Predictive AI Workshop Summary

UTS Student Partnership in AI

UTS CRICOS 00099F
UTS TEQSA PRV12060

Edited by: Simon Buckingham Shum (Director, Connected Intelligence Centre)

Approved by:

- Workshop team
- Workshop participants
- Nour Al Hammouri (President, Students Association)

Contents

Acknowledgements to our student partners:

Alexander Cook
Amogh Guragai
Amy Chia
Bharti Bulchandani
Bri-anne Lake
Duong Sinh Hoang
Edward Cameron
Florian Huebscher
Hanli Smith
Jet Hodgson
Joshua Pye
Leanne Aldridge
Lorenzo Canu
Mark Brewer
Olivia Dinnar
Pegah Jalali
Raphaella Revis
Rex Siu Xiao
Soudabeh Niroomand
Vaishali Raja

Acknowledgements to the workshop experts:

Gregory Martin (Associate Dean for Teaching & Learning, Faculty of Social Sciences)
Kirsty Kitto (Associate Professor, Connected Intelligence Centre)
Simon Buckingham Shum (Professor, Connected Intelligence Centre)
Susan Gibson (Head of Data Analytics & AI)

Introduction

The UTS Student Partnership gives students a meaningful voice in shaping how the university runs. The **Student Partnership in AI** initiative was convened by Prof. Kylie Readman (Deputy Vice-Chancellor for Education & Students), with the support of Craig Napier (Chief Data Officer), Susan Gibson (Head, Data Analytics & AI). Working in partnership with the Students Association, these workshops are coordinated by Prof. Simon Buckingham Shum (Director, Connected Intelligence Centre) with support from Dr. Jan McLean (Director, Institute for Interactive Media in Learning), whose teams designed and ran the workshops. Gregory Martin (Assoc. Dean Teaching Learning, Faculty of Arts and Social Sciences) and team played a key role in helping to conceptualise the proposed pilot of predictive AI, discussed in the workshop (including TLSU Manager, Teaching & Learning Coordinators, and First & Further Years Experience Coordinator).

The 2023 workshops used the principles of **Deliberative Democracy** pioneered at UTS in 2021 as a successful process for student/staff consultation around the ethics of analytics and AI in educational technology. Principles proposed by that team have since been formally adopted as part of the UTS AI Operations Policy.

Recruitment: Out of 154 expressions of interest responding to online advertisements linking to a website, participants for two workshops (20 each, 24-25 May 2023) were recruited through stratified sampling to maximise the diversity of voices, balancing as far as possible students' faculty, gender, undergraduate/ postgraduate, Indigenous status, domestic/international, and early/mid/late stage of study. All faculties were represented at both UG and PG level, and included 4 students identifying as non-binary, 6 identifying as Aboriginal and/or Torres Strait Islander people, and 28 speaking a language other than English at home. The 4 hours students committed to preparation, the workshop and report review were compensated with a gift voucher.

One workshop focused on **Predictive AI** (this report), and the other on **Generative AI** and Turnitin.

Workshop

Following pre-workshop reading and online discussion of predictive AI scenarios both fictional and factual (Appendix 1), a half-day workshop convened face-to-face on campus. UTS experts gave introductory briefings (Appendix 2) to inform student thinking:

- introducing predictive modelling of student outcomes.
- ethical questions to consider.
- a proposed pilot at UTS.

Participants then split into groups and were asked to review Predictive AI against the UTS AI ethics criteria, using the template developed by the AI Operations Board (overleaf). This was an accelerated version of the analysis that UTS AIOB conducts, informing the students how an AI ethics audit can be conducted, in order to surface their perspectives. The groups' analyses are aggregated in this report and reproduced in Appendix 3. Students were permitted to add to their analyses for a few days after the workshop, and the Teams online discussion remains open.

UTS AI Ethics Principles template: <*AI System*>

Principle			Alignment	Rating 🟩🟨🟥
1.	**Community benefit**	AI should deliver the best outcome for human users, in this case, the UTS community, and provide key insights into decision making. AI must be the most appropriate solution for a service delivery or a policy problem, considered against other analysis and policy tools.	***Student views…***	
2.	**Fairness**	Use of AI will include safeguards to manage data bias or data quality risks. The best use of AI will depend on data quality and relevant data as well as careful data management to ensure potential data biases are identified and appropriately managed.	***Student views…***	
3.	**Privacy and security**	AI will include the highest levels of assurance. The UTS community must have confidence that data is used safely and securely in a manner that is consistent with privacy, data sharing and information access requirements.	***Student views…***	
4.	**Transparency**	Review mechanisms will ensure that the UTS community can challenge and question AI-based outcomes and will have access to an efficient and transparent review mechanism if there are questions about the use of data or AI-informed outcomes.	***Student views…***	
5.	**Accountability**	While AI is recognised for analysing and looking for patterns in large quantities of data, undertaking high-volume routine process work, or making recommendations based on complex information, AI-based functions and decisions must always be subject to human review and intervention. AI system owners and business owners are responsible for the management of their AI systems.	***Student views…***	

Risk	Cause(s)	Controls to reduce risk	Rating
			Critical
			High
			Medium
			Low

Key Themes

Drawing on these materials and a closing plenary discussion, the following themes emerged.

