Poster: *Rethink*: A Personalized Regret Analyzer for Social Media Posts

William Namgyal
Northeastern University
william.rnam.rgyal@gmail.com

Tianshi Li
Northeastern University
tia.li@northeastern.edu

***Abstract*—Social media users often experience regret after impulsively sharing sensitive content, leading to personal and professional consequences. Existing static nudges lack awareness of the varied cultural backgrounds, individual preferences, and social contexts that influence posting behavior and regret. We present Rethink, an LLM-powered browser extension that provides real-time, tailored feedback on potential regrets before posting on Facebook. Rethink analyzes post content while examining social context through personal details, group memberships, and friend information. The extension also incorporates user feedback for continuous personalization. We present examples showcasing Rethink's ability to generate relevant regret analyses based on the user's social context and user feedback.**

I. Introduction

Social media platforms like Facebook enable users to share content instantly, but this often leads to impulsive, regrettable posts which can result in personal embarrassment, reputational damage, or even job loss [1]. These regrettable posts are driven by factors like emotional impulsivity, social validation, misjudgment of audience and social norms. Under such conditions, it is challenging for users to manually assess their potential audience and the impact of their posts in real-time.

Prior work identified two challenges on assessing sensitive and regrettable posts: *individual user differences* and *contextual factors*. Every user's unique personality and cultural backgrounds may have significant influence on choices to post certain content and level of regret experienced [2], [3]. Posting behavior is also influenced by social context—users often self-censor content and refrain from posting due to the inability to precisely target their intended audience, highlighting the necessity for selective sharing mechanisms [4].

Privacy nudge systems have been developed to remind users to reconsider before posting impulsively. Wang et al. [5] evaluates two privacy nudges on Facebook: an audience nudge that displays the profile pictures of potential viewers, and a timer nudge that delays posting by 10 seconds. Their static nudges were perceived differently by users, with some finding the nudges unnecessary. The varying user opinions are attributed to different posting behaviors and personalities, which the system does not account for due to limited context awareness. These limitations underscore the need for tailored nudges that enhance context awareness and continuously adapts to the user's preferences.

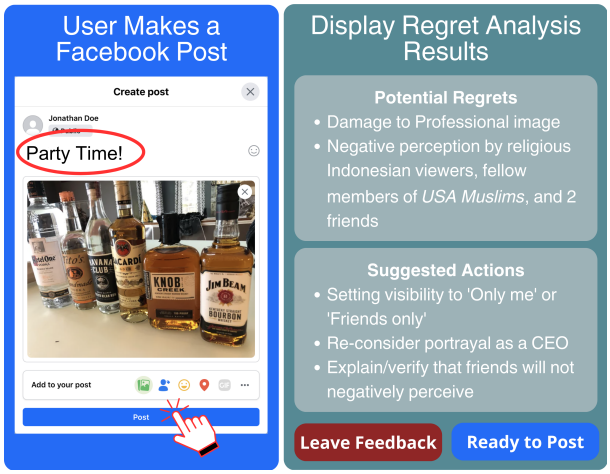


Fig. 1. Example of Rethink's regret analysis workflow: A Facebook post (left) is intercepted to display potential regrets and suggested actions (right). Users can leave feedback to improve future analyses or proceed with posting.

To address these challenges, we introduce Rethink, an LLM-powered browser extension that provides real-time, personalized regret analysis to promote cautious content-sharing. Rethink factors in post media, captions, and user social context to generate tailored reports with potential regrets, recommended actions, and severity ratings. Rethink incorporates social context to provide personalized privacy recommendations. It utilizes a continuous feedback system, refining and tailoring the regret analyses to the individual over time.

II. Implementation

Rethink is a Google Chrome extension for Facebook that intercepts posts to provide real-time regret analysis. It evaluates three main factors: image, caption, and social context, while incorporating user feedback for personalization.

A. *Regret Analysis*

1) Image and Caption Extraction: Rethink intercepts the images and captions to generate a post description using GPT-4, focusing on people, objects, interactions, mood, and setting.

Network and Distributed System Security (NDSS) Symposium 2025
24-28 February 2025, San Diego, CA, USA
ISBN 979-8-9894372-8-3
https://dx.doi.org/10.14722/ndss.2025.[23|24]xxxx
www.ndss-symposium.org

2) Social Context Analysis: Rethink uses a Selenium script to gather user details for personalized recommendations:

- "About" information: Education, current location, hometown, and relationship status. This helps infer local cultural and religious norms (e.g., in Figure 1, the user from Indonesia has a post flagged for including alcohol)
- Groups/communities the user is apart of, which will help identify potential audience backlash.
- Information about friends and their public "About" details. We focused on the top five most and least interacted-with friends. This considers how posts may affect close friends, who are more likely to evoke regret, versus distant friends, with varied perceptions due to weaker ties.

