ENPM703 Fundamentals of AI and Deep Learning (Fall 2025)

Phishing Email Detection with Multimodal Deep Learning

The Power of Fusion: Detecting Phishing from Every Angle

Group 3
Vishal Patil, Akash Vora, Srihari Narayan, Anila Sai Namburi

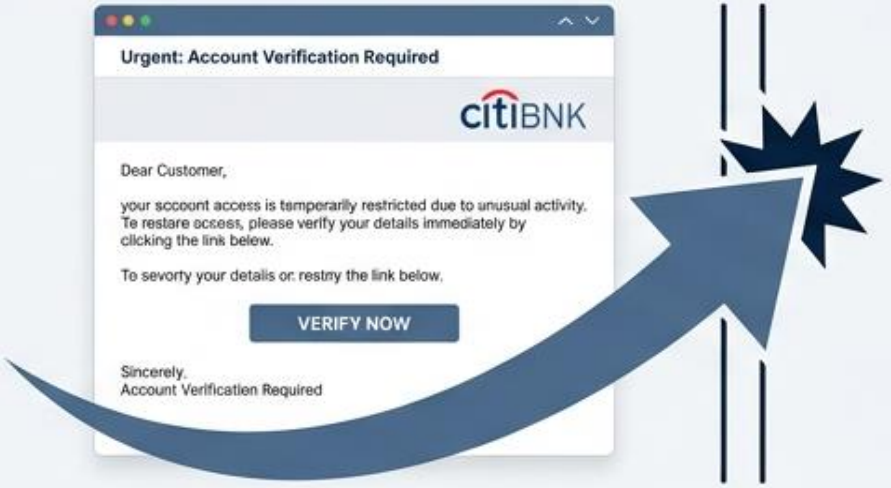
THE PROBLEM: AN EVOLVING, MULTIMODAL THREAT

Responsible for billions in financial losses annually.

Traditional Defenses

Traditional text-only systems fail to analyze the complete picture

Modern Attacks



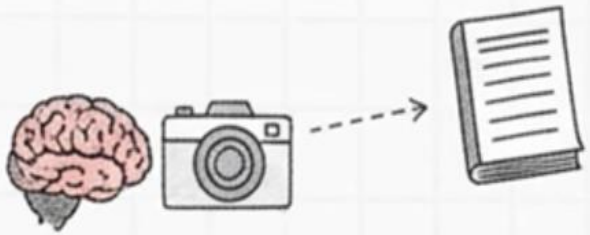
Attackers combine deceptive language with legitimate-looking brand logos.

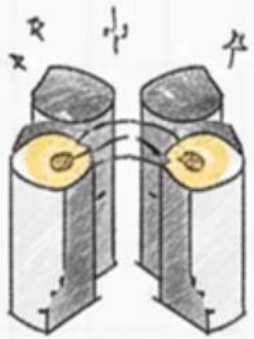
Our Goal: A system that analyzes both what an email *says* and what it *shows*.

Input: Raw email content (text, images, HTML structure).

Output: Binary classification: Phishing or Legitimate.

Four Project Phases





Phase 1

Pre-training the text and image 'specialist' models independently.

Phase 2

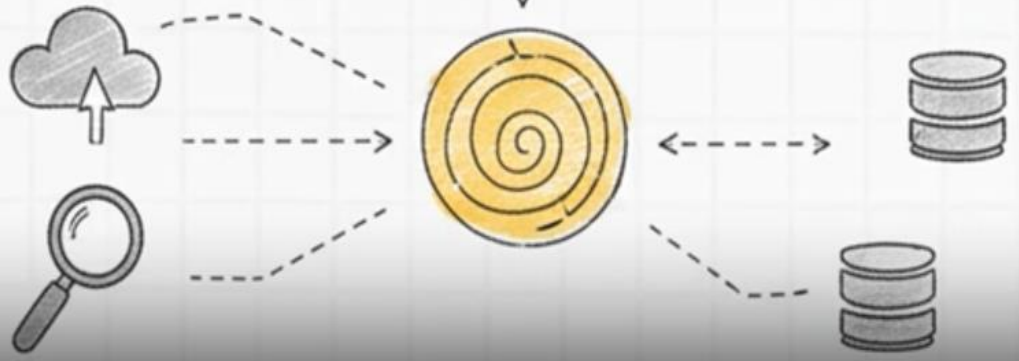
Integrating all data into a single 'golden dataset'.

