Goal:

* Create a tool that scans the internet to best fill missing email fields given a list of information about the subject.

Process

1)

* Email pattern generator: the first step is to create as wide of a universe of emails as possible, where the user’s name will be remixed into the most popular email patterns for a search
* Reverse discovery

2)

* Search and scrape: Via the SERP API (100 free searches per month here), scrape all visible text and context around emails
* A

3)

- Semantic filtering: SBERT to compute embeddings for candidate emails and anchor sentences

- Output cosine similarity and semantic match score

4)

Architecture

* Inputs entry: this will allow the user to input basic information about the subject (Salesforce fields), which will form the basis of scraping and SBERT verification