What does it mean to ask about "the human use of machine learning“?

On positioning (not only) for engineers

Bettina Berendt
Dept. Computer Science, KU Leuven

https://people.cs.kuleuven.be/~bettina.berendt/

The Human Use of Machine Learning: An Interdisciplinary Workshop,
Dec. 16, 2016, Venice

We’re asking our students to think about ethics …

- … for example by role-playing

GAMASUTRA
The Art & Business of Making Games

GAME JOBS | UPDATES | BLOGS | EVENTS | CONTRACTORS | NEWSLETTER | STORE | SEARCH GO

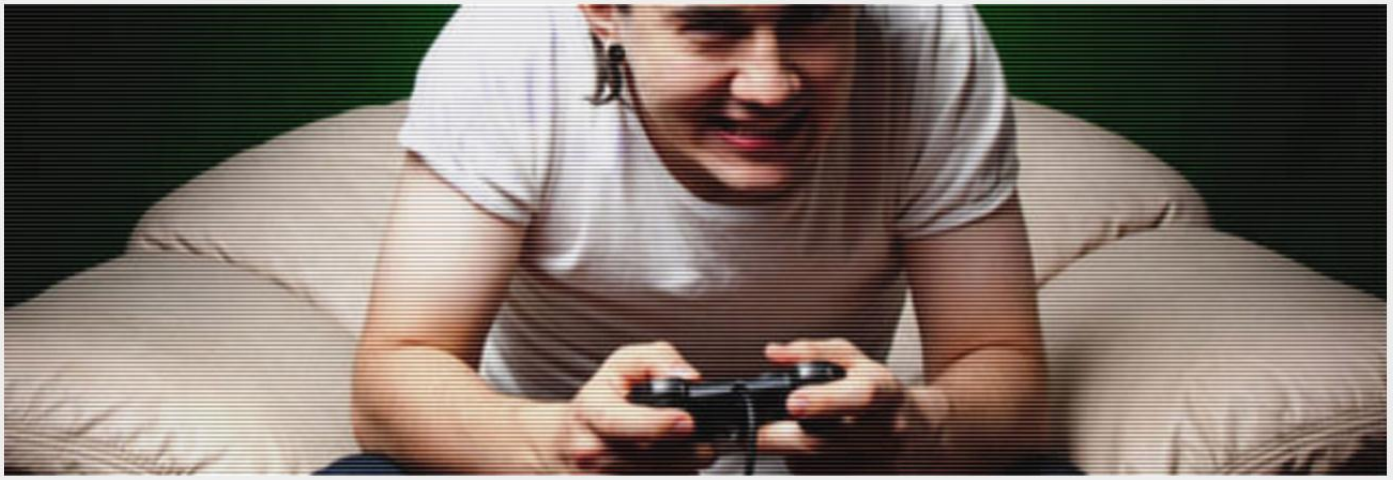
ALL | CONSOLE/PC | SMARTPHONE/TABLET | INDEPENDENT | VR/AR | SOCIAL/ONLINE | GAME DEVELOPER ON GAMASUTRA

Member Login
Email:
Password:
Login
Forgot Password? Sign Up

PROGRAMMING
ART
AUDIO
DESIGN
PRODUCTION
BIZ/MARKETING

'Gamers' don't have to be your audience. 'Gamers' are over.

EXCLUSIVE



I often say I'm a video game culture writer, but lately I don't know exactly what that means. 'Game culture' as we know it is kind of embarrassing -- it's not even culture. It's buying things, spackling over memes and in-jokes repeatedly, and it's getting mad on the

August 28, 2014 | By Leigh Alexander

291 comments

More: Console/PC, Social/Online,

http://www.gamasutra.com/view/news/224400/Gamers_dont_have_to_be_your_audience_Gamers_are_over.php; Rockwell & Berendt (2016)

We’re asking our students to think about ethics …

- … for example by role-playing
- … and we find ethical dilemmas
  - (well, that’s part of the didactic plan)

We’re asking our students to think about ethics …

- … for example by role-playing
- … and we find ethical dilemmas
  - (well, that’s part of the didactic plan)

- But what about ourselves?

