Ling/CSE 472: Introduction to Computational Linguistics

5/25/17

Ethics & NLP 1

Overview

- Why ethics and NLP?

- Value Sensitive Design & Value Scenarios

- Envisioning Cards

- Reading Questions

Why Ethics in NLP?

- NLP technology is mature enough to have broad impact

- "The L in NLP is Language, language means people" (Tyler Schnoebelen)
  https://medium.com/@TSchnoebelen/ethics-and-nlp-some-further-thoughts-53bd7cc3ff69

- Growing interest from the field

  - EACL 2017 workshop

  - Many related workshops

  - More: http://faculty.washington.edu/ebender/2017_575/

Things to consider

- Exclusion/discrimination

- Bias (treating language behavior as ground truth)

- Privacy

- Dual use

- User misapprehension of infallibility

Overview

- Why ethics and NLP?

- Value Sensitive Design & Value Scenarios

- Envisioning Cards

- Reading Questions

Value Sensitive Design & Value Scenarios

- A family of techniques for informing the design process such that human values are supported by technology that is developed

- Better not best / progress not perfection

- vsdesign.org

VSD methods (Friedman & Henry, to appear)

- Direct and indirect stakeholder analysis

- Value source analysis

- Co-evolve technology and social structure

- Value scenarios

- Value sketch

- Value-oriented structured interview

VSD methods (Friedman & Henry, to appear)

- Scalable assessments of information dimensions

- Value-oriented coding manual

- Value-oriented mock-up, prototype or field deployment

- Ethnographically informed inquiry on values and technology

- Model for informed consent online

- Value dams and flows

Value Scenarios: Dimensions to Consider

- Stakeholders

- Pervasiveness

- Time

- Systemic effects

- Value implications

RQ: Could you run through the stakeholders, pervasiveness, time, systemic effects, and value implications of Google Translate?

Overview

- Why ethics and NLP?

- Value Sensitive Design & Value Scenarios

- Envisioning Cards

- Reading Questions

Reading questions

- Are there any specific/famous examples of projects or propositions that are known to have changed direction drastically due to the consideration of possible future 'side effects'?

- Could value scenarios include criteria as to prevent unethical technologies from being developed? During an enactment, certain qualities (who and how they affect people, environmental impact, etc.) about a technology become obvious. Before a platform or product is finalized, are there any regulatory steps like this?

Reading questions

- It wouldn't be too difficult for the developers to defend against potentially dangerous and unethical practices, only if the words "dangerous" and "unethical" were properly defined, which is not the case. How could the developers determine what is "unethical", if it is not properly defined and people just have various understanding of this term?

- How do we define the word "ethical" when developing? If a new technology will have foreseeable great negative effects in relatively short period, lets say about 10-20 years, but it is very obvious that this new technology benefit the whole human beings in the following hundreds of years?

Reading questions

- Sourour claimed that developers are the last lines of defense against potentially dangerous and unethical practices, but how much power does a developer actually have if what the company is doing doesn't break any laws? They could simply be fired and replaced with the next developer. At best, you've only set the project back a couple weeks or months and meanwhile you've probably lost your job and possible references.

- There seems to be a separation that allows for "It isn't my problem to look out for that."/"I'm just doing what I'm told.". If this is human nature, ingrained deeply, how can we prevent it? If most people aren't even aware it is something to watch out for?

Reading questions

- It's impossible for us to predict every possible outcome of creating some technology; couldn't it be said that the cases we can't predict could be more important than anything glaringly obvious?

- And then, there's the opposite extreme: with any given technology, it would probably be possible to sit around all day coming up with possible ways it could go wrong in some particular circumstance. You also start picking up on issues that aren't even really related to the original idea--for example, the Geminoid example brings up economic inequality and prejudice, but those aren't specific to the product and affect anything that can be purchased with money.

- I guess I'm basically questioning how thorough these value scenarios can really be--too much or too little both make them pretty useless, and at some point inbetween they can be a useful reference, but how would you ever know where that is?

Reading questions

- You could be super pessimistic and come up with a bad value scenario for any sort of technology, but who's to say how realistic it is? This seems to be more about considering what a technology will be primarily used for (instead of trying to come up with edge cases where it will be abused), but that is hard/impossible to predict accurately.

- The paper on value scenarios portrays rather pessimistic views on potential technology. History has shown that our predictions about technology are often very inaccurate. Could focusing on these depressing potential futures dissuade people from using technology that would actually improve their lives?

Reading questions

- Presumably more goes into constructing value scenarios than a priori speculation. What kinds of research and analysis might inform a value scenario?

- Are there any guidelines for creating value scenarios? It seems like it would be difficult to predict the widespread effects of a piece of technology, at least in some cases, and the prediction would definitely be informed off of what specific use for the technology the designers or whoever have in mind, which may not actually end up being correct.

Reading questions

- As the scenarios presented in both readings illustrate, the potential for technology (and by extension those coding it) for misuse or possible negative effects is something to be considered in its preliminary stages. Relating this more towards language and linguistics, how have the preliminary stages of NLP development been organized or otherwise planned to avoid negative outcomes? What scenarios do CompLinguists foresee as undesired consequences of their tech?

- Interestingly I did a short group presentation in another class concerning ASR technologies and its impact on Language policies, is this a topic on the ethics table in the CL field?