

Internsl^oth

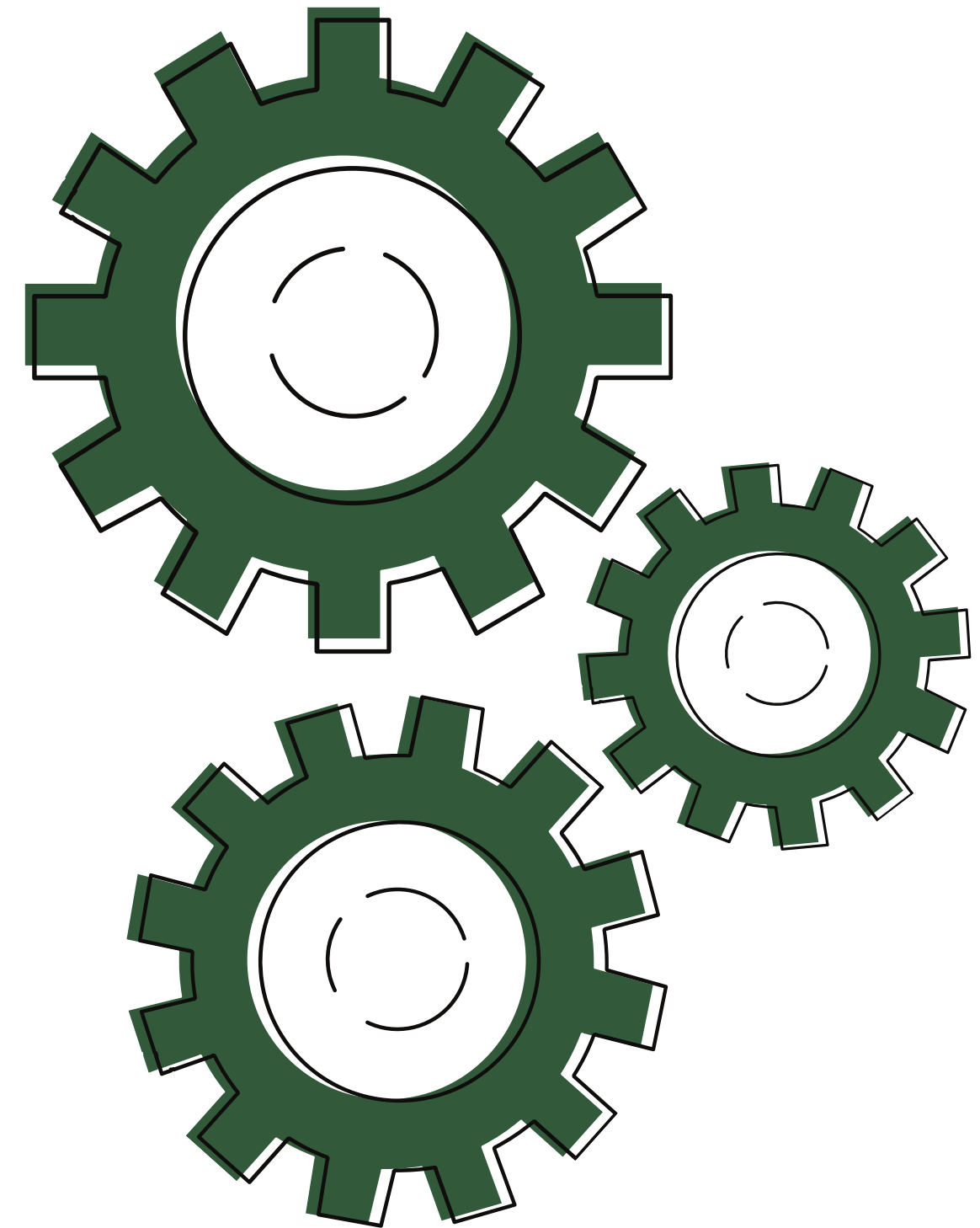
AI-assisted Internship and Job Matchmaker



