ASI METHODOLGY

A random sample of people online is collected using **Conditional Independence Coupling (CIC).** CIC is an algorithm that crawls online networks and creates a sample of people that is representative of the overall population. In all statistical respects, the sample generated by CIC is identical to Random Digit Dialing used in traditional polling, with one important exception — the sample produced is many tens of thousands to hundreds of thousands of people.

CIC can be adjusted to sample from a geographic location, like a country, state or province, and municipality. This is done by restricting the network CIC is crawling to people that reside in the desired geography. Two points of information are used by CIC to determine a person's geography: geo location information and self-reported location. The demographics (gender, age, race) for individuals are then assigned by a collection of Artificial intelligences (AIs).

A major difference between traditional polling and ASI's approach is that **no actual questions are asked** of the individuals in the sample. Asking questions may introduce bias by either forcing a person to take a stance on an issue they have previously been un-interested in, or biasing the response through wording of the question.

ASI avoids these issues through a combination of **topic detection and stance detection.** Topic detection is an AI that determines if a person is discussing the topic under study. Stance detection is an AI that determines if a person supports or opposes the topic. Topic detection starts with the AI learning the topics around an issue. Just like traditional polling, the AI starts with a selection of articles that summarize the topics around the issue under study: arguments for and against, hot button items, and current events around the issue. The AI learns the topics using an algorithm called Latent Dirichlet Allocation (LDA) from the provided articles.

The topic model the AI learns is very similar to the questions a pollster might develop. They include topics that are framed for and against the issue, topics that are neutral to the issue, and topics focused on current events around the issue.

For each person in the sample, the AI examines all of their online communication within the study period to determine if it matches one of the learned topics. Based on this analysis, the AI assigns a probability if a person is discussing the topic or not. Stance detection is a separate AI that determines if a person opposes or supports the given topic.

Traditional polls refer to the response rate within the sample. The response rate for ASI's method is technically 100%. An alternative measure of the response rate is engagement: a measure of how many people were actually engaged with the topic and discussing the topic online. While posing a question prompts a survey taker to respond, the engagement rate indicates how interested the population is in the topic without being asked.

NOTE: All mined data is taken from public online sites where each user has agreed to allow open access to their accounts and data. Mining of the data is done in compliance with each site's terms and conditions. If an individual on the site has indicated they wish their information to be private and not mined, ASI respects the wish and does not mine their data.

Privacy is preserved using two methods: data anonymity, k-anonymity and differential privacy.