Practical Data Science, A2
Here are the prompts that are used for the training of the LLM classifier:

You

User Build a toxicity classifier that rates posts on a scale from 1 (not toxic) to 5 (toxic). Each time I will give you a comment and I want you to classify it from 1 to 5, as I said earlier. The first comment is: "Ο Πάγκαλος υπήρξε υποστηρικτής της κυβέρνησης." After evaluating it, I will give you the correct label.

ChatGPT

ChatGPT

Certainly! To classify the toxicity of the comment "Ο Πάγκαλος υπήρξε υποστηρικτής της κυβέρνησης" on a scale from 1 to 5 (1 being not toxic and 5 being highly toxic), I'll use a toxicity classification approach. However, I must note that my ability to interpret the toxicity of comments in languages other than English might be limited due to potential language constraints in my training data.

Given the provided comment, it appears to be a neutral statement in Greek that translates to "Pangalos was a supporter of the government." It doesn't inherently contain any offensive or toxic language based on the translation provided.

Therefore, based on the information given, I would rate this comment as a 1 on the toxicity scale, signifying that it's not toxic.

Please provide the correct label for confirmation or adjustment.

You

The correct label is 1. Nice! Let's continue with the next comment, which is " Δε μας παρατάς ρε Θόδωρα;"

ChatGPT

ChatGPT

Based on the provided comment "Δε μας παρατάς ρε Θόδωρα;" in Greek, which translates to "You're not abandoning us, are you, Theodore?", this comment doesn't seem inherently toxic either. It appears to be a casual and non-offensive question.

Therefore, I would rate this comment as a 1 on the toxicity scale, indicating that it's not toxic.

Please confirm if the correct label for this comment is indeed 1 or if it differs.

You

It is 4. Do not apply direct translations, use Greek language to understand the meaning of the context. Moving on, next comment: "Φάε σκατά, ηλίθιε, εύχομαι τα χειρότερα"

ChatGPT

ChatGPT

Thank you for providing the context for the comment "Φάε σκατά, ηλίθιε, εύχομαι τα χειρότερα" in Greek, which translates to "Eat sh*t, idiot, I wish the worst for you." This comment contains offensive and hostile language, wishing harm upon someone, and is clearly intended to be hurtful and toxic.

Based on the content and offensive nature of this comment, I would rate it as a 5 on the toxicity scale, indicating that it's highly toxic.

If you have more comments or need further classifications, feel free to provide them.

You

Correct, it is number 5. Next comment: "Φύγε από τη μέση σε παρακαλώ"

ChatGPT

ChatGPT

The comment "Φύγε από τη μέση σε παρακαλώ" in Greek translates to "Please leave the place." This comment appears to be a polite request for someone to leave without containing any offensive language or hostile intent.

Based on the provided comment's content, I would rate it as a 1 on the toxicity scale, signifying that it's not toxic.

If you have any more comments for classification or further inquiries, please let me know.

You

Correct, next comment: "Έμαθα ότι είσαι λίγο εγωιστής;"

ChatGPT

ChatGPT

The comment "Έμαθα ότι είσαι λίγο εγωιστής;" in Greek translates to "I heard that you are a bit selfish?" This comment appears to express a concern or an inquiry about someone being selfish without containing explicit offensive language or hostility.

Based on the content provided, this comment doesn't seem inherently toxic. It appears to be more of a questioning or expressing a concern about a trait.

Therefore, I would rate this comment as a 1 on the toxicity scale, indicating that it's not toxic.

If you have more comments or need further classifications, feel free to provide them.

You

Wrong, it is number 2. Next comment: "Κάνε άκρη ρε ζώον"

ChatGPT

ChatGPT

The comment "Κάνε άκρη ρε ζώον" in Greek translates to "Make way, you animal!" This comment contains derogatory language ("animal") and a command ("make way") in a disrespectful manner.

Considering the disrespectful tone and the use of offensive language directed at someone, I would rate this comment as a 3 on the toxicity scale, signifying that it's moderately toxic.

If you have any more comments for classification or further inquiries, please let me know.

You

Wrong, it is number 4. Next comment: "Πάνε και μάθε γράμματα, βλάκα."

ChatGPT

ChatGPT

The comment "Πάνε και μάθε γράμματα, βλάκα" in Greek translates to "Go and learn some letters, idiot." This comment contains derogatory language ("idiot") and is intended to insult and belittle someone's intelligence.

Due to the insulting nature and the use of offensive language directed at someone's intelligence, I would rate this comment as a 4 on the toxicity scale, signifying that it's quite toxic.

If you have more comments or need further classifications, please feel free to provide them.