Phase 3

Training the final dual-tower fusion model with all data types.

Phase 4

Deploying the model for real-time inference on raw HTML emails.



OUR APPROACH VS. RELATED WORK

Previous Works (e.g., Doshi et al.)

- ✓ Content-based features & URL analysis.
- ✗ Modalities treated independently (no true fusion).

Other works focus only on text (Zhang & Wallace's CNNs) or only on images.

Our Approach

- ✓ Deep text analysis via Custom CNN.
- ✓ Deep logo analysis via Custom CNN.

Key Innovation:

A jointly trained, multimodal fusion architecture that learns to dynamically weigh text, image, and metadata signals in a single end-to-end system.

OUR FOUNDATION: A DIVERSE, LARGE-SCALE DATASET

Email Text & Metadata

76,346

Emails

From 4 public sources (CEAS_08, Enron, Nazario, etc.)
Balanced 50/50 Phishing-Legitimate Split

20-dim

Engineered Metadata Vector

Brand Logos

72,652

Images

From OpenLogo Dataset
Resized to 224×224 pixels and normalized.

352

Brand Classes

Loss Function: Cross-Entropy Loss | **Evaluation Metrics:** Accuracy, F1-Score, and ROC-AUC

Single-Modality Detectors Have Clear Limitations

Text-Only Analysis

Custom Text CNN

98.96%

Strong, but can be fooled by carefully crafted language.

Logo-Only Recognition

Custom Image CNN

76.30%

A useful signal, but insufficient on its own.

Building the Foundation for Fusion

1. Train Specialist Models

Text CNN

Trained on 76,346 emails (CEAS_08, Enron).

Image CNN

Trained on 72,652 logos (OpenLogo).

2. Engineer a Unified Dataset

The Golden Dataset

- Integrated text, logo IDs, and labels.
- Engineered 20 metadata features (URL counts, sender patterns, etc.).

