Emotion-aware text simplification of user generated content using LLMs

Anastasiia Bezobrazova
Centre for Translation Studies
University of Surrey, UK
a.bezobrazova@surrey.ac.uk

Daria Sokova
Centre for Translation Studies
University of Surrey, UK
d.sokova@surrey.ac.uk

Constantin Orăsan
Centre for Translation Studies
University of Surrey, UK
c.orasan@surrey.ac.uk

Abstract

Digital inclusion increasingly supports adults with intellectual disabilities (ID) to participate online, yet social media posts can be difficult to understand, particularly when they contain strong emotions, slang, or non-standard writing. This paper investigates whether large language models (LLMs) can simplify social media texts to improve cognitive accessibility and preserve emotional meaning. Using an accessibility-oriented prompt based on existing guidance, posts are simplified and emotion preservation is assessed. The results suggest that many simplified posts retain the same emotions, though changes occur, especially when emotions are weakly expressed or ambiguous. Qualitative analysis shows that simplification improves fluency and structure but can also shift perceived emotion through changes to tone, formatting, and other affective cues common in social media text. The research has also revealed that different LLMs produce very different outputs.

1 Introduction

Digital technologies have become a central part of everyday life, reshaping how people communicate, search for information and use services. Organisations across the UK have introduced programmes to help people with intellectual disabilities (ID) get online and participate in digital life, so they are not excluded from the digital society (Triantafyllopoulou et al., 2025). According to recent estimates from Office for Health Improvement & Disparities (OHID) (2025), approximately 1.3 million people in England are having an ID, underlining both the scale and the policy importance of digital inclusion for this population.

ID involve significant difficulties with learning, understanding information and managing everyday tasks independently (American Psychiatric Association, 2013). Nevertheless, many adults with ID use the internet to maintain social connections, access information and seek entertainment (Glencross et al., 2021; Chadwick et al., 2022). An England-wide survey of adults with ID shows that 72.2% used the internet daily and 79.1% used social media (48% daily) (Triantafyllopoulou et al., 2025). Also, people with ID commonly experience reading difficulties, and emotionally charged social media posts can be especially hard to understand.

Syntheses of the field since 2020 emphasise both benefits of online participation such as belonging, identity work, autonomy and wellbeing and persistent structural and cognitive barriers that shape its quality (Anderson et al., 2023; Chadwick et al., 2022). A recent systematic review identifies four recurrent motivations for social internet use among adults with ID: *fitting in/belonging*, *maintaining connections*, *making new connections*, and *autonomy and empowerment* (including self-expression and self-determination) (van Alem et al., 2025). The same review underscores literacy-related and support-dependent barriers and a persistent tension between autonomy and safeguarding, signalling the need for tailored supports (van Alem et al., 2025; Caton et al., 2022; Triantafyllopoulou et al., 2025).

This paper investigates whether large language models (LLMs) can simplify social media posts in ways that improve cognitive accessibility for adults with ID while preserving the original emotional content. Our paper focuses on the emotion preservation as a key indicator of simplification quality. To evaluate this systematically, an automatic emotion classifier is trained on social media texts and used to compare the emotions assigned to the original and simplified versions. A linguistic analysis of the simplified posts is also carried out to gain insights into how LLMs simplify the posts.

The structure of the paper is as follows. Section 2 reviews background on Easy-to-Understand (E2U) practices, social media accessibility guidelines, and emotional processing in people with learning disabilities. Section 3 introduces the GoEmotions-based dataset and our emotion classi-

The Proceedings for the 15th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2026), pages 107–122
March 29, 2026

fier. Section 4 presents the prompt design, simplification experiments and cross-model comparison, including analyses of emotion preservation. The paper finishes with a discussion and conclusions.

2 Background information

2.1 Text Accessibility Guidelines

Recently, E2U practices have emerged to make texts easier to read, with Plain Language (PL) and Easy Language (EL) as the main approaches (Deleanu et al., 2024). PL bridges professional–public communication in health, law, administration and personal finance, helping adults with limited literacy navigate information and make decisions (European Commission, 2012; NHS England, 2017; United States Congress, 2010). EL, first designed for people with learning disabilities is now applied more broadly and is routinely paired with layout conventions such as legible sans-serif fonts, left alignment and generous spacing (Misako Nomura and Tronbacke, 2010; Scope Australia, 2015; Perego, 2020; Hansen-Schirra and Maaß, 2020).

Although labels vary, the underlying guidance is similar: keep vocabulary familiar and define unavoidable terms; avoid metaphors and idioms; and maintain strict consistency in terminology (Scope Australia, 2015). Syntactic recommendations call for short, single-idea sentences (around 15–20 words), clear Subject–Verb–Object ordering, minimal punctuation and the use of numerals rather than number words, while favouring splitting complex sentences and using verbs instead of abstract nouns (Inclusion Europe, 2010; Hertfordshire County Council, 2018).

This core set of rules aligns with recent academic work that situates E2U within a wider accessibility agenda, distinguishes PL, EL and related types, and examines trade-offs between ease of understanding and social acceptability (Hansen-Schirra and Maaß, 2020; Perego, 2020). Policy has reinforced this shift: in the UK, guidance from the Office for Disability Issues[1] helped embed inclusive communication and shaped NHS publishing policy (NHS England, 2017). At EU level, the Web Accessibility Directive (2016/2102) and the European Accessibility Act (2019/882) set accessibility duties for public-sector content and key products and services, placing E2U within a broader regulatory framework (European Union, 2016, 2019).

Despite extensive guidance, challenges in text accessibility remain. Across standards there is agreement on core practices but less on thresholds such as sentence length, treatment of complex numerals, use of grammar, and procedures for terminology control (when to introduce terms, how often to repeat them, and how to maintain consistency) (Mencap, 2000; Change, 2016). Overall, the guidelines converge on short sentences, clear structure, familiar vocabulary and consistent layout as key features that make public texts easier to understand.

2.2 Guidelines for Accessibility for Social Media

Most major organisations now provide guidance on making social media posts accessible. The main focus across these accessibility guidelines is on images, video and visual design, while written text still receives comparatively little detailed attention.

A common core of recommendations concerns alternative formats for non-text content. The need to add descriptive *alt* text to images and to provide captions or transcripts for audio and video is emphasised in many guidelines (University of Edinburgh, 2022; UK Association for Accessible Formats (UKAAF), 2020; Sprout Social, 2024). They also stress accessible typography and layout, such as using legible fonts, ensuring sufficient colour contrast and avoiding text embedded in images.

Text-level guidance is more fragmented. Most documents call for "plain language" or "clear English", with generic advice to keep posts concise, avoid jargon and unexplained acronyms, favour active voice and avoid ALL CAPS (Harvard University, 2023; University of Edinburgh, 2022; University of Reading, 2023). They rarely explain how to adapt emotionally charged or noisy user-generated text for readers with ID. Mencap and the Government Communication Service offer more detail, recommending posts of around 25 words, avoiding non-standard symbols, not squeezing too much text into one graphic and testing content with assistive technologies (Government Communication Service, 2021; Mencap, 2022). All sources stress careful use of hashtags and emojis, suggest CamelCase hashtags (#LearningDisabilityWeek rather than #learningdisabilityweek) (Mencap, 2022), limiting hashtags, using emojis sparingly at the end of posts and never as word substitutes as this confuses screen readers (Sprout Social, 2024).

Overall, existing guidelines for social media accessibility provide a baseline on visual aspects and

[1] https://www.gov.uk/government/organisations/office-for-disability-issues

offer only high-level instructions for accessible writing. They converge on specific conventions for hashtags and emojis, but the level of detail and linguistic precision is nowhere near that found in established PL and EL guidance. As a result, there is still limited practical advice on how to rewrite short, informal and emotionally laden posts for people with learning disabilities.

