Team 7: AI-generated Content/Fake News Detector

Names: Aryan Rakshit, Dhiya Pereira, Jack White, Seung Hyeon (Leah) Lee, Raj Penmetcha, Seth DeWhitt

Problem Statement

AI is very powerful, but that comes as a double-edged sword. With few official regulations on its use, social media users of all age groups regularly struggle to identify AI-generated content. Most pre-existing AI-detecting tools function as a "copy-paste" system, relying on users to go out of their way to tab out of their social media platform in order to verify if text/images are AI-generated. In contrast, our extension offers in-site identification by providing automatic identification in the context of a feed, timeline, or scroll.

Background Information

(a) Explain background information about the problem, the domain, and targeted users.
(b) Mention whether there are any applications or systems that are similar to your planned work.
(c) Discuss the limitations of other solutions and how you address each limitation.

People of all ages are being fooled on social media platforms by AI-generated content. There exist many web platforms that can detect AI-generated images and/or text, like GPTZero, Winston AI, or CopyLeaks for example, but they require users to manually upload or paste content. Our project seeks to streamline this process to make it user-friendly via a browser extension that can automatically ingest social media content and determine its legitimacy.

Requirements

(a) Divide this section into two subsections, "Functional" and "Non-Functional".
(b) Format your user stories properly (e.g. "As a , I would like to so that .") and you may mark some of the user stories with "(if time allows)" tag, when appropriate. Provide as much detail for each user story as you can right now.
(c) In the Non-Functional section, please include performance requirements such as response time, scalability, usability, security, and etc. and quantify your requirements, e.g. within 500ms, 24 hours per day, 10,000 simultaneous requests, and etc. This can be either in a "user story" format with detailed explanations or just a discussion for each requirement.
(d) Create as many user stories as you can. Even if there is not enough time to finish all of them, it is better to have too many user stories than too few.
(e) All user stories but the ones marked as "(if time allows)" should contain enough work for this semester so that each team member will spend around 10 hours/week for the project. If not, points will be deducted and the team will be asked to resubmit the backlog with adequate amount of work.

Functional Requirements

1. As a general user, I would like to scroll through my social media feed quickly without noticeable performance drops due to the extension so that the extension serves as an enhancement with no drawbacks.
2. As a general user, I would like detection results to update automatically as I scroll so that newly loaded content is analyzed in real time.
3. As a general user, I would like detection indicators to be visually distinct but non-intrusive so that they do not obstruct my ability to read posts.
4. As a potential user, I would like a smooth and intuitive website design, with clear and concise instructions regarding installation.
5. As a potential user with doubts regarding the extension, I would like for there to be an FAQ page so that I can likely answer my questions without needing to reach out to customer support.
6. As a general user, I would like in-extension onboarding tips so that I can understand how detection results are presented.
7. As a general user, I would like to hover over a detection indicator to view more detailed information so that I can understand results without leaving the page.
8. As a general user, I would like to have recent (last 24 hours) detections cached so that I will not have to wait for the extension to re-detect when viewing a post again.
9. As a general user, I would like to view a history of the cached analyzed posts so that I can revisit past detection results and sites where they were detected.
10. As a non-chrome user, I would like to run the extension on Firefox, Edge, or Safari so that I am not limited to only using this extension on Chrome.
11. As a general user, I would like the extension's detection and analysis functionality to be consistent across different social media platforms so that I have a unified experience.
12. As a general user, I would like to disable detection on specific websites so I can control where the extension runs.
13. As a general user, I would like for there to be user statistics so that I can visibly observe the effectiveness of the extension.
14. As a general user, I would like the system to be accurate at detecting AI-generated content so that I am not being misled by the extension itself.
15. As a general user, I would like to see a confidence percentage which indicates how likely a post is AI-generated so I can better judge credibility
16. As a general user, I would like the system to be able to ingest video content so that I can detect AI use in videos.
17. As a general user, I would like a short explanation of why a post was flagged as AI-generated so I can judge the credibility of the result.
18. As a general user, I would like confidence ranges or uncertainty indicators so that I know when results are less reliable.
19. As a general user, I would like to report any incorrect detections so that the system can improve over time.
20. As a general user, I would like to access verified/reliable supporting materials for non-flagged content so that if I want to I can further verify it myself.

