Chapter 3: Deliverables and Evaluation

**Introduction**

This chapter outlines the deliverables produced during the development of the AI-generated text ,image , audio detection system and describes the methods used to evaluate its effectiveness. The chapter includes sections on the user manual, testing procedures, and the evaluation through user experiments. These deliverables and evaluations are critical in ensuring that the system is user-friendly, reliable, and accurate in distinguishing between human-written and AI-generated text.

**User Manual for "Catch the AI" Website**

**Introduction**

The user manual is a crucial component of our project, serving as a comprehensive guide for users to navigate the AI detection platform. It will provide step-by-step instructions on utilizing various features such as user authentication, media submission, subscription management, and profile management. Our goal is to ensure that users can effectively utilize the platform to detect AI-generated content and differentiate it from human-created content.

**Getting Started**

**Requirements**

* **Internet Browser:** Latest versions of Google Chrome, Mozilla Firefox, Safari, or Microsoft Edge.
* **Internet Connection:** A stable internet connection is required for accessing the website and performing analyses.

**Accessing the Website**

1. **Open your browser:** Launch your preferred web browser.
2. **Enter URL:** Go to the website by typing https://www.catchtheai.com in the address bar.
3. **Homepage:** You will land on the homepage, which provides an overview of the platform and its features.

**User Authentication**

**Sign Up**

1. **Navigate to Sign Up:** Click on the "Sign Up" button on the top right corner of the homepage.
2. **Fill in Details:** Enter your name, email address, and create a password.
3. **Submit:** Click the "Create Account" button to register.

**Login**

1. **Navigate to Login:** Click on the "Login" button on the top right corner of the homepage.
2. **Enter Credentials:** Enter your registered email address and password.
3. **Submit:** Click the "Login" button to access your account.

**Password Recovery**

1. **Navigate to Login Page:** Click on the "Login" button.
2. **Forgot Password:** Click on the "Forgot Password?" link.
3. **Enter Email:** Provide your registered email address.
4. **Receive Email:** Check your email for a password reset link.
5. **Reset Password:** Follow the instructions in the email to reset your password.

**Media Submission**

**Supported File Formats**

* **Text:** .txt, .docx, .pdf
* **Images:** .jpg, .png, .bmp
* **Audio:** .mp3, .wav, .ogg

**File Size Limitations**

* Maximum file size: 50MB for text, images, and audio.

**Upload Procedure**

1. **Navigate to Upload Section:** Click on the "Upload" tab in the main navigation bar.
2. **Select Media Type:** Choose the type of media (text, image, or audio) you wish to upload.
3. **Upload File:**
   * Click the "Choose File" button and select the file from your device.
   * Click the "Upload" button to start the upload process.
4. **Start Analysis:** Once the file is uploaded, click the "Analyze" button to begin the analysis.

**Subscription Management**

**Viewing Subscription Plans**

1. **Navigate to Subscription:** Click on the "Subscription" tab in the main navigation bar.
2. **View Plans:** Browse through the available subscription plans.

**Managing Payment Methods**

1. **Payment Information:** Go to your profile and click on "Manage Payment Methods."
2. **Add/Update Payment:** Enter or update your payment details and save the changes.

**Changing Subscription Plan**

1. **Current Plan:** View your current subscription plan under the "Subscription" tab.
2. **Upgrade/Downgrade:** Select a new plan and follow the prompts to change your subscription.

**Accessing Billing History**

1. **Billing History:** Click on "Billing History" in your profile settings.
2. **Review Statements:** View and download past billing statements.

**Profile Management**

**Viewing and Editing Profile**

1. **Profile Page:** Click on your username in the top right corner and select "Profile."
2. **Edit Information:** Update your personal information such as name, email, and password.

**Notification Settings**

1. **Navigate to Settings:** Go to the "Settings" section in your profile.
2. **Update Preferences:** Customize your notification preferences for updates and alerts.

**Deactivating Account**

1. **Account Settings:** In your profile, go to "Account Settings."
2. **Deactivate/Delete:** Follow the instructions to temporarily deactivate or permanently delete your account.

**Interpreting Results**

**Output Metrics**

* **Confidence Scores:** Indicates the likelihood that the content is AI-generated.
* **Classification Labels:** Identifies whether the content is human-generated or AI-generated.

**Visual Aids**

* **Confusion Matrix:** Displays the performance of the model in a tabular format, showing true positives, false positives, true negatives, and false negatives.
* **Classification Report:** Provides precision, recall, and F1 score for the analysis.

**Example Outputs**

* **Text:** Displays the analyzed text with highlighted sections indicating AI-generated parts.
* **Images:** Shows the analyzed image with markers on suspected AI-generated regions.
* **Audio:** Lists segments of the audio flagged as AI-generated.

**Troubleshooting**

**Common Issues**

* **Data Upload Errors:** Ensure the file format and size are within the supported limits.
* **Analysis Failures:** Check if all required dependencies are installed and the model is correctly configured.

**Error Messages**

* **File Not Supported:** Ensure your file type is supported and try re-uploading.
* **Analysis Timeout:** The server might be busy; try analyzing a smaller file or wait for a few minutes and try again.

**Support**

* **Contact Information:** For further assistance, contact support at ahmed01094652515@gmail.com.
* **Documentation:** Refer to the project documentation available on the website for more detailed technical information.

**Conclusion**

This user manual provides all necessary instructions for accessing, using, and troubleshooting the "Catch the AI" website. By following this guide, users can effectively utilize the platform to detect AI-generated content across various media types. For additional support, refer to the troubleshooting section or contact technical support.

**Model Training and Evaluation**

Several models were trained and evaluated, including BERT, DistilBERT, RoBERTa, and DeBERTa:

* **Training Process:**
  + **Parameters:** Models were trained for 2 epochs, with evaluation every 100 steps, a batch size of 128, using the AdamW optimizer with a learning rate of 3e-5, and a cosine learning rate scheduler.
  + **Reporting:** Training progress was tracked using Weights & Biases (wandb).
* **Data Preparation:**
  + **Tokenization:** Text data was tokenized using the respective model tokenizers, with a maximum token length of 512.

**Performance Metrics**

Each model’s performance was evaluated using classification reports and confusion matrices, focusing on key metrics such as accuracy, precision, recall, and F1 score.

* **BERT:**

**A screenshot of a computer screen

Description automatically generated**

A screenshot of a graph

Description automatically generated

* **DistilBERT:**

A screenshot of a computer screen

Description automatically generated

A screenshot of a graph

Description automatically generated

* **RoBERTa:**

A screenshot of a computer

Description automatically generated

A blue squares with white text

Description automatically generated

* **DeBERTa:**

A screenshot of a computer screen

Description automatically generated

A blue squares with white text

Description automatically generated

**Evaluation (User Experiment)**

User experiments were conducted to assess the system’s effectiveness and usability:

* **Objective:** Evaluate the system’s accuracy in distinguishing between human-written and AI-generated text and assess user satisfaction.
* **Participants:** A diverse group of users, including technical and non-technical individuals.
* **Procedure:** Participants used the system to analyze a predefined set of texts and provided feedback on its ease of use, accuracy, and any issues encountered.
* **Metrics:** Key performance indicators included precision, recall, F1 score, and user satisfaction ratings, along with qualitative feedback to identify areas for improvement.

**Summary**

This chapter has detailed the deliverables of the AI-generated text , image , audio detection project, including the comprehensive user manual, rigorous testing procedures, and thorough evaluation through user experiments. These elements collectively ensure the system is reliable, accurate, and user-friendly, providing valuable insights into its performance and guiding future enhancements.