Presented By:
Ashish Namdeo, Mohith Kanthamneni, Kaung Khant

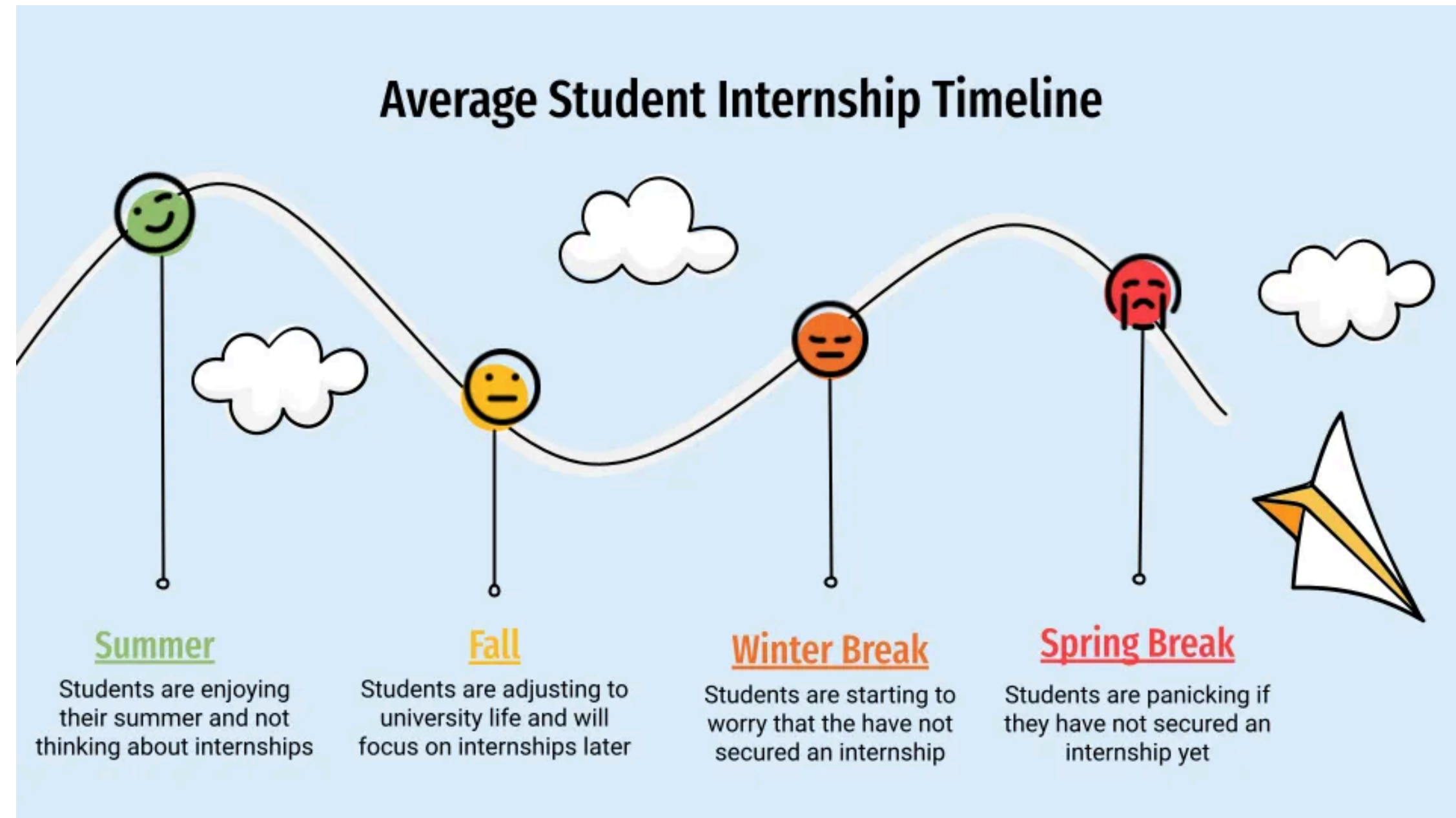
Agenda

- Introduction
- Methodology
- Challenges
- Dataset
- System Architecture
- Result and Evaluation
- Live Demo
- Future Work