SPION

Ero Balsa
Seda Gürses
Claudia Diaz
Bart Preneel

Dave Clarke
Frank Piessens
Rula Sayaf

KATHOLIEKE UNIVERSITEIT LEUVEN

COSIC

Bram Lievens
Jo Pierson
Ralf De Wolf

KATHOLIEKE UNIVERSITEIT LEUVEN

DistriNet RESEARCH GROUP

Vrije Universiteit Brussel

SMIT STUDIES MEDIA INFORMATION TELECOMMUNICATION

KATHOLIEKE UNIVERSITEIT LEUVEN

Interdisciplinary Centr for Law & ICT

DTAI

UNIVERSITEIT GENT

Bettina Berendt

Bo Gao

Vakgroep onderwijskunde

Tammy Schellens
Martin Valcke
Ellen Vanderhoven

Carnegie Mellon University Heinz College

Alessandro Acquisti
Fred Stutzman

Jos Dumortier
Eva Lievens
Brendan Van Alsenoy

Two (BIG) sample problems

- Privacy / privacy violations
  - De Wolf, Vanderhoven, Berendt, Pierson, & Schellens (2016)
- Non-discrimination, fairness / discrimination
  - With (esp.) S. Preibusch and G. Rockwell since ~ 2011

Q1. Which actors are involved in formulating the problem?

- Privacy
  - Security experts?
  - Users?
- Non-discrimination / fairness
  - Data miners?
  - Formal modellers?
  - Lawyers?
  - Sociologists?
- Answers co-determine who will “accept” the “solution” (and even how seriously the problem will be taken)

Gürses & Diaz (2013); Berendt & Preibusch (2014, under revision)

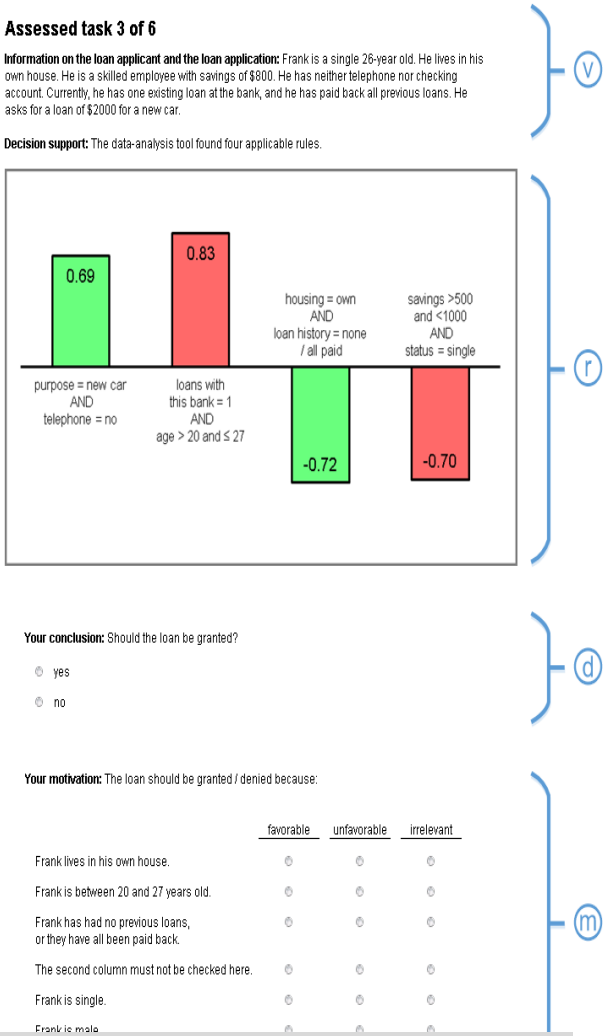
Q2. What type of right?

- Not only “what is …?”, also “what should … be?”
- Privacy
  - Human right?
  - Property right?
  - …?
- Non-discrimination
  - The “risk-utility tradeoff”: “but if it’s the truth?”

  → What is the underlying idea of justice?

Q3. Is information always a good thing?

- Privacy
  - Is informing users of privacy dangers always a good thing?
- Non-discrimination
  - Start ignoring or keep emphasizing the classification?
  - Sanitize or flag up discriminatory patterns? → HCI



Berendt & Preibusch (2014, under revision)

Q4. Do we want to influence (e.g. users´) attitudes and behaviours?