2.3 People with Learning Disabilities and Emotional Content on Social Media

People with ID often find it hard to identify their emotions (Davies, 2013), and research on alexithymia (difficulty identifying and describing feelings) shows that they have limited emotional insight (Mellor and Dagnan, 2005). Emotion recognition is linked to IQ and receptive language, so people with lower intellectual ability perform less well on emotion-recognition tasks (Scotland et al., 2015), making understanding emotional content in faces, voices, pictures or text challenging.

Most empirical investigations have focused on how people with learning disabilities recognise emotions from photographic facial stimuli rather than textual content. Across multiple studies, participants frequently identify basic expressions such as happiness with reasonable accuracy, yet demonstrate significantly lower performance than typical users when tasks incorporate a broader range of emotions or more nuanced expressions (Scotland et al., 2015). Owen and Maratos (2016) reported that adults with ID exhibited lower accuracy than typical users in labelling both basic and subtle emotional expressions, with the greatest challenges observed for neutral and low-intensity emotions.

In contrast, less research examines how people with ID understand emotional meaning in written communication, including social media content. Research shows that social media can support belonging, social connection and autonomy, but says little about how users decode emotional nuance in text. This gap matters because emotions and attitudes on social media are often expressed through figurative language, sarcasm, irony, memes, emojis and other non-literal cues that people with ID find difficult to interpret.

3 Data

In this work we use the GoEmotions dataset (Demszky et al., 2020), a manually annotated corpus of 58k English Reddit comments labelled for 27 fine-grained emotion categories plus a Neutral label. The dataset is also available in a reduced taxonomy based on Ekmans six basic emotions (anger, disgust, fear, joy, sadness, surprise) plus neutral (Ekman, 1992). This is the version used in this research. The comments are sampled from popular subreddits and carefully curated to reduce toxicity, demographic bias and sentiment skew through subreddit filtering, length constraints, sentiment and emotion balancing, and masking of sensitive identity and religion terms (Demszky et al., 2020). Compared to other emotion datasets based on news headlines, posts and other domains (Strapparava and Mihalcea, 2007; Mohammad et al., 2018; Bostan and Klinger, 2018), GoEmotions is, to the authors knowledge, the largest human-annotated emotion dataset with multiple labels per instance and demonstrates robust inter-rater agreement (Demszky et al., 2020).

Emotion	Precision	Recall	F_1-score
anger	0.54	0.71	0.61
disgust	0.61	0.67	0.64
fear	0.44	0.95	0.61
joy	0.88	0.81	0.84
neutral	0.75	0.54	0.63
sadness	0.55	0.80	0.65
surprise	0.54	0.79	0.64
macro avg	0.62	0.75	0.66
weighted avg	0.74	0.71	0.71

Table 1: Ekman-level results of XLM-RoBERTa classifier

For the purposes of this paper, we randomly split the Ekman-level subset into two disjoint parts: 90% of the data is used to train the emotion classifier described in Section 3.1, and the remaining 10% is reserved for the simplification experiments, where we apply the classifier to predict emotions for the original and simplified posts (see Section 4.2).

3.1 Emotions-Classifier

We fine-tuned an XLM-RoBERTa-based classifier on the GoEmotions dataset (Demszky et al., 2020), following the Kaggle implementation for Ekman-level labels [2]. On the test split, the model achieves an accuracy of 0.71 and a macro-averaged F_1 of 0.66, with macro-precision 0.62 and macro-recall 0.75 (see Table 1). Our scores are slightly higher than the results reported by Demszky et al. (2020) for the same Ekman taxonomy as seen in Table

[2] https://www.kaggle.com/code/anassouzaouit/fine-tuning-xlm-roberta-on-go-emotions-dataset

A.3 (in the appendix). This difference is likely due to the more recent model used. The most marked trade-off appears for *fear*, where we obtain very high recall (0.95 vs. 0.76) at the cost of much lower precision (0.44 vs. 0.61), indicating that fear is frequently over-predicted. The confusion-matrix heatmap (see Figure A.2) shows that neutral instances are often misclassified as *anger*, *joy* or *surprise*, and it also highlights the strong class imbalance (e.g., 1,712 instances of *joy* vs. only 57 of *disgust*), which likely drives some of the remaining confusion patterns across classes. This classifier is used in the next section to determine the emotion in the simplified version of a post.

4 Experiment and results

4.1 Prompt design for accessible posts simplification

For the post simplification stage, we designed a task-specific prompt to guide language models in producing accessible rewrites. The model receives the following instruction:

> *Simplify the posts so that people with learning disabilities can easily understand it. Keep the same meaning and facts. Preserve the same emotion. Do not soften or exaggerate the emotion. Make the feelings clear and simple. Do not add new facts or advice. Do not judge the person. Use common words and active voice. Keep emojis only if they add meaning, and also name the feeling in words. Use CamelCase for hashtags. For example, instead of #learningdisabilityweek, write #LearningDisabilityWeek.*

This prompt was written using existing guidelines on accessible social media from Mencap (2021); Button (2021); Rowell (2021); Sprout Social (2024), which all emphasise clear language, consistent formatting and consideration of cognitive access needs. The first part of the prompt specifies the target audience and communicative goal, encouraging the model to prioritise understanding for people with learning disabilities rather than generic style improvement. The next group of instructions constrains how content and emotion may be changed. It requires the model to keep the same meaning and facts, preserve the emotion, and avoid adding advice or moral judgement. This reflects ethical recommendations that accessible versions should respect the writers voice while making the emotional content easier to follow.

The remaining parts of the prompt convert general plain-language and formatting guidance into specific, actionable rules for the model. Accessibility guidance for social media recommends short sentences, everyday vocabulary and active voice to reduce reading effort and support screen-reader users (Button, 2021; Rowell, 2021; Sprout Social, 2024). Many guidelines recommend CamelCase hashtags and advise using emojis sparingly and never as substitutes for words (see Section 2.2).

We decided to keep the prompt simple and naive, rather than using complex multi-step prompting or detailed role specifications. This choice was meant to approximate a realistic instruction that non-expert practitioners (e.g., support workers or family members) could reuse with minimal prompt-engineering experience.

4.2 Evaluation of automatic emotion detection

We examined how well automatically detected emotions are preserved after simplification using the prompt introduced in the previous section using GPT-4o via the OpenAI API [3]. For each instance in our evaluation subset (4,947 items), we consider three labels: (i) the Ekman-level GoEmotions label (*gold*), (ii) the prediction of our XLM-RoBERTa classifier on the original text (*pred_orig*), and (iii) the prediction of the same classifier on the simplified version produced by GPT-4o (*pred_simp*).

We compared *pred_orig* and *pred_simp* directly, assuming that any systematic errors of our classifier are likely to affect the original and simplified versions in similar ways. This means that changes in the assigned labels provide a conservative signal of genuine shifts in perceived emotion. Across the full set, *pred_orig* and *pred_simp* are identical for 3,588 out of 4,947 items (72.5%), so in roughly three quarters of the posts the classifier assigns the same emotion label before and after simplification. As shown in Figures A.3 and A.4, stability is highest for *joy* and *surprise*, moderate for *sadness* and *fear*, and lower for *anger*, *disgust* and *neutral*.

We also compared the models prediction on the simplified texts by comparing *pred_simp* with *gold*. The agreement between the two labels is 0.59. This indicates that, more of the original emotion is lost during the simplification. Given that the gold labels were assigned by annotators to the original post and the *pred_simp* is assigned by the automatic

[3] All models used in this paper were prompted in November 2025.

classifier to the simplified post, we consider this comparison less reliable for the emotion shifts.