3) Feedback Mechanism: To incorporate individual preferences and avoid over-sensitivity, Rethink includes a feedback mechanism. It is designed as a free-form comment system to capture detailed, specific user preferences that can be most effectively factored into our LLM prompt, as compared to a simple thumbs up/down or rating system which will leave our LLM to infer upon why the certain ratings exist.

4) Regret Analysis Prompt: A comprehensive prompt is crafted to the GPT-4 LLM, where the image analysis's resulting description, caption, and gather social context details are inputted. In addition, previously inputted user feedback, if any, is added to the end of the prompt to personalize the regret analysis and unnecessary or inaccurate suggestions. We use in-context learning, providing three examples in the prompt to guide the LLM to account for the contextual nuances and generate actionable suggestions in the regret analysis.

B. *User Interface*

The regret analysis back-end processes begin after the user clicks "Post" (Figure 1). While the process runs, a progress bar appears, followed by the regret analysis. The user can toggle the *Ready to Post* button to proceed to Post, or click "Close" on the display and edit the post content. If the user finds that certain presented regrets are not relevant to themselves, they may leave feedback through the "Leave Feedback" button. The future regret analysis will reflect the factoring of the feedback.

III. Example Use Cases

We present two example use cases. Note they differ from the in-context learning examples provided in the prompt.

A. *Social Context*

Sarah, a nurse at Toronto General Hospital living in Toronto, decided to post a picture of anti-vaccine protesters outside her workplace with a caption expressing her frustration: "These idiots need to get a clue! #VaccinesSaveLives #AntiVaxxers." Before posting, the Rethink extension intercepted her post for analysis. Considering her professional background, group affiliations ("Vaccines Save Lives," "Muslims Community in USA," "Anti-Vax Moms"), and location in Toronto, the analysis highlighted several potential regrets. It noted that her sensitive caption could provoke backlash from anti-vaccine group members (Severe), confuse her audience due to conflicting group memberships (Mild), appear insensitive during a health crisis in Toronto (Mild), and possibly spark controversy within her religious community group (Mild). Rethink advised her to revise the caption, ensure message consistency, limit post visibility to *Friends Only* or *Only Me*, and prepare for respectful dialogue if she posted. Sarah heeded this advice, adjusted her post, and set its visibility to "Friends only."

B. *Feedback Mechanism*

John is a married man and current Cybersecurity Engineer with his hometown listed as Jerusalem, Israel. He posted an image of multiple bottles of alcohol lined up with the caption "Party Time!." John is also a part of "The Drinking Club" group. Rethink's initial regret analysis flagged several potential issues including "promoting daytime drinking might be inappropriate for his audience from Jerusalem especially those with conservative or religious views (Severe)" However, after John submitted feedback clarifying that he and his family are not Jewish, have moved to the US decades ago, and that his Facebook audience consists of American friends with no ties to religious groups against drinking, Rethink removed the first potential regret with concern to a religious audience.

IV. Conclusion and Future Work

By leveraging LLMs and user-specific details, Rethink provides tailored, adaptive guidance. Despite these advancements, Rethink has limitations. For example, a regret analysis costs $0.37 on average due to API usage, assuming an optimal caption length of 50 characters[1], average of 338 friends[2], and average of five groups joined[3]. Longer captions or more social connections can increase these costs. Future research will explore strategies to reduce expenses, including optimized API usage and alternative LLM models.

References

[1] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor, ""i regretted the minute i pressed share" a qualitative study of regrets on facebook," in *Proceedings of the seventh symposium on usable privacy and security*, 2011, pp. 1–16.

[2] K. Moore and J. Mcelroy, "The influence of personality on facebook usage, wall postings, and regret," *Computers in Human Behavior*, vol. 28, pp. 267–274, 01 2012.

[3] Y. Li, *Cross-Cultural Privacy Differences*. Cham: Springer International Publishing, 2022, pp. 267–292. [Online]. Available: https://doi.org/10.1007/978-3-030-82786-1_12

[4] M. Sleeper, R. Balebako, S. Das, A. L. McConahy, J. Wiese, and L. F. Cranor, "The post that wasn't: exploring self-censorship on facebook," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ser. CSCW '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 793–802. [Online]. Available: https://doi.org/10.1145/2441776.2441865

[5] Y. Wang, P. Leon, A. Acquisti, L. Cranor, A. Forget, and N. Sadeh, "A field trial of privacy nudges for facebook," 04 2014.

