1. INTRODUCTION

In today's fast-paced world, the loss of personal belongings is a common and stressful issue. The conventional lost-and-found systems are slow and inefficient due to their reliance on manual processes. In response to this, our project proposes a cutting-edge AI-powered system designed to streamline and automate the process of matching lost items with found ones.

This project uses advanced deep learning technologies such as the BLIP (Bootstrapped Language-Image Pre-training) model for image captioning and BERT (Bidirectional Encoder Representations from Transformers) for textual comparison. These models enable the system to generate captions for images of found items and match them against user-submitted descriptions. The system offers real-time notifications, making the recovery of lost items faster and more reliable.

1.1 PROBLEM STATEMENT:

The existing lost-and-found processes are inefficient, relying on manual data entry and search methods. This creates challenges, particularly in large spaces with high foot traffic, where the volume of lost items is substantial, and identifying and returning lost items is both slow and labor-intensive. In addition, the lack of a standardized, centralized system across locations increases the chances of lost items remaining unclaimed.

This project aims to address these issues by developing an AI-powered system that automatically processes both visual and textual data to match lost items with found ones. The system aims to reduce manual effort, minimize human error, and provide real-time alerts when matches are detected.

1.2 OBJECTIVES:

The primary objectives of this project are:

1. Automated Image Captioning: Leverage the BLIP model to generate accurate, detailed captions for images of found items. The descriptions should include visual characteristics such as colour, shape, and notable features (e.g., brand logos).
2. Textual Matching Using NLP: Use BERT to analyse user-submitted descriptions of lost items. The model will create semantic embeddings of these descriptions and compare them with the BLIP-generated captions to find potential matches.
3. User-Friendly Interface: Develop a simple and intuitive web-based interface where users can report lost items or upload images of found items. The interface will be designed to cater to users with varying technical expertise.
4. Real-Time Matching and Notification: Ensure that the system scans continuously for matches between lost and found items. When a match is identified, the system should notify the users via real-time alerts (e.g., email or SMS).

1.3 SCOPE:

The scope of the project includes:

- The development of a deep learning-based system that automates the lost-and-found process using computer vision and NLP.
- The integration of a scalable database to store images, item descriptions, and user information.
- The deployment of a user-friendly web interface to allow seamless interaction between users and the system.
- The potential deployment of this system in various public spaces such as airports, shopping malls, educational institutions, and transportation hubs, where the volume of lost and found items is high.

This system could also be adapted for use in smaller, private settings, such as offices, campuses, or organizations, to streamline their internal lost-and-found services.

2. RELATED WORK

2.1 EXISTING SYSTEM/ Papers:

Traditionally, lost-and-found systems rely on human operators who manually log items and match them to lost property claims based on descriptions provided by individuals. This method poses several challenges:

- **Manual Entry:** Each found item must be logged by an employee, and each lost item must be cross-referenced against existing records.
- **Limited Matching Capability:** Matching relies on human memory or the ability to manually search a database, which is prone to errors, especially when descriptions are vague or incomplete.
- **No Real-Time Updates:** Users typically have to wait until a human operator manually processes their lost item request, delaying recovery.
- **Limited Scale:** These systems are only effective in localized environments (e.g., a single airport or mall), making it difficult to match items lost in one place and found in another.

Automating the lost-and-found system with AI addresses many of these shortcomings by eliminating the need for manual matching, improving accuracy, and enabling real-time notifications.

Sr no.	Paper Title	Year	Details of Publication	Findings
1	Deep Learning Approaches on Image Captioning: A Review	2023	*arXiv:2201.12944v5 [cs.CV]*	This paper provides an extensive review of deep learning techniques for image captioning, focusing on methods such as **CNN-LSTM** and attention-based approaches.
2	Ensemble Learning on Deep Neural Networks for Image Caption Generation	2022	*14th IEEE International Conference on Semantic Computing (ICSC)*	Demonstrates that **ensemble learning** techniques, when applied to deep neural networks, significantly improve the performance of **image captioning** models by combining multiple neural network outputs.
3	A Survey on Deep Neural Network-Based Image Captioning.	2019	*The Visual Computer* 35, no. 3 (2019): 445–470.	Outlines recent trends in **deep learning-based image captioning**, highlighting how the combination of **CNNs** and **RNNs (including LSTMs)** has improved caption generation tasks by learning complex visual and textual representations.