In many cases, the emotion is clearly maintained in the simplified post. For instance, *joy* is preserved when "If that's ice cream, then honestly I eat ice cream from a cup at home too lmao." is simplified to "If that's ice cream, I eat it from a cup at home too. 😂". Likewise, *fear* is preserved when "[NAME] is pretty fucking scary" becomes "[NAME] is really scary."; profanity is removed, but the core fear emotion is unchanged.

In other cases, the emotional framing shifts. An originally angry comment, "Talk about a fucking hot take. Quality shit post." (gold and *pred_orig* = *anger*), is rewritten as "Wow, that's a strong opinion! Great funny post about this.", which the classifier interprets as *joy*: the simplifier softens and positively reframes the post, so the original anger is effectively lost. Similarly, an originally neutral statement, "It's how the government treats them." (gold and *pred_orig* = *neutral*), is simplified to "The government treats them badly.", and the classifier now assigns *sadness*; the negative evaluation is made explicit ("badly"), which may be clearer for readers but shifts the stance from neutral description to a sad or critical tone. A different neutral post, "Should have been a and 1 tbh, [NAME] smacked him in the face." (gold and *pred_orig* = *neutral*), is simplified to "[NAME] hit him in the face. It should have been a foul.", the classifier labels now is *anger*, because the simplification foregrounds the sense of unfairness more strongly.

These examples support the quantitative picture: our prompt-based simplification generally maintains the overall emotional profile of the texts, especially for prototypical emotions such as *joy* and *surprise*, but can introduce subtle shifts for borderline or weakly expressed emotions, particularly *neutral*, *anger* and *disgust*. This situation was also noticed with machine translation of user generated content (Saadany et al., 2023).

4.3 Analysis of GPT-4o simplifications

To better understand how our accessibility prompt shapes the output, we analysed the simplified posts produced by GPT-4o. We considered running automatic evaluation metrics to assess the quality of the simplification. However, this was not possible due to the absence of gold reference simplifications for the user-generated social media posts. Reference-based metrics were therefore not used: BLEU relies on n-gram overlap with reference texts, and SARI explicitly compares the system output to the input and to reference simplifications (Papineni et al., 2002; Xu et al., 2016). Learned "reference-free" metrics such as SIERA and ARTS were also not applied because, despite not requiring references at evaluation time, they still depend on supervised resources to train or calibrate an evaluator (e.g., aligned original-simplified pairs for SIERA and simplicity-labelled or pairwise-judgement datasets for ARTS) (Yamanaka and Tokunaga, 2024; Engelmann et al., 2024). Finally, BERTScore is defined as candidate-reference similarity; treating the original post as a proxy reference would mainly reward closeness to the source rather than successful simplification (Zhang et al., 2019).

In light of the limitations of the measures presented above, we attempted to assess the readability of the produced text using existing readability measures. The traditional readability formulas were developed for edited, continuous texts and estimate difficulty from simple surface features such as sentence length and word length (Flesch, 1948; Kincaid et al., 1975; Chall and Dale, 1995). These metrics are less reliable for social media, where short fragments, informal punctuation, hashtags and emojis can disrupt tokenisation and make scores unstable (Redish, 2000). We report there the results of four widely used measures that can be computed consistently on short texts: FleschReadingEase, Kincaid[4], ARI [5], and DaleChallIndex [6]. Posts were also pre-processed to remove emojis. We calculated the readability scores using the *readability 0.3.2* package[7]. As shown in Table 2, all the texts were deemed easy to read, with the simplified posts scored even "easier" on average. This is due to the fact that many posts are short and use common words. However, these apparently "easy" scores mask difficulties typical of user-generated content, including non-standard or missing grammar, slang, and dense abbreviations or initialisms. For this reason, traditional readability scores provide only limited information for social-media texts and should be interpreted with caution, in particular, they do not guarantee that posts are accessible for people with ID.

[4]https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/

[5]https://readable.com/readability/automated-readability-index/

[6]https://readable.com/readability/new-dale-chall-readability-formula/

[7]https://pypi.org/project/readability/

Metric	Original	Simplified
FleschReadingEase	99.47	105.59
Kincaid	1.76	0.41
ARI	2.56	0.93
DaleChallIndex	4.07	3.37

Table 2: Average readability scores for original vs. simplified posts.

Our analysis below focuses on qualitative examination of how GPT-4o rewrites the posts under the accessibility prompt showing in detail how the text itself changes after simplification. It is based on the same 4,947-item subset described above, comparing original posts to its GPT-4o simplification. GPT-4o almost always rewrites the input rather than leaving it unchanged, with only a few posts remaining identical to the original. These are typically very short, already accessible messages, such as "I like Tom and Kato.", "It's cool." or "Thank you [NAME].", where the model reproduces the input verbatim. However, we did not detect any pattern to indicate when these short posts are going to be rewritten and when not. In some instances, GPT-4o rephrases already short posts without adding real clarity. For example, "Lmao quality." is simplified as "Haha, this is great quality.", and "Lol I'm glad" becomes "Haha, I'm happy.".

A noticeable pattern observed concerns the insertion of emojis and hashtags, despite the prompt instructing that emojis should be kept only when they add meaning and that hashtags should use CamelCase. In the original dataset, there are 164 emojis in 90 posts (less than 2% of the posts) and 12 hashtags in 11 posts (less than 0.2% of the posts). Although the prompt did not encourage adding new emojis or hashtags, GPT-4o often introduces both in the simplified versions. In total, the simplified outputs contain 696 emojis in 632 posts (nearly 13% of posts) and 706 hashtags across 692 posts (nearly 14% of posts).

Typical examples of inserted hashtags include a gratitude hashtag, as in "Great thanks for the advice!" becoming "Thanks a lot for the advice! *(#Grateful)*", even though the model was not instructed to add hashtags. The hashtags **always** follows the formatting guidance (using CamelCase and clearer tags) and comply with recommendations that advise placing them at the end of a sentence. (University of Reading, 2023).

The way existing hashtags are treated is unpredictable. Sometimes they are left as they are e.g., "Happy Daily Peko #270!", sometimes explained in the running text, for instance, "Fried Egg is my #1 since cricket cafe stopped doing breakfast sandwiches." becomes "Fried Egg is my favorite now because Cricket Cafe stopped making breakfast sandwiches.", and sometimes replaced by new, sentiment-laden tags such as "If that's ice cream, then honestly I eat ice cream from a cup at home. It's great for portion control." is rewritten as "If that's ice cream, I eat it from a cup at home. It helps me eat the right amount. *#IceCreamLove*". They partly follow formatting guidance (using CamelCase and clearer tags).

A similar pattern can be seen with emojis. Typical examples include adding a new emoji to mark a feeling, as in "account got suspended lmao" becoming "My account got suspended. 😂 #Funny". Sometimes emojis that are already present are simply preserved and the text around them is expanded, for example "Sigh, that was beautiful 😢" becomes "Wow, that was beautiful. I'm sad 😢". In other posts, GPT-4o removes or replaces emojis: "Im literally shaking right now 😫" is simplified to "I am shaking right now. I feel upset.", where the emoji is dropped but the feeling is spelled out in words, and "omg [NAME] and his dad walking out together is so cute 😍" becomes "Wow, [NAME] and his dad walking together is so cute. Heart eyes emoji." The prompt instructs the model to "keep emojis only if they add meaning", yet GPT-4o often introduces new emojis in the simplified posts. This is not consistent with accessibility guidance, which recommends using emojis sparingly, placing them at the end of a sentence, using widely recognised emojis and not replacing text with emojis (AbilityNet, 2023; Readability Guidelines, 2020). This behaviour could be as a result of the large number of social media posts used to train the LLM.