[1] https://buzzsumo.com/blog/ultimate-guide-facebook-engagement-2017/

[2] https://www.pewresearch.org/short-reads/2014/02/03/what-people-like-dislike-about-facebook/

[3] https://www.facebook.com/community/whats-new/facebook-communities-insights-survey/

Poster: Rethink - A Personalized Regret Analyzer for Social Media Posts

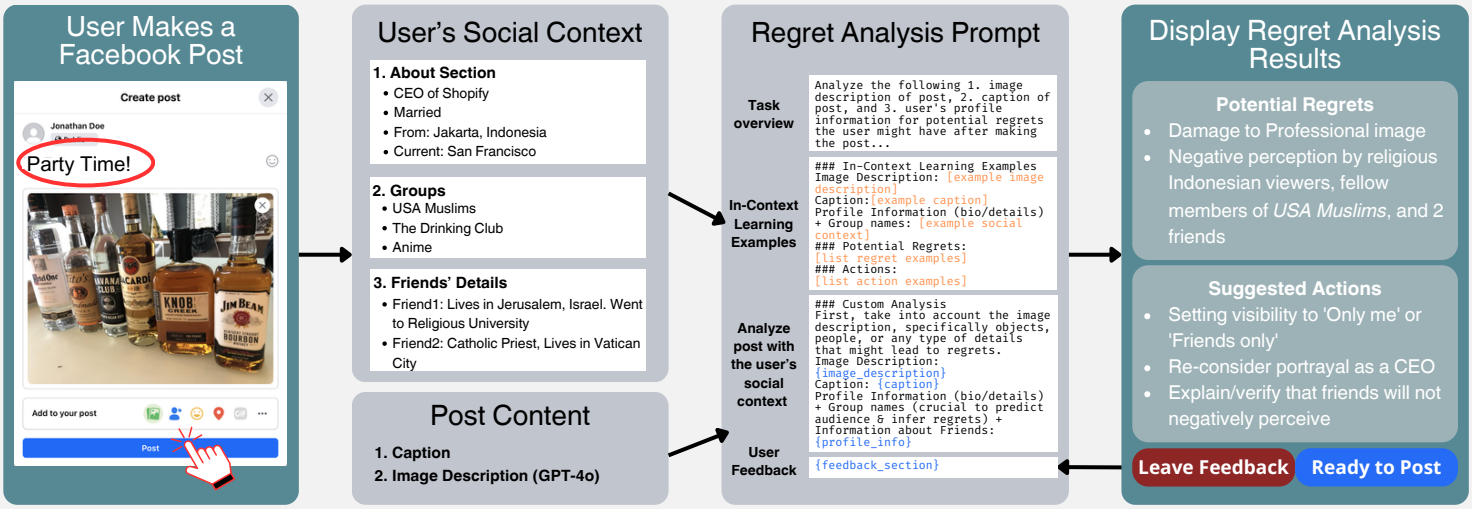
William Namgyal, Tianshi Li
Northeastern University

Introduction:

- Impulsive social media posts can lead to severe backlash, reputational damage, or job loss
- Spanish actress Karla Sofía Gascón recently faced severe backlash after past derogatory and controversial tweets resurfaced, leading to her removal from Netflix's promotional materials and public apologies.
- Why Rethink?
  - Static privacy nudges for social media posts fail to adapt to users' cultural and social contexts
  - Introducing *Rethink*, a personalized, real-time regret analyzer powered by large language models (LLMs)
  - Provides tailored regret analysis based on user's post content and social context
  - Uses NLP and continuous feedback for personalized recommendations

Overview of Rethink

Rethink's process for analyzing Facebook posts is demonstrated by the shown scenario where a user creates a post by uploading an image and writing a caption (first image).



1. When the user clicks "Post," Rethink intercepts the action, extracts the image and caption, and uses the Selenium package to scrape the user's social context, including the user's "About" section, group memberships, and friends' details.
2. These inputs are combined into a prompt for GPT-4, generating a regret analysis report with potential regrets and suggested actions.
3. The results are displayed to the user in a popup (right screenshot), showing potential regrets and recommendations.
4. User feedback provided through the "Leave Feedback" button is used to refine future analyses.

Example Use Case

Sarah is a nurse living in Toronto who is apart of the groups:
Vaccines Save Lives, Muslims Community in USA, Anti-Vax Moms

Potential Regrets:

- Backlash from anti-vaccine group members (Severe).
- Confusion due to conflicting group memberships (Mild).
- Appearing insensitive during a health crisis (Mild).
- Controversy within her religious community group (Mild).

Suggested Actions:

- Revise the caption for consistency.
- Limit visibility to "Friends Only" or "Only Me."
- Prepare for respectful dialogue if posted.

Limitations & Future Directions

High Costs & Delays:

- Increased API costs from relying on GPT-4
- Slow social context scraping with Selenium, especially for users with extensive networks.
- Future work will optimize scraping algorithms and explore cost-efficient LLM models.

Limited Platform Support:

- Currently restricted to Facebook.
- Future efforts will expand Rethink to other platforms like Twitter, Instagram, and LinkedIn.

Broader Content Analysis:

- Limited to text and images.
- Future enhancements include multi-modal analysis for videos, audio, and live streams.