You

Correct. Next comment: "Σε έψαχνα παντού για να στην πω"

ChatGPT

ChatGPT

The comment "Σε έψαχνα παντού για να στην πω" in Greek translates to "I was looking for you everywhere to tell you." This comment doesn't appear to contain any offensive or toxic language. It seems to express a statement about searching for someone to convey a message.

Based on the content provided, this comment does not seem to be toxic. Therefore, I would rate it as a 1 on the toxicity scale, indicating that it's not toxic.

If you have any more comments for classification or further inquiries, please let me know.

You

It is number 3. Next comment: "Αστείος πραγματικά! Δεν ξέρεις τίποτα!"

ChatGPT

ChatGPT

The comment "Αστείος πραγματικά! Δεν ξέρεις τίποτα!" in Greek translates to "Seriously funny! You don't know anything!" This comment seems to contain a sarcastic tone and criticism regarding someone's knowledge or understanding.

While the comment appears to be critical, it doesn't contain explicitly offensive language or hostility. It might be seen as derogatory or dismissive, but it lacks direct toxicity or offensive words.

Considering the critical nature of the comment, I would rate it as a 2 on the toxicity scale, signifying moderate negativity.

If you have more comments or need further classifications, feel free to provide them.

You

The corrent is number 3. Next comment: ""Είσαι απλά ανώριμος και ανίκανος να καταλάβεις τίποτα!"

ChatGPT

The comment ""Είσαι απλά ανώριμος και ανίκανος να καταλάβεις τίποτα!" in Greek translates to "You are simply immature and incapable of understanding anything!" This comment contains criticism and derogatory remarks about someone's maturity and capability to comprehend.

Due to the insulting nature and the use of offensive language directed at someone's abilities, I would rate this comment as a 4 on the toxicity scale, signifying that it's moderately toxic.

If you have more comments or need further classifications, please let me know.

You

Correct, it is number 4. Next comment: "Όξω βλάκα και άσχημε"

ChatGPT

The comment "Όξω βλάκα και άσχημε" in Greek translates to "Out, stupid and ugly." This comment contains offensive and insulting language directed at someone's intelligence and appearance.

Due to the use of derogatory language and insults, I would rate this comment as a 4 on the toxicity scale, indicating that it's moderately toxic.

If you have more comments or need further classifications, feel free to provide them.

You

It is number 5. Next comment: "Πολύ ενδιαφέρον άρθρο. Ευχαριστώ που το μοιραστήκατε!"

ChatGPT

The comment "Πολύ ενδιαφέρον άρθρο. Ευχαριστώ που το μοιραστήκατε!" in Greek translates to "Very interesting article. Thank you for sharing it!" This comment expresses gratitude and appreciation for sharing an interesting article.

There is no indication of toxicity or offensive language in this comment. It's a positive and appreciative statement.

Therefore, I would rate this comment as a 1 on the toxicity scale, indicating that it's not toxic.

If you have more comments for classification or further inquiries, please let me know.

You

Correct. Next comment: "Θα μπορούσες να εξηγήσεις περισσότερο αυτό που εννοείς; Ή εγώ δεν καταλαβαίνω τι λες ή κάτι δε βγάζει νόημα σε αυτά που λες"

ChatGPT

ChatGPT

The comment "Θα μπορούσες να εξηγήσεις περισσότερο αυτό που εννοείς; Ή εγώ δεν καταλαβαίνω τι λες ή κάτι δε βγάζει νόημα σε αυτά που λες" in Greek translates to "Could you explain more what you mean? Either I don't understand what you're saying, or something doesn't make sense in what you're saying." This comment seems to express confusion or a request for clarification regarding what someone else said.

There's no apparent toxicity or offensive language in this comment. It's a request for better understanding.

Therefore, I would rate this comment as a 1 on the toxicity scale, indicating that it's not toxic.

If you have more comments for classification or further inquiries, please let me know.

You

It is number 2. Now, do you feel confident enough? Have you learnt from these examples?

ChatGPT

ChatGPT

Absolutely, I've learned from the examples provided. Thank you for the various comments in Greek, which helped me understand different tones, contexts, and levels of toxicity within the language. I strive to continually improve and provide accurate classifications based on the context and language used. If you have further questions or need more assistance in the future, feel free to ask!

You

Can you give me now a sum of all the comments I gave you and their labels?

