Refocusing on Relevance: Personalization in NLG

@shirandudy shiran.dudy@colorado.edu

Shiran Dudy, Steven Bedrick, and Bonnie Webber



who is on Tiktok"

A tiktoker

Jewish heritage

The user's perspective may be crucial to consider when working on language generation tasks, therefore

Goal: Can we generate relevant text

Contributions:

- 1.Discuss why it's time to work on personalization in NLG
- 2.Describe the potential <u>harms</u> of personalization in NLG
- 3.Discuss high level but also concrete approaches and examples other fields on how can we start working on it

Potential harms of personalization

Unintended data disclosure

For instance: health conditions can be leaked to health insurances, which in turn would increase coverage rates that user

Deliberate targeting

When the system exploits user's data against their interests

Filter bubbles

Providing a particular point of view may harm understanding other people's experience

Misleading advice

May occur when the system's assumptions on personal data are wrong.

Automation bias

Overly trusting the automation process

How can we design to avoid these harms?

Value-Sensitive design

"in designing tools we are designing ways of being"

Accessibility

Personalizing text to accommodate age, expertise, character through register/style, but also through content (1)

Agency

Yield the control over the type of metadata/ text to be used when generating text (2)

Trust

Indirect and direct aspects of trust. What are the types of trust a user may have in text? How to build it? (4)

Safety

Especially in interactive systems, the generated text can be harmful/ misleading/biased (3)

- Scarton et al. 2018b, Text simplification from professionally produced corpora, In Proceedings of the Eleventh International Conference on Language Resources and Evaluation
- 2. Synofzik et al. 2008, Beyond the comparator model: a multifactorial two-step account of agency. Consciousness and cognition
- 3. Dinan et al. 2021, Anticipating safety issues in E2E Conversational AI: Framework and tooling.
- 4. arXiv
- Ribeiro et al. 2016, "Why should I trust you?" Explaining the predictions of any classifier, In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining

2. Li and Liang. 2021 proposed a prefix to be attached to a source text to control the generated text. Our example: conditioning the generated text on a particular <u>user</u>

3. A framework of federated learning can offer insights on how to work locally on data while preserving privacy. See Kulkarni et al. 2020

Personalization in the medical field

To work on highly sensitive data privacy is a priority value in the medical field

1. MIMIC- III medical record dataset (Pollard et al. 2016)

Access restriction (legal restriction)

- Go through a training module
- Sign a data use agreement of use and distribution
- 2. The National COVID Cohort Collaborative (N3C) (Haendel et al. 2020)

No access because it is a data-enclave (technical restriction)

means more values/examples in the paper

Final conclusions and thoughts

- 1. Personalization can be crucial in NLG
- 2. Personalization bears harms
 - 3. Through high level and concrete approaches we can start working on personalization tasks
 - 4. Reach out to collaborate on creating personalization benchmarks or setting up a workshop
 - 5. Thank you!!

Concrete approaches for personalization

1. Khalifa et al. 2021: an architecture to control a language model. Example: to make all predictions to relate to sports, and 50% of them reflect female characters.

Our Example: to make all predictions to relate to a certain user, and 50% of them reflect the current query.