|  |  |  |  |
| --- | --- | --- | --- |
| **Table 5. Confusion matrix for the 1000 segmentation outputs evaluated by Canary.** All outputs are evaluated by experienced human experts to assess whether they are actually ready for production or not. 29 out of the 58 problematic segmentation outputs could be automatically fixed by canary. | | | |
|  | **Predicted Applicable** | **Predicted Problematic** | **Total** |
| **Actual Applicable** | 699 | 241 | 940 |
| **Actual Problematic** | 2 | 58 (29 auto-corrected) | 60 |
| **Total** | 701 | 299 | 1000 |