There was a spread of reaction to the pre-workshop scenarios (fictional and factual), with some students feeling uneasy, and others seeing it as an appropriate use of technology by a university. This workshop did not provide enough time to dig very deeply into the different viewpoints (e.g., to see if disagreements could be resolved). Future workshops could open up the issues in more depth.

Themes 1-5 reflect the five principles in the ethics audit template. It will be clear that there are many interdependencies between them.

Community benefit

Principle: AI should deliver the best outcome for human users, in this case, the UTS community, and provide key insights into decision making. AI must be the most appropriate solution for a service delivery or a policy problem, considered against other analysis and policy tools.

Some students welcomed the idea that UTS would be taking care of them in this way. The view is along the lines that analytics/AI are used in every sector, and education should be no exception. Potential benefits if predictive AI is used well, including relieving load on current student support services by making timely interventions, improving retention, and more informed decision making. Informed consent was the precondition for reaping these benefits (discussed further). The pre-workshop example discussed by students of the Open University demonstrated for these students care for student wellbeing and success, as long as personal information about students is handled professionally.

Other students were uncomfortable with this kind of surveillance, even if intended to be only in their best interests. Some students felt that this was inappropriate at university, where adult students should take responsibility to reach out for help if they need it. Not having given consent to be tracked and contacted was deeply problematic.

Online activity tracking should be clearly stated as restricted to named UTS platforms, and not extend to social media, etc.

It was suggested that some students may feel more confident talking to a conversational chatbot than their tutor, which would be another source of data (but students would need to know that this data might be used by a student support service).

Fairness

Principle: Use of AI will include safeguards to manage data bias or data quality risks. The best use of AI will depend on data quality and relevant data as well as careful data management to ensure potential data biases are identified and appropriately managed.

UTS must not lose the trust of students that data is being collected for ethical purposes in their interests.

If the predictive model can function well based on student activity in UTS platforms, without including demographic attributes, this helps mitigate concerns and risks about bias from known sources of (or proxies for) historical inequities (ethnicity; postcode; gender; first-in-family at uni; etc.). We do not want past injustices to cast a shadow over a student and mark them as "high risk" before they even start at UTS.[1]

There were different levels of trust in UTS, reflecting the diversity of the group. More time would be needed to clarify if there are important patterns here. Anything with UTS branding on it should be trustworthy.

There was a strong sense that this should not be done without consent. To give informed consent, well designed communications are needed that explain to a diverse audience how and why UTS uses predictive AI.[2]

---

1 Note of clarification: Even if demographic data is excluded from predictive models, UTS can still use such data in other systems to monitor other social justice indicators (e.g., imbalances in enrolment or retention; if we are closing gaps in outcomes).

2 Note of clarification: All UTS websites provide a Privacy Notice explaining that users' data is being logged, and may be used to "for quality improvement and planning, including strategic planning, in relation to our admission processes, courses and marketing strategies, and to evaluate overall student outcomes through to course completion". This is, however, not a document that many students view, and specific analytics/AI approaches are not mentioned. A potential action from this consultation is the design of a more visibly promoted statement, similar to this example from The Open University UK: Using information to support student learning.

The question of who sees and acts on the predictive model's output is important and spans multiple principles. Some students expressed concerns that some staff might misuse the insights. The question of staff training and which staff have access to the models therefore arises.

Privacy and security

Principle: AI will include the highest levels of assurance. The UTS community must have confidence that data is used safely and securely in a manner that is consistent with privacy, data sharing and information access requirements.

The following questions were raised:

- *Is my data being used for predictive modelling of my progress?*
- *What data does that cover?*
- *Who can see the data, and how are access/edits controlled?*
- *Can I view and correct data about myself?*
- *How reliable is the predictive model?*
- *Will only people (not AI) be making decisions based on the model?*
- *What decisions could be taken?*
- *How does UTS monitor what decisions were taken, or not taken?*
- *Who will contact me? (Did I consent for them to have my mobile number?)*
- *Can I ask why was I contacted?*
- *What happens if there is a data breach?*
- *How long are the models predictions stored?*
- *Is any of this data shared outside UTS, e.g., with AI vendors whose products we use?*
- *Is the data sold externally?*
- *Can I opt out?*

The last question has variations, such as permit the university to include my data but opt out of being contacted or indicate how would prefer to be contacted. Tracking and respecting such preferences have technical implications that will need to be investigated.[3]

The clear student communications that were called for around predictive AI should provide answers to these questions.

[3] The modelling experts pointed out that permitting students to opt out could also lead to biases in the data, although exactly what the nature and seriousness of these would be is an empirical question.

Transparency

Principle: Review mechanisms will ensure that the UTS community can challenge and question AI-based outcomes and will have access to an efficient and transparent review mechanism if there are questions about the use of data or AI-informed outcomes.

Many comments on this principle overlapped with Privacy and are already recorded above.

Students made it clear that understanding how the prediction is arrived is important. In machine learning (as proposed for the pilot, and in the ethical pre-workshop example) it is statistical likelihood (your online activity traces match those of others who fail to submit their next assignment). In one of the preparatory, unethical examples discussed by students, it was deeply suspect inferences about likelihood of success based on survey responses during Yr.1 orientation. This must be clearly explained in the communications already discussed.

