Algorithmic Transparency Group Working Notes

First draft <u>based on</u> flipchart notes arising from a discussion involving: Anne-Marie Scott, Ian Dolphin, Niall Sclater, Daniel BIedermann, Adam Cooper. Subsequent clarifications by the same group. We acknowledge that our background (largely HE, exclusively European, largely UK) shapes the following.

Limited scope of "learning analytics": the discussion largely focussed on learning analytics where a student is the subject. This is not to say that there are not similar concerns when considering matters of personalised learning, or LA which focuses on learning or curriculum design or aspects of the learning environment.

"Algorithm"

Scope

We take this to include all steps from source data through to the final prediction or descriptive result which is presented to the system users. This includes the transformation steps and the definition of derived quantities.

Although our initial discussion focussed on this data/IT processing perspective we felt that the term "algorithmic transparency" is being used by stakeholders to refer to information about the validity and reliability of the outputs from learning analytics algorithms. We agreed to extend scope to this because we're interested in matters of practice, not theory.

Level of Description

An algorithm may be hierarchical, embedding inner units which are considered 'black boxes'. There are different levels of description, providing incremental levels of detail. Which level is appropriate depends on the audience.

Audiences

1. Individuals in teams concerned with institutional due diligence in the procurement or development of learning analytics. Such teams should include someone with sufficient data science expertise to assess a level of description which would satisfy a data scientist.

2. The subject of the LA, the student.
3. Staff responsible for student pastoral care, "academic advising" etc
4. Teachers (i.e. staff with a primarily subject-specific role)
5. IT people and Educational Technologists
6. Researchers using the data
7. Ethics committees

Motivations

A brainstorm of why we might be interested in algorithmic transparency:
1. As a basis for informed consent under GDPR. i.e. to be sufficiently informed to give consent, some understanding of what is being done is required.
2. Institutional quality assurance (etc) both a) as "acceptance" during procurement or development, before deployment into use and b) for monitoring ongoing operations to give assurance that the running system remains within bounds of "acceptable".
3. To satisfy broader educational aims, helping stakeholders to better understand how the LA system works, that it is not "magic". This is a matter of data literacy.
4. As a means of building trust between students, staff, institutional management/leadership and suppliers
5. To enable ethical action; a more clear understanding of the limitations, making the consumers of LA more accurate in their understanding.
6. To support action; transparency to inform planning of what to do, explaining the basis of prediction etc. The point here is to provide more than e.g. a prediction or the description of a social network, to better support theorising about how to make difference through action.
7. Assessment of validity and reliability of predictions, SNA models, clusters, etc…

Audience-Motivation Matrix

We undertook an informal ("quick and dirty") mapping of what we thought the "to be" state should look like, rating each non-empty cell as being Desirable or Essential.

	Legal	QA	Data literacy	Trust	Ethical	Action planning	Validity
Purchasing	E	E		D	D		E
Students	E		D	E		D	
Teachers	E						
Tutors	E		D	E	D	E	

IT/Learning Tech	E		E	E	E		E
Ethics committees	E	E	E	E	E		E
Researchers	E		D	E	E		

What Might Algorithmic Transparency Look Like?

A brainstorm of some ideas. Day 2 activity will look to expand on these in the context of one or more audience stereotypes and different motivating use cases.

1. For an individual prediction, who how various attributes contributed to that prediction being made.
2. Description [text or visual] of what data sources were used, and the statistics derived from them
3. Indication of the specific data used in building the statistical model - e.g. data from given academic years of activity for specified courses or subset of the student population
4. Indication of which attributes are more or less associated with outcomes of interest. I.e. a description of the statistical model
5. Allowing users to choose stereotype "personas" or change attributes to see what effect that has on the results of the LA, the predicted likelihood, SN graph, etc
6. Making intermediate stages of data preparation accessible, not trapped in the LA software.
7. Exposing predictive model performance metrics, goodness-of-fit statistics, etc for assessment of reliability and validity.

Other Resources

We did not discuss these during the session but note the following relevant resources:
- Talk by Simon Buckingham Shum at UCL about Algorithmic Accountability [worth noting that Adam's hackathon challenge pitch (see slides in github) asked about difference between "accountability" and "transparency"]
  https://utscic.edu.au/algorithmic-accountability-learning-analytics/
- World Wide Web Foundation report on algorithmic accountability
  http://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf

- Web site for Governing Algorithms conference http://governingalgorithms.org/
- LA Google Group discussion on AA arising from LASI-Denmark
  https://groups.google.com/forum/#!topic/learninganalytics/Yt2siQnlYtY
- How the Machine 'Thinks:' understanding opacity in machine learning algorithms, Burrell, J 2016
  https://www.ischool.berkeley.edu/research/publications/2016/how-machine-thinks-understanding-opacity-machine-learning-algorithms
- Australian National Statement on Ethical Conduct in Human Research
  https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72_national_statement_may_2015_150514_a.pdf