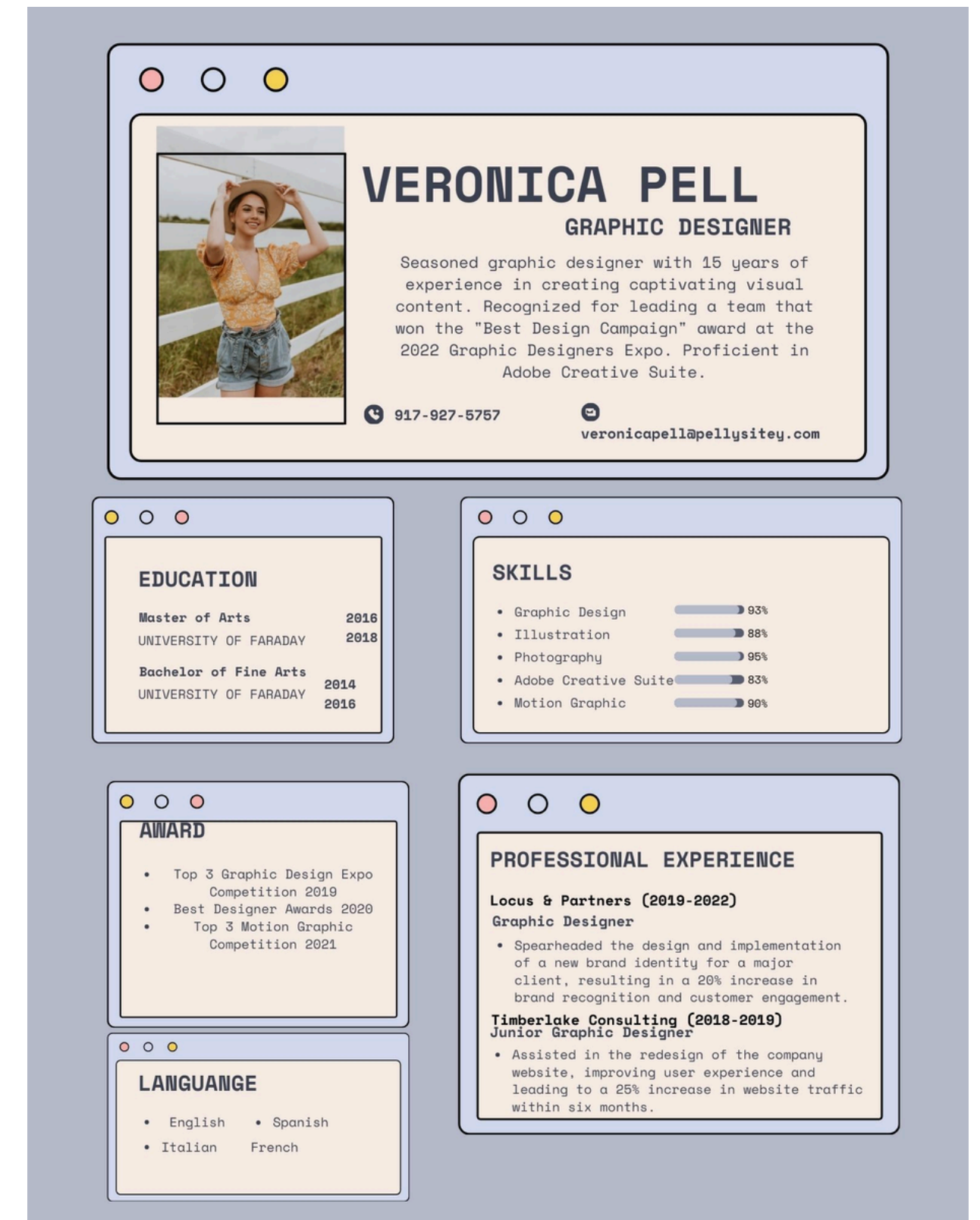
Introduction

- Manual job and internship searching is tiring, time-consuming, and often leads to the wrong roles.
- **InternSloth** simplifies these processes by matching relevant skills on a Resume to the most pertinent **real-time** internship or job positions

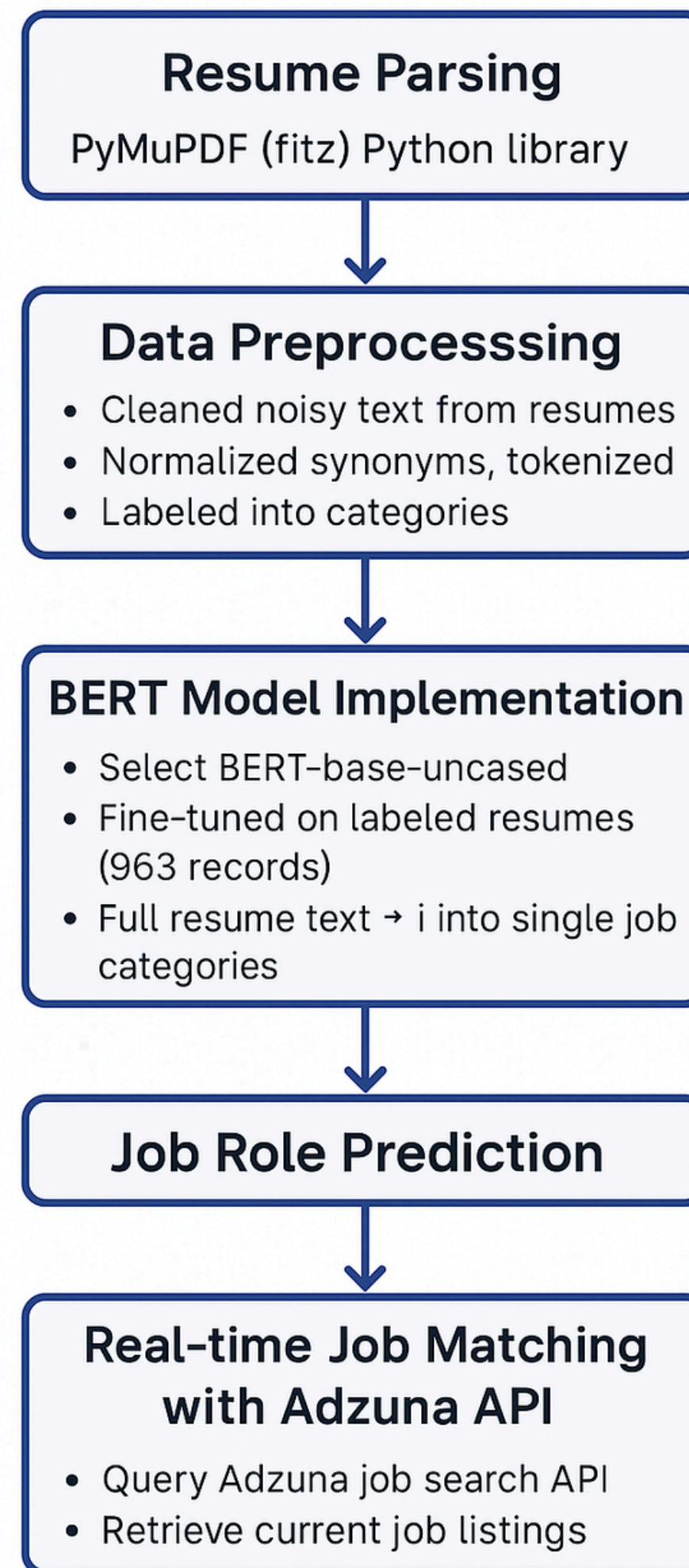


Major Challenges

- **Messy Resume Text:** Resumes come in many formats and styles, making it hard to clean and prepare the text for analysis.
- **Training BERT is Tricky:** Fine-tuning BERT needs powerful GPUs, and it's easy to make mistakes that can slow down or crash your system.
- **Similar Words Confuse the Model:** Without cleaning and standardizing similar words, BERT may treat them as completely different, which can hurt accuracy.