Our prompt explicitly says that the post should be simplified for people with learning disabilities in a hope that it will be successfully tackle abbreviations and slang. However, the handling of these phenomena is inconsistent. Common abbreviations such as *tbh*, *idk*, *imo* or *lmao* are usually removed or paraphrased rather than explicitly expanded. For example, "Should have been a and 1 tbh, [NAME] smacked him in the face." is simplified to "[NAME] hit him in the face. It should have been a foul and 1 point.", completely discarding *tbh*, leading to information loss. In some cases, the simplified version is still unclear to readers who are

not familiar with the context of the post. One illustrative example is "Holy shit that SSP was beautiful", becomes "Wow, that SSP was amazing!"; the profanity is softened, but the unexplained abbreviation *SSP* is preserved, so the core referent remains unclear for non-expert readers. Laughter markers also fluctuate: *LOL* and *Lmao* may be rewritten as "Haha" in some posts but retained as *lol* in others, and sequences such as *hahaha* can be normalised to *lol*, again without a consistent pattern.

Filtering of offensive and sensitive language is more systematic. Across the dataset, 314 original posts contain swear words or sensitive terms; the most frequent items include *fuck* (110 occurrences), *shit* (50), *stupid* (33) and *kill* (22). In the simplified outputs, only 31 posts contain any of these terms, with just 1 instance of *fuck*, no instances of *shit*, 1 instance of *kill* and 15 instances of *stupid*. GPT-4o removes or paraphrases the vast majority of such content. For example, "[NAME] is pretty fucking scary" becomes "[NAME] is really scary"; "kills me" is often rewritten as "makes me feel upset"; and "Holy shit" is frequently reduced to "Wow". More explicit violent phrasing such as "kill someone" can be paraphrased as "end someone's life". This kind of automatic detoxification and softening of aggressive or offensive language mirrors broader trends in safety-tuned language models, where filtering and controlled generation are used to reduce toxic content in model outputs (Xu et al., 2021). By contrast, neutral or identity-related terms such as *gay*, *sex*, and *porn* are generally preserved, suggesting a distinction between aggressive swearing and descriptive references to sexuality. As a result, emotion preservation becomes less predictable: removing or weakening strong profanity can reduce the intensity or nuance of the original affect, even when the core propositional content is retained.

The model also tends to make emotions more explicit, sometimes going beyond what is stated in the original post. A clear example is "I miss you [NAME] 😭", which is simplified to "I miss you [NAME] and I feel sad. 😭". In this case, GPT-4o preserves the original wording but adds an explicit statement of sadness, aligning with the instruction to "make the feelings clear and simple". Unfortunately, this explicit emotion labelling is not applied consistently across posts. Similar expansions occur with congratulations messages: original "Happy cake day" posts are often changed to "Happy birthday" or "Happy birthday to you", sometimes with an added birthday-cake emoji.

Overall, GPT-4o improves style, producing grammatical, fluent sentences and often correcting hashtags to CamelCase. However, it inconsistently adds new hashtags, emojis and explicit emotion statements, systematically softens offensive language and reframes offensive content in a polite tone, sometimes introducing emotional and stylistic cues that do not match the original post.

4.4 Comparing different models on the same prompt

In addition to experiments presented in the previous section, we carried out a comparison of several large language models on a smaller selection of posts that covered the main phenomena of interest: swear words and offensive language, emojis and hashtags, abbreviations and initialisms, and already short, apparently accessible posts. We applied the same accessibility prompt (Section 4.1) to a manually selected set of posts. Our experiments reveal consistent differences in how the models respond to the same accessibility prompt. However, they also introduced additional behaviours. For ChatGPT 5, explicit first-person feeling statements were added in 58.3% of the simplified posts. DeepSeek often shifted from simplification to meta-commentary, in 37% of cases the output described the original post (e.g., "They are saying...", "This tweet...") instead of providing a self-contained simplified version. Gemini showed a similar case in 35% of simplified posts, it switched into explanation mode, providing commentary or interpretation instead of a direct simplification.

ChatGPT-4o vs. ChatGPT 5 behave very similarly on this prompt. Both usually preserve the basic facts and overall emotion and produce fluent, grammatical rewrites. However, they systematically make feelings explicit, even when the original post already conveys them. For instance, "[NAME] is pretty fucking scary" is simplified as "[NAME] is really scary. I feel afraid.". The core meaning is preserved, but the models add first-person emotion statements that was only implicit in the source. A similar pattern appears in more abstract posts: "As long as blind luck exists, there is no upper limit on stupidity." is rendered as "While blind luck exists, people can still do very stupid things. I feel annoyed.", which improves syntactic clarity but does not explain the idiom "blind luck".

Both ChatGPT models apply safety and politeness norms to offensive or potentially discrimina-

tory content. For example, “Brown woman bad” is rewritten as “They are saying a brown woman is bad. I feel angry and upset.”, which shifts from a direct racist statement to a meta-commentary on that statement, explicitly judging it. ChatGPT-4o and ChatGPT 5 improve fluency and make emotions explicit, but they tend to add extra cues (feelings sentences, emojis, softening or reframing of attitudes) and offer limited help with implicit references, abbreviations or idiomatic language.

DeepSeek V3.2 shows a noticeably different pattern, especially around swear words and toxicity. It rarely repeats strong offensive words and instead rephrases it in terms of emotional states. For example, posts that contain *fuck* or similar words are often rewritten as short first-person statements of feeling, such as “Fuck my life” becoming “Feeling hopeless. Everything is going wrong for me.” or “Move bitch get out the way.” being rendered as “The person is angry and frustrated. They are shouting: “Move! Get out of my way!””. In more complex hostile content, like “And everybody clapped! Fuck this loser!”, DeepSeek suppresses the insult (“They are saying that a story someone told is not true. They think the person is lying to seem important. The feeling is anger and disbelief.”). Similarly, “Brown woman bad” is turned into “I am angry and upset. A woman with brown skin is being called a bad person.”. Across the examples, DeepSeek is more aggressive than GPT-4o in filtering swear words and slurs, replacing them with descriptions of anger, disgust or frustration and often adding an angry emoji in the end of the sentence. This behaviour aligns with a strong safety layer and may be preferable for reducing exposure to offensive vocabulary, but it further distances the output from the original emotion and can blur the distinction between reporting a harmful statement and expressing the models own stance.

Gemini 2.5 Flash is less well aligned with the prompt. On many instances it switches from rewriting to explaining or commenting, or it asks for more context instead of producing a self-contained simplified post. For instance, when given the very short insult “An ugly fuk”, this model first responds that the post seems incomplete and asks for the full text, then offers a meta-description such as “the meaning is a person is calling someone else an ugly curse word” before giving the simplified text. For “fucking fuck fuck”, it explains what the sequence of swear words means and finally suggests “I am very angry.” as a replacement. In the case of “Holy shit that SSP was beautiful”, the reference to “SSP” is lost and the post is reduced to “That food was really good.”, which removes both the swear word and the specific object of evaluation and changes the meaning completely. In other examples, Gemini produces relatively long paraphrases that merge simplification with interpretive commentary (e.g. spelling out why something is sexist or unfair). These behaviours indicate that, under our prompt, Gemini treats the task as explanation and moral evaluation rather than rewriting.

