

?????? Reviewers ??? ?? glossApi

???? guide ?? ????????

- [Αναλυτικός οδηγός reviewers](#): πλήρης έκδοση με κριτήρια, καλές πρακτικές και παραδείγματα.
- [One-page guide](#): σύντομη έκδοση για γρήγορη αναφορά.
- [Οδηγός για μη τεχνικούς reviewers](#): απλή έκδοση σε καθημερινή γλώσσα.

?????? ??? review

Ο ρόλος του reviewer δεν είναι απλώς να “διορθώσει κείμενα”. Κάθε record που γίνεται `accept`, `edit` ή `reject` επηρεάζει άμεσα την ποιότητα του dataset που θα χρησιμοποιηθεί για fine-tuning. Ένα λάθος `accept` μπορεί να περάσει θόρυβο στο training. Ένα σωστό `reject` μπορεί να αφαιρέσει προβληματικά παραδείγματα. Ένα σωστό `edit` μπορεί να μετατρέψει ένα μέτριο παράδειγμα σε χρήσιμο training sample.

Στόχος είναι κάθε τελικό ζευγάρι `question` και `answer` να είναι:

- σωστό με βάση το `source_text`
- καθαρό και φυσικό στα Ελληνικά
- χρήσιμο για εκπαίδευση μοντέλου
- συνεπές με τα υπόλοιπα reviewed records

??? ?? ????????????

Για την τρέχουσα παραγωγική ροή:

- Argilla URL: `https://argilla.glossapi.gr`
- Workspace: `default`

Επιλέξτε το review dataset που σας έχει ανατεθεί και ξεκινήστε να δουλέψετε.

Αν υπάρχουν φίλτρα στο UI, προτιμήστε να δουλεύετε πρώτα στα records που είναι ακόμη pending ή δεν έχουν review response.

?? ?????? ? reviewer ?? ????? record

Το review dataset έχει τρία read-only πεδία και τρία review πεδία.

Read-only πεδία:

- `source_text`: το αρχικό κείμενο από το οποίο πρέπει να προκύπτουν η ερώτηση και η απάντηση
- `generated_question`: η ερώτηση που παρήγαγε το μοντέλο
- `generated_answer`: η απάντηση που παρήγαγε το μοντέλο

Review πεδία:

- `review_action`: `accept`, `edit`, `reject`
- `corrected_question`: διορθωμένη ερώτηση, μόνο όταν χρειάζεται
- `corrected_answer`: διορθωμένη απάντηση, μόνο όταν χρειάζεται

?????? ????? ??????????????

Ο reviewer πρέπει να στηρίζεται μόνο στο `source_text`.

Μην χρησιμοποιείτε:

- εξωτερική γνώση
- προσωπικές υποθέσεις
- γνώσεις για το μάθημα ή το βιβλίο που δεν εμφανίζονται ρητά στο κείμενο
- συμπλήρωση κενών με εικασία

Αν μια πληροφορία δεν υποστηρίζεται από το κείμενο, δεν πρέπει να παραμείνει στο τελικό training sample.

??? ?????????? ??? record

1. Διαβάστε πρώτα ολόκληρο το `source_text`.
2. Διαβάστε το `generated_question`.
3. Ελέγξτε αν η ερώτηση είναι σαφής, μία και μόνο μία, και απαντιέται από το κείμενο.
4. Διαβάστε το `generated_answer`.
5. Ελέγξτε αν η απάντηση είναι σωστή, πλήρης και αυστηρά τεκμηριωμένη από το κείμενο.

6. Επιλέξτε `Accept as-is`, `Edit` ή `Reject`.
7. Αν επιλέξετε `Edit`, συμπληρώστε μόνο τα πεδία που χρειάζονται πραγματικά αλλαγή.
8. Κάντε έναν τελευταίο έλεγχο ότι η τελική ερώτηση και η τελική απάντηση ταιριάζουν μεταξύ τους.
9. Υποβάλετε το review response.