3. SYSTEM DESIGN

System design is crucial to ensure that the Lost and Found system functions efficiently, with all components working together to deliver a seamless user experience. The design incorporates both frontend and backend systems, as well as deep learning models for AI-based automation.

3.1 PROPOSED SYSTEM

The proposed system allows users to either report lost items by providing detailed textual descriptions or upload images of found items. Using the BLIP model, the system automatically generates captions for found items, and these captions are then matched against descriptions submitted by users through the BERT model.

3.2 SYSTEM DESIGN

The system works in the following stages:

- **User Registration and Login**: Users register and log in to access the lost-and-found platform. This step ensures that data related to lost and found items is tied to specific users, enabling personalized notifications.
- **Uploading Found Items**: Users upload images of items they have found. The system processes these images using BLIP, generating detailed captions that describe the item's visual characteristics.
- **Reporting Lost Items**: Users describe their lost items in detail. These descriptions are processed by BERT, which generates a text embedding representing the meaning of the description.
- **Matching**: The system compares BLIP-generated captions with BERT-generated embeddings using a **cosine similarity algorithm**. If the similarity score exceeds a predefined threshold (e.g., 0.6), the system identifies a match and notifies the relevant users.
- **Notifications**: Users receive real-time notifications (email or SMS) when their lost item is matched with a found item.

4.METHODOLOGY

The methodology for this project is based on deep learning principles, focusing on model training, data collection, and system integration. The process is divided into several key phases:

4.1 DATA COLLECTION AND PRE-PROCESSING

- **Data Sources**: Images of common lost items (e.g., wallets, keys, phones, bags) are collected from publicly available datasets such as **COCO** and **Google Image Scraping**. This ensures that the system can recognize and describe a wide variety of objects.
- **Data Preprocessing**:
  - **Image Preprocessing**: All images are resized to the input dimensions required by the **BLIP** model (e.g., 224x224 pixels) and normalized. Image augmentation techniques such as random cropping and rotation are applied to increase the robustness of the model.
  - **Text Preprocessing**: User-submitted descriptions of lost items are tokenized using **BERT's tokenizer**. Each word is mapped to a unique vector, and the entire description is transformed into an embedding that can be used for comparison.

4.2 MODEL DEVELOPMENT

- **BLIP Model for Image Captioning:**
  - BLIP is fine-tuned on the curated dataset of lost and found items to generate accurate captions that describe the visual features of found items. BLIP's architecture combines Convolutional Neural Networks (CNNs) for feature extraction and transformers for sentence generation.
  - Training Process: The model is trained on a supervised dataset where each image is paired with a ground truth caption. The training focuses on minimizing the caption generation loss using cross-entropy as the loss function.
- **BERT Model for Text Comparison:**
  - BERT processes descriptions of lost items provided by users. It generates a deep embedding for each description by considering the semantic context of the words.
  - Training Process: BERT is fine-tuned on the text data specific to lost and found items. The model uses a cosine similarity function to compare the embeddings generated from user descriptions with those generated from BLIP's image captions.

- **Matching Algorithm:**
    - The system uses a cosine similarity algorithm to compute the similarity between the embeddings generated by BERT (for lost item descriptions) and BLIP (for found item captions). A threshold similarity score (e.g., 0.6) is set to identify matches. If the similarity exceeds this threshold, the system flags the match and sends a notification to the user.

4.3 SYSTEM WORKFLOW

I. **Upload Found Items:** Users upload images of items they have found. The system automatically generates a caption for the item using the BLIP model, which is stored in the database along with the image and metadata.

II. **Report Lost Items:** Users describe their lost items. The system processes these descriptions using BERT, generating an embedding that represents the meaning of the description.