Methodology



Dataset

- Dataset Source: **Hugging Face**
- Number of Dataset records: **962 samples**
- Number of Features: **2 (column 1 -Category, column 2 - Resume)**
- Training : **80%**, Testing : **20%**

Dataset – Intern Categories and Resume

Mechanical Engineer	Education Details June 2014 to June 2018 BE Mechanical Engineering Pune, Maharashtra Savitribai Phule Pune University Mechanical Design Engineer Mechanical design engineer Skill Details...
Mechanical Engineer	* I'm hard working person. * I'm self confident and can mould myself to all work environments.Education Details January 2016 B.E MECHANICAL ENGINEERING ALAMURI RATNAMALA...
Sales	Education Details Bachelor's Bachelor's Commerce India Guru Nanak high school Sales Manager Skill Details Data Entry- Exprience - Less than 1 year months Cold Calling- Exprience - Less...
Arts	I Other Skills Course/Skill Name Board Year Of Passing Grade Intermediate Grade Drawing Art Examination Committee, 2011 B Examination Maharashtra State I Academic Programme Detail Nationa...
Web Designing	Technical Skills Web Technologies: Angular JS, HTML5, CSS3, SASS, Bootstrap, Jquery, Javascript. Software: Brackets, Visual Studio, Photoshop, Visual Studio Code Education Details January 2015...
Web Designing	Education Details B.C.A Bachelor Computer Application Pune, Maharashtra Pune University H.S.C. Pune, Maharashtra Pune University S.S.C. Pune, Maharashtra Pune University Web Designing and...
Health and fitness	Education Details January 2009 P.G. Sports science Dr. BMN College of Home Science January 2008 BSc Food Science & Nutrition Dr. BMN College of Home Science January 2004 HSC Central Railway's...
Health and fitness	Education Details May 2014 Diploma Nutrition Education Bengaluru, Karnataka IGNOU University June 2004 Bachelor of Science Clinical Nutrition and Dietetics Bengaluru, Karnataka Smt. VHD...
Civil Engineer	Education Details B.E in Civil Engineering Civil Engineering Chennai, Tamil Nadu Anna University QA/QC Civil Engineer QA/QC Civil Engineer Skill Details Company Details company - Ray...

Resume Parsing using PyMuPDF Python Library

TON NGUYEN NGOC DUNG

github.com/ngocdung99 linkedin.com/in/ngocdung99 ngocdung.contact@gmail.com
344, 2/9 street, Da Nang, Viet Nam (915)-768-423

SUMMARY

I graduated Bachelor's program in Computer Science from the University of Science and Technology and have had a solid year of experience in full spectrum of AI Engineer. My strength includes searching, building, and designing the artificial intelligence responsible for machine learning and maintaining and improving existing artificial intelligence systems.

OBJECTIVE

To work for an organization which provides me the opportunity to improve my skills and knowledge to grow along with the organization objective.

EDUCATION

University of Science and Technology, University of Da Nang Sep 2017 - May 2022
Bachelor of Computer Science
GPA: 3.6/4.0

Le Quy Don High School July 2014 - June 2017
Specialization: Physics
Good qualification

SKILLS

Languages: English, Chinese, Korean
Programming: C++, Python, Javascript, SQL, Object Oriented Programming
Technologies: FastAPI, Django, Postgresql, ReactJs, Tensorflow, PyTorch, AWS, Google Cloud
Soft Skills: Presentation, Public Speaking, Creative Problem-Solving, Communication, Time Management

WORK EXPERIENCE

Orient Software Development Corp. Nov 2021 - Present
AI Engineer
- Identify and choose right AI or cognitive computing technologies for solving problems and formulate AI recipes for development.
- Develop required machine learning models or prototype applications applying formulated AI recipes and verify the problem/solution fit.
- Involve in development of AI Platform and AI-nization projects.

Neurond AI Sep 2021 - Nov 2021
Associate Data Engineer
- Create and maintain optimal data pipeline architecture.
- Assemble large, complex data sets that meet functional / non-functional business requirements.
- Identify, design, and implement internal process improvements: automating manual processes, optimizing data delivery, re-designing infrastructure for greater scalability, etc.

PyMuPDF (imported as fitz) is a text extraction library in Python used for working with PDF files.

```
for page in doc:
    text = page.get_text()
    print(text)
```

TON NGUYEN NGOC DUNG

github.com/ngocdung99 linkedin.com/in/ngocdung99 ngocdung.contact@gmail.com
344, 2/9 street, Da Nang, Viet Nam (915)-768-423

SUMMARY

I graduated Bachelor's program in Computer Science from the University of Science and Technology and have had a solid year of experience in full spectrum of AI Engineer. My strength includes searching, building, and designing the artificial intelligence responsible for machine learning and maintaining and improving existing artificial intelligence systems.

