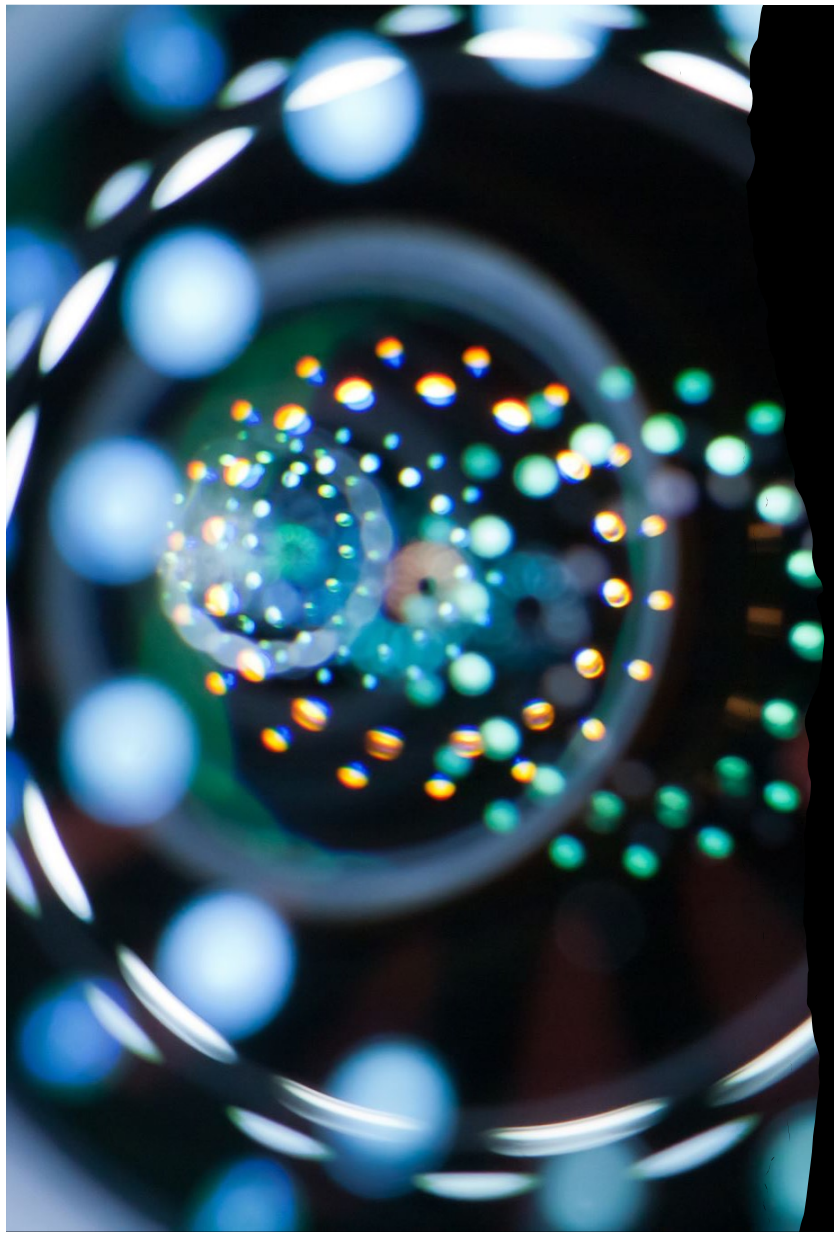


Is "good enough" good enough?

Ethical and responsible development of sign language technologies

Maartje De Meulder
HU University of Applied Sciences Utrecht

1st International Workshop on Automatic Translation for Signed and Spoken Languages

20th August 2021

Aim of the paper

- Identify some common and specific pitfalls in the development of sign language technologies targeted at deaf communities
- Interrogate some of the ideologies behind technologies + issues of ethical and responsible development
- Technology very promising so far, but critical that it is voluntarily adopted by end users
- Discussion usually lags behind scientific innovation (lag time becomes long now)


How are data being collected to make machines learn?
Who evaluates the outcomes, and how?
Who invents the technologies, and what is their motivation for developing them?
Is there an actual demand from the speech communities?
How will language rights keep pace with the development of language technologies?
Who benefits from these technologies, and who is at risk of being left behind?
What are the ideologies behind these technologies?

Outline

1. Data sets and bias
2. User feedback
3. Applications
4. Take away points

1. Data sets en bias

- Sign languages as “low resourced languages”: lack of available training data and fragmentation of efforts in resource development (Sayers et al. 2021)
- SL dictionaries developed for human use
- SL corpora confronted with problems regarding
  - Size and representativeness: mainly white ‘native’ signers
  - Absence of (semi-)automated annotation

Data sets and bias

- Machines will be trained on those corpora
- Context of data scarcity: machines also trained on readily available data like interpreted datasets
- Risk for bias: machines trained on amalgam of data sets produced by either primarily white, native signers or interpreters
- SL interpreters *already* language models to which deaf people have to adapt
- Signing avatars in SLI training programs ⚠

2. User feedback

- Most deaf people have life-long experiences understanding different signing styles (e.g. Friedner 2016)
- Understanding out of appreciation for developers (Woodcock 2020)
- "Uncanny valley" (Mori et al. 2012)
- Lack of testing in real-world settings

3. Applications

✅ Highly constrained and predictable domains (but do deaf people want this?)