???? ?? ??????? Accept as-is

Επιλέξτε `Accept as-is` όταν ισχύουν όλα τα παρακάτω:

- η ερώτηση είναι κατανοητή και σαφής
- η ερώτηση βασίζεται στο `source_text`
- η ερώτηση δεν ζητά πληροφορία που δεν υπάρχει στο κείμενο
- η απάντηση απαντά ακριβώς στην ερώτηση
- η απάντηση υποστηρίζεται από το κείμενο χωρίς εικασίες
- η απάντηση δεν περιέχει πρόσθετες, μη τεκμηριωμένες λεπτομέρειες
- η γλώσσα είναι επαρκώς φυσική και καθαρή
- δεν υπάρχει ουσιαστικό ορθογραφικό, νοηματικό ή πραγματολογικό πρόβλημα

Κάντε `Accept as-is` ακόμη και αν θα μπορούσατε θεωρητικά να το γράψετε λίγο καλύτερα, όταν η υπάρχουσα διατύπωση είναι ήδη σωστή και χρήσιμη για training. Δεν χρειάζεται edit για καθαρά προσωπικές προτιμήσεις ύφους.

????? ?? ??????? Accept as-is

Το `Accept as-is` είναι σωστή επιλογή όταν το παράδειγμα είναι ήδη καλό, γιατί:

- διατηρείται η ταχύτητα του review
- αποφεύγονται περιττές ανθρώπινες παρεμβάσεις
- μειώνεται ο reviewer-induced drift στο ύφος του dataset
- παραμένει μεγαλύτερος ο όγκος καλών samples χωρίς άσκοπο rewording

???? ?? ??????? Edit

Επιλέξτε `Edit` όταν το record είναι χρήσιμο αλλά χρειάζεται διόρθωση για να είναι training-quality.

Τυπικές περιπτώσεις για `Edit`:

- η ερώτηση είναι σωστή ως ιδέα αλλά θέλει πιο καθαρή διατύπωση
- η ερώτηση είναι υπερβολικά ασαφής ή αμφίσημη

- η απάντηση είναι μερικώς σωστή αλλά ελλιπής
- η απάντηση είναι σωστή αλλά περιέχει μία μικρή μη τεκμηριωμένη προσθήκη
- υπάρχει ορθογραφικό ή γραμματικό πρόβλημα που επηρεάζει την ποιότητα του sample
- η ερώτηση και η απάντηση δεν ταιριάζουν ακριβώς μεταξύ τους, αλλά διορθώνονται εύκολα
- η διατύπωση είναι αφύσικη ή μη κατάλληλη για εκπαιδευτικό QA, αλλά το record σώζεται εύκολα

???????? ???? ?????? Edit

- Αν αλλάζει μόνο η ερώτηση, συμπληρώστε μόνο το `corrected_question`.
- Αν αλλάζει μόνο η απάντηση, συμπληρώστε μόνο το `corrected_answer`.
- Αν αλλάζουν και τα δύο, συμπληρώστε και τα δύο.
- Αν αλλάξετε την ερώτηση, επιβεβαιώστε ότι η απάντηση εξακολουθεί να την απαντά σωστά.
- Αν αλλάξετε την απάντηση, επιβεβαιώστε ότι η ερώτηση εξακολουθεί να είναι η σωστή ερώτηση για την απάντηση αυτή.

Πολύ σημαντικό:

- Αν επιλέξετε `Edit` αλλά αφήσετε και τα δύο `corrected` fields κενά, στο export θα γίνει fallback στο αρχικό generated pair. Άρα το record θα μείνει ουσιαστικά αδιόρθωτο.

???? ?? ?????? Reject

Επιλέξτε `Reject` όταν το record δεν πρέπει να περάσει στο training set και δεν αξίζει να σωθεί με review edit.

