

Black Box Learning Analytics? Beyond Algorithmic Transparency

Simon Buckingham Shum

Professor of Learning Informatics
Director, UTS Connected Intelligence Centre
@sbuckshum • Simon.BuckinghamShum,net



I was speaking at this event

code acts in education

learning through code/learning to code



I asked Siri to find the conference website...



"Find code acts in education"

I asked Siri to find the conference website...





https://www.youtube.com/watch?v=Dlr401aEJvl&sns=em

Societal impact of data science is now reaching mass audiences



<u>https://weaponsofmathdestructionbook.com</u> • <u>http://datasociety.net</u> • <u>http://artificialinteligencenow.com</u>
 <u>https://obamawhitehouse.archives.gov/blog/2016/05/04/big-risks-big-opportunities-intersection-big-data-and-civil-rights</u>

thought experiment

A student warning system, somewhere near you...

Student: "Being told by the LMS every time I login that I'm at grave risk of failing is stressful. I already informed Student Services about my disability and recent bereavement, and I'm working with my tutor to catch up..."

Student	Int.1	Int 2	Int 3	Int-4	Int 5	Int 6	Int.Z	Int.8	Int
💱 Student A		0	0	•	•	•	•	•	
Student B	•	•	•	•	•	•		۲	0
Student C	•	•	•	•	•	•	•		
Student D	•	•	۲		•	•		•	
Student E	•	•	•	•	•	•	•	•	
Student F	•	•	•	•	•	•			
Student G	•	•	•	•	•	•		•	
Student H		•	•	•		•			
😡 Student I	•			•		•	•		
Student J	•			•		•	•	•	0

A student warning system, somewhere near you...

Student: "Being told by the LMS every time I login that I'm at grave risk of failing is stressful. I already informed Student Services about my disability and recent bereavement, and I'm working with my tutor to catch up..."



A social network visualisation, somewhere near you...



A social network visualisation, somewhere near you...



What would it mean for learning analytics to be accountable?

...there's intense societal interest right now in algorithms

Algorithms are generating huge interest in the media, policy, social justice, and academia

Governing Algorithms

A conference on computation, automation, and control

In an increasingly algorithmic world [...] What, then, do we talk about when we talk about "governing algorithms"?

http://governingalgorithms.org

Be afraid: unaccountable algorithms are counting our money – and our other assets



http://www.lse.ac.uk/newsAndMedia/videoAndAudio/channels/publicLecturesAndEvents/player.aspx?id=3350

#AlgorithmicAccountability

Meaning 1

Algorithms that make you Accountable

more objectively, efficiently and rewardingly than a human can

Meaning 2

Making Algorithms that make you Accountable Accountable



We must be able to open the black box (Frank Pasquale)



A topic now engaging diverse expertises... **Statisticians / Data Miners** Programmers **Computer Scientists Social Scientists** Designers **Historians** Lawyers **Policy Makers Business Entrepreneurs**



What are Algorithms?

abstract rules for transforming data

which to exert influence require

programming as executable CODE

operating on data structures running on a platform in an increasingly distributed architecture

Paul Dourish: The Social Lives of Algorithms. Lecture, 23 Feb. 2016, University of Melbourne. http://www.eng.unimelb.edu.au/engage/events/lectures/dourish-2016

What makes algorithms opaque?

intentional secrecy technical illiteracy complexity of infrastructure

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society 3(1). http://doi.org/10.1177/2053951715622512

Algorithms are not to be found 'in' Pasquale's black box



'Algorithms' are abstractions invented by computer scientists, mathematicians, etc. They may be impossible to reverse engineer from a live, material, distributed system.

