1. Tool is neutral, human is polite:
   1. Syntactic features not too strong. Tools rate are in borderline (.5-.6) [not a big deal]
   2. Some indirection of politeness - the tool gets totally wrong. E.g. polite criticism. [hard to solve with short training]
   3. Positive lexicons don’t have enough weight. E.g. congrats, thanks, LGTM, +1 [should get improved by training]
   4. Emojis don’t have enough weight.
   5. Some subtle features are missing---- a. Like agreement- “he’s right”
   6. What about greetings-? Polite due to greeting is a bit inconsistent wven with

b) tool polite, human neutral:

i) syntactic features not too strong determined by human raters, “and also long” so the rating goes higher [not a big deal]

c) human impolite, tool neutral:

i) **impoliteness is much more implicit within the context than politeness.** Hard to guess features even for human raters.

ii) short text. Even in polite case. The tool gives rating somewhere within the middle range.

iii) missing features --- a. ‘Need to do this” b. “Do not add”, “see above”, “replace”

iv) the tool have less idea on the subtle features of impoliteness

Common --

i) different structure is making the tool confused. Can’t point to one sole intention.

ii) some just plain wrong [not a lot]

iii) some features might be wrong --- “I don’t think”

iv) confusing use of polite/impolite features in an opposite direction

v) too long/too short

Interesting- why the polite bot comments getting low rating?