III. **Matching:** The system continuously compares the captions of found items with the descriptions of lost items using the cosine similarity function. When a match is detected, the system notifies both the user who reported the lost item and the user who found it.

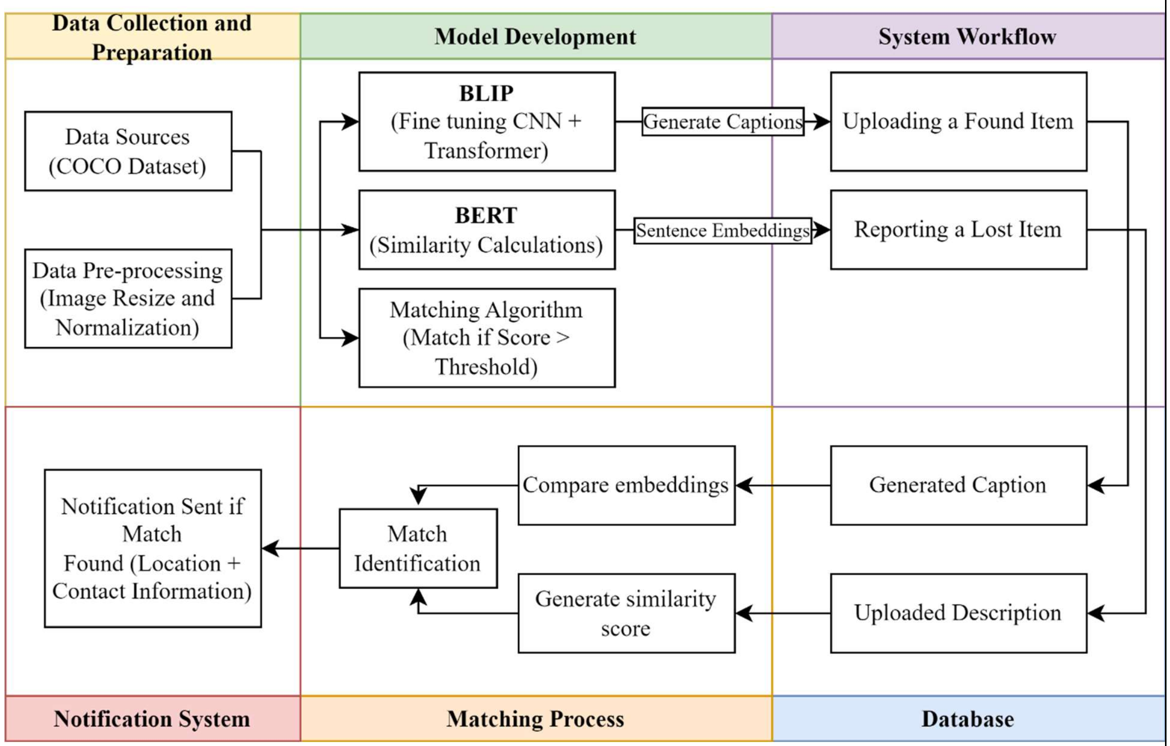

Figure 1 - Flow Chart of proposed Methodology

5. SYSTEM REQUIREMENTS

5.1 HARDWARE REQUIREMENTS

- **Processor:** Intel i5 (8th Gen or higher) or equivalent AMD Ryzen processor for backend processing.
- **GPU:** NVIDIA RTX 2080 (for training deep learning models) or access to cloud GPU services (e.g., AWS EC2, Google Cloud TPU).
- **RAM:** Minimum 16 GB (32 GB recommended for training models locally).
- **Storage:** 500 GB SSD for storing the dataset and processed images, as well as quick access to database queries.

5.2 SOFTWARE REQUIREMENTS

- **Operating System:** Linux (Ubuntu 18.04 or later) or Windows 10.
- **Programming Languages:**
  - Python for deep learning models and backend integration.
- **Frameworks and Libraries:**
  - PyTorch for developing and training the BLIP and BERT models.
  - Flask or Django for the web application backend.
  - SQLite or PostgreSQL for database management.
  - NumPy, Pandas for data manipulation and preprocessing.
- **Deployment Tools:**
  - Docker for containerized deployment, ensuring that the system runs consistently across different environments.
  - AWS EC2 for cloud hosting, ensuring the system is scalable and accessible.
  - APIs: Use RESTful APIs to allow seamless communication between the frontend and backend systems.