21. As a user installing the extension for an older generation, I wish the interface to be age-inclusive or have a setting to enable (turn on) age-inclusive, low vision accessible interface (control font size).
22. As a system, I would like to apply rate limiting per user or per session so the backend services are not overwhelmed.
23. As a general user, I would like the extension to detect AI-generated content in comments and replies so that online discussions can also be evaluated.
24. As a general user, I would like to manually trigger detection on a specific post so that I can recheck content.
25. As a general user, I would like detection to happen on sponsored posts and advertisements so that promotional content is also evaluated.
26. As a general user, I would like to be able to collapse detection indicators when on a webpage, so that the information does not clutter the screen.
27. As a general user, I would like to receive a notification of detection when it has finished for content I may have scrolled past so that I do not miss results.
28. As a general user, I would like the extension to distinguish between AI-generated, AI-assisted, and human-written content for more specific results.
29. As a general user, I would like the extension to also work in private/incognito mode so that I can browse securely.
30. As a general user, I would like the extension to support dark mode on my computer/mobile device so that it integrates visually with my browser settings.
31. As a general user, I would like the extension to pause detection automatically on low battery mode so that device resources are conserved.
32. As a general user I would like the extension to avoid reanalyzing the same content multiple times in the same session.
33. As a general user, I would like the extension to prioritize posts that are only currently visible on the screen.
34. As a general user, I would like the extension to show an error message when a detection fails.
35. As a general user, I would like the extension to retry failed detections automatically up to a limit.
36. As a general user, I would like to enable or disable text detection and image detection independently so I can customize what is analyzed.
37. As a general user, I would like the extension to remember preferences such as enabling or disabling certain websites, dark/light mode, and turning off certain websites even when the browser is restarted.
38. As a general user, I would like the extension to detect the language of a post so the behaviour of detection can adapt.
39. As a general user, I would like the extension to indicate when a language is not supported.
40. As a general user, I would like the extension to avoid flagging content that is labeled satire or parody so nothing is misclassified.
41. As a general user, I would like the extension to display a disclaimer reminding that detection results are based on probability.

42. As a general user, I would like to search for detection history by keyword.
43. As a general user, I would like the extension to highlight specific segments of the text that triggered the AI detection.
44. As a general user, I would like the extension to have a “Manage Disabled Websites” bashboard/settings so that I can easily view and manually add or remove websites from my exclusion list.
45. As a user with old devices, I would like the extension to automatically pause when I am browsing on a different tab or an application to conserve CPU and RAM.
46. As a general user, I would like to have “Scan entire page” to force the extension to analyze every visible post at once if the automatic scroll detection fails.
47. As a user with technical difficulty, I would like a “Simple mode” toggle which would change the AI detection text to hide percentages and instead use “Likely AI”(orange) and “Likely Human” (green) with large text and icons.
48. As a Google user, I would like the extension to apply to AI-made webpage descriptions, articles, and the “AI Overview”.
49. As a Reddit user, I would like for the extension to work on Reddit so that I can detect AI content on Reddit.
50. As a LinkedIn user, I would like for the extension to work on LinkedIn so that I can detect AI content on LinkedIn.
51. As a X/Twitter user, I would like for the extension to work on X/Twitter so that I can detect AI content on X/Twitter.
52. As a Facebook user, I would like for the extension to work on Facebook so that I can detect AI content on Facebook.
53. As an Instagram user, I would like for the extension to work on Instagram so that I can detect AI content on Instagram.
54. As a general user, I would like for the extension to have fact-checking capabilities so that I can verify if social media content is accurate, whether it is AI-generated or not.
55. As a Spanish-speaking user, I would like for the extension to have a Spanish language option so that I can use the extension in my native language.

Non-Functional Requirements

- Detect AI use within 20 seconds for 800*600 images and 10 seconds for 500 words of text
    - Have some visual “loading” display during processing
    - within 45 seconds for 1920x1080p images
- Handle up to 1000 simultaneous requests on the server side.
- Detect AI use with >80% accuracy
- As a privacy-conscious user, I would like all stored detection data to be automatically deleted in <24 hours.
- All requests for detections should have the content metadata be anonymous before transmission
- As a generalized user, I would like false positive rates to be <= 10% so that legitimate content is not continuously mislabeled.

- Have a low memory footprint so that the extension doesn’t slow down the user’s device.
- If the user has a bad internet connection, I wish that the extension would timeout gracefully and show a ‘retry’ option.
- UI interactions shall respond within < 100ms
- High-Contrast, large-text mode should be available for low-vision users
- Onboarding instructions should be easily understandable by non-technical users
- As a privacy-conscious user, I would like assurance that no personally identifiable information is stored so that my own data remains secure.
- As a privacy concerned user, I would like a “learn how your data is stored” link.
- As a privacy conscious user, I would like the ability to clear all cached detections manually so that I control stored data.