✅ Sharing of sensitive/confidential information, anonymity issues

Applications

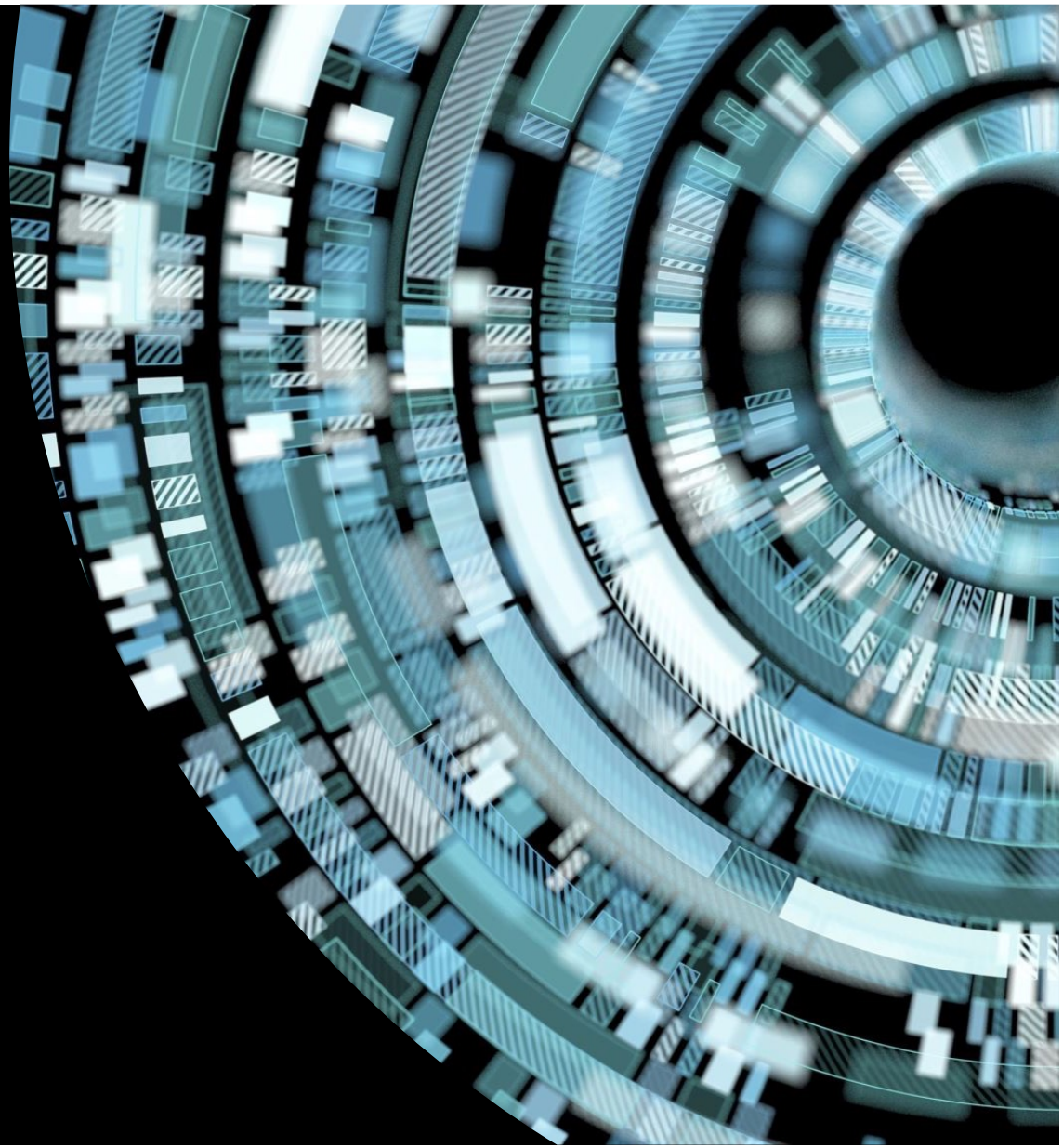
- Yet: applications in healthcare domain already happening despite explicit statements by e.g. WFD and WASLI
- Practice of SLI largely accepted (and even normative) in healthcare situations – avatar as logical next step?

