principle not collect data which can be traced back to an individual user, such data is also not shared with Mistral
- For more on processing and privacy policy: Mistral AI.
- For more on. sustainability policy: Mistral AI
- For more on Responsible A.I. measures: Mistral AI

Quality assurance, limitations and statement generation

Given that Civinc uses an external Large Language Model (LLM), we rely on quality assurance testing to make sure that data outputs are of adequate quality. As with all LLM's, the AI output may, however, contain inaccuracies. Civinc data reporting always features other data insights based on user responses to statements and feedback to survey questions, so the AI component is part of a larger whole and should be understood as such.

Lastly, Civinc admins can invite the help of AI to formulate statements for Civinc sessions. We use the same LLM's for this, where only the admin input/request is shared with the processor (no personal or other information).