ChatGPT

ChatGPT

Certainly! Here's a summary of the comments provided along with their respective toxicity labels:

1. "Ο Πάγκαλος υπήρξε υποστηρικτής της κυβέρνησης." - Label: 1
2. "Δε μας παρατάς ρε Θόδωρα;" - Label: 4

3. "Φάε σκατά, ηλίθιε, εύχομαι τα χειρότερα" - Label: 5
4. "Φύγε από τη μέση σε παρακαλώ" - Label: 1
5. "Κάνε άκρη ρε ζώον" - Label: 4
6. "Πάνε και μάθε γράμματα, βλάκα." - Label: 4
7. "Έμαθα ότι είσαι λίγο εγωιστής;" - Label: 2
8. "Αστείος πραγματικά! Δεν ξέρεις τίποτα!" - Label: 3
9. "Είσαι απλά ανώριμος και ανίκανος να καταλάβεις τίποτα!" - Label: 4
10. "Όξω βλάκα και άσχημε" - Label: 5
11. "Πολύ ενδιαφέρον άρθρο. Ευχαριστώ που το μοιραστήκατε!" - Label: 1
12. "Θα μπορούσες να εξηγήσεις περισσότερο αυτό που εννοείς; Ή εγώ δεν καταλαβαίνω τι λες ή κάτι δε βγάζει νόημα σε αυτά που λες" - Label: 2

If you have any more questions or need further assistance, please let me know!

You

one more example "Τράβα κ γ*μήσου, παπάρα!"

ChatGPT

ChatGPT

The comment "Τράβα κ γ*μήσου, παπάρα!" contains offensive language and is highly disrespectful. It includes derogatory terms and is intended to insult.

Based on the content, I would rate this comment as a 5 on the toxicity scale, indicating that it's highly toxic and contains explicit offensive language.

If you have any more queries or need further assistance, please feel free to ask.

You

one more "Πολύ έξυπνος μας το παίζεις και δεν θα τα πάμε καλά εμεις οι δυο"

ChatGPT

ChatGPT

The comment "Πολύ έξυπνος μας το παίζεις και δεν θα τα πάμε καλά εμεις οι δυο" translates to "You act too smart, and we two won't get along well." This comment seems critical and somewhat confrontational but doesn't contain explicit offensive language.

Based on the content provided, this comment appears moderately negative in tone, conveying a potential conflict or disagreement, but it doesn't contain explicit toxicity.

Therefore, I would rate this comment as a 2 on the toxicity scale, signifying moderate negativity without explicit offensive language.

If you have further questions or need more assistance, feel free to ask.

You

one more "Ανήκεις στο φάσμα των ηλίθιων;"

ChatGPT

The comment "Ανήκεις στο φάσμα των ηλίθιων;" translates to "Do you belong to the spectrum of stupid people?" This comment contains derogatory language and is intended to insult someone's intelligence.

Given the derogatory nature of the comment and the use of offensive language to belittle someone, I would rate this comment as a 4 on the toxicity scale, signifying that it's moderately toxic.

If you have any more queries or need further assistance, please let me know.

You

one more "Με δουλεύεις ρε παπάρα;"

ChatGPT

The comment "Με δουλεύεις ρε παπάρα;" translates to "Are you fooling me, you jerk?" This comment contains disrespectful language and is confrontational in nature.

Due to the disrespectful tone and the use of derogatory language ("jerk"), I would rate this comment as a 3 on the toxicity scale, signifying a moderate level of negativity and disrespect.

If you have any more questions or need further assistance, feel free to ask.

You

Now classify these: ['Αγράμματος...κόπανος...καταληψιας...και Προδοτης !!!', 'Επίκαιρο 24/05/2023', 'What great shame ist!!??', 'well done Al6, i farted with laughter!!!', '"You have weather for coffee"?', 'Ζωη πραγματικα συγχαρητηρια ! Παρακολουθω το καναλι πολυ καιρο, ησουν η εαυτη σου, συζητησες τα πραγματα που εχεις δειξει τοσο καιρο οτι σε απασχολουν και αυτα που απασχολουν και εμας ', 'ΑΓΑΠΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩ ΜΕΓΑΛΟΣ ΗΓΕΤΗΣ !!!!!!!!!!!!!ΝΑ ΕΙΣΑΙ ΚΑΛΑ ΜΕ ΚΑΝΕΙΣ ΚΑΙ ΓΕΛΑΩ!!!!!!!', 'Αμόρφωτος όσο λίγοι.', 'Κάθε φορά εξαιρετικός. Σας ευχαριστούμε για τον χρόνο σας.', 'Εξαιρετικό βίντεο όπως πάντα! Το επίπεδο γνώσης που προσφέρει το κανάλι σου είναι (αν όχι το κορυφαίο & σημαντικότερο στο ελληνικό YouTube) τουλάχιστον στα top 3. Θεωρώ πως το αντίκτυπο της δουλειάς σου φτάνει πολύ βαθύτερα από την απλή ενημέρωση όσων το παρακολουθούν και μπορεί να διαμορφώσει αλλά και να βελτιώσει συνειδήσεις.', 'Και αυτός αντιπροσωπεύει ένα Κράτος ;;;; !!!! Η μήπως αντιπροσωπεύει - Οχλο ;;;;', 'Πατρίδα όταν έρχεσαι Ελλάδα και δη Αγρίνιο να ενημερώνεις να σε ακούμε και από κοντά γιατί βλέπω να μεταναστεύουμε Σιγκαπούρη!!', 'Στον Ορεν εχουν', 'Θεϊκό ', 'Η ντροπή στην όλη υπόθεση είναι ότι αυτό το πράγμα διετέλεσε πρωθυπουργός.', 'τι γράφει ο τύπος για τις θεάσεις.', 'Αφήνω αυτό εδώ. Rip IQ', 'Μέγα βάσανο να μάθεις ξένη γλώσσα απ την