Τυπικές περιπτώσεις για `Reject`:

- το `source_text` είναι τόσο θορυβώδες, κατεστραμμένο ή ελλιπές που δεν επιτρέπει αξιόπιστη ερώτηση-απάντηση
- η ερώτηση και η απάντηση είναι σοβαρά εκτός κειμένου
- η απάντηση περιέχει σοβαρή hallucination ή αντίφαση με το κείμενο
- το generated pair θα χρειαζόταν πλήρη αναδημιουργία και όχι απλή διόρθωση
- το κείμενο είναι τόσο fragmentary που δεν μπορεί να υποστηρίξει σαφές και χρήσιμο QA pair
- το record είναι άδειο, πρακτικά μη αξιοποιήσιμο ή προφανώς defective
- υπάρχει περιεχόμενο που δεν θα έπρεπε να χρησιμοποιηθεί για fine-tuning ακόμη και αν μπορούσε να διορθωθεί πρόχειρα

Πρακτικός κανόνας: αν δεν μπορείτε να υπερασπιστείτε το τελικό sample αποκλειστικά με βάση το `source_text`, μην το κρατήσετε.

????? ?? ??????? Reject

Το `Reject` δεν είναι τιμωρία του μοντέλου. Είναι μηχανισμός ποιοτικού ελέγχου. Χρησιμοποιείται για να:

- απομακρυνθεί θόρυβος από το training set
- αποφευχθεί η εκπαίδευση του μοντέλου σε λάθος ή αβέβαια patterns
- προστατευτεί η αξιοπιστία του dataset
- μην ξοδεύεται reviewer time σε samples που δεν σώζονται αποδοτικά

???? ?? ????????????? Edit ????? ?? Reject

Προτιμήστε `Edit` όταν:

- η βασική ιδέα της ερώτησης είναι σωστή
- το κείμενο υποστηρίζει ξεκάθαρα ένα σωστό QA pair
- η διόρθωση είναι συγκεκριμένη και υπερασπίσιμη
- το sample χρειάζεται περιορισμένη ή μέτρια παρέμβαση, όχι πλήρη επανασχεδίαση

Προτιμήστε `Reject` όταν:

- δεν είστε βέβαιοι ποια θα ήταν η σωστή διόρθωση
- θα έπρεπε να ξαναγράψετε από την αρχή και την ερώτηση και την απάντηση
- το κείμενο δεν παρέχει σταθερή βάση για καλό sample

????? ????????????? ??? ??????? ???????????? review

- Να είστε αυστηροί με την πιστότητα στο κείμενο.
- Να είστε συντηρητικοί με εξωτερικές ερμηνείες.
- Να κάνετε τη μικρότερη δυνατή αλλαγή που φέρνει το record σε αποδεκτή ποιότητα.
- Να κρατάτε μία μόνο ερώτηση ανά record.
- Να προτιμάτε σαφείς και συγκεκριμένες ερωτήσεις αντί για αόριστες.
- Να προτιμάτε απαντήσεις πλήρεις αλλά όχι φλύαρες.

- Να αποφεύγετε άσκοπες επαναλήψεις του source text λέξη προς λέξη, εκτός αν αυτό είναι ο πιο καθαρός τρόπος να αποδοθεί σωστά η απάντηση.
- Να αποφεύγετε απαντήσεις που εισάγουν νέο περιεχόμενο που δεν τεκμηριώνεται στο κείμενο.
- Να κρατάτε συνέπεια σε παρόμοιες περιπτώσεις από record σε record.
- Να θυμάστε ότι ο στόχος είναι training usefulness, όχι λογοτεχνική τελειότητα.

????? ?????????? ??????? ????
 ?????????? ? OCR-like ??????????

Σε αυτό το dataset είναι πιθανό να υπάρχουν κείμενα με OCR artifacts, παλιά τυπογραφικά σύμβολα, κομμένες φράσεις ή θόρυβο μορφοποίησης.