To understand them as phenomena, and perhaps to call them to account, we need to ask how they were conceived, with what intent, who sanctioned their implementation, etc... (Paul Dourish)

Paul Dourish: The Social Lives of Algorithms. Lecture, 23 Feb. 2016, University of Melbourne. http://www.eng.unimelb.edu.au/engage/events/lectures/dourish-2016

#LearningAnalytics

See the Society for Learning Analytics Research: http://SoLAResearch.org

#AlgorithmicAccountability

#AlgorithmicAccountability

Learning Analytics System Integrity

Algorithms don't act, it's the whole system that counts "Algorithmic Assemblages" are more useful units of analysis

	Aggregation inferring associations	Classification inferring boundaries	Temporality inferring when to act
algo code			
work practices			
cultural norms			

Ananny, M., 2016. Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology & Human Values*, 41(1), pp.93–117

Stakeholders in a Learning Analytics system

Some accountability relationships in an analytics system

Forms of Algorithmic Accountability in Learning Analytics

Human-Centred Informatics Lenses **Ethics of Technology**

Computer Science

Data Science

Human-Data Interaction

User-Centred Design

Learning Technology

Legal Accountability...?

Accountability in terms of: Ethics of Technology

Accountability in terms of: Ethics of Technology

Deontological critique: do the analytics produce results that satisfy current duties/rules/policies/principles, and/or correspond with how we already classify the world?

Ethics of Technology Teleological critique: do the analytics lead to consequences that we value?

Virtues critique: What are the values implicit/explicit in the design (at every level), and do stakeholders perceive and use the system as intended? (cf. HCI Claims Analysis)

...but for an emerging field, esp. with personalisation, D & T can be challenging to apply

Ananny, M., 2016. Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology & Human Values*, 41(1), pp.93–117

Accountability in terms of: Computer Science

Accountability in terms of: Computer Science

Algorithmic Integrity: does the running code (code+data+platform) operationalise the algorithm with integrity? Is it possible to reverse engineer the algorithm from the system?

Source code availability: to what extent can developers inspect and test the source code?

Source code maintainability: how easily can a developer modify the source code?

System output intelligibility: can we understand why the implemented system generates its outputs under different conditions? (a particular issue with machine learning)

Accountability in terms of: Data Science

Accountability in terms of: Data Science

Data Science

Training Data Integrity: for machine learning, does the training set embody discriminatory bias?

Model Integrity: for machine learning, do the target variables and class labels embody discriminatory bias?

Feature Selection Integrity: does the discriminatory power of features do justice to the complexity of the phenomenon?

Proxy Integrity: do apparent proxy features for a target quality embody discriminatory bias?

Barocas, S. and A. Selbst (In Press). Big Data's Disparate Impact. *California Law Review* 104. Preprint: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899

Algorithmic shortlisting for interview

Eleni Kalorkoti

TheUpshot

ROBO RECRUITING

M Email

f Share

Tweet

Save

A More

Can an Algorithm Hire Better Than a Human?

Claire Cain Miller @clairecm JUNE 25, 2015

Hiring and recruiting might seem like some of the least likely jobs to be automated. The whole process seems to need human skills that computers lack, like making conversation and reading social cues.

But people have biases and predilections. They make hiring decisions, often unconsciously, based on similarities that have nothing to do with the job requirements — like whether an applicant has a friend in common, went to the same school or likes the same sports.

That is one reason researchers say traditional job searches are broken. The question is how to make them better.

http://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html?_r=0

Obviously no university would vet student applications by algorithm...

(fictional scenario)

Ascilite 2011 Keynote: Slide 77: http://people.kmi.open.ac.uk/sbs/2011/12/learning-analytics-ascilite2011-keynote

Obviously no university would vet student applications by algorithm...

For one British university, what began as a time-saving exercise ended in disgrace when a computer model set up to streamline its admissions process exposed - and then exacerbated - gender and racial discrimination.

As detailed here in the British Medical Journal, staff at St George's Hospital Medical School decided to write an algorithm that would automate the first round of its admissions process. The formulae used historical patterns in the characteristics of candidates whose applications were traditionally rejected to filter out new candidates whose profiles matched those of the least successful applicants.

Stella Lowry & Gordon Macpherson, *A Blot on the Profession*, 296 BRITISH MEDICAL J. 657 (1988). http://www.theguardian.com/news/datablog/2013/aug/14/problem-with-algorithms-magnifying-misbehaviour

Returning to algorithmic staff recruitment...