OBJECTIVE

To work for an organization which provides me the opportunity to improve my skills and knowledge to grow along with the organization objective.

EDUCATION

University of Science and Technology, University of Da Nang
Sep 2017 - May 2022
Bachelor of Computer Science
GPA: 3.6/4.0

Le Quy Don High School
July 2014 - June 2017

Specialization: Physics
Good qualification

SKILLS

Languages: English, Chinese, Korean
Programming: C++, Python, Javascript, SQL, Object Oriented Programming
Technologies: FastAPI, Django, Postgresql, ReactJs, Tensorflow, PyTorch, AWS, Google Cloud
Soft Skills: Presentation, Public Speaking, Creative Problem-Solving, Communication, Time Management

WORK EXPERIENCE

Orient Software Development Corp.

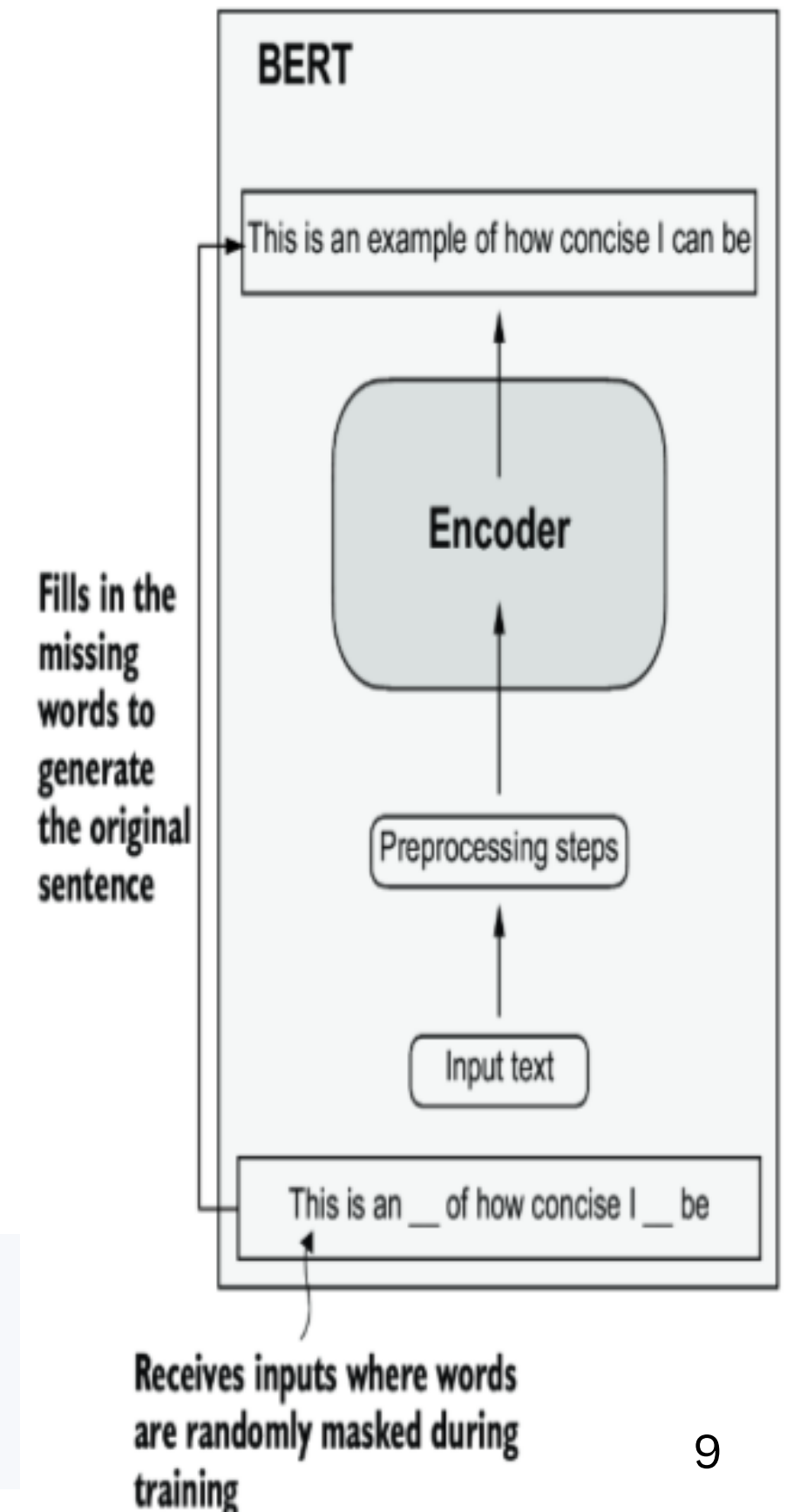
Natural Language Processing Model - BERT

BERT (Bidirectional Encoder Representations from Transformers)

Two stages of **BERT**: **Pre-training** and **Fine-tuning**

- **Pre-trained mode** - A large model (24-layer Transformer, 340 M parameters) on a large corpus (Wikipedia + BookCorpus) for a long time (1M update steps), used **BERT-base-uncased**
- **Fine-tuning** - During training,
 - Feed in the **resume_content** text and its **correct job_role**.
 - BERT fine-tunes its weights and learns to associate certain kinds of resumes with specific job roles.