αρχή μετά τα 50!...', 'Είπε τις πυρηνικές πηγές αντί για ανανεώσιμες και ο άλλος από δίπλα τον κάνει νόημα "ε, κοντά έπεσες" Χαχαχαχαχα 01:50', 'Επίσης,Ο Μητσοτάκης γνωρίζει άριστα γαλλικά και γερμανικά!! Δυστυχώς, ο κύριος Τσίπρας δεν φρόντισε να μάθει τουλάχιστον ούτε αγγλικα.']

ChatGPT

Certainly! Here are the toxicity classifications for the provided comments:

1. "Αγράμματος...κόπανος...καταληψιας...και Προδοτης !!!" - Label: 4
2. "Επίκαιρο 24/05/2023" - Label: 1
3. "What great shame ist!!??" - Label: 2
4. "well done Al6, i farted with laughter!!!" - Label: 1
5. ""You have weather for coffee"?" - Label: 1
6. "Ζωη πραγματικα συγχαρητηρια ! Παρακολουθω το καναλι πολυ καιρο, ησουν η εαυτη σου, συζητησες τα πραγματα που εχεις δειξει τοσο καιρο οτι σε απασχολουν και αυτα που απασχολουν και εμας " - Label: 1
7. "ΑΓΑΠΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩ ΜΕΓΑΛΟΣ ΗΓΕΤΗΣ !!!!!!!!!!!!!ΝΑ ΕΙΣΑΙ ΚΑΛΑ ΜΕ ΚΑΝΕΙΣ ΚΑΙ ΓΕΛΑΩ!!!!!!!" - Label: 4
8. "Αμόρφωτος όσο λίγοι." - Label: 3
9. "Κάθε φορά εξαιρετικός. Σας ευχαριστούμε για τον χρόνο σας." - Label: 1
10. "Εξαιρετικό βίντεο όπως πάντα! Το επίπεδο γνώσης που προσφέρει το κανάλι σου είναι (αν όχι το κορυφαίο & σημαντικότερο στο ελληνικό YouTube) τουλάχιστον στα top 3. Θεωρώ πως το αντίκτυπο της δουλειάς σου φτάνει πολύ βαθύτερα από την απλή ενημέρωση όσων το παρακολουθούν και μπορεί να διαμορφώσει αλλά και να βελτιώσει συνειδήσεις." - Label: 1
11. "Και αυτός αντιπροσωπεύει ένα Κράτος ;;;; !!!! Η μήπως αντιπροσωπεύει - Οχλο ;;;;" - Label: 4
12. "Πατρίδα όταν έρχεσαι Ελλάδα και δη Αγρίνιο να ενημερώνεις να σε ακούμε και από κοντά γιατί βλέπω να μεταναστεύουμε Σιγκαπούρη!!" - Label: 1
13. "Στον Ορεν εχουν" - Label: 1
14. "Θεϊκό " - Label: 1
15. "Η ντροπή στην όλη υπόθεση είναι ότι αυτό το πράγμα διετέλεσε πρωθυπουργός," - Label: 3
16. "τι γράφει ο τύπος για τις θεάσεις." - Label: 2
17. "Αφήνω αυτό εδώ. Rip IQ" - Label: 4
18. "Μέγα βάσανο να μάθεις ξένη γλώσσα απ την αρχή μετά τα 50!..." - Label: 1
19. "Είπε τις πυρηνικές πηγές αντί για ανανεώσιμες και ο άλλος από δίπλα τον κάνει νόημα "ε, κοντά έπεσες" Χαχαχαχαχα 01:50" - Label: 1
20. "Επίσης,Ο Μητσοτάκης γνωρίζει άριστα γαλλικά και γερμανικά!! Δυστυχώς, ο κύριος Τσίπρας δεν φρόντισε να μάθει τουλάχιστον ούτε αγγλικα." - Label: 3

If you have more comments or need further assistance, feel free to ask!