Σε αυτές τις περιπτώσεις:

- Αν το νόημα του κειμένου παραμένει σαφές, μπορείτε να κάνετε ή , ανάλογα με την ποιότητα του QA pair.
- Αν το νόημα είναι μόνο μερικώς σαφές, προτιμήστε μόνο όταν η διόρθωση είναι βέβαιη.
- Αν το νόημα δεν είναι σαφές χωρίς εικασία, προτιμήστε .
- Μην μαντεύετε ονόματα, αριθμούς, ημερομηνίες ή ορισμούς όταν το source text είναι αλλοιωμένο.

Checklist ??? ??? ??????????

Πριν κάνετε submit, ελέγξτε αν η τελική ερώτηση:

- είναι μία και μόνο μία ερώτηση
- είναι σαφής και όχι διφορούμενη
- είναι απαντήσιμη από το
- δεν απαιτεί εξωτερική γνώση
- δεν περιέχει μη τεκμηριωμένες παραδοχές
- είναι φυσική στα Ελληνικά

Checklist ??? ??? ?????????????

Πριν κάνετε submit, ελέγξτε αν η τελική απάντηση:

- απαντά ακριβώς στην ερώτηση

- είναι σωστή με βάση το `source_text`
- δεν περιέχει hallucinated στοιχεία
- δεν παραλείπει κρίσιμη πληροφορία που χρειάζεται για να είναι πλήρης
- δεν είναι άσκοπα μακροσκελής
- είναι καθαρή και φυσική στα Ελληνικά

?????? ???? ??? ?????? ??
 ??????????????

- `Accept as-is` ενώ η απάντηση περιέχει μία λεπτομέρεια που δεν υπάρχει στο κείμενο.
- `Edit` μόνο και μόνο επειδή ο reviewer προτιμά διαφορετικό ύφος.
- `Edit` με κενά και τα δύο `corrected` fields.
- Αλλαγή της ερώτησης χωρίς έλεγχο ότι η απάντηση παραμένει σωστή.
- Αλλαγή της απάντησης χωρίς έλεγχο ότι εξακολουθεί να απαντά τη συγκεκριμένη ερώτηση.
- Χρήση εξωτερικής γνώσης για να διορθωθεί ένα record.
- `Reject` σε record που χρειαζόταν μόνο μικρή και καθαρή διόρθωση.

????????? ?????????? ???????????

- Αν το sample είναι ήδη σωστό και χρήσιμο, κάντε `Accept as-is`.
- Αν το sample σώζεται με σαφείς, περιορισμένες και τεκμηριωμένες αλλαγές, κάντε `Edit`.
- Αν το sample είναι αναξιόπιστο, αβέβαιο ή απαιτεί πλήρη αναδημιουργία, κάντε `Reject`.

????????????????? ????????

Παράδειγμα για `Accept as-is`:

- Η ερώτηση είναι σαφής, η απάντηση είναι σωστή και δεν υπάρχει κανένα μη τεκμηριωμένο στοιχείο. Ακόμη κι αν θα μπορούσε να γίνει ελαφρώς πιο κομψή διατύπωση, δεν υπάρχει λόγος για edit.

Παράδειγμα για `Edit`:

- Η ερώτηση είναι καλή, αλλά η απάντηση προσθέτει μία λεπτομέρεια που δεν αναφέρεται ρητά στο κείμενο. Διορθώνετε μόνο το `corrected_answer`.

Παράδειγμα για `Reject`:

- Το `source_text` είναι τόσο αποσπασματικό ή θορυβώδες που δεν είστε βέβαιοι ποιο είναι το σωστό νόημα. Οποιαδήποτε διόρθωση θα απαιτούσε εικασία. Το record δεν πρέπει να περάσει στο training set.

