**Ideas**

* (From meeting) – see what happens if mask offensive words from dataset – can we still accurately tell toxic language based on context? (might be innocent if between friends)