Eleni Kalorkoti

"It is dangerously naïve, however, to believe that big data will automatically eliminate human bias from the decisionmaking process.

...inherit the prejudices of prior decisionmakers

...reflect the widespread biases that persist in society

...Discover ...regularities that are really just preexisting patterns of exclusion and inequality."

Written Testimony of Solon Barocas: *U.S. Equal Employment Opportunity Commission,* Meeting of July 1, 2015 - EEOC at 50: Progress and Continuing Challenges in Eradicating Employment Discrimination. <u>http://www1.eeoc.gov//eeoc/meetings/7-1-15/barocas.cfm?renderforprint=1</u>

Returning to algorithmic staff recruitment...

Eleni Kalorkoti

"...Adopted or applied without care, data mining can deny historically disadvantaged and vulnerable groups full participation in society."

"...it can be unusually hard to identify the source of the problem, to explain it to a court, or to remedy it technically or through legal action."

Written Testimony of Solon Barocas: *U.S. Equal Employment Opportunity Commission,* Meeting of July 1, 2015 - EEOC at 50: Progress and Continuing Challenges in Eradicating Employment Discrimination. <u>http://www1.eeoc.gov//eeoc/meetings/7-1-15/barocas.cfm?renderforprint=1</u>

Accountability in terms of: Human-Data Interaction

Accountability in terms of: Human-Data Interaction

"Accountable Data Transactions": does the data infrastructure have clear protocols for personal data requests, permission, and audit?

Social Data Infrastructure: as interactions around data are formalised and automated, can this infrastructure be held to account?

Collective Data Ownership: is collectively owned data handled in ways that preserve individual rights?

User Agency: to what degree do citizens have control over who accesses their data, and can they comprehend what analysis will be done, and by whom?

Crabtree, A. and R. Mortier (2015). Human Data Interaction: Historical Lessons from Social Studies and CSCW. *Proc. European Conference on Computer Supported Cooperative Work*, Oslo (19-23 Sept 2015), Springer: London. Preprint: <u>http://mor1.github.io/publications/pdf/ecscw15-hdi.pdf</u>

Accountability in terms of: User-Centred Design

Accountability in terms of: User-Centred Design

User Sensemaking: who are the target end-users, do they understand the analytic output, and can they take appropriate action based on it?

User-Centred Design Intelligibility: what, if any, explanation can a user ask the analytic system to provide about its behaviour, can they understand the answers, and can they give feedback?

Participatory Design: to what extent are stakeholders (e.g. academics; instructors; students) involved in the system design process, and are there interfaces for them to modify the deployed system's behaviour?

Accountability in terms of: Learning Sciences & Educational Technology

Accountability in terms of: Learning Sciences & Educational Technology

Conceptual Integrity: does the algorithm implement the intended constructs (theory/model/rubric...) with integrity?

Learning Sciences & Educational Technology **Conceptual Integrity**: is the data congruent with other sources of educational evidence?

Improved Learning: in what ways do learners benefit from using the system?

Improved Teaching: in what ways do educators benefit from using the system?

worked example 1 analytics for professional, academic, reflective writing

Buckingham Shum, S., Á. Sándor, R. Goldsmith, X. Wang, R. Bass and M. McWilliams (2016). Reflecting on Reflective Writing Analytics: Assessment Challenges and Iterative Evaluation of a Prototype Tool. *6th International Learning Analytics & Knowledge Conference (LAK16)*, Edinburgh, UK, April 25 - 29 2016, ACM, New York, NY. http://dx.doi.org/10.1145/2883851.2883955

Automated formative feedback on writing (Civil Law)

