

DATA ANNOTATION PROJECT

QUALITY ASSURANCE FOR E-
COMMERCE DATASET

Project Scope

TEXT ANNOTATIONS:

- 40 customer support messages
- 5 intent categories (cancellation, technical_support, billing_inquiry, feature_request, general_question)
- Balanced distribution (17-22% per class)

VISUAL ANNOTATIONS:

- 40 customer support messages
- 5 intent categories (cancellation, technical_support, billing_inquiry, feature_request, general_question)
- Balanced distribution (17-22% per class)

Deliverables

01

Annotation guidelines (text + visual)

02

Annotated datasets (CSV + COCO JSON)

03

QA tracking system (Google Sheets)

04

Python validation scripts

Quality Metrics

TEXT ANNOTATIONS

- Total: 40 annotations
- Approval rate: 100%
- High confidence: 92.5%
- Edge cases: 3 identified
- Error rate: 7.5%

VISUAL ANNOTATIONS

- Total: 78 bounding boxes
- Approval rate: 96.9%
- Needs review: 2 tasks (3.1%)
- Small objects: 2 detected
- Large backgrounds: 3 flagged

OVERALL QUALITY SCORE

- Average QA Score: 98.3%
- Issues corrected: 11/13 (84.6%)
- Error rate: 16.9 per 100 annotations

Key Learnings & Process

ANNOTATION CHALLENGES ENCOUNTERED

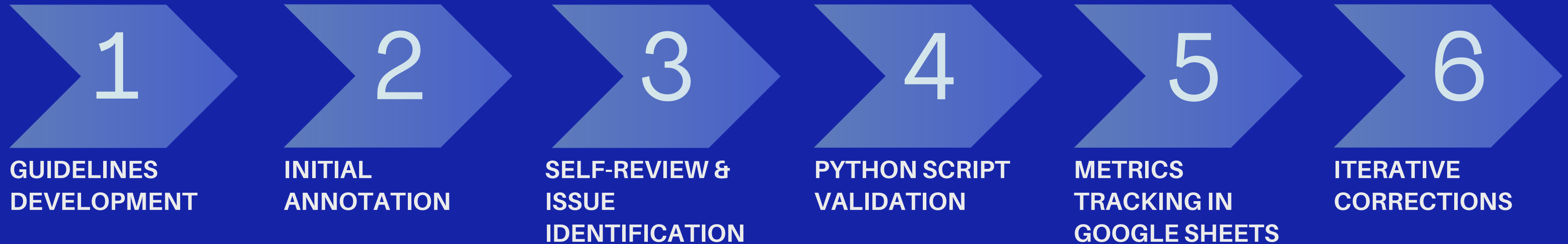
EDGE CASES

- Ambiguous text intents (e.g., pricing inquiry vs billing complaint)
- Large background boxes in product photography (50-84% coverage)
- Very small accessory items (<1% of image area)

SOLUTIONS IMPLEMENTED

- Clear decision criteria documented in guidelines
- QA tracking with reasoning for edge case decisions
- Automated validation scripts for consistency checks

Quality Assurance Workflow



Tools & Technologies

roboflow

 **Google Sheets**

 **python**TM

Results & Impact

DATASET QUALITY

- High accuracy: 98.3% average QA score
- **Balanced classes:** Even distribution across all categories
- **Well-documented:** All edge cases tracked and justified
- **ML-ready:** COCO JSON format for object detection

KEY ACHIEVEMENTS

- Identified and resolved 13 annotation issues
- Developed reusable validation scripts
- Created comprehensive QA documentation
- Maintained 96.9% approval rate

NEXT STEPS

- Scale annotation process to larger datasets
- Implement inter-annotator agreement testing
- Develop automated quality checks pipeline

THANK YOU

PORTFOLIO:
[HTTPS://GITHUB.COM/KAPLIAIRYNA/DATA-ANNOTATION-PORTFOLIO](https://github.com/kapliairyna/data-annotation-portfolio)

CONTACT:
[KAPLYA.PO4TA@GMAIL.COM](mailto:kaplya.po4ta@gmail.com)