A SYSTEMATIC EVALUATION ACROSS SIX ARCHITECTURES

Traditional ML Baselines

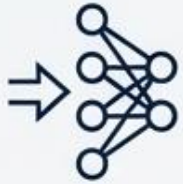
K-Nearest Neighbors: **81.71%**

Logistic Regression: **80.00%**

Unimodal Deep Learning

Custom Text CNN: **98.96%**

Custom Image CNN: **76.30%**



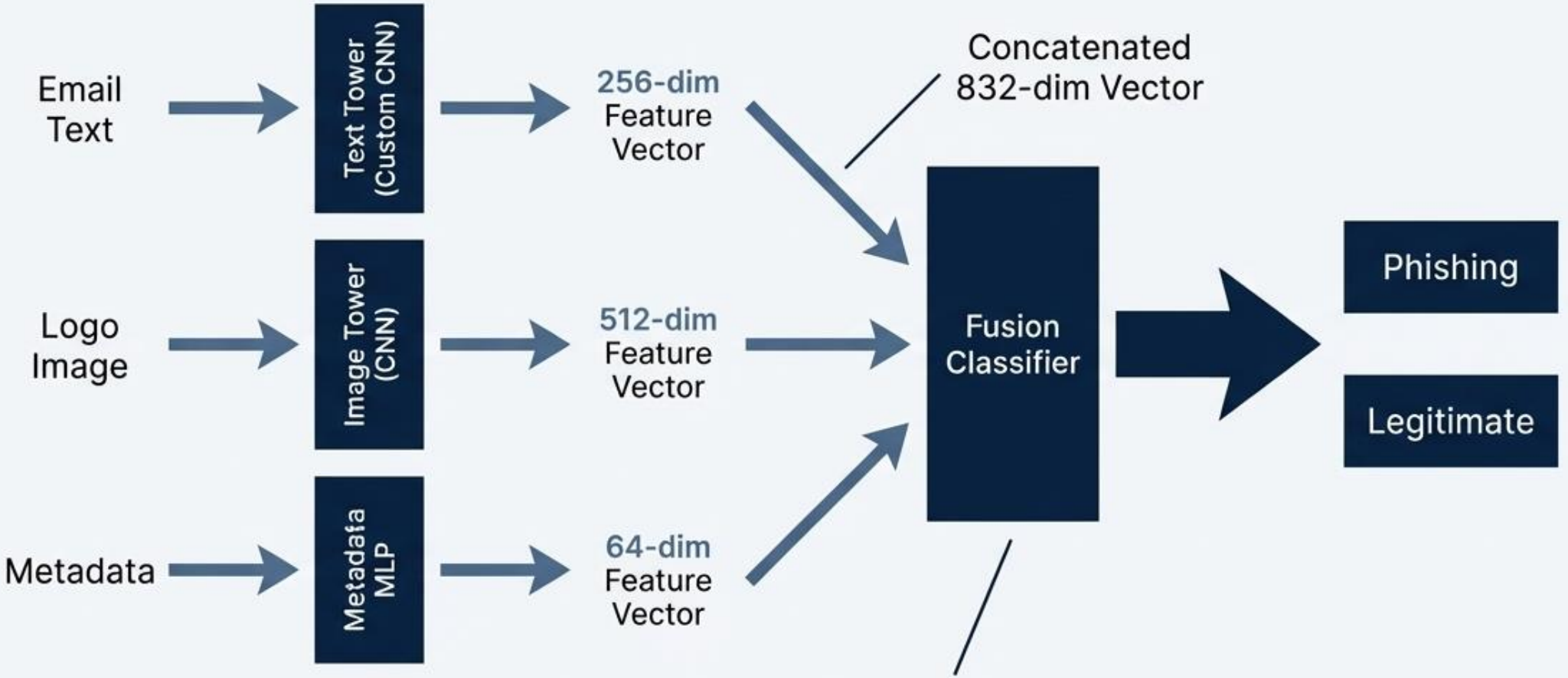
Transfer Learning