???????? ???? submit

Πριν ολοκληρώσετε κάθε record, ελέγξτε τα εξής:

- Έχω επιλέξει το σωστό `review_action`.
- Αν έβαλα `Edit`, έχω συμπληρώσει τουλάχιστον ένα corrected field.
- Η τελική ερώτηση και η τελική απάντηση ταιριάζουν μεταξύ τους.
- Η τελική απάντηση στηρίζεται στο `source_text`.
- Δεν έχω κρατήσει πληροφορία που δεν μπορώ να υπερασπιστώ από το κείμενο.

?? ?????????????

Όταν έχετε αμφιβολία:

- μην κάνετε `Accept as-is` αν δεν είστε βέβαιοι για την τεκμηρίωση της απάντησης
- κάντε `Edit` μόνο αν ξέρετε ακριβώς ποια είναι η σωστή διόρθωση
- κάντε `Reject` όταν η αβεβαιότητα είναι δομική και όχι απλώς stylistic

Ο σωστός reviewer δεν προσπαθεί να “σώσει” κάθε sample. Προσπαθεί να διασφαλίσει ότι ό,τι μένει στο dataset είναι πραγματικά κατάλληλο για fine-tuning.

????? ??? ?????????????

Τα παρακάτω παραδείγματα είναι ενδεικτικά. Ο στόχος τους είναι να δείξουν όχι μόνο τι είναι σωστό, αλλά και ποιο είναι το συνηθισμένο λάθος reviewer decision σε κάθε περίπτωση.

????????? 1: ?????? `Accept as-is`

`source_text`:

"Το νερό βράζει στους 100 βαθμούς Κελσίου σε κανονική ατμοσφαιρική πίεση."

`generated_question`:

"Σε ποια θερμοκρασία βράζει το νερό σε κανονική ατμοσφαιρική πίεση;"

`generated_answer`:

"Το νερό βράζει στους 100 βαθμούς Κελσίου σε κανονική ατμοσφαιρική πίεση."

Λάθος reviewer απόφαση:

- `Edit`, μόνο και μόνο για να αλλάξει τη διατύπωση σε κάτι πιο κομψό.

Σωστή reviewer απόφαση:

- `Accept as-is`

Γιατί:

- Η ερώτηση είναι σαφής.
- Η απάντηση είναι σωστή και πλήρως τεκμηριωμένη.
- Δεν υπάρχει κανένας ουσιαστικός λόγος παρέμβασης.

?????????? 2: ?????? `Accept as-is`, ??????

`Edit`

`source_text`:

"Η Αθήνα είναι η πρωτεύουσα της Ελλάδας."

`generated_question`:

"Ποια είναι η πρωτεύουσα της Ελλάδας;"

`generated_answer`:

"Η Αθήνα είναι η πρωτεύουσα της Ελλάδας και η μεγαλύτερη πόλη της χώρας."

Λάθος reviewer απόφαση:

- `Accept as-is`

Γιατί είναι λάθος:

- Το δεύτερο μισό της απάντησης δεν τεκμηριώνεται από το `source_text`.
- Ακόμη κι αν είναι αληθινό γενικά, εδώ είναι εξωτερική γνώση.

Σωστή reviewer απόφαση:

- `Edit`
- `corrected_answer`: "Η Αθήνα είναι η πρωτεύουσα της Ελλάδας."

Γιατί:

- Το sample σώζεται εύκολα.
- Η βασική ερώτηση είναι σωστή.
- Χρειάζεται μόνο αφαίρεση της μη τεκμηριωμένης προσθήκης.

??????????? 3: ????? Edit ???? ???? ????????

source_text:

"Ο ποιητής γεννήθηκε το 1911 στη Θεσσαλονίκη."

generated_question:

"Τι έγινε τότε;"

generated_answer:

"Ο ποιητής γεννήθηκε το 1911 στη Θεσσαλονίκη."