Input: the man went to the [MASK1] . he bought a [MASK2] of milk.
Labels: [MASK1] = store; [MASK2] = gallon



How BERT is used in Internsloth?

- Reads the resume **bidirectionally** left <-> right, meaning it understands the relationship between skills, experience, and roles no matter where they appear in text.
- Understand the **full context** of the resume — not just predict next words unlike GPT transformers.
- Captures context and synonyms, aligning **resume phrases** with **job-post wording** far better than **TF-IDF** (or) **Word2Vec** (e.g. JS <-> Java Script, ML <-> Machine Learning)
- **Single label classification, Input:** one full resume (as plain text). **Output:** one label (e.g., "Backend Developer", "Data Scientist", etc.).

Job Search API (Adzuna)



- A tool that allows apps (or) websites to search for job listings from all over the internet.
- Use the Predicted role from the BERT NLP model to search the relevant job results using Adzuna API.

API End Point -

[https://api.adzuna.com/v1/api/jobs/us/search/5?](https://api.adzuna.com/v1/api/jobs/us/search/5?app_id=${application_id}&app_key=${key_value}&results_per_page=10&title_only=${predicted_role})

[app_id=\\${application_id}&app_key=\\${key_value}&results_per_page=10&title_only=\\${predicted_role}](https://api.adzuna.com/v1/api/jobs/us/search/5?app_id=${application_id}&app_key=${key_value}&results_per_page=10&title_only=${predicted_role})

Results

- **Small classes** (Advocate = 20) hit **1.0/1.00/1.00** for **precision, recall** and **f-1 score**
 - Reality check: one extra mistake **would drop F1 to 0.95** so the perfect scores still mean **vulnerable**; more data would stabilize them
- **Largest class** (Java Developer = 84) still holds **0.99 F1 score** which means credible and reliable for large examples

	precision	recall	f1-score	support
Advocate	1.00	1.00	1.00	20
Arts	1.00	1.00	1.00	36
Automation Testing	0.95	0.69	0.80	26
Blockchain	1.00	1.00	1.00	40
Business Analyst	1.00	1.00	1.00	28
Civil Engineer	1.00	1.00	1.00	24
Data Science	1.00	1.00	1.00	40
Database	1.00	1.00	1.00	33
DevOps Engineer	1.00	0.96	0.98	55
DotNet Developer	1.00	1.00	1.00	28
ETL Developer	0.98	1.00	0.99	40
Electrical Engineering	1.00	1.00	1.00	30
HR	1.00	1.00	1.00	44
Hadoop	1.00	1.00	1.00	42
Health and fitness	1.00	1.00	1.00	30
Java Developer	0.98	1.00	0.99	84
Mechanical Engineer	0.95	1.00	0.98	40
Network Security Engineer	1.00	1.00	1.00	25
Operations Manager	1.00	1.00	1.00	40
PMO	1.00	1.00	1.00	30
Python Developer	1.00	1.00	1.00	48
SAP Developer	1.00	1.00	1.00	24
Sales	1.00	1.00	1.00	40
Testing	0.95	1.00	0.97	70
Web Designing	1.00	1.00	1.00	45

Evaluation

Metric	Score & Plain Meaning
Accuracy	0.99 — the model gets the right job label on 99 out of 100 résumés.
Macro-average F1	0.99 — even the small, hard-to-find job categories score about 99 %; the model doesn't ignore rare roles.
Weighted-average F1	0.99 — when we give big categories more weight (to mirror real data), performance is still 99 %, so common roles are covered just as well.