Pretrained ResNet18 (Image): **97.43%**

Multimodal Fusion

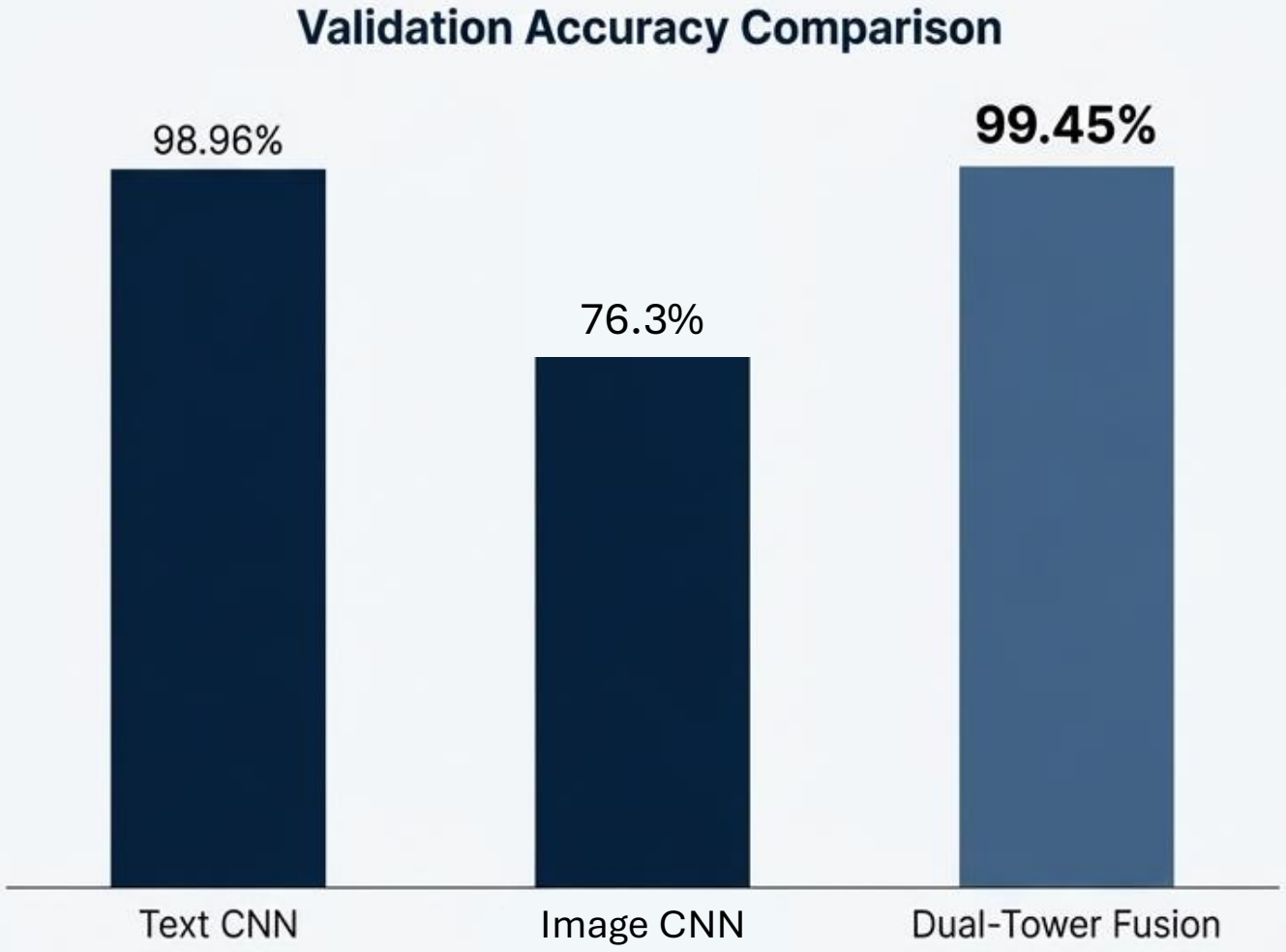
Dual-Tower Fusion Model: 99.45%

THE INNOVATION: A DUAL-TOWER FUSION ARCHITECTURE



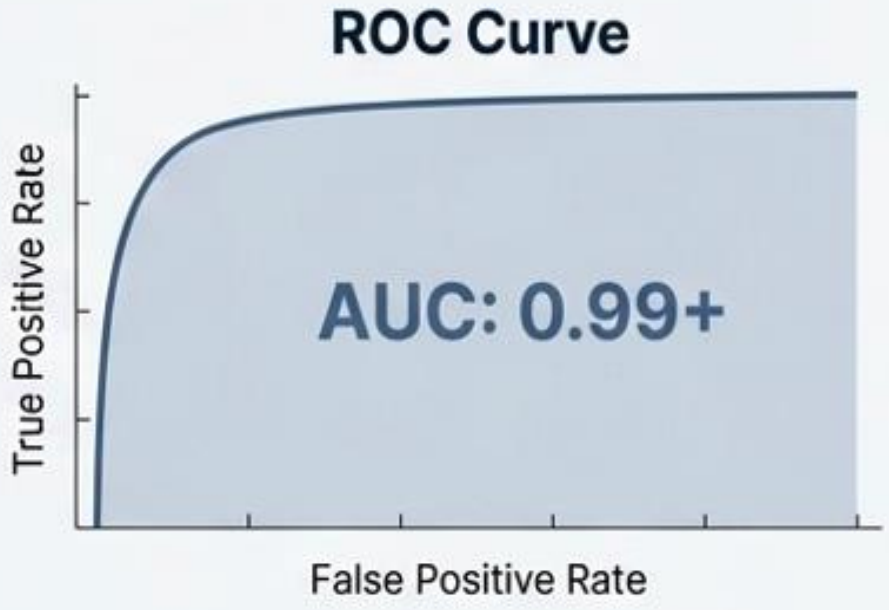
Catches inconsistencies a single-modality approach would miss, such as a mismatched sender domain and brand logo.

