

# Readme

This project is a LinkedIn data extraction tool that can fetch profile information based on either a person's name or their LinkedIn profile URL.

## Installation:

1. Create a virtual environment named 'extractor': `python -m venv extractor`
2. Activate the virtual environment:
  - For Windows: `extractor\Scripts\activate`
  - For macOS/Linux: `source extractor/bin/activate`
3. Download `generator.py` , `helper.py` , `name_parser.py` and `main.py` and put it in `extractor` directory
4. Install the required packages from the 'requirements.txt' file: `pip3 install -r requirements.txt`

## Usage:

Note : Give your openai key in (`generator.py`) and proxycurl api key in (`helper.py`)

To run the extraction tool, execute the following command:

- If you have a person's name, run the command with the `--name` option: `python main.py --name "Stella Biderman"`
- If you have a LinkedIn profile URL, run the command with the `--profile` option: `python main.py --profile "https://www.linkedin.com/in/jonwray/"`

Ensure that you replace the example name and profile URL with the actual name or URL you want to extract data for.

The extracted data will be displayed in the console or written to a output folder named "output" ( created automatically ), depending on the implementation in 'main.py'.

## Note :

The profiles which has only linkedin profile works perfectly , However , where there was only names it only extracts some of them.

- Reason - All the 3rd Party data providers like ( hunter.io, fullcontact) need more context such as company name or email address
- Currently, in the solution created, i generate linkedin profile url's. If it exists in proxycurl we extract them else it doesnot
- We can improve it by getting atleast one more metadata feild