Take away points for developers and research projects

- Technology developers: co-engineer and engage in interdisciplinary collaborations (and look at Deaf Studies work)
- Data sets: be mindful of biases that creep into data sets, and consider long-term implications

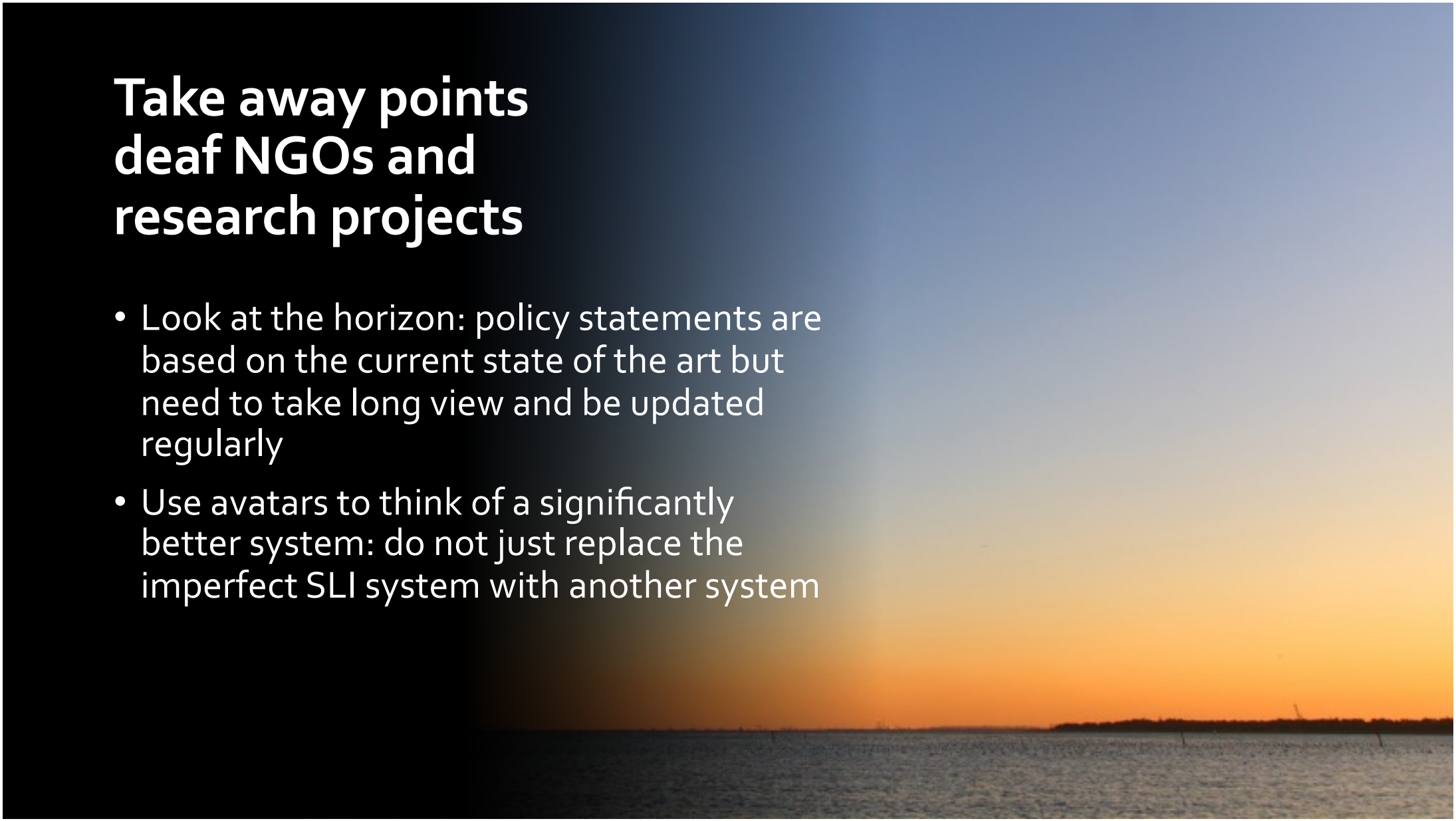
Take away points for developers, research projects, policy actors

- Prioritize application domains: (early) identify research agendas which are problematic while leaving space for those that are not
- Prioritize *within* application domains: not all applications within one domain are similar



Take away points deaf NGOs and research projects

- Look at the horizon: policy statements are based on the current state of the art but need to take long view and be updated regularly
- Use avatars to think of a significantly better system: do not just replace the imperfect SLI system with another system



Presentation slides:
https://maartjedemeulder.be/slides/

Paper:
https://aclanthology.org/2021.mtsummit-at4ssl.2/

References

Friedner, M. (2016). Understanding and Not-Understanding: What Do Epistemologies and Ontologies Do in Deaf Worlds? *Sign Language Studies*, 16(2), 184-203.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*, 98-100.

Sayers, D., Sousa-Silva, S., & Höhn. (2021). The Dawn of the Human-Machine Era: A forecast of new and emerging language technologies. *Report for EU COST Action CA19102 'Language in the Human-Machine Era'*. https://doi.org/10.17011/jyx/reports/20210518/1

Woodcock, K. (2020). Paved with good intentions. https://safeandsilent.wordpress.com/category/technology/