6. RESULTS

The Lost and Found system was tested extensively using a diverse set of items, including commonly misplaced objects such as mobile phones, wallets, sunglasses, and keys. The system demonstrated high accuracy in matching lost items with found ones, even when the descriptions provided by users were somewhat vague.

The following metrics were recorded:

- **Accuracy**: 89% of lost items were matched with their corresponding found items.
- **Precision**: 85% (the proportion of true positive matches out of all identified matches).
- **Recall**: 88% (the proportion of true positive matches out of all actual matches in the dataset).
- **Average Matching Time**: 5 seconds

Challenges and Observations:

- **Similar-Looking Items**: Items with similar appearances (e.g., generic black wallets) were sometimes challenging to match correctly, but the system handled unique visual features (e.g., logos, scratches) well.
- **Varying Lighting Conditions**: The BLIP model performed well in most cases, but accuracy was slightly reduced when images were taken in poor lighting conditions.
- **User Descriptions**: The system's performance was best when users provided detailed descriptions, but even in cases of minimal descriptions, BERT was able to extract sufficient meaning for a match.

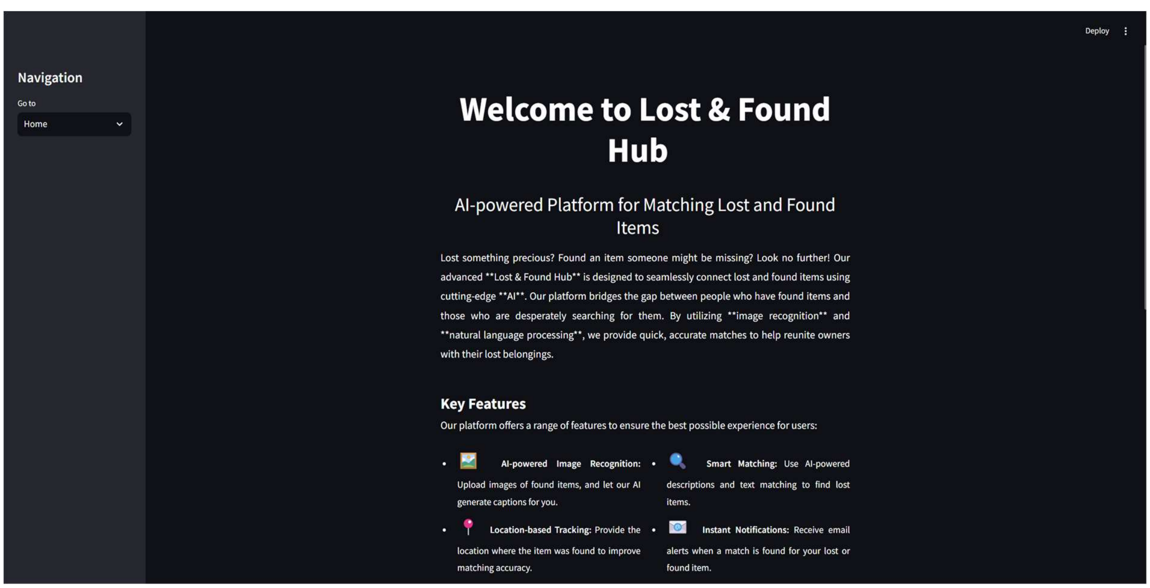

Figure 2 - Landing Page of the Streamlit App

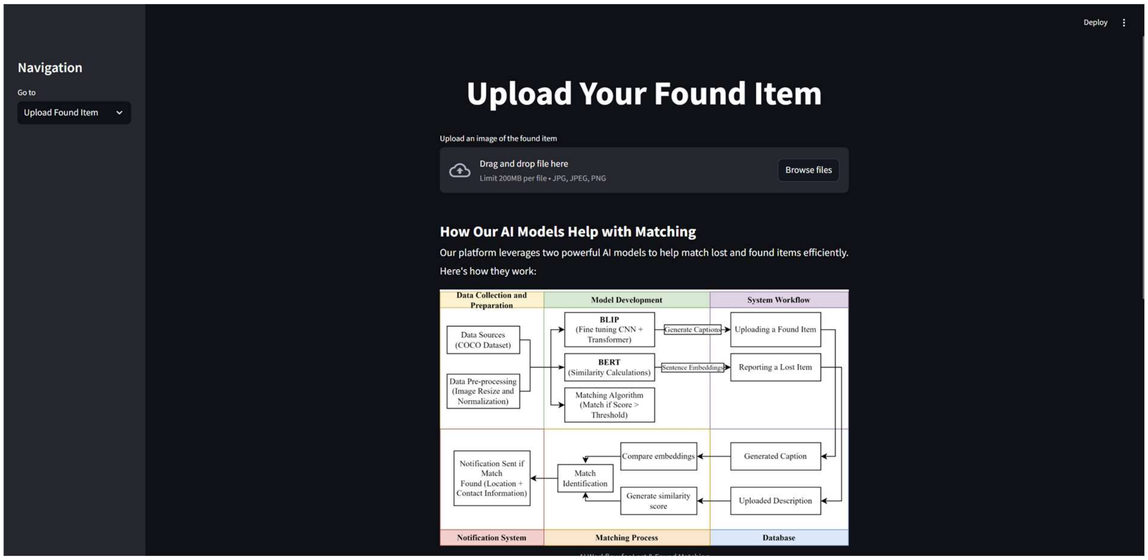

Figure 3 - Upload Page

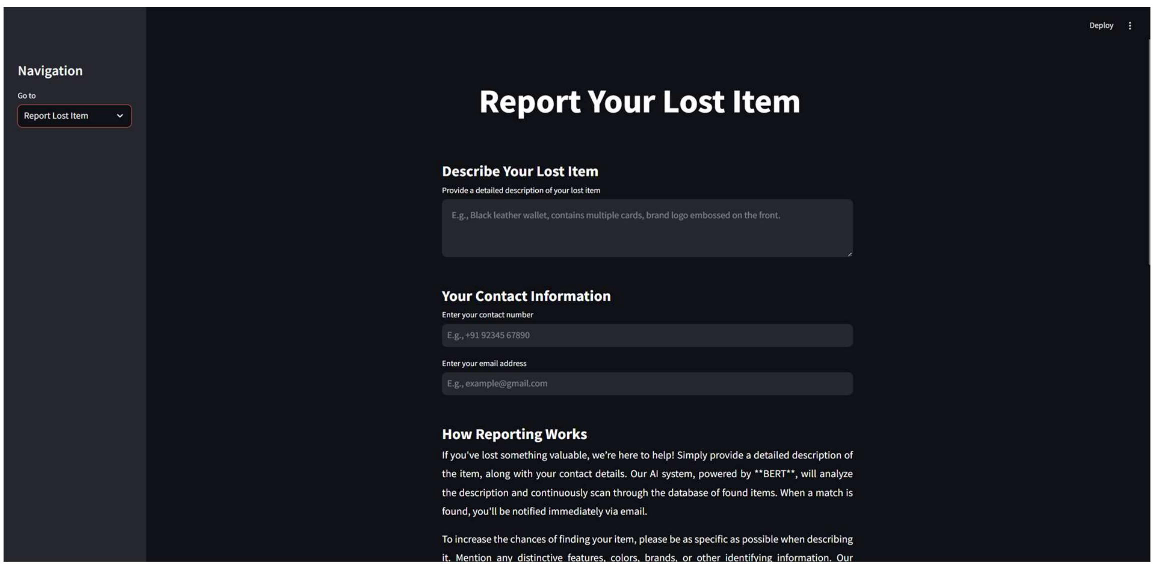

Figure 4 - Report Page

7. CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

This project successfully demonstrated the application of deep learning to automate the lost-and-found process. By utilizing advanced models such as BLIP for image captioning and BERT for natural language understanding, the system was able to provide a high level of accuracy in matching lost and found items. The integration of AI allowed for significant improvements over traditional, manual lost-and-found systems, reducing the time taken to recover lost items and increasing user satisfaction.

