Cleaned data:

participants_data.csv:

This dataset contains all the information needed to recreate the results of the paper. It contains 893 rows of data, with each row corresponding to one participant in the note-writing experiment.

Variable	Type	Explanation
code	String	The unique ID that belongs to the participant
tweet	String	The unique ID that belongs to the post that is shown to the participant. This ID can be used to find the post on x.com
tweet_part	String	The partisanship of the post. D indicates that the post is presenting Democratic ideology, and R indicates that the post is presenting Republican ideology.
gpt_note	String	The feedback provided by the GPT model.
partisanship	String	The self-reported partisanship of the participant. D indicates Democrat, and R indicates Republican.
partisanship_cont	int	Users were asked to identify their partisanship on a scale from strongly Democrat to strongly Republican. This number refers to the position participants chose. -4 is strongly Democrat, and +4 is strongly Republican. Participants who chose 0 (neutral) were screened out of the experiment.
initial_note	String	The initial note that was written by the participant.
initial_tag	String	The initial tag that the participant chose for the post, before receiving the feedback. (Not_Misleading, Misleading)
treatment	String	The type of feedback that was given by the GPT model. (neutral, support, argue)
grouping	String	The label of the source of the feedback that was shown to the participants. (AI, human)
mh_Second_1 to **mh_Second_10**	int	After the experiment, participants were asked about their comfort level with technology. The answers are stored in these fields. Participants were given 6 different choices.

		From “Extremely uncomfortable” (1) to “Extremely comfortable” (5), with the additional “I am not familiar with the technology” (6). The questions are listed in the Supplemental Materials.
duration	float	The number refers to how long each participant was in the experiment (in minutes).
attention_passed	binary	The variable refers to whether or not the participant has passed the attention check. All the participants in this dataset have passed the attention check
sim_score	float	The number indicates the similarity score between the initial and the final note. The higher the number, the higher the similarity. Notes with a higher similarity score than 0.96 are considered similar (see *Similar*).
initial_length	int	The number of characters in the initial note. Participants were told to write less than 280 characters.
final_length	int	The number of characters in the final note. Participants were told to write less than 280 characters.
initial_links	int	The number of hyperlinks used in the initial note.
final_links	int	The number of hyperlinks used in the final note.
initial_std_level	int	The number refers to the sophistication score given to the initial note based on the United States education system. More information is available in the Methods section.
final_std_level	int	The number refers to the sophistication score given to the final note based on the United States education system. More information is available in the Methods section.
initial_stats	binary	This variable indicates whether or not a number was used in the initial note. (1 if true and 0 if false.)
final_stats	binary	This variable indicates whether or not a number was used in the final note. (1 if true and 0 if false.)
mh_level	float	This variable indicates the level of comfort the participant has with technology, based on the post-experiment survey. 1 means that the

		participant indicated that they are extremely comfortable with all the listed technologies.
length	int	The difference between the final and the initial note lengths.
links	int	The difference between the number of links used in the final and the initial note.
std_level	int	The difference between the final and the initial note sophistication score.
stats	int	The difference between the final and the initial note stats variable. (-1,0,1)
similar	binary	The value indicates if the similarity score between the initial and the final note is higher than 0.96.
initial_note_id	int	The unique ID of the initial note.
final_note_id	int	The unique ID of the final note.
helpfulness_d_initial	float	The average helpfulness score of the initial note given by the Democrat raters.
helpfulness_d_final	float	The average helpfulness score of the final note given by the Democrat raters.
helpfulness_r_initial	float	The average helpfulness score of the initial note given by the Republican raters.
helpfulness_r_final	float	The average helpfulness score of the final note given by the Republican raters.
d_bias_by_d_initial	float	The average bias score favouring Democrats of the initial note given by the Democrat raters.
r_bias_by_d_initial	float	The average bias score favouring Republicans of the initial note given by the Democrat raters.
d_bias_by_r_initial	float	The average bias score favouring Democrats of the initial note given by the Republican raters.
r_bias_by_r_initial	float	The average bias score favouring Republicans of the initial note given by the Republican raters.
d_bias_by_d_final	float	The average bias score favouring Democrats of the final note given by the Democrat raters.
r_bias_by_d_final	float	The average bias score favouring Republicans of the final note given by the