- *Can* the scientist devolve themselves?
- Do we *want* to?
- *Should* we? (Autonomy!)
- “Privacy is for papers”?
- “Privacy begins at home”?

Q5. Who is the target audience?

- Vulnerable audiences?
- Ethics of care?
  - BUT: Caring instead of ensuring fundamental rights?

Rockwell & Berendt (2016); Wittkower (2016)

Q6. What is the notion of understandability that we deem relevant?

- Claim:
  - "deep learning is un-understandable" is not the issue!
  - "Neural networks are function approximators" (and that's it)
- Instead/in addition: we need to understand the context *(for want of a better term)* of learning
  - E.g. Lum & Isaac (2016)
  - "Big Data's Disparate Impact" (Barocas & Selbst, 2016)

Q7. What assumptions are baked into the discourse?

- "Have you ever noticed that […] we talk incessantly about <Anerkennung> [recognition, appreciation, acknowledgement, acceptance, respect] and diversity, but hardly ever any more about social inequality?
- The obsession with which, in rich societies, for example the <Anerkennung> of even the most peculiar sexual orientation is struggled for, is symptomatic of a setting in which one should *not any more* talk about social, i.e. *changeable* inequality.
- This is not any more about the abolition of inequality, but only about the <Anerkennung> of diversity, and all of this in a morally hypercharged discourse.“ (Welzer, 2016)

- What do you believe are the two most self-reported grounds of discrimination in Germany in 2016?

  (as reported in Berghahn et al., 2016)

Q7.’ What assumptions are baked into the methodology / technology?

- Cf. data mining’s essentialism
  - “features”
  - Even harder to change modelling approaches than in legal concepts
  - “race” → “racialization” → “racist discrimination”, “racist attribution”, “so-called race”
    - See proposals in (Berghahn et al., 2016)

Q8. What can and should be the role of ethics codes, ethics boards, ethics reviews?

“We’re the good ones.”

Example ERC

http://grantsaccess.ethz.ch/fileadmin/content/downloads/150910_Presentation_Veronica_Martinez_Ocana.pdf

The Ethics Review Process

ERC Proposals

erc

Legal requirement

– *Article 19 of the H2020, article 13 of Rules for Participation, which states that any proposal which contravenes fundamental ethical principles shall not be funded*

What could be an ethics issue? Main examples:

- Privacy and personal data

- Dual Use
  - *Potential military/terrorist application*
  - *Council Regulation (EC) No 428/2009*

- Misuse

- *Aim of the research: Possible distortion of the facts by the methodology used*
- *Discrimination and stigmatisation: During research and/or dissemination of results*
- *Misinterpretation of results: Preventing wrong political use of results*

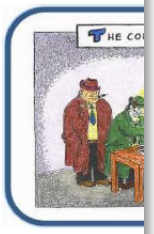
- Involvement of Human participants

Medical studies

- *Declaration of Helsinki*
- *Oviedo Bioethics Convention*
- *Regulation No 536/2014 of the European Parliament,*

Children involved in research

- *Information sheet and assent*
- *Reconsenting when turning adults*



Social or Human Sciences Studies

- *Oral and written consent*
  - *Illiterate individuals*
  - *Oral tradition countries*
  - *Risks of written consent*

Vulnerable populations

- *Free will of university students as research subjects*
- *Financially vulnerable populations and use of incentives*

The APA case

- “[The APA] in 2002 amended its ethics code to permit psychologists to follow “governing legal authority” even if it went against other aspects of the code. Then, in 2005, following the first major revelations of detainee abuse in Abu Ghraib and CIA interrogations, the APA convened a special task force, that, while condemning torture, affirmed that psychologists could supervise and conduct research as part of national security interrogations.”
- Some psychologists appear to have been thus involved.
- “We must use critical thought to distinguish what is ethical from what is lawful and to consider what it means to be a professional. Therefore, we must continually question and re-question authority, whether it is the law or a code of ethics, or else we may be doomed to serve the interests of those who crafted the code, not necessarily the interests of those who need to embody the code or use it to guide their practice. Just because a principle is codified does not make it ethical. “

Currier (2015); Risen (2015); Patel & Elkin (2015)

Thank you!