In addition, it was pointed out that there are many students at UTS who can understand predictive modelling. Why not share model details openly as open-source code, and if not, what would the risks be? [4] We need to consider mitigating for unintended consequences such as trying to trick models. An example given in the workshop recounted one experiment in which students became aware that library data was utilised in the predictive model and gamed by going repeatedly through the library gates. This would then likely cause inaccuracies in the model.

Accountability

Principle: While AI is recognised for analysing and looking for patterns in large quantities of data, undertaking high-volume routine process work, or making recommendations based on complex information, AI-based functions and decisions must always be subject to human review and intervention. AI system owners and business owners are responsible for the management of their AI systems.

This principle connects to all of the others, and for students, centres on how they can hold UTS to account for implementing the above principles and responding to questions. The Students Association is currently the best route, since they have a seat on the AI Operations Board. Future forums/workshops could provide other ways.

[4] Notes of clarification:

- Publishing the models (algorithms) does not require releasing student data.
- This clearly connects to *Community benefits* (e.g., could the models be improved, perhaps even in student projects?) and *Accountability* (review is impeded without full technical details).
- Open-source code is part of open sharing, but the details of algorithms must be shared mathematically, and summarised in prose, which assists non-technical readers.

Making contact with students in the appropriate way

This theme is not one of the 5 principles in the UTS AI Operations Policy but came through in the workshop.

Central to predictive AI is the concept of timely intervention: someone is going to act on the basis of a prediction. At present, this would most likely be contact from someone, to see how they're going.

Students spoke a lot about what would make this contact feel supportive:

- Many felt that receiving an unsolicited call, text or email would be surprising and stressful.
- There should be clear communications at start of each session that predictive AI is being used, how students might be contacted, and by whom. An email notifying a phone call coming next day would increase the chances of the call (from an unknown number?) being taken.
- Can students express preferences about how they are contacted (text; phone; email…), and by whom (student volunteer; tutor; subject coordinator…)? Indigenous and Torres Strait Islander students might appreciate being contacted by Jumbunna noting demographic data would need to be utilised to triage appropriately.
- Could a student ask on what grounds they had been contacted, and be informed in accessible terms what the AI model was saying? [5]

Summary

Given the diversity of the students selected for this workshop, it is not surprising that there is no single student perspective. What does come through very clearly, is that students would value clear communications, and the option to explicitly consent, or decline, to being part of student support programs using predictive AI. This would recognise the range of experiences that students have had with data, analytics and AI, and their personal preferences, and avoid undermining a very precious quality — students' trust in UTS — by imposing predictive AI with no student agency.

These principles informed the design and implementation of the predictive AI pilot, whose outcomes are disseminated in later reports.

[5] Note for clarification: "Explainability" is an important technical property that varies widely depending on the kind of AI. The product used by UTS provides technical explanations underlying its outputs, but these must be translated into accessible terms for the lay person. The workshop did not have time to explore this in more depth, or the question of whether informing a student that the AI model had flagged them would in fact be counter-productive if they are in a distressed state. On the other hand, if the use of predictive AI has been clearly communicated, and the student has opted in, arguably this would not be an issue.

Aggregated Risk Analyses

This table aggregates all the risk analyses from the student groups, verbatim.

Note: **bold** indicates responses from the UTS staff team to points raised by students.

Risk	Cause(s)	Controls to reduce risk	Rating
Unintentional sharing/ public peeking of student's data when staff (and student representatives) accesses the predictive AI system outside of office etc. in WFH arrangements.	They (relevant authorities) may leave their screens unattended at home/outside when taking breaks from assessing students' performance. This could lead to unnecessary leakage of sensitive student data (like their performance rates, demographic background, and/or attendance) which may bring rise perpetual social injustices or possible misuse of contact details.	Consider implementing the following policies to limit access and privatise information shared by the AI model. Tracking who, when and why they are accessing (e.g., student representatives vs staff controls) Access only given on specific computers or during regular office hours. Staff team: Data management practices are in place to ensure that only appropriate staff have access to confidential information. Staff have a responsibility for ensuring that they do not leave screens unattended. Pilots envisage using the same teams of trained students (who call all 1st Year students) to see how they're going. They follow a script, and refer to a Faculty member of staff, or the Student Support Unit, should the student wish to speak with someone. Contacting students and staff to be done only during office hours as well, to prevent any misconduct or possible exploitations. Staff team: We can confirm that students would only be called during office hours.	Critical
Scaring and making students uncomfortable	Approached based on socio-economical background (or other sensitive data).	Interacting with specific depts. (Jumbunna, Financial system, CSJI).	Critical

Misuse/misinterpretation of data (by tutors or staff) Human bias Incorrect data being used. When reviewing data input this could be distressing No one trusts the UTS review systems.	Lack of education or knowledge of consequences UTS doesn't support right to self ID Gender People's experiences with plagiarism reviewal process etc **Staff team: Turnitin's AI-writing detector is currently being evaluated and will not be deployed until we are satisfied that the risk of errors is low enough to be acceptable.**	Education, training, disciplinary actions Education on a more tech-neutral perspective	Critical
Data is hacked	Scam	Update the firewall regularly **Staff team: All AI models and related data are implemented in systems that have undergone UTS cybersecurity and solution design approvals**	Critical
Enabling or installing predictive AI technology into student applications, by tracking online activities without consent and ethical consideration re. surveillance and personal preferences/wish Biased human intervention, editing the data according to their opinions, causing model to skew	Academics may overlook the gravity or sensitivity out of the good intentions to support students. "There is nothing malicious from us - we are only using the technologies to better support you." Lack of informed communication or awareness of the use of predictive AI. As humans we all have our own biases, feel something is important and something is not, they should be not able to edit the information based on what they "feel" is appropriate which might put various students at disadvantage while also skewing the model	- Concise and intermittent communication among the cohort during the course about AI use and faculty's intention - Always provide opt-out options - Consent-first basis • Proof to be submitted for claim being made • No direct edits made unless vetted by panel of experts who are in charge of the integrity of data **Staff team: The technical feasibility of opting out will need to be investigated.**	High