	Summary			
Highlighted sentences are colour- coded according to their broad type	Important			
Kingdom, Australia has remained stagnant in its development of third party liability for knowing assistance. This paper	Both			
seeks to argue that the High Court's preferential use of precedents over legal and equitable principles has hindered the development of third party liability in the knowing assistance of trust or fiduciary duties. This over-refinement of the two				
	C Contrast			
iduciaries such as lawyers cannot. Ac CONTRAST: Disagreement, tension, cessorial liability of a third party, which				
which could ever apply to relieve the : options, inconsistency	Novelty			
It appears that the plainly wrong finding in respect of Bell was based, primarily, on: CDrummond AJA's incorrect interpretation of the dishonest and fraudulent design requirement as articulated in Farah ; and the lack of careful	P Position			
formulation of the Bell Test, which fails to appreciate the inconsistent or unsound practical effect of imposing a test -	Q Question			
separate and, in some ways, dissimilar context. The explication of the NSWCA's deterror concerning Bell provides				
Sentences with Function Keys have more precise functions (e.g. Novelty)	Trend			

Automated formative feedback on writing (Civil Law)

consequence of the plainly wrong f Barnes v Addy claim.	EMPHASIS: Additional emphasis to highlight importance	alia for the application of a second lin		
above, it was not strictly necessary		was plainly wrong . However, Leeming		
A held it was necessary to do so (see	[X] above). EHis Honour's application of	the plainly wrong test with respect to		
Bell is crucial to this discussion.				
nterpretation is plainly wrong.	QUESTION: question or missing			
Significantly, the High Court in Farah h	knowledge	tatutory law . As noted by Leeming		
JA, there are at least two complexities	with the above pass Je. WI w J hat upos it	mean to be convinced that an		
nterpretation is plainly wrong? And to	what decisions does it apply? And as e	expressed by Basten JA, the legal basis		
or such a principle is unknown, and co	onsequently, has created uncertainties in	the application of the rule. For		
established by separate inquiries) the	OUESTION: question or missing	ly, a person seeking relief under s		

1318 is not required to show they acte 1318, it will be fail to enliven protectio QUESTION: question or missing knowledge ly, a person seeking relief under s onduct will attract protection under s Leeming JA:

What then is sufficient, according to Bell, to engage second limb Barnes v Addy liability? ONot lightly would I conclude that the High Court has reformulated the law in a way which is not well defined.

COMPARISON OF HUMAN VS. MACHINE

human highlighting

automated highlighting

Example accountability analysis: writing feedback

Ethics: Does the system output results that match human analysts? Does instant feedback lead to novel benefits, or is it in fact damaging to students? What is the motivation behind the focus on reflection (e.g. rather than factual accuracy)?

Computer Science: Does the NLP platform implement the linguistic features with integrity?

Data Science: Do the linguistic features have sufficient discriminatory power? Do they ignore important qualities of value? Do they bias against certain kinds of student?

Human-Data Interaction: Do students consent to their writing being analysed?

User-Centred Design: Were students and educators involved in the design of the system? Can they make sense of the user interface and act on output?

Learning Technology: What is the educational basis for the parser rules? Does the algorithm implement the theory/rubric with integrity? Educator/student reaction?

Legal: Could a student sue the university for parser errors, or discrimination?

worked example 2 building and visualizing a student social network

Kitto, K. *et al.* (2016). The Connected Learning Analytics Toolkit. *6th International Learning Analytics & Knowledge Conference (LAK16)*, Edinburgh, UK, April 25 - 29 2016, ACM, New York, NY. <u>http://beyondlms.org</u>

Social Network Visualisation of student activity

Click on your node to view the messages you have posted. Click on a link (or relationship) to view the messages that were shared or commented on by another user.

Legend:

- White Nodes: Mentioned users not registered with CLAToolkit

- Maroon Nodes: Teaching Staff

Connected Learning Analytics Toolkit: http://beyondlms.org

⁻ Blue Nodes: Students

Example accountability analysis: SNA visualisation

Ethics: Do users validate the social networks? Does the system lead to novel benefits? What is the motivation behind the design of the system?

Computer Science: Does the social network tool implement SNA metrics correctly?

Data Science: Is the dataset biased or incomplete in important respects? Does the social network visualization bias against certain kinds of student?

