COMP 550 - RA4
Namdar Kabolinejad 260893536

The core argument of the paper is that language meaning cannot be learned from form alone. The authors point out how, in some recent publications, the loose use of words like “understanding” and “comprehending” leads to mischaracterization and unwarranted hype. They claim that while large LMs are undoubtedly useful, they are not complete solutions to the challenge of Human-analogous NLU. Focusing on whether large pre-trained LMs “understand” language, they argue that “a system trained only on form has a priori no way to learn meaning”. The paper also contains several insightful thought experiments, including one they call “the octopus test”, that try to illustrate what understanding truly is. Finally, the authors call for a more top-down approach to future research in NLP and propose some best practices on how to navigate the way. They argue that genuine progress in the field depends on maintaining clarity around big picture notions such as “understanding” and “form”. A clear advantage of the paper is how it critiques the current path the field is on, which has caused vast discussions on social media. Another advantage is the examples used to illustrate the arguments. A disadvantage of the paper is that they don’t provide constructive technical solutions/ideas on how to start climbing the “new hill” they have in mind. Furthermore, I believe the paper is being lopsided and not acknowledging the clear success and potential of the current path the field is taking.

I do agree with the central premise of the paper that true meaning can’t be learned by form alone and that pre-trained models don’t understand language or the meaning of it, based on the definition given in the paper. However, I believe this should be clear to see given that the field of ML is very much based on statistics and “real-world communicative intent” can’t be incorporated into the method that easily. It seems to me that the authors are basing their whole argument on a few misused words in some papers and blog posts and disregarding the rest of the community that understands the true nature of the current LMs and are trying to gradually improve it, to maybe eventually understand the “meaning” of a language. Furthermore, I don’t agree that we are climbing the wrong hill or that we necessarily need a top-down approach to succeed. This is not to say that we shouldn’t have a remote end goal, but that we should continue working at the current hill that we are climbing, which is yielding significant results but adjust our path by keeping in mind the end goal that we are trying to reach. We should definitely not abandon the encouraging work that has been done in the promise of something that might never happen. I believe eventually the end goal of “General Linguistic Intelligence” will be reached by modifying, working, or at least learning from the current LMs we have.

The authors define meaning to be the relation between the form and something external to language, or $M \subseteq E \times I$, (e,i) where e represents expressions and i is their communicative intent. And form to be any observable realization of language. However, in the course, we have defined meaning to be the relation of linguistic expression to the real world or to each other. Defining meaning as the relation of linguistic expression to the real world is quite similar to the definition of meaning in the paper, given that they both related the form of an expression to something external to the language. However, if we define meaning to be the relations of linguistic expression to each other, then that will have some contrast with what we have in the paper given that relationships between expressions can be learned and worked on by current LMs and external factors don’t play a role in it.
The paper defines Understanding as retrieving intent (i) given expression (e). I’m not sure if understanding has been explicitly defined in the course, but during the course, we have mentioned that a machine “understands” if it takes in any expression, for example, “What is the weather forecast for tomorrow?” and then generates an appropriate response for the expression which is quite similar to the definition in the paper.