Poor communication by university	• Use of institutional language and "you must trust me" • Lack of sensitivity towards language barriers	Making sure that key stakeholders are aware of the use of AI • Students should be aware of how their data is kept safe – Universities should communicate appropriately	**High**
Students attempting to influence the model	• Not being aware on the consequences and detriments that this can have to the students themselves as the data collection is for their benefit	– Opt-in basis but having a disclaimer that allows students to be aware of what happens when they opt-out (i.e., not having such a curated model, etc.)	**High**
Misuse of personal data by staff	•	– Controlling who has access to the data or people only have access to specific data **Staff team: UTS has stringent data management and privacy policies and procedures around ensuring appropriate utilisation and access of data**	**High**
Access to data is not centralised to one person or group of people	• Information can be edited by people singularly without any repercussions but need to be pre-emptive	Ensuring that information, especially administrative rights are distributed across multiple people/ entities to ensure democratic decisions **Staff team: UTS has stringent data management and privacy policies and procedures around ensuring appropriate utilisation and access of data**	**Medium**
The potential of missing data/ messy data	• For some reason there's previous data that hasn't been accounted for or for some reason it isn't accurate/ noisy data, then it could create inaccurate decisions	Would have to ask students for consent to collect data Store all data in a data warehouse to make sure its raw data that can be data processed/ cleaning to create information which can lead to knowledge/ insights	**Medium**
			Low

Appendix 1: Pre-workshop activities

https://cic.uts.edu.au/projects/ai-ethics-consultation-2023

UTS Student Partnership in AI Workshop:

Predictive analytics to improve student outcomes

Preparatory work

Wednesday 24th May, 2023, 10am-12pm + lunch

Connected Intelligence Centre [GMap]

Thanks for committing to an hour's prep so you hit the ground running at the workshop!

1. Please read these two stories, and post your thoughts on at least one in the Workshop Team, where we encourage you to also respond to others.
2. Take a look at the AI Ethics template that we'll be using in the workshop.
3. Please bring your laptop to contribute on the day to the shared Workshop Notes.

Any questions, email the Connected Intelligence Centre (CIC) cic@uts.edu.au.

Story 1: Your degree has been going fairly well, getting the grades you hoped for, and exceeding them a couple of times. However, you hit a bad patch, which just gets worse as you head for Christmas. Family stresses. Your best friend has switched to another university. To cap it all, you had to find a new rental place, and the internet is mediocre at best.

It feels like the coursework just got tougher. You know you're not staying on top of it. You're tired and low, and miss a couple of Zoom and face-to-face sessions. You got confused when the last deadline was, and only downloaded the assignment the day before it was due. Your friends Zach and Millie are doing their best to be helpful, but they're cruising through.

With 2 weeks to go before the next deadline, you get a phone call from your tutor asking how it's going, since you don't seem to have been your usual efficient self, and would you like to have a chat? Maybe there are some things they can help with? You're impressed! This is really helpful.

But hang on. Are they calling every one of the 800 students, or just you? Is this level of attention a bit weird? You check in with Zach and Millie: turns out they didn't get phone calls, but emails encouraging them to keep up the good work since they look on course for HDs. But Mimi, one of your other friends who's hit a tough patch, was also phoned by her tutor.

In order to provide differentiated communications like this, UTS has been tracking students' online activity, and an AI model recommends which may be in need of extra support. Universities using predictive AI regard it as an obligation to use every available tool to support students.

Do you agree?

Do you want to know more about how this is done?

Do you trust that UTS is doing this responsibly?

Story 2a and 2b. Ethical and unethical use of predictive models in two universities.

Ethical (The Open University, UK). "An example of how a teacher successfully used the predictive analytics dashboard is discussed below. This teacher was able to use the dashboard to provide timely support to a female engineering student from a Black and Minority Ethnic (BME) background with no prior higher education experience, and enable her to succeed. Prior to this, the student received 100% on the first assignment (a quiz) and 86% on her second assignment. However, in week 10, the dashboard flagged the student as unlikely to submit the third assignment. Upon further inspection by the teacher, it emerged that the student had not accessed the VLE after submitting the previous assignment three weeks earlier. When the teacher contacted the student, it became apparent that the student's lack of activity on the VLE was due to the birth of her third child. The student had not disclosed her pregnancy as she was unsure whether the university would allow her to carry on with her studies. Not only did the teacher resolve the misunderstanding, but also provided support enabling the student to get back on track.

Subsequent monitoring of the student's performance helped the teacher identify another occasion when the student had limited VLE activity and was likely to fail to submit her next assignment. Again, the teacher was able to prevent the student from giving up by identifying the problem she was facing and providing timely support. The student eventually completed the course with an average score of over 80%."