I‘ll be more than happy to hear your

Q1. Which actors are involved in formulating the problem?
Q2. What type of right?
Q3. Is information always a good thing?
Q4. Do we want to influence attitudes and behaviours?
Q5. Who is the target audience?
Q6. What is the notion of understandability that we deem relevant?
Q7. What assumptions are baked into the methodology / technology?
Q8. What can and should be the role of codified & institutionalised ethics?

References

Barocas, S. & Selbst, A.D. (2016). Big Data's Disparate Impact. 104 California Law Review 671 (2016). http://ssrn.com/abstract=2477899

Berendt, B., Büchler, M., & Rockwell, G. (2015). Is it research or is it spying? Thinking-through ethics in Big Data AI and other knowledge sciences. *Künstliche Intelligenz, 29(2),* 223-232. https://people.cs.kuleuven.be/~bettina.berendt/Papers/berendt_buechler_rockwell_KUIN_2015.pdf

Berendt, B. & Preibusch, S. (2014). Better decision support through exploratory discrimination-aware data mining: foundations and empirical evidence. *Artificial Intelligence and Law, 22 (2),* 175-209. http://people.cs.kuleuven.be/~bettina.berendt/Papers/berendt_preibusch_2014.pdf

Berendt, B. & Preibusch, S. (under revision). Towards accountable non-discriminatory data mining: The importance of keeping the human in the loop – and under the looking-glass.

Berghahn, S., Egenberger, V., Klapp, M., Klose, A., Liebscher, D., Supik, L.,, & Tischbirek, A. (2016*). Evaluation des Allgemeinen Gleichbehandlungsgesetzes [Evaluation of the German General Anti-discrimination Law]*, erstellt im Auftrag der Antidiskriminierungsstelle des Bundes. http://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/AGG/AGG_Evaluation.pdf?__blob=publicationFile&v=14

Currier, C. (2015). Emails Reveal Close Relationship Between Psychology Group and CIA. *The Intercept*, 30 April 2015. https://theintercept.com/2015/04/30/emails-show-close-relationship-psychology-group-cia/

De Wolf, R., Vanderhoven, E., Berendt, B., Pierson, J., & Schellens, T. (2016). Self-reflection on privacy research in social networking sites. *Behaviour & Information Technology*. DOI: 10.1080/0144929X.2016.1242653. https://people.cs.kuleuven.be/~bettina.berendt/Papers/dewolf_vanderhoven_berendt_pierson_schellens_2016.pdf

Gürses, Seda, and Claudia Diaz.(2013). Two Tales of Privacy in Online Social Networks. *IEEE Security & Privacy, 11 (3),* 29–37.

Lum, K. & Isaac, W. (2016). To predict and serve? *Significance, 13 (5),* 14-19. http://dx.doi.org/10.1111/j.1740-9713.2016.00960.x

Patel , N.A., & Elkin, G.D. (2015). Professionalism and Conflicting Interests: The American Psychological Association's Involvement in Torture. *AMA Journal of Ethics, 17(10),* 924-930. http://journalofethics.ama-assn.org/2015/10/nlit1-1510.html

Risen, J. (2015). American Psychological Association Bolstered C.I.A. Torture Program, Report Says. *The New York Times*, 30 April 2015. http://www.nytimes.com/2015/05/01/us/report-says-american-psychological-association-collaborated-on-torture-justification.html

Rockwell, G. & Berendt, B (2016). *Information wants to be free: Thinking-through Respect by Design*. Göttingen Dialog for Digital Humanities. University of Göttingen. 9 May 2106. http://people.cs.kuleuven.be/~bettina.berendt/Talks/rockwell_berendt_2016_05_09.pdf

Welzer, H. (2016). *Die smarte Diktatur. Der Angriff auf unsere Freiheit. [The Smart Dictatorship. The Attack on our Freedom].* Frankfurt am Main: S. Fischer Verlag.

Wittkower, D.E. (2016). Lurkers, creepers, and virtuous interactivity: From property rights to consent and care as a conceptual basis for privacy concerns and information ethics. *First Monday, 21(10),* Oct. 2016. http://firstmonday.org/ojs/index.php/fm/rt/printerFriendly/6948/5628