Transparency for Human-Centered AI

Vision Statement for
SIGCHI Italy HCI4AI Workshop

B.L. William Wong

Principal Scientist • Genetec, Inc.
Professor of Human-Computer Interaction • Middlesex University London

7 July 2020

What we want: AI/ML to augment Human

- FP7 VALCRI: Next gen criminal intelligence analysis and investigation system
- Human-machine teams that use AI/ML (McDermott et al, 2018) e.g. HCAI
  - designed to operate on the basis of
  - Human-machine symbiosis (Licklider, 1960)
    - to enable
  - Joint Cognitive System (Hollnagel and Woods, 2005)
    - that will have the
  - Requisite Variety (Ashby, 1958)
    - to manage the variety of demands presented by the work environment
  - "Human decides, machines do heavy lifting"
- Black box automation, algorithmic opacity
  => 'automation surprise'; lack system visibility
  - accidental discrimination, the mosaic effect
- Unexpected outcomes
  - what is the system doing? why is the system giving contradicting information? or why is the AI system recommending that black men be given longer prison sentences that are longer than white men for the same kinds of crimes?

Northpointe's COMPAS* tool: *Offenders could not challenge the risk assessments produced by the proprietary algorithms*

Q2. What are examples of failures of AI systems due to poor knowledge of HCI theories, principles, and methodologies?

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Prediction Fails Differently for Black Defendants

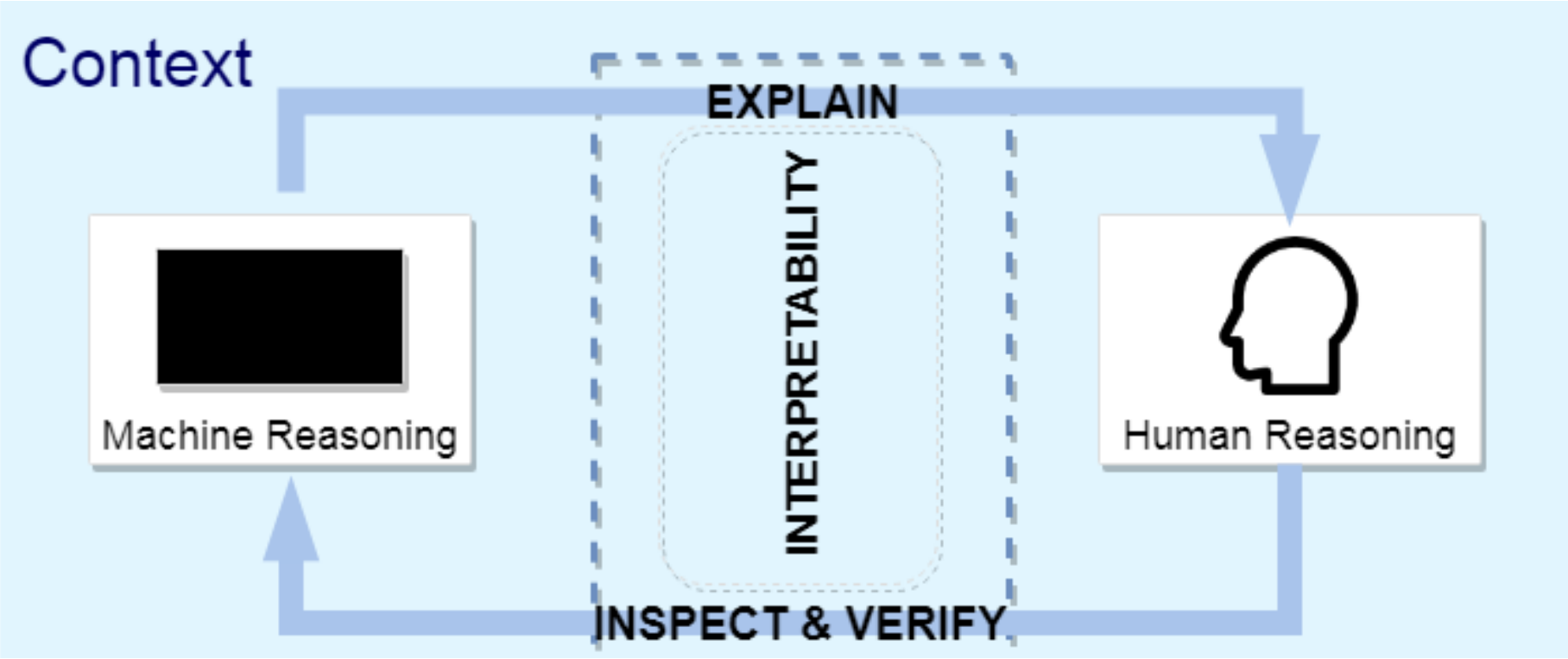
Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes.

Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks.
*by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica*May 23, 2016
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

* Hight consequence tool

Algorithmic Transparency

Users should be able to **inspect, challenge, ascertain, verify** black box outputs without significantly increasing cognitive workload



Hepenstal, S., Kodagoda, N., Zhang, L., Paudyal, P., & Wong, B. L. W. (2019). Algorithmic Transparency of Conversational Agents. In *Proceedings of ACM IUI ATEC '19, Algorithmic Transparency in Emerging Technologies, 20 March 2019, Los Angeles, USA.* (pp. Accepted for publication). New York, NY: ACM Press.

What students need to know? Possible HCAI Curriculum

Q1. What design, development and user testing methods and practices … could be adopted in AI?
Q3. What interaction paradigms /modalities/metaphors for AI systems that best support interaction with users?

- **Some Theories and Concepts for HCAI**
  - Human-machine symbiosis – joint cognitive systems – human-machine teams.
  - Algorithmic Transparency – design to enable transparency, overcome machine bias
  - Ethics, Legal & Privacy by Design – human rights, how to implement GDPR in design
  - Resilience – recoverability e.g. COCOMO (Hollnagel)
- **Methods for understanding and design:** Cognitive Engineering
  - *Methods for modelling the algorithms* e.g. *cognitive work analysis (Vicente, 1999)*
  - *Methods that describe how a user thinks and reasons* e.g. *cognitive task analysis (Klein, et al, 1989)*
  - *Methods to represent the cognitive work* on a user interface, e.g. *Representation Design* (Bennett & Flach, 2011), Ecological Interface Design (Burns & Hajdukiewicz, 2004)

Conclusion

- Different analysis and design methods serve different purposes
- The methods we use depends on the kind of AI system we want to build e.g. degree of autonomy vs smart tool
- System visibility should be a design goal
- Cognitive engineering analysis and design methods could complement HCI and AI system analysis, design and development methods in a HCAI syllabus