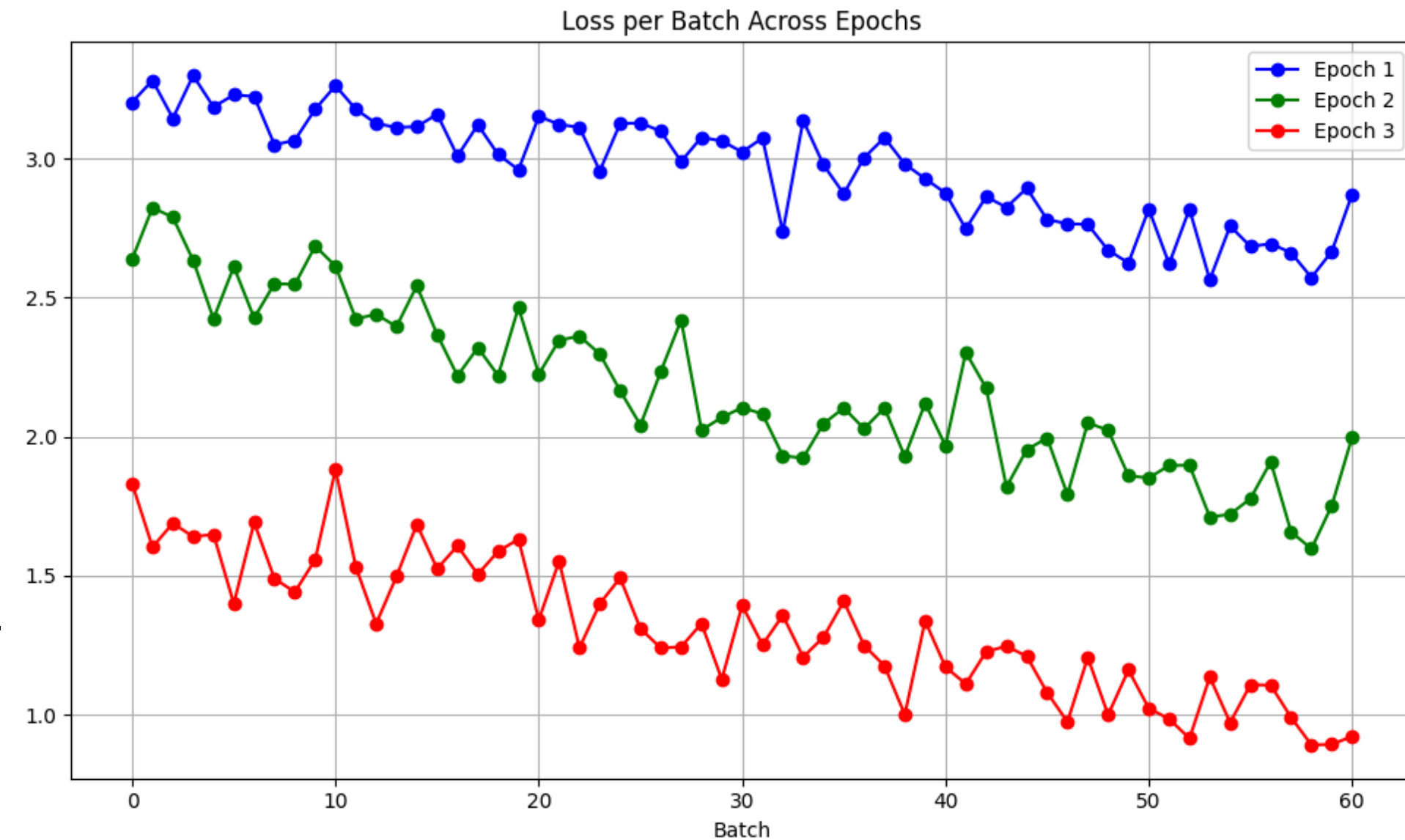
	precision	recall	f1-score	support
accuracy			0.99	962
macro avg	0.99	0.99	0.99	962
weighted avg	0.99	0.99	0.99	962

Evaluation

Loss and Accuracy

- **Loss** - running error score (0 means perfect, higher = more error in accuracy)
- Sharp fluctuations (loss-spikes) - get smaller each epoch, showing the model is learning tricky resume patterns—not just the easy ones
- Curve is flattening out, meaning extra epochs would give tiny gains but risk over-fitting