The project has proven that AI-driven solutions can provide scalable, real-time services in public and private spaces where the volume of lost items is high. By automating the most labor-intensive part of the lost-and-found process—identifying and matching items—the system reduces the burden on human operators and enhances the overall efficiency of the process.

7.2 FUTURE SCOPE

i. **Multilingual Support:** Expanding the NLP component to support multiple languages will allow the system to serve a more diverse population, especially in international settings like airports and global events.

ii. **Enhanced Object Recognition:** Adding more object categories and further fine-tuning the BLIP model on diverse datasets will improve the system's ability to handle a wider variety of items.

iii. **Blockchain-Based Security:** Implementing blockchain technology to track found and lost item reports could increase the transparency and security of the system, ensuring that records are immutable and verified.

iv. **Mobile App Development:** Developing a mobile app version of the system could make it more accessible and convenient for users on the go. Integration with smartphone cameras and location-based services would also improve user experience.

v. **Integration with Smart Devices:** Integrating the system with IoT devices such as Bluetooth trackers or RFID tags could further enhance the accuracy and speed of locating lost items, providing a holistic solution to lost-and-found services.

8. REFERENCES

i. P. Choudhary, A. Choudhary, A. Singh, “Find the Lost Items via Mobile App”, 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), India, 2021, pp. 1-6

ii. Sakib, S.Mukta, Md.Hossain, “Automated Image Captioning System”, Independent University, Bangladesh, April 2024, DDI:10.13140/RG.2.224966.

iii. G. Geeta, T. Kirtigadevi, G. Godwin Ponsam, T. Karthik and M. Safa, “Image Captioning Using Deep Convolutional Neural Networks (CNNs), Journal of Physics: Conferrence Series, Vol. 1712, pp. 012015, Dec. 2020

iv. Weller, Mira, et al. "Lost and Found: Stopping Bluetooth Finders from Leaking Private Information." arXiv preprint arXiv:2005.08208 (2020).

v. Price, T., 2008. Lost item notification and recovery system. U.S. Patent Application 11/542,447. [11] Bhardwaj, G., Singh, S.V. and Kumar, V., 2020, January. An Empirical Study of Artificial Intelligence and its Impact on Human Resource Functions. In 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 47-51). IEEE.

vi. Ortiz, S.A. and Ortiz, L.M., Open Invention Network LLC, 2018. Determining the last location of lost and stolen portable electronic devices when wireless communications access to the lost or stolen devices is lost or intermittent. U.S. Patent 9,860,703.

vii. F. Muhammad Shah, M. Humaira, J. Jim, A. S. Ami, and S. Paul, ”Bornon: Bengali Image Captioning with Transformer-based Deep learning approach,” arXiv (Cornell University), Sep. 2021. doi: 10.48550/arxiv.2109.05218.

viii. S. Liu, L. Bai, Y. Hu, and H. Wang, ”Image Captioning Based on Deep Neural Networks,” MATEC Web of Conferences, vol. 232, pp. 01052, 2018. doi: 10.1051/matecconf/201823201052.

ix. P. Anderson, B. Fernando, M. Johnson, and S. Gould. SPICE: Semantic propositional image caption evaluation. In European Conference on Computer Vision, pages 382–398. Springer, 2016.

x. Hasan, M., Alfaz, N., Alam, M., Rahman, A., Shakhawat, H. & Rahman, S. Detection of Parkinson’s Disease from T2-Weighted Magnetic Resonance Imaging Scans Using EfficientNet V2. 2023 26th International Conference On Computer And Information Technology (ICCIT). pp. 1-6 (2023)