Source: **Learning Analytics in Open and Distance Higher Education: The Case of the Open University UK**

Unethical (Mount St. Mary's University, USA). In this example, it was not an AI model, but a student survey issued during Orientation: "This year, we are going to start the Veritas Symposium by providing you with a very valuable tool that will help you discover more about yourself. This survey has been developed by a leadership team here at The Mount, and it is based on some of the leading thinking in the area of personal motivation and key factors that determine motivation, success, and happiness. We will ask you some questions about yourself that we would like you to answer as honestly as possible. There are no wrong answers."

The university's President hit the headlines when he proposed using the data to kick out the weakest looking students, in order to boost the university's retention statistics. His infamous quote that they needed to "drown the bunnies" is now an archetypal example of how in the wrong hands, predictive models could be used to victimise rather than support students.

Sources: The Mountain Echo / Washington Post / Minding the Campus

What responses do these two examples provoke in you?

Appendix 2: Workshop schedule and briefing slides

We will have Student Association representatives observing, and the following UTS experts will be presenting/discussing:

- Greg Martin (Associate Dean for Teaching & Learning, Faculty of Social Sciences)
- Kirsty Kitto (Associate Professor, Connected Intelligence Centre)
- Simon Buckingham Shum (Professor, Connected Intelligence Centre)
- Susan Gibson (Head of Data Analytics & AI)

10.00 Welcome & Overview (Simon, Susan & Student)

- Student Partnership
- EdTech Ethics
- CIC & DAIU
- Workshop Notes

10.10 Who's in the room: 30 sec intros

10.20 Predictive Analytics 101 (Kirsty)

- Core concepts underpinning machine learning-based predictive models of student outcomes

10.30 Q&A

10.35 AI Ethics (Simon)

- Some critical questions to think about

10.45 Q&A

11.00 Piloting at UTS (Greg & Susan)

- First steps for UTS: the opportunity
- This is to explain what is being planned as a first step, but not to walk through our ethics audit template in order to declare that we think everything's fine (!) — we want to see what the students come up with themselves

11.10 Q&A

11.15 Grps x5: Applying the AI Operations Policy principles

- Each group assigned one of the 5 ethics principles in AIOP: discussion and notes in Workshop Notes
- Split between Studio and adjoining room

11.35 Plenary discussion (Greg)

- Identify key themes, cross-connections
- All supporting if questions arise that need expert input
- Blaise take notes in Workshop Notes

11.58 Next steps (Simon)

- Teams: discussion continues and the chance to ask further questions which we will answer
- Workshop Report will be circulated for comment on 2nd June
- Final report will be taken into consideration by the Faculties, and Predictive Analytics team, who will also consult with the Student Advisory Group to the Deputy Vice-Chancellor for Education & Students (reps here)
- Student Association will present report to AI Operations Board on 5th July.

12.00 Lunch, chat and stickies: *I like, I wish, I wonder…*

- Free lunch! Stick around, chat, and post more feedback stickies on the 3 zones: *I like, I wish, I wonder…*

Student Partnership in AI Workshop

Predictive AI

Student Partnership in AI Workshop

Welcome & Overview

Simon Buckingham Shum (Director, Connected Intelligence Centre)
Nour Al Hammouri (President, UTS Students Association)
Susan Gibson (Head of Data Analytics & AI)

Acknowledgment of Country

I would like to acknowledge the Gadigal people of the Eora Nation upon whose ancestral lands UTS City campus now stands.

I would also like to pay respect to the Elders both past and present, acknowledging them as the traditional custodians of knowledge for this land.

Student Partnership Agreement

"Successful partnerships, as fostered in this agreement, depend on mutual respect, integrity, meaningful interaction, open collaboration and an agreement on common goals and values, acknowledging that diversity is a strength."

Student Partnership Agreement
→ for the responsible use of AI

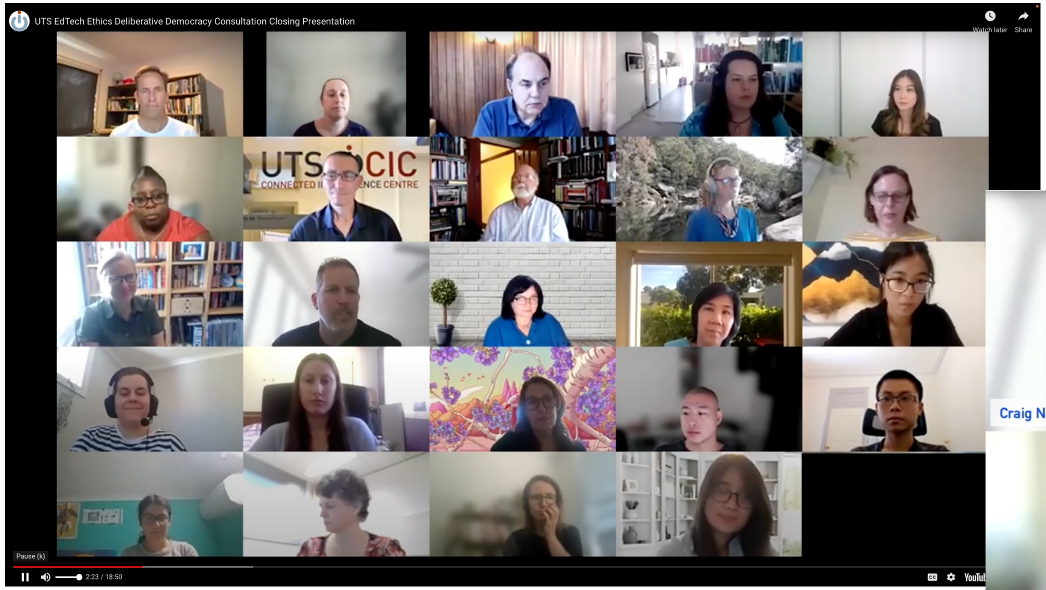
"Successful partnerships, as fostered in this agreement, depend on mutual respect, integrity, meaningful interaction, open collaboration and an agreement on common goals and values, acknowledging that diversity is a strength."