THE VERDICT: MULTIMODAL INTEGRATION ACHIEVES SUPERIOR PERFORMANCE



99.45 %

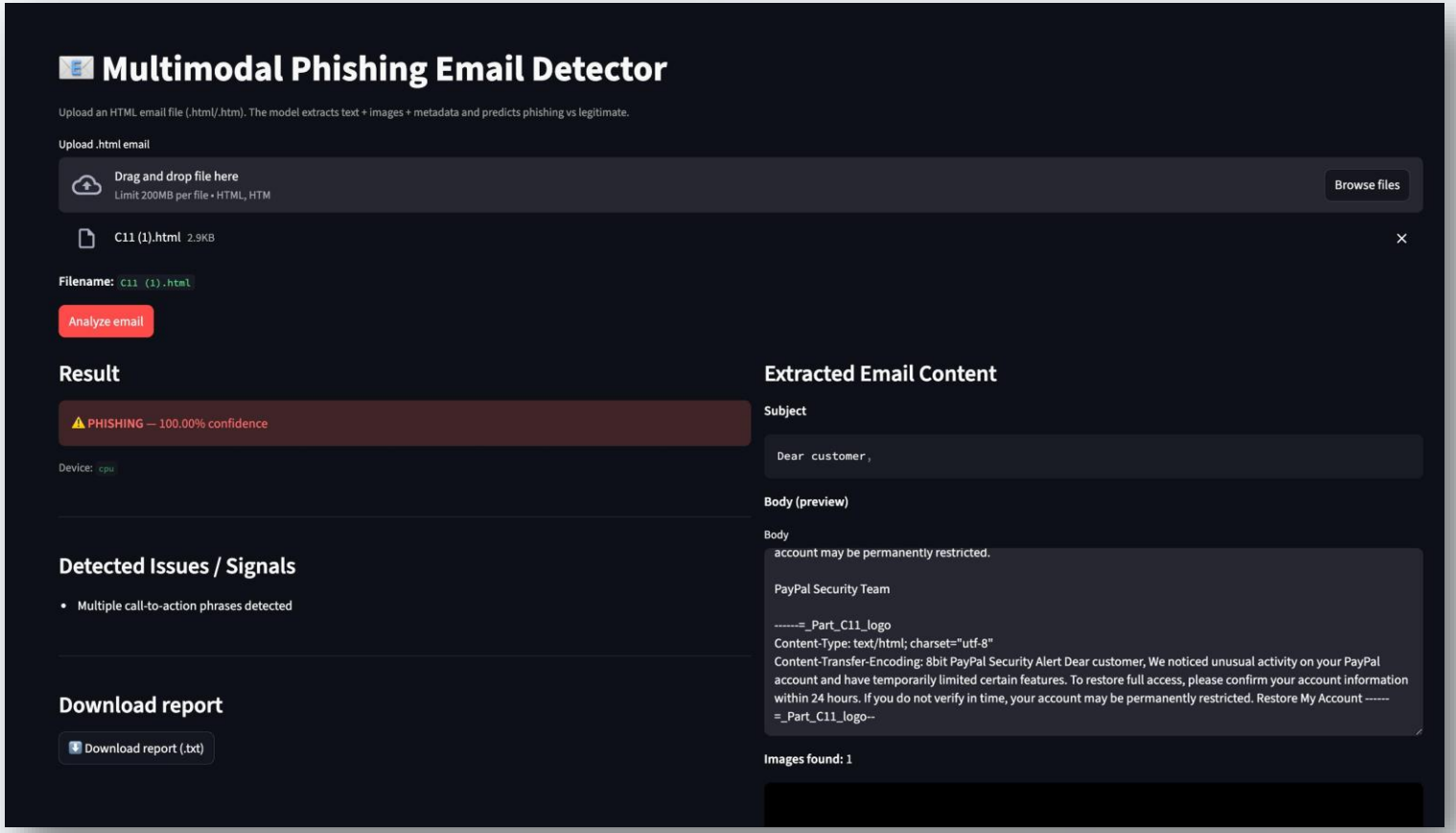
Fusion Model Accuracy



FROM MODEL TO APPLICATION: A REAL-WORLD DEPLOYMENT

The Pipeline

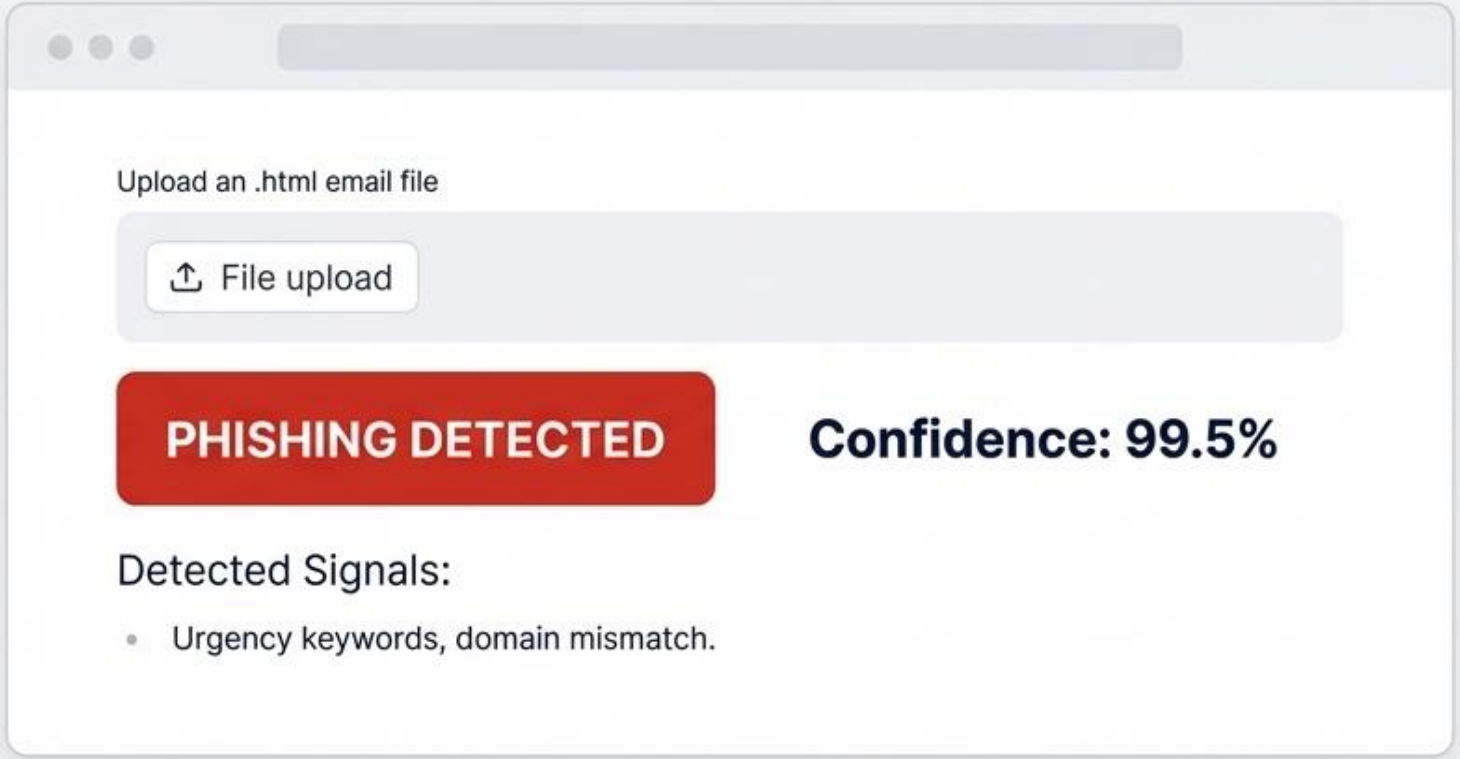
The Application



Successfully identified **100%** of 20 synthetic phishing emails (with and without logos), with **zero** false positives on legitimate test emails.

Live System Demo

https://huggingface.co/spaces/anilawork/phish-detection-ui-final



An end-to-end inference pipeline deployed as an interactive web application on Hugging Face.

IMPACT, LIMITATIONS, AND THE PATH FORWARD

BROADER IMPACT

- Provides an open-source framework for multimodal threat detection.
- Demonstrates a practical, end-to-end deployable system.

LIMITATIONS

- Static datasets may not reflect new zero-day attacks.
- Scarcity of real-world phishing HTML samples for training.
- Class imbalance exists in the logo dataset for less common brands.

FUTURE WORK

- Evaluate the system on a live feed of phishing emails.
- Incorporate advanced text embeddings like BERT.
- Expand classification to distinguish between spam and phishing.

Thank You

Group 3
Vishal Patil, Akash Vora, Srihari Narayan, Anila Sai Namburi

Phishing Email Detection with Multimodal Deep Learning