

# Smart Lead Gen

*AI system for personalized outreach*

## Team Members

Anirundh Penumatcha

Banoth Pavan

I Varun Sai

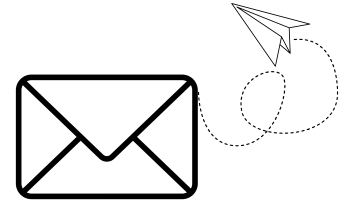
Daaris Ameen

Sumit Gangwani

## Course Name

Datascience and AI Lab (BSD4001)

# Introduction



- Companies frequently need to send emails to potential clients to market their products, or to candidates for recruitment.
- To reach a large audience, they rely heavily on mass email outreach, but this approach prevents meaningful personalization, a key factor in converting leads.
- Manually personalizing each email requires researching the recipient, understanding their context, and crafting a tailored message. This process is slow, labor-intensive, and not scalable for large sales or hiring pipelines.
- As a result, most organizations fall back on generic outreach but, generic email campaigns yield less than a 2% response rate, leading to poor engagement, lost leads, and missed business opportunities for the company.

## Problem Statement

Companies struggle to deliver personalized, high-impact outreach at scale, resulting in poor engagement and lost opportunities.

### Key Points:

- Crafting personalized outreach requires deep manual client profile research and custom writing.
- As outreach volume increases, maintaining personalization becomes impossible to scale.
- Lack of personalization reduces trust, credibility, and response likelihood from potential clients or candidates.

# Existing Market Solutions

Segment	Tools	Limitation
CRM Tools	Outreach.io	They automate outreach using static templates but limited personalized.
Social Media Solutions	Linkedin Talent Solutions	They help to find prospects on their platform but they do not generate messages.
LLM Models	GPT-4	They do not have data ingestion pipeline

# Identified Gaps

- There is no integrated data to draft pipeline, current outreach tools handle only isolated parts of the workflow. Sales automation platforms send sequences, and lead-finding platforms collect profiles, but none of them convert profile data directly into a personalized message draft.
- There is weak profile information extraction, existing systems fail to leverage rich, unstructured professional signals such as recent LinkedIn posts, interviews or achievements, company announcements, role transitions and milestones, etc. Due to this, automated messages remain generic.
- There are no domain specific outreach systems

## Opportunity for Innovation

Our project introduces an end-to-end AI outreach system that enables hyper-personalized emails at scale by combining structured professional data with a fine-tuned generative AI model.

## Key Points

- To bridge the personalization gap, we develop a domain-specific fine-tuned model that generates personalized outreach messages tailored to each recipient's background, context, and intent.
- We have integrated data to draft pipeline, our system extracts career history, recent activity, and company signals, converts them into structured data, and feeds them directly into the model for personalized output.
- Context-aware, conversion optimized messaging, our generative model uses both structured data and explicit outreach intent (e.g., "book a demo") to create messages that are more relevant, persuasive, and conversion-focused.
- Eliminating the "Spray & Pray" Problem, by automating high-quality personalization at scale, we help companies move beyond mass generic outreach and achieve meaningful engagement with every lead.
- Improving efficiency and scalability: The system reduces manual workload, speeds up outreach creation, and enables thousands of personalized messages to be generated instantly, that is something impossible through manual methods.

# Project Motto

Develop an AI system capable of generating hyper personalized outreach emails at scale, taking structured professional data as input and a fine-tuned generative model for text generation.

## Key Objectives

### Build a scalable data ingestion pipeline

Create a data pipeline that collects, parses, and structures professional data (career history, posts, company updates, achievements) into a machine-readable JSON format that can be used directly by email generation model.

### Create a multi-input model architecture

Develop an architecture where the LLM receives:

1. Structured profile data
2. User intent (e.g., book a demo, invite for hiring, initiate partnership)
3. Classified client Personality

This ensures messages are personalized and optimized for the specific outreach goal

### Finetune an LLM on a outreach dataset for email generation model

Train a base model Mistral on the dataset of high-performing outreach messages, enabling the system to learn tone, style, persuasion patterns, and personalization cues relevant to professional communication.