→ **UTS-SA representation on the AI Operations Board**

→ **These workshops on AI ethics**

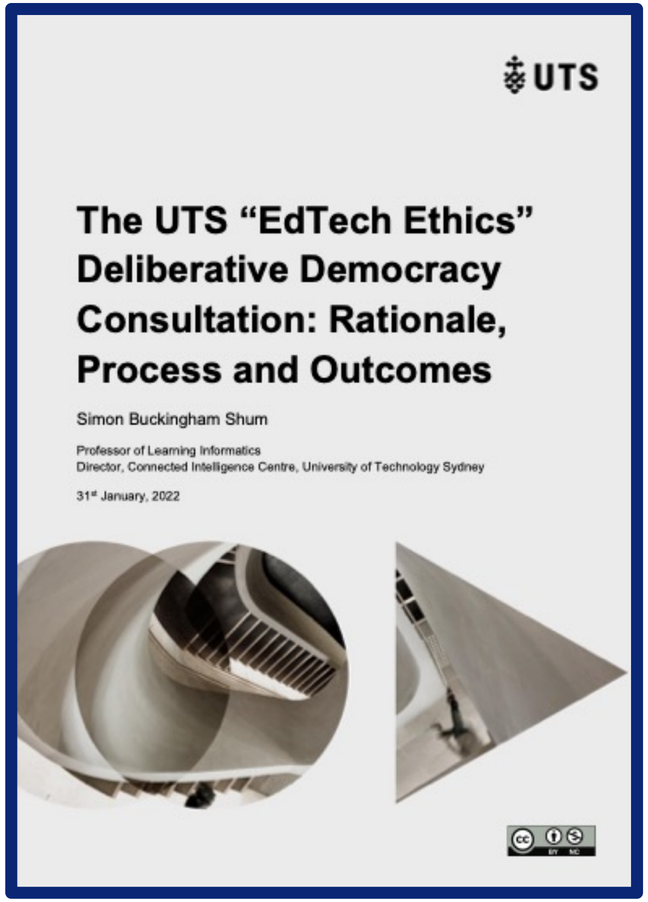
2021: In the spirit of the Student Partnership Agreement

Deliberative Democracy consultation on EdTech Ethics



2021: In the spirit of the Student Partnership Agreement

Deliberative Democracy consultation on EdTech Ethics


UTS

The UTS "EdTech Ethics" Deliberative Democracy Consultation: Rationale, Process and Outcomes

Simon Buckingham Shum

Professor of Learning Informatics
Director, Connected Intelligence Centre, University of Technology Sydney

31st January, 2022

"I did not have any experience with being tasked with such a big responsibility to come up with principles that would affect everyone at the University. All the stakeholders. So, it was a genuinely proud moment when we finished, but I'm just interested in how this conversation goes on, moving forward, and as we discussed in the final meeting, we would really like it not to be a full stop; rather, an ongoing conversation."

Who’s in the room?

30 second intro:

Name + preferred pronouns

Degree program

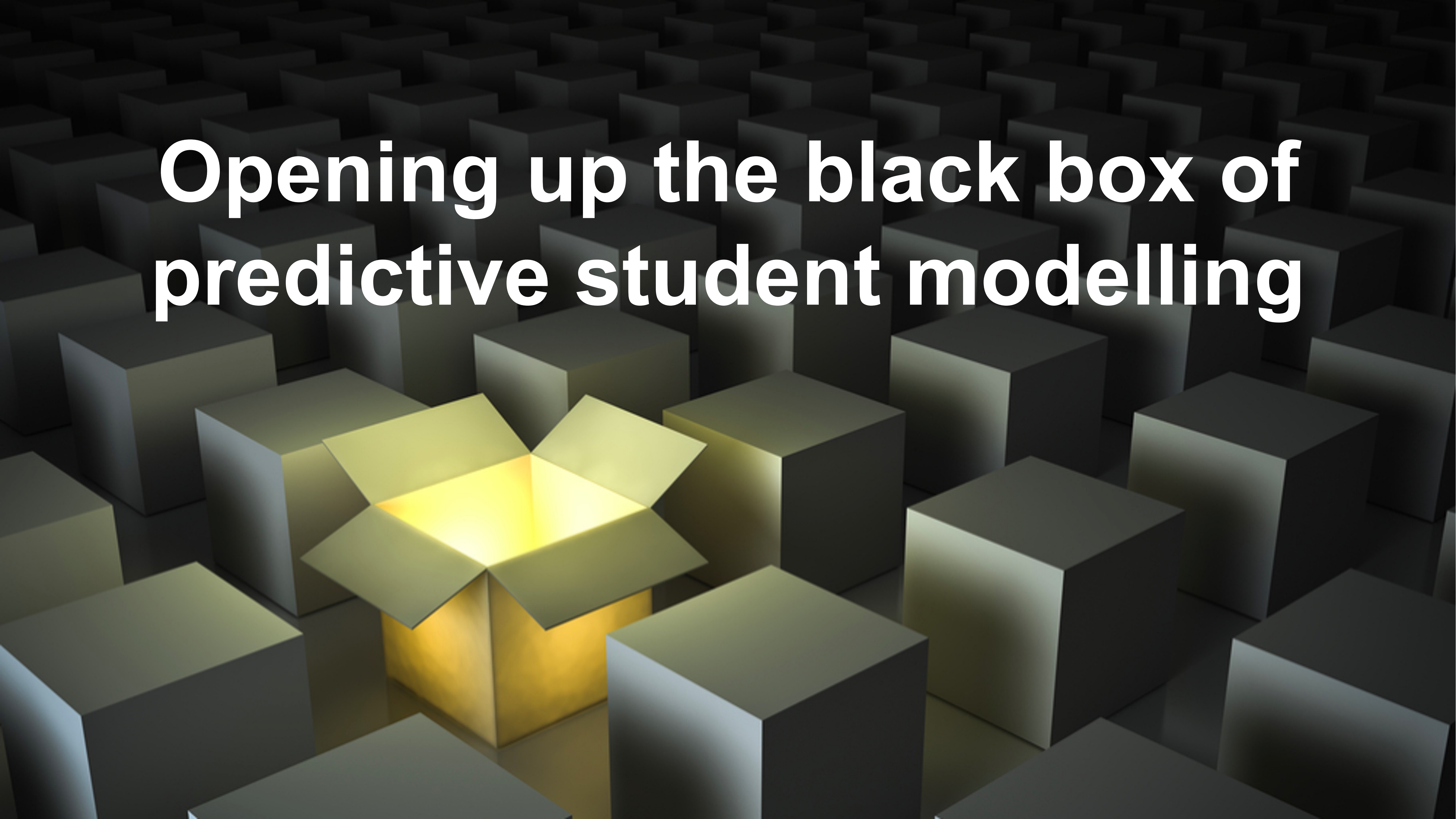
Why I’m here!