		Democrat raters.
d_bias_by_r_final	float	The average bias score favouring Democrats of the final note given by the Republican raters.
r_bias_by_r_final	float	The average bias score favouring Republicans of the final note given by the Republican raters.
helpfulness_d	float	The relative difference between helpfulness_d_final and helpfulness_d_initial
helpfulness_r	float	The relative difference between helpfulness_r_final and helpfulness_r_initial
d_bias_by_d	float	The relative difference between d_bias_by_d_final and d_bias_by_d_initial
d_bias_by_r	float	The relative difference between d_bias_by_r_final and d_bias_by_r_initial
r_bias_by_d	float	The relative difference between r_bias_by_d_final and r_bias_by_d_initial
r_bias_by_r	float	The relative difference between r_bias_by_r_final and r_bias_by_r_initial
d_bias_by_d_improvement	int	This variable indicates whether or not the bias score favouring Democrats given by the Democrat raters has improved after receiving the feedback. More information on how this variable is calculated is available in the Methods section of the paper. (-1, 0, 1)
d_bias_by_r_improvement	int	This variable indicates whether or not the bias score favouring Democrats given by the Republican raters has improved after receiving the feedback. More information on how this variable is calculated is available in the Methods section of the paper. (-1, 0, 1)
r_bias_by_d_improvement	int	This variable indicates whether or not the bias score favouring Republican sgiven by the Democrat raters has improved after receiving the feedback. More information on how this variable is calculated is available in the Methods section of the paper. (-1, 0, 1)
r_bias_by_r_improvement	int	This variable indicates whether or not the bias score favouring Republicans given by the Republican raters has improved after receiving the feedback. More information on how this variable is calculated is available in the Methods section of the paper. (-1, 0, 1)

improvement_d	int	This variable indicates whether the note has improved or declined in helpfulness according to the Democrat raters. More information on how this variable is calculated is available in the Methods section of the paper. (-1, 0, 1)
improvement_r	int	This variable indicates whether the note has improved or declined in helpfulness according to the Republican raters. More information on how this variable is calculated is available in the Methods section of the paper. (-1, 0, 1)
gpt_sim_score	float	The number indicates the similarity score between the feedback and the final note. The higher the number, the higher the similarity. This variable is referred to as **the feedback acceptance rate** in the paper.

notes_data.csv:

This dataset contains all the unique notes gathered throughout the note-writing experiment alongside the evaluations given to the notes by raters. The dataset contains 2001 rows, each row refers to a unique note. Notes that were considered similar only appear once and were evaluated once.

Variable	Type	Explanation
noteId	String	The unique ID that belongs to the participant who has written this note.
id	String	The unique ID of the note.
note	String	The text of the note.
tweet	String	The unique ID that belongs to the post that is shown to the participant. This ID can be used to find the post on x.com
note_type	String	The type of the note. (initial_note, final_note)
examined_by_R	int	The number of times this note was evaluated by Republican raters.
examined_by_D	int	The number of times this note was evaluated by Democrat raters.
helpfulness_list_d	List of floats	The normalised helpfulness scores given to the notes by Democrat raters. More

		information on how this variable was calculated is available in the Methods section.
helpfulness_list_r	List of floats	The normalised helpfulness scores given to the notes by Republican raters. More information on how this variable was calculated is available in the Methods section.
helpfulness_d	float	The average helpfulness score of the note given by the Democrat raters. More information on how this variable was calculated is available in the Methods section.
helpfulness_r	float	The average helpfulness score of the note given by the Republican raters. More information on how this variable was calculated is available in the Methods section.
dict_of_helpfulness	A dictionary with the key an integer and the value a list of floats	Raters were asked to identify their partisanship on a scale from strongly Democrat to strongly Republican. The key in the dictionary refers to the position the rater chose. -4 is strongly Democrat, and +4 is strongly Republican. The values indicate the normalised helpfulness score given to the note from raters with that level of partisanship.
d_bias_list_r	List of floats	The normalised bias scores favouring Democrats that are given to the notes by Republican raters.
r_bias_list_r	List of floats	The normalised bias scores favouring Republicans that are given to the notes by Republican raters.
d_bias_list_d	List of floats	The normalised bias scores favouring Democrats that are given to the notes by Democrat raters.
r_bias_list_d	List of floats	The normalised bias scores favouring Republicans that are given to the notes by Democrat raters.
r_bias_by_r	float	The average bias score favouring Republicans given to the note by Republican raters.
r_bias_by_d	float	The average bias score favouring Republicans given to the note by Democrat raters.
d_bias_by_r	float	The average bias score favouring Democrats given to the note by Republican raters.
d_bias_by_d	float	The average bias score favouring Democrats

		given to the note by Democrat raters.

Raw files:

The raw files from the note writing and the note evaluation experiments are available in “note_writing_raw_file.csv” and “evaluation_raw_file.csv”, respectively.