### Develop a personality-classification model for deeper personalization

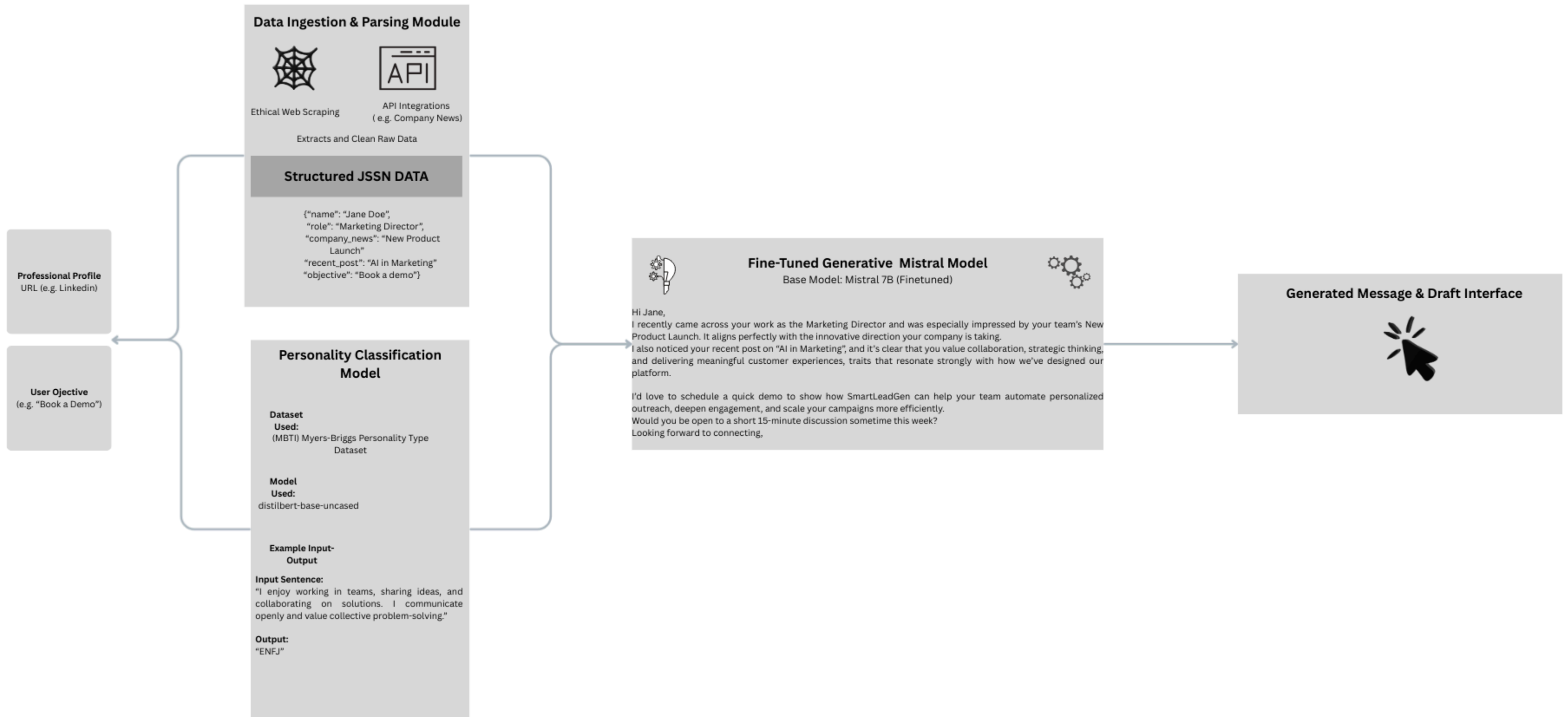
Build a model that infers a recipient's personality from profile information and posts. This personality vector is then passed as an additional input to the email generation model, enabling tone-matched messaging.

# Target Audience

A fully automated SaaS platform that helps marketing teams create hyper-personalized outreach at scale, improving campaign relevance and conversion rates.

- Sales & Marketing Teams – For personalized client outreach and demo booking.
- Recruiters – To contact high-value candidates with tailored messages.
- Founders & Entrepreneurs – For networking, partnerships, and investor outreach.
- Job Seekers & Freelancers – To send customized messages to hiring managers or clients.

# Solution Architecture(High-Level)



# Detailed Workflow

## 1. User Provides Input

The user enters a profile URL (e.g., LinkedIn) and selects an outreach objective such as “Book a Demo”, “Invite for an Interview”, or “Start a Discovery Call”. These two inputs trigger the entire pipeline.

## 2. Data Ingestion & Parsing Module

The system performs ethical web scraping and/or API-based data extraction to gather publicly available information from the profile. The extracted details include; Name, Role & experience, Company news, Recent posts or activities, Achievements or milestones. The Raw data is processed, cleaned, and organized into a structured JSON format. This JSON acts as the primary input to both the personality model and the generative model.

## 3. Personality Classification Module

The JSON extracted from the user’s profile is passed into the personality model. Using a model trained on the MBTI dataset (e.g., distilbert-base-uncased), the system predicts the recipient’s communication personality type. The predicted personality label (e.g., ENFJ) is added to the structured input data, providing an additional signal for tone and writing style.

## 4. Multi-Input Fine-Tuned Generative Model

The fine-tuned generative model (e.g., Mistral 7B ) receives:

- Structured JSON profile data
- Outreach objective (intent)
- Personality type

The model synthesizes all inputs to produce a context-aware, Highly personalized,Tone-aligned, Persuasive outreach message

## 5. Message Generation

The generative model produces a draft outreach email tailored to the recipient’s: Professional background, Recent activities, Company context,Communication personality, User-defined objective,The output is refined for clarity, relevance, and persuasion.

## 6. Draft Interface Delivery

The final personalized email draft is displayed in the message interface, where the user can:

- Review
- Edit
- Copy
- Use directly in outreach campaigns

This completes the end-to-end, AI-powered personalization workflow.

# Feasibility

- **Technically Feasible:**

Uses open-source LLMs and PEFT methods (LoRA/QLoRA), making fine-tuning efficient and lightweight.

- **Data Feasible:**

Synthetic datasets remove ethical scraping issues while providing diverse, high-quality training examples.

- **Computationally Feasible:**

QLoRA enables training on low-cost GPUs; inference runs smoothly on modest hardware.

- **Operationally Feasible:**

Simple inputs (profile URL + intent) make the system easy to use and integrate into a demo workflow.

- **Economically Feasible:**

Low cost due to open-source models, minimal cloud usage, and synthetic data generation.

## Conclusion

- Personalized outreach is essential for effective sales, hiring, and professional communication, yet current methods are either generic and low-impact or manual and unscalable.
- Our project solves this gap by creating an end-to-end AI system that transforms professional profile data into hyper-personalized outreach emails at scale.
- By combining a structured data pipeline, a personality classification model, and a fine-tuned generative LLM, we enable companies to produce messages that are context-aware, persuasive, and aligned with communication style.
- The system reduces manual workload, improves engagement, and increases conversion potential—allowing organizations to move beyond “spray & pray” outreach.