



BLANKING OUT NAMES AND ADDRESSES

Information explosion is reality of today, and abuse of personal information for malicious purposes is an issue being grappled with by all companies, stakeholders, IT departments and CISO's. Governments, on their part, have shifted onus of protecting PII (Personally Identifiable Information) to enterprises that collect them for legitimate business reasons and for all the right reasons. Imagine a customer reaching out with a mail that reads ' I am on a holiday in Malibu and my son Chris was playing with my XST pad and it just went blank. Unable to reach you on phone. Could you let me know whom I can contact and how best to fix it. My XST number is W3J 4623/45T and credit card number is 3456 9100 0033'. Agents attending to this message do not really need to know where the customer is holidaying, his son's name or credit card number to start solving this issue and your company may not want that data accessible or transmitted openly. Identifying and masking over credit card numbers and SSNs are no brainers as they are deterministic in nature. Real challenge in privacy protection lies in taking out names and addresses. Enterprises have long struggled with cleansing records of PII, and mostly given up trying to use their rich information. Names and addresses, with unstructured formats are particularly difficult for

autonomous filters to catch and mask. You can tell a system that "John" need to be blanked, but then "Jon" pops up, followed by a "Sean", "Ian" and a "Giovanni", not to speak of "Jane" which the system has no way of recognizing with its preprogrammed rules and dictionaries.

Sainapse makes this process seamless and painless.

While tighter compliances to ensure PII in data is rapidly spreading, Sainapse can now protect enterprises to lessen their burden. Contextual Identification of entities (CIE) is an integral part of Sainapse's data curation process. Sainapse trains itself on data involving labeled entities such as person or location. Given unseen data, it can virtually flawlessly predict if set of entities occurring in data meet the criteria of those that need to be redacted. It can, in fact, train on any customizable entity from adequate data and labels. Sainapse achieves accuracy of 92%+ while training machine learning model in less than 15 minutes.