Human-Data Interaction: Should students be permitted to control their degree of visibility on different platforms? Does disclosing peer data to a student violate peers' rights?

User-Centred Design: Were students and educators involved in the design of the system? Can they make sense of the user interface and act on output?

Learning Technology: What is the educational rationale for making social ties visible? Do the selected user actions implement this with integrity? Do students find it helpful?

Legal: Could a student sue for discrimination due to the network map?

approaches to designing for Learning Analytic System Integrity

Participatory design methods that build trust

Carlos G. Prieto-Alvarez, Roberto Martinez-Maldonado, Theresa Anderson (In Press). Co-designing in Learning Analytics: Tools and Techniques. In: Jason Lodge, Jared Cooney Horvath & Linda Corrin (Eds.), *From data and analytics to the classroom: Translating learning analytics for teachers.*

Communicating the algorithm to educators and students?

Design features that make assessment analytics trustworthy: a case study

LaK DESIGN, March 2017, Vancouver, Canada

	LEVEL 1 READER	LEVEL 2 CONSUMER OF INSTRUCTION	LEVEL 3 SELF-REGULATED LEARNER	LEVEL 4 COLLABORATIVE LEARNER	LEVEL 5 RECIPROCAL TEACHER
Attending class	Accesses MOOG_once	Accesses MOOC in several weeks	Accesses MOOC most weeks		Views forums in multiple weeks
Engaging widely	Watches videos	Tries in-video exercises	Uses in-video exercises routinely Reads some course readings	Accesses the range of course resources	Intensive forum user
Monitoring/producing		Tries practice quizzes	Repeats practice quizzes	Posts longer than a tweet	Posts at paragraph length
Perspective-taking		Cursory forum visitor	Occasionally uses forums	Explores forums	Reads many threads
Focusing			Revisits a thread Re-reads some course resources	Visits a thread more than twice Systematically repeats in-video exercises Re-reads many course resources	Visits a thread more than three times Re-reads many course resources Follows discussions in forums
Reaching out			Posts in forums	Posts in multiple threads	Posts in multiple weeks
Shaping debate				Posts in popular threads Votes up another's posts	Creates threads/new discussions Votes up other's posts more than once Posts in depth
Generating value				Receives a peer vote on a post	Receives peer votes on multiple posts

Milligan, S. and Oliveira, E. (2017). Design features that make assessment analytics trustworthy: a case study. *DesignLAK17 Workshop*, LAK17, Vancouver, March 2017. <u>https://sites.google.com/site/designlak17</u>

ETHICAL DESIGN CRITIQUE (EDC) PROCESS

CIC team Pro-VC Education Equity & Diversity Unit Planning & Quality Unit IT Division Director, Risk Director, Teaching Innovation Data Privacy Officer Indigenous Students Centre

EDC: ETHICAL DESIGN CRITIQUE

The EDC Template

<insert screenshot of dashboard element>

EDC team is invited to comment on:

Strengths

 e.g. Data is at an appropriate level that it does not disclose inappropriate information about a cohort or individuals

EDC: ETHICAL DESIGN CRITIQUE

Indicate concerns e.g.

- Data is displayed from incomplete sources, or has been filtered in ways, that users might not expect (Specify why this violates expectations)
- Data is shown for which there is no apparent reasonable use (Recommend removal)
- There is scope for misinterpretation due to poor design (Suggest improvement)
- This data is useful but could be inappropriately used (Caution and specify examples of inappropriate usage)

Conclusion: a new career in the Learning Analytics profession?

"Certified LASI Engineer"

Trained to audit Learning Analytics System Integrity at multiple levels, via multiple lenses

-> New principles to guide design, vendor selection, academic peer review (and maybe legal deliberation)

Learning Analytics Summer Institute, Ann Arbor, June 2017

Black Box Learning Analytics? Beyond Algorithmic Transparency

Simon Buckingham Shum

Professor of Learning Informatics Director, UTS Connected Intelligence Centre @sbuckshum • Simon.BuckinghamShum.net

cic.uts.edu.au