This cross-model comparison shows that the underlying model and safety configuration substantially influence how the same prompt is handled in practice. GPT-4o and ChatGPT 5 are more likely to follow the prompt and produce fluent, well-formed rewrites, but they systematically add explicit emotion labels and sometimes extra emojis or hashtags, while leaving many abbreviations, idioms and culture-specific references unexplained. DeepSeek V3.2 places more emphasis on removing or softening offensive language and reduces lexical toxicity but can obscure the original post. Gemini 2.5 Flash, by contrast, frequently shifts into explanatory or advisory mode and occasionally loses important details, making its outputs unsuitable as simple, accessible substitutes for the original posts. However, the Gemini 2.5 Flash model is smaller than the OpenAI’s models tested in this paper.

4.4.1 Results on alternative prompts

In addition to the main accessibility prompt, we experimented with several alternative prompts across all models. These variants were also applied to a small subset of posts and were motivated by specific problems observed with the original prompt: over-production of “I feel X” sentences, addition of new emojis and hashtags, and lack of explanation for abbreviations, slang and idiomatic expressions.

One group of alternatives targeted the explicit emotion clause using the prompt presented in Figure A.5. Removing the instruction to “also name the feeling in words”, or adding a prohibition such as “do not add feelings or any assumptions about how a person feels”, reduced but did not fully eliminate first-person emotion statements in ChatGPT 5 and ChatGPT-4o. In some runs, the models switched from, for example, “I feel disgusted.” to more implicit intensifiers (e.g. “Fear”, “Feeling:amused”, “(Emotion: anger)”), but in few in-

stances they still added sentences such as "I feel happy for you". This shows that the model does not strictly obey prompts. Its behaviour is shaped by context and the underlying safety-tuned policy, not only by the prompt (Kung and Peng, 2023).

The second group of prompts focused on emojis and hashtags (Figure A.6). We removed or softened the original instructions (for example, omitting the CamelCase clause or changing the wording about adding new hashtags or emojis). In some cases, expressions such as "lol" were still replaced by emojis, or new emojis were introduced even when the original post did not contain any. However, when the part of the prompt referring to hashtags was removed entirely, the models typically did not introduce new hashtags at all, but sometimes used an emoji instead of a hashtag.

Finally, we tested a prompt that explicitly asked the model to explain or expand abbreviations, famous people or events (Figure A.7). These variants sometimes produced helpful expansions such as spelling out the meaning of "lmao" or clarifying some events or places, but the behaviour was inconsistent: in many cases, compressed jokes, memes and culture-specific references remained unexplained, or were paraphrased only partially. We plan to experiment with more advanced prompts that can produce better explanations for the posts.

Overall, the alternative prompts helped diagnose which aspects of the behaviour are prompt-sensitive and which are largely determined by the underlying model. They show that some issues can be mitigated, for example, slightly fewer emojis or more literal paraphrases, but that core tendencies, for instance, adding explicit emotion statements and using emojis persist across prompt variants. This shows that prompt design can steer, but not fully control, accessible post simplification, and that model choice and safety configuration remain crucial factors.

5 Discussion and conclusion

This paper explores the use of LLMs for simplifying social media posts. Our experiments show that LLM-based simplification can often preserve the perceived emotion of social media posts, but preservation is not guaranteed and varies with the LLM. Comparison between the emotion in the original and the simplified versions shows that in 72.5% cases rewrites retain the same emotion category, especially for frequent classes such as *joy* and *surprise*. Stability is lower for *anger*, *disgust*, and particularly *neutral*, which aligns with qualitative observations that simplification can often shift a neutral description towards a negative emotion. Whilst distortion of emotion changes the meaning, a preliminary analysis revealed there are also cases where the meaning is changed due to the fact that information is added or removed without having an impact on the overall emotion. Moreover, in several cases it was difficult to decide whether the information was preserved, as the lack of context made the original post hard to interpret. In future work, we plan to conduct a larger and more systematic analysis to better understand how to design prompts that preserve not only emotional content, but also the full informational meaning.

The cross-model comparisons we carried out indicate that model choice and safety configuration affect outcomes. GPT-4o and ChatGPT 5 behave similarly under the same instructions, whereas DeepSeek V3.2 appears more sensitive to hostile content, and Gemini 2.5 Flash often shifts into an explanatory register. This suggests that LLM-based accessibility rewriting is not a uniform capability: even with the same prompt, different models can produce outputs that vary in faithfulness to the source and handling of offensive language, hashtags or emojis. Since the differences between GPT-4o and ChatGPT 5 were not critical for the main analyses, the more cost-effective option was used for large-scale experiments.

We also run ChatGPT-4o and ChatGPT 5 several times using the same prompt on the same posts in order to assess how stable the results were. We noticed that the simplified posts did not differ too much from run to run which gives us confidence that the results presented in this paper are reliable and robust, suggesting that the observed patterns are not artifacts of randomness in model sampling.

Prompt-based LLM simplification shows clear potential to make emotionally charged social media posts easier to read. However, it should not be treated as a fully reliable solution without additional control. Emotion preservation is not consistently reliable across models and settings. Safety configurations and default rewriting behaviours can introduce subtle changes in wording and tone that shift how a post is interpreted. More advanced approaches such as using a cascade of LLMs which simplify and assess the content, or fine-tuning will be explored in future research.

References

AbilityNet. 2023. Four Ways to Make Emojis Accessible. https://abilitynet.org.uk/news-blogs/four-ways-make-emojis-accessible.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, 5th edition. American Psychiatric Publishing.

Sian Anderson, Tal Araten-Bergman, and Gillian Steel. 2023. Adults with intellectual disabilities as users of social media: A scoping review. *British Journal of Learning Disabilities*, 51(4):544–564.

Laura-Ana-Maria Bostan and Roman Klinger. 2018. An analysis of annotated corpora for emotion classification in text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2104–2119, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Jo Button. 2021. Learning to Make Twitter Content More Accessible. https://digital.canada.ca/2021/03/12/learning-to-make-twitter-content-more-accessible/. Canadian Digital Service blog.

Sue Caton, Chris Hatton, Amanda Gillooly, Edward Oloidi, Libby Clarke, Jill Bradshaw, Samantha Flynn, Laurence Taggart, Peter Mulhall, Andrew Jahoda, Roseann Maguire, Anna Marriott, Stuart Todd, David Abbott, Stephen Beyer, Nick Gore, Pauline Heslop, Katrina Scior, and Richard P Hastings. 2022. Online social connections and internet use among people with intellectual disabilities in the united kingdom during the covid-19 pandemic. *New Media & Society*, 26(5):2804–2828.

Darren Chadwick, Kristin Alfredsson Ågren, Sue Caton, Esther Chiner, Joanne Danker, Marcos Gómez-Puerta, Vanessa Heitplatz, Stefan Johansson, Claude L Normand, Esther Murphy, and 1 others. 2022. Digital inclusion and participation of people with intellectual disabilities during covid-19: A rapid review and international bricolage. *Journal of Policy and Practice in intellectual Disabilities*, 19(3):242–256.

Jeanne Sternlicht Chall and Edgar Dale. 1995. Readability revisited : the new dale-chall readability formula. In *Readability Revisited: The New Dale-Chall Readability Formula*.

Change. 2016. How to make information accessible: A guide to producing easy read documents. Technical report, CHANGE People.

Bronwen Davies. 2013. *Emotional perception and regulation and their relationship with challenging behaviour in people with a learning disability*. PhD dissertation, Cardiff University.

Andreea Maria Deleanu, Constantin Orasan, and Sabine Braun. 2024. Accessible communication: a systematic review and comparative analysis of official english easy-to-understand (e2u) language guidelines. In *Proceedings of the 3rd Workshop on Tools and Resources for People with REAding DIfficulties (READI)@ LREC-COLING 2024*, pages 70–92.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*.