Three epochs at high accuracy with efficient training - good stopping point

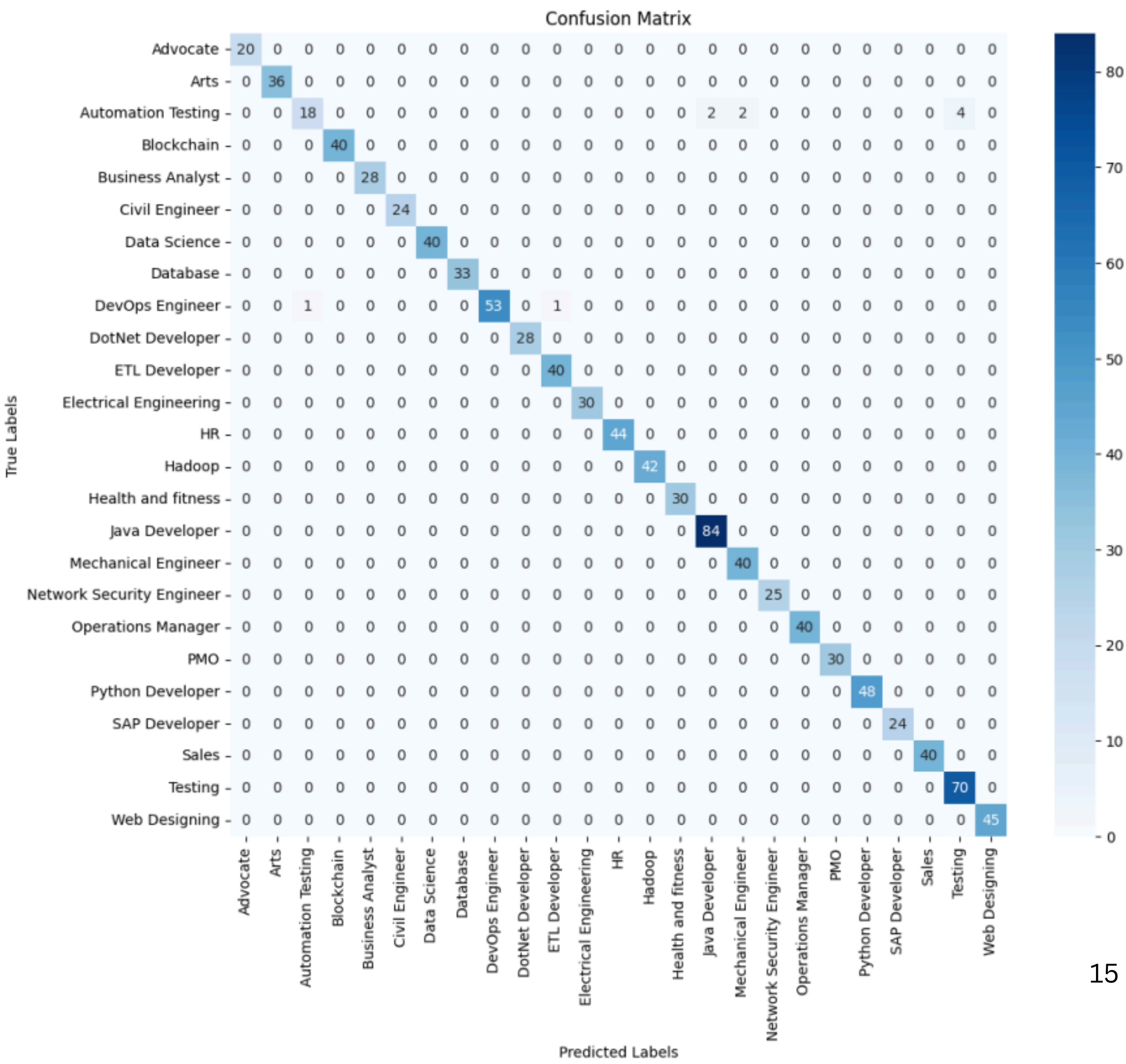


Confusion Matrix

- Automation Testing’s keywords similar to **Java Developer** and **Testing** which are large classes in Dataset
- Model picks the more common label when signals collide

How to Solve?

Automatic Testing - to target with extra data or multi-label training

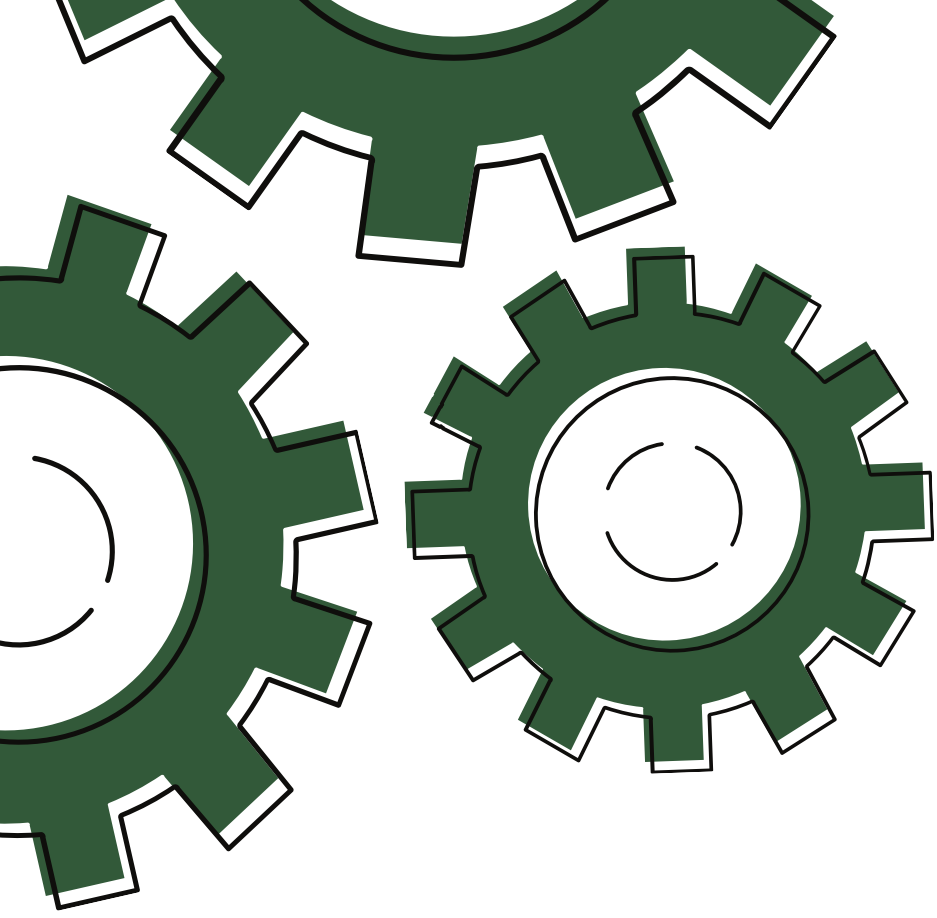


Live Demo



Future Work

- **Automatic Application Submissions:** Without manually applying, the system will apply all the real-time job positions on platforms (Potential **AI ethics** considerations and Data Privacy concerns)
- **Refining as Intelligent Career Assistance System**
 - **Reinforcement Learning** for adaptation to learning advanced features
 - **Web-Scrapping** - gather live internship or job postings and feed them into your matching model
 - Integration of **Personalized Resume Analysis** with Skill Gap Radar and **AI Interview-Coach** with current system



Thank you
Q&A