Λάθος reviewer απόφαση:

- Accept as-is

Γιατί είναι λάθος:

- Η ερώτηση είναι υπερβολικά ασαφής.
- Δεν είναι καλό training signal, παρότι η απάντηση είναι κοντά στο νόημα του κειμένου.

Σωστή reviewer απόφαση:

- Edit
- corrected_question: "Πότε και πού γεννήθηκε ο ποιητής;"

Γιατί:

- Η απάντηση ήδη καλύπτει σωστά την πληροφορία.
- Το πρόβλημα είναι μόνο η κακή ποιότητα της ερώτησης.

????????????? 4: ????? Edit ??? ??? ??? ??????

source_text:

"Το μουσείο λειτουργεί από Τρίτη έως Κυριακή, από τις 09:00 έως τις 17:00."

generated_question:

"Πότε είναι ανοιχτό το μουσείο τη Δευτέρα;"

generated_answer:

"Το μουσείο είναι ανοιχτό από τις 09:00 έως τις 17:00."

Λάθος reviewer απόφαση:

- `Accept as-is`

Γιατί είναι λάθος:

- Η ερώτηση ρωτά για Δευτέρα, αλλά το κείμενο αναφέρει λειτουργία από Τρίτη έως Κυριακή.
- Η απάντηση δεν απαντά σωστά στην ερώτηση που δόθηκε.

Σωστή reviewer απόφαση:

- `Edit`
- `corrected_question`: "Ποιες ημέρες και ώρες λειτουργεί το μουσείο;"
- `corrected_answer`: "Το μουσείο λειτουργεί από Τρίτη έως Κυριακή, από τις 09:00 έως τις 17:00."

Γιατί:

- Το record μπορεί να σωθεί.
- Χρειάζεται ευθυγράμμιση και της ερώτησης και της απάντησης με το `source_text`.

?????????? 5: ?????? `Reject`

source_text:

"...Μακρ... 1821 ... σχολ... αγών... πρόσωπ..."

generated_question:

"Ποιος ήταν ο βασικός στόχος του σχολείου κατά την Επανάσταση του 1821;"

generated_answer:

"Ο βασικός στόχος ήταν η εθνική αφύπνιση των μαθητών."

Λάθος reviewer απόφαση:

- `Edit`, προσπαθώντας να ξαναγράψετε από την αρχή ένα πιθανό σωστό pair.

Γιατί είναι λάθος:

- Το `source_text` είναι υπερβολικά θορυβώδες και αποσπασματικό.

- Δεν υπάρχει ασφαλής βάση για τεκμηριωμένη ερώτηση και απάντηση.

Σωστή reviewer απόφαση:

-

Γιατί:

- Οποιαδήποτε διόρθωση θα βασιζόταν σε εικασία.
- Το sample δεν είναι αξιόπιστο για fine-tuning.

?????????? 6: ?????? , ??????

"Ο ήρωας επιστρέφει στο χωριό του μετά από δέκα χρόνια απουσίας."

"Μετά από πόσα χρόνια επιστρέφει ο ήρωας στο χωριό του;"

"Ο ήρωας επιστρέφει στο χωριό του μετά από πολλά χρόνια."

Λάθος reviewer απόφαση:

-

Γιατί είναι λάθος:

- Το sample δεν είναι κατεστραμμένο.
- Η σωστή πληροφορία υπάρχει καθαρά στο κείμενο.
- Χρειάζεται μόνο μικρή διόρθωση στην απάντηση.

Σωστή reviewer απόφαση:

-
- : "Ο ήρωας επιστρέφει στο χωριό του μετά από δέκα χρόνια απουσίας."

Γιατί:

- Το record είναι πλήρως αξιοποιήσιμο μετά από απλή, σαφή διόρθωση.

Revision #1
Created 2026-06-16 08:18:20 UTC by Ευστάθιος Ιωσηφίδης
Updated 2026-06-16 08:19:02 UTC by Ευστάθιος Ιωσηφίδης