Paul Ekman. 1992. Are there basic emotions? *Psychological review*, 99(3).

Björn Engelmann, Christin Katharina Kreutz, Fabian Haak, and Philipp Schaer. 2024. ARTS: Assessing readability & text simplicity. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 14925–14942, Miami, Florida, USA. Association for Computational Linguistics.

European Commission. 2012. How to write clearly.

European Union. 2016. Directive EU 2016/2102 on the accessibility of the websites and mobile applications of public sector bodies. Official Journal of the European Union.

European Union. 2019. Directive EU 2019/882 on the accessibility requirements for products and services (european accessibility act). Official Journal of the European Union.

Rudolf Flesch. 1948. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–233.

Sarah Glencross, Jonathan Mason, Mary Katsikitis, and Kenneth Mark Greenwood. 2021. Internet use by people with intellectual disability: Exploring digital inequalitya systematic review. *Cyberpsychology, Behavior, and Social Networking*, 24(8):503–520.

Government Communication Service. 2021. Planning, creating and publishing accessible social media campaigns. https://www.communications.gov.uk/guidance/digital-communication/planning-creating-and-publishing-accessible-social-media-campaigns/.

Silvia Hansen-Schirra and Christiane Maaß, editors. 2020. *Easy language research: text and user perspectives*, volume 2 of *Easy Plain Accessible*. Frank & Timme.

Harvard University. 2023. Social media accessibility best practices. https://www.harvard.edu/in-focus/the-accessible-world/social-media-accessibility-best-practices/.

Hertfordshire County Council. 2018. Easy read guidance and checklist. Technical report, Hertfordshire County Council.

Inclusion Europe. 2010. Information for all: European standards for making information easy to read and understand. Guideline document.

J. P. Kincaid, R. P. Fishburne, R. L. Rogers, and B. S. Chissom. 1975. Derivation of new readability formulas for navy enlisted personnel. Technical report, Naval Technical Training Command.

Po-Nien Kung and Nanyun Peng. 2023. Do models really learn to follow instructions? An Empirical Study of Instruction Tuning. *arXiv preprint arXiv:2305.11383*.

Karen Mellor and Dave Dagnan. 2005. Exploring the concept of alexithymia in the lives of people with learning disabilities. *Journal of Intellectual Disabilities*, 9(3):229–239.

Mencap. 2000. Am I Making Myself Clear? Mencap's Guidelines for Accessible Writing. Technical report, Mencap.

Mencap. 2021. Let's Make Social Media More Accessible. https://www.mencap.org.uk/blog/lets-make-social-media-more-accessible.

Mencap. 2022. Mencap social media accessibility guidelines. https://www.mencap.org.uk/resource/mencap-social-media-accessibility-guidelines.

Gyda Skat Nielsen Misako Nomura and Bror Tronbacke. 2010. Ifla guidelines for easy-to-read materials.

Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. SemEval-2018 task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 1–17, New Orleans, Louisiana. Association for Computational Linguistics.

NHS England. 2017. Personalised health and care: Information for people and families. Guidance document. Integrated Personal Commissioning (IPC).

Office for Health Improvement & Disparities (OHID). 2025. Learning disability Applying All Our Health. https://www.gov.uk/government/publications/learning-disability-applying-all-our-health/learning-disabilities-applying-all-our-health. Updated 6 January 2025.

Sara Owen and Frances A Maratos. 2016. Recognition of subtle and universal facial expressions in a community-based sample of adults classified with intellectual disability. *Journal of Intellectual Disability Research*, 60(4):344–354.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Elisa Perego. 2020. *Accessible communication: A cross-country journey*, volume 4 of *Easy Plain Accessible*. Frank & Timme.

Readability Guidelines. 2020. Using emojis. https://readabilityguidelines.co.uk/images/emojis/.

Janice Redish. 2000. Readability formulas have even more limitations than klare discusses. *ACM J. Comput. Doc.*, 24(3):132137.

Eleanor Rowell. 2021. Accessibility for all: 8 ways to make your social media content more accessible. https://blogs.edgehill.ac.uk/learningedge/2021/07/06/accessibility-for-all-8-ways-to-make-your-social-media-content-more-accessible/. Edge Hill University Digital Learning blog.

Hadeel Saadany, Constantin Orasan, Rocio Caro Quintana, Felix Do Carmo, and Leonardo Zilio. 2023. Analysing mistranslation of emotions in multilingual tweets by online MT tools. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 275–284, Tampere, Finland. European Association for Machine Translation.

Scope Australia. 2015. Clear written communications: The easy english style guide. Technical report, Scope (Aust) Ltd.

Jennifer L Scotland, Jill Cossar, and Karen McKenzie. 2015. The ability of adults with an intellectual disability to recognise facial expressions of emotion in comparison with typically developing individuals: a systematic review. *Research in developmental disabilities*, 41:22–39.

Sprout Social. 2024. 10 guidelines to make social media posts more accessible. https://sproutsocial.com/insights/social-media-accessibility/.

Carlo Strapparava and Rada Mihalcea. 2007. SemEval-2007 task 14: Affective text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 70–74, Prague, Czech Republic. Association for Computational Linguistics.

Paraskevi Triantafyllopoulou, Jessie Newsome, Winnie Tsang, Michelle McCarthy, and Karen Jones. 2025. Safer online lives: Internet use and online experiences of adults with intellectual disabilitiesa survey study. *Journal of Applied Research in Intellectual Disabilities*, 38(3):e70061.

UK Association for Accessible Formats (UKAAF). 2020. G028: Social media guidance. https://www.ukaaf.org/wp-content/uploads/2021/03/G028-UKAAF-Social-Media-Guidance-December-2020.pdf.

United States Congress. 2010. Plain writing act of 2010. Public Law 111–274.

University of Edinburgh. 2022. Social media accessibility guidance. https://information-services.ed.ac.uk/help-consultancy/accessibility/creating-materials/social-media-accessibility-guidance.

University of Reading. 2023. Accessibility tips: Social media posts. https://www.reading.ac.uk/digital-accessibility/blog/blog-2023/social-media-posts.

Johanna LL van Alem, Noud Frielink, and Petri JCM Embregts. 2025. Social internet use by people with intellectual disabilities: A systematic review and thematic synthesis of qualitative studies. *Journal of Intellectual Disability Research*, 69(4):243–264.

Albert Xu, Eshaan Pathak, Eric Wallace, Suchin Gururangan, Maarten Sap, and Dan Klein. 2021. Detoxifying language models risks marginalizing minority voices. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2390–2397, Online. Association for Computational Linguistics.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415.

Hikaru Yamanaka and Takenobu Tokunaga. 2024. SIERA: An evaluation metric for text simplification using the ranking model and data augmentation by edit operations. In *Proceedings of the 3rd Workshop on Tools and Resources for People with REAding DIfficulties (READI) @ LREC-COLING 2024*, pages 47–58, Torino, Italia. ELRA and ICCL.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

A Appendix

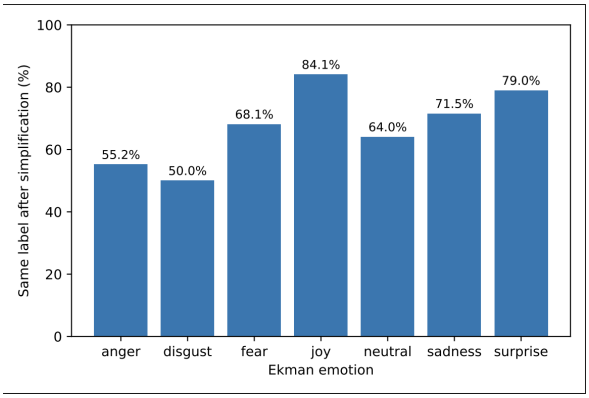

Figure A.1: Label stability between *pred_orig* and *pred_simp*: percentage of items for which the classifier assigns the same Ekman emotion before and after simplification.

