**Key Components of the Script**

1. **Data Loading**:
   * The script loads two datasets: emails\_normal and emails\_phishing. These datasets contain email messages that the script will process to extract various features.
   * It uses pandas, a powerful data manipulation library in Python, to handle these datasets.
2. **URL Sanitization Function (sanitize\_url)**:
   * This function cleans URLs found within the email texts. It removes trailing characters such as periods, brackets, and commas that could corrupt URL parsing.
   * It also removes square brackets around non-IP address sections of the URL, ensuring that only valid URLs are processed.
3. **Feature Extraction Function (extract\_advanced\_features)**:
   * Extracts several advanced features from the email content:
     + **URLs**: Counts the number of URLs.
     + **Domains**: Extracts and counts unique domains, checking if they are IP addresses.
     + **HTML tags**: Checks for the presence of HTML tags within the content.
     + **JavaScript code**: Detects embedded JavaScript within the emails.
   * Uses regular expressions and the urlparse function from the Python standard library to parse and validate URLs.
4. **Email Text Processing (process\_email\_text)**:
   * Extracts the subject of the email using a regular expression.
   * Cleans the email body by removing headers and footers specified by the pattern "Message-ID:" to "X-FileName:".
5. **General Feature Extraction (extract\_features\_from\_message)**:
   * Calls the process\_email\_text function to preprocess the email.
   * Extracts basic features such as the length of the email, the number of links, special characters, and the presence of certain keywords.
   * Combines these basic features with the advanced features extracted from the email content.
6. **Application of Feature Extraction**:
   * Applies the feature extraction to both the normal and phishing email datasets.
   * Processes each email and stores the extracted features in new columns within the original data frames.
7. **Data Conversion and Storage**:
   * Converts the lists of features into pandas DataFrame objects for easy manipulation and analysis.
   * Saves the resulting DataFrames to CSV files for potential use in machine learning models or further analysis.
8. **Output**:
   * Prints confirmation messages to the console indicating that the features have been successfully saved to CSV files.

**Summary**

This script is a comprehensive tool for processing email data to extract meaningful features that can help in identifying phishing attempts. It utilizes Python’s text processing capabilities to analyze and extract data from structured email formats. The script is well-structured and modular, making it adaptable to various types of email analysis tasks, especially in cybersecurity contexts.

4