Ekman Emotion	Precision	Recall	F_1-score
anger	0.50	0.65	0.57
disgust	0.52	0.53	0.53
fear	0.61	0.76	0.68
joy	0.77	0.88	0.82
neutral	0.66	0.67	0.66
sadness	0.56	0.62	0.59
surprise	0.53	0.70	0.61
macro-average	0.59	0.69	0.64
std	0.10	0.11	0.10

Table A.3: Ekman-level BERT baseline on GoEmotions (from Demszky et al. (2020)) for comparison with our XLM-RoBERTa classifier.

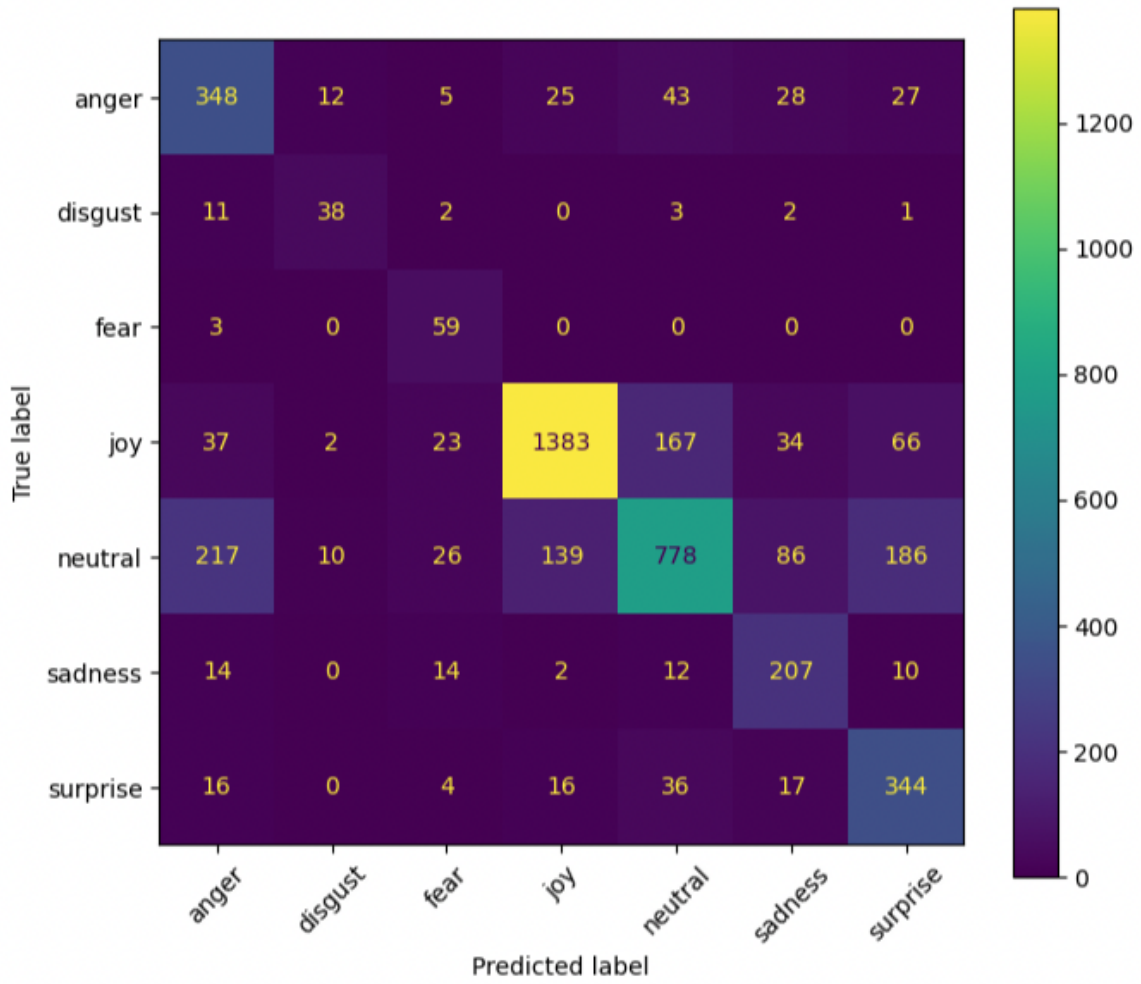

Figure A.2: Confusion-matrix heatmap for the XLM-RoBERTa classifier on the GoEmotions dataset.

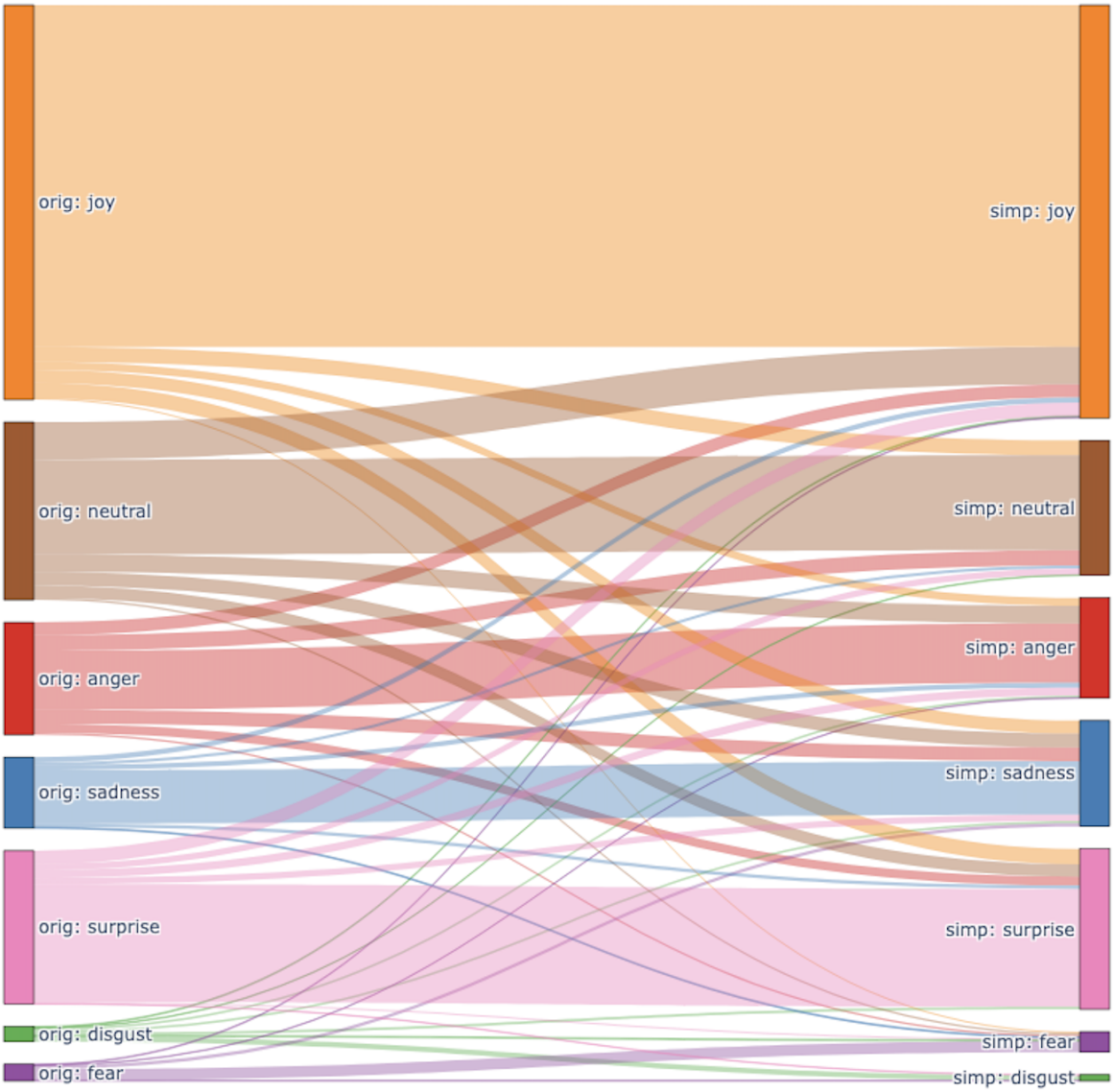

Figure A.3: Alluvial diagram showing flows from emotion labels predicted on the original posts (*pred_orig*, left) to labels predicted on the simplified posts (*pred_simp*, right).

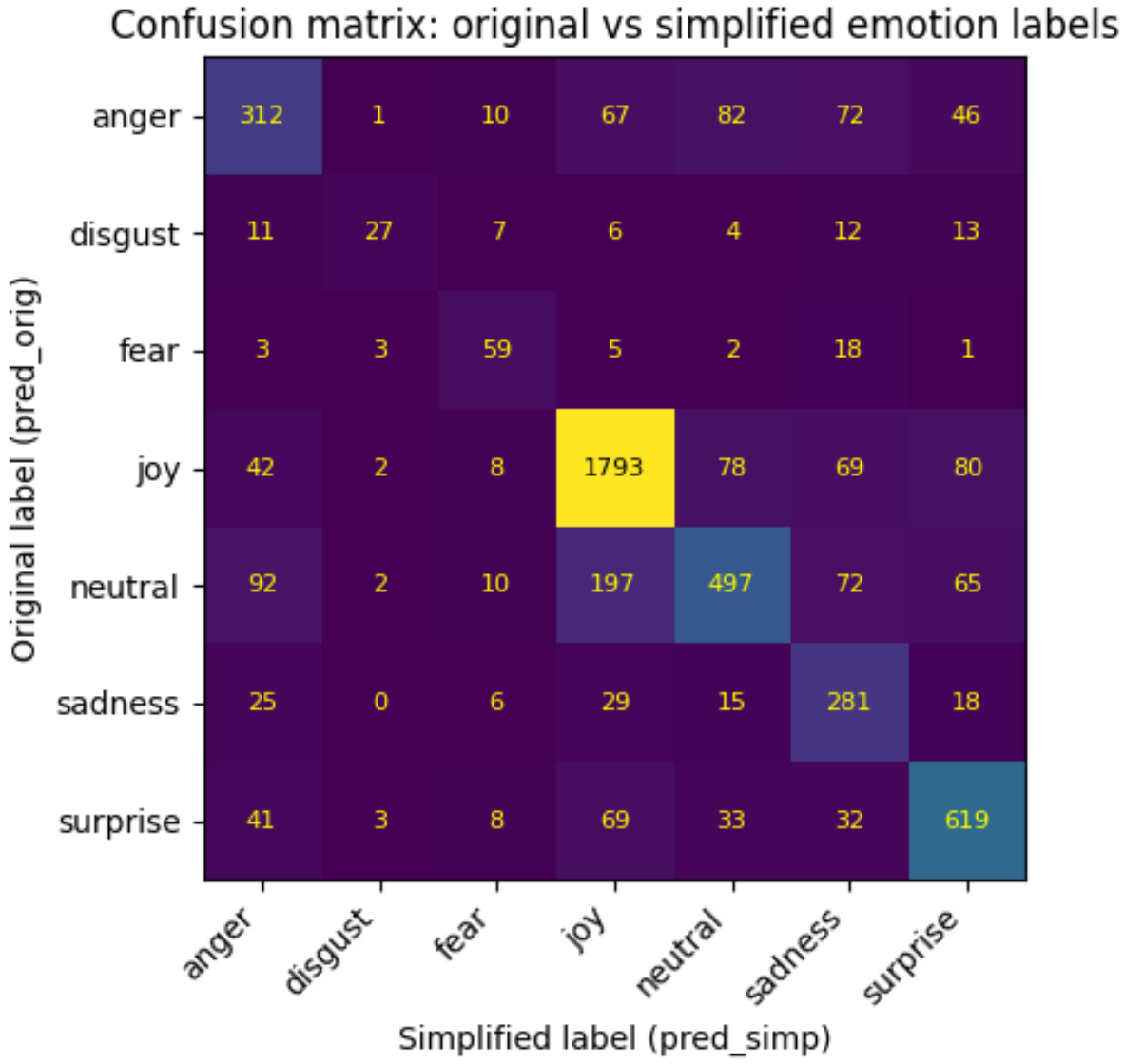

Figure A.4: Confusion matrix showing counts of emotion labels predicted on the original posts (*pred_orig*) versus the simplified posts (*pred_simp*).

Simplify the post so that people with learning disabilities can easily understand it. Keep the same meaning and facts. Preserve the same emotion. Do not soften or exaggerate the emotion. Do not add new facts or advice. Do not judge the person. Use common words and active voice. Keep emojis only if they add meaning. Use CamelCase for hashtags. For example, instead of #learningdisabilityweek, write #LearningDisabilityWeek. Do not add feelings or any assumptions about how a person feels. Keep simple posts as they are, even though they contain swear words. Explain all abbreviations, famous people, events or any other entities. Do not add hashtags or emojis

Simplify the post so that people with learning disabilities can easily understand it. Keep the same meaning and facts. Preserve the same emotion. Do not soften or exaggerate the emotion. Make the feelings clear and simple. Do not add new facts or advice. Do not judge the person. Use common words and active voice. Keep emojis only if they add meaning. Use CamelCase for hashtags. For example, instead of #learningdisabilityweek, write #LearningDisabilityWeek. Do not add "I feel". Keep simple posts as they are, even though they contain swear words. Explain all abbreviations, famous people, events or any other entities. Do not add hashtags or emojis.

Figure A.5: Prompt variant removing the instruction to "also name the feeling in words"

.

Simplify the post so that people with learning disabilities can easily understand it. Keep the same meaning and facts. Preserve the same emotion. Do not soften or exaggerate the emotion. Make the feelings clear and simple. Do not add new facts or advice. Do not judge the person. Use common words and active voice. Keep emojis only if they add meaning, and also name the feeling in words. Do not assume how the person feels.

Simplify the post so that people with learning disabilities can easily understand it. Keep the same meaning and facts. Preserve the same emotion. Do not soften or exaggerate the emotion. Make the feelings clear and simple. Do not add new facts or advice. Do not judge the person. Use common words and active voice. Keep emojis only if they add meaning. Do not assume how the person feels.

Figure A.6: Prompt variant removing the instruction about hashtag or emoji.

You are an accessibility editor for social media.
GOAL: Make the posts easy to read without changing the original emotion or facts.
CONSTRAINTS: (1) Short sentences; one idea per sentence. (2) Keep names, numbers, links; keep hashtags (CamelCase). (3) No advice, opinions, or extra facts. (4) Do not judge the person. (5) Keep emojis only if they were in the original text and keep emojis at the end of the message (6) Explain all abbreviations, famous people, events or any other entities.
EMOTION: Keep the same emotion and intensity. Make the feelings clear and simple. Do not add feelings or any assumptions about how a person feels.
OUTPUT: Only the simplified text.

Figure A.7: Different prompt with additional instructions