Student Partnership in AI Workshop

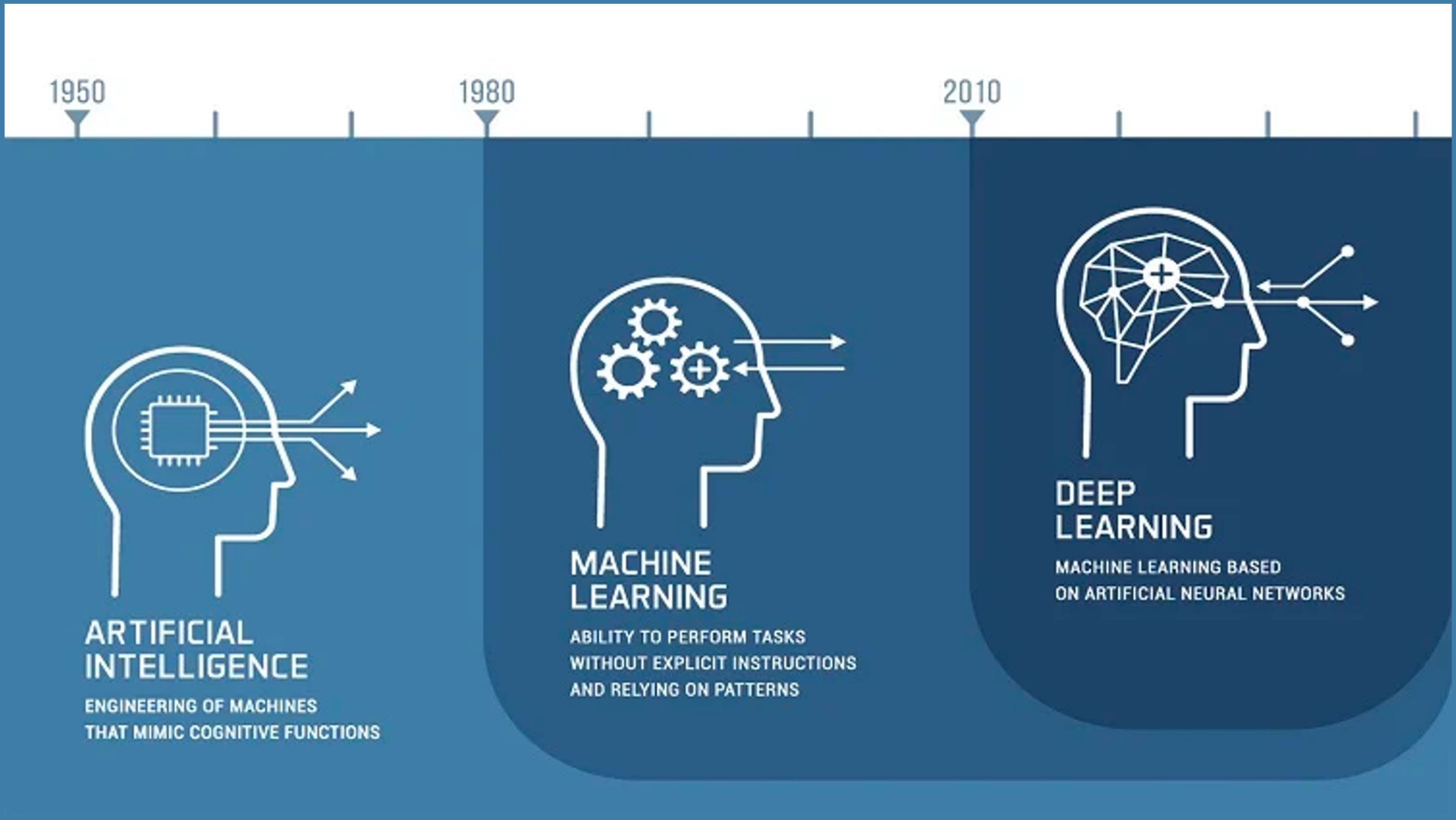
Predictive Analytics 101

Kirsty Kitto
Associate Professor of Data Science, Connected Intelligence Centre

Opening up the black box of predictive student modelling

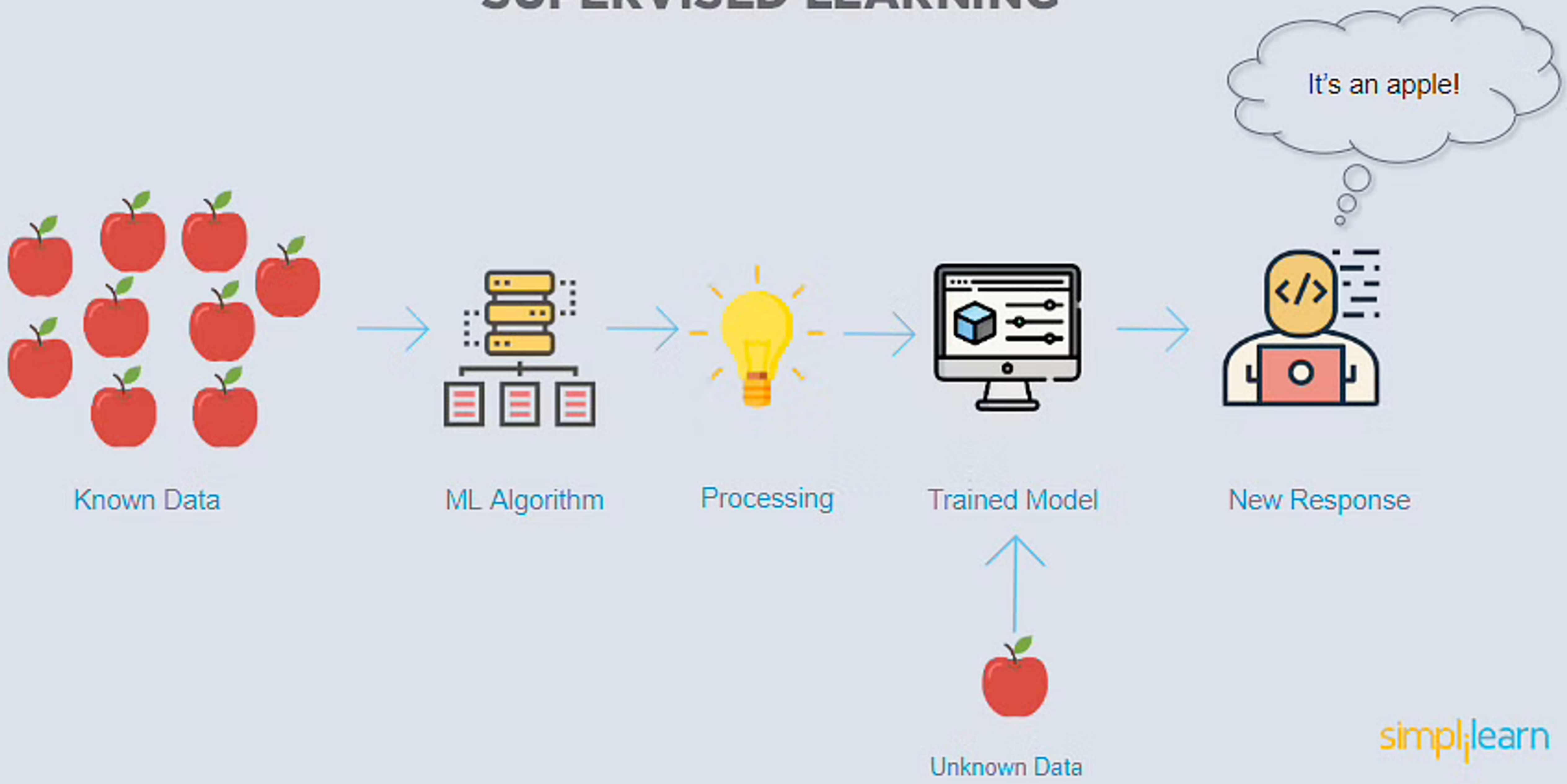


Predictive student models tend to use methods from machine learning



and this is most commonly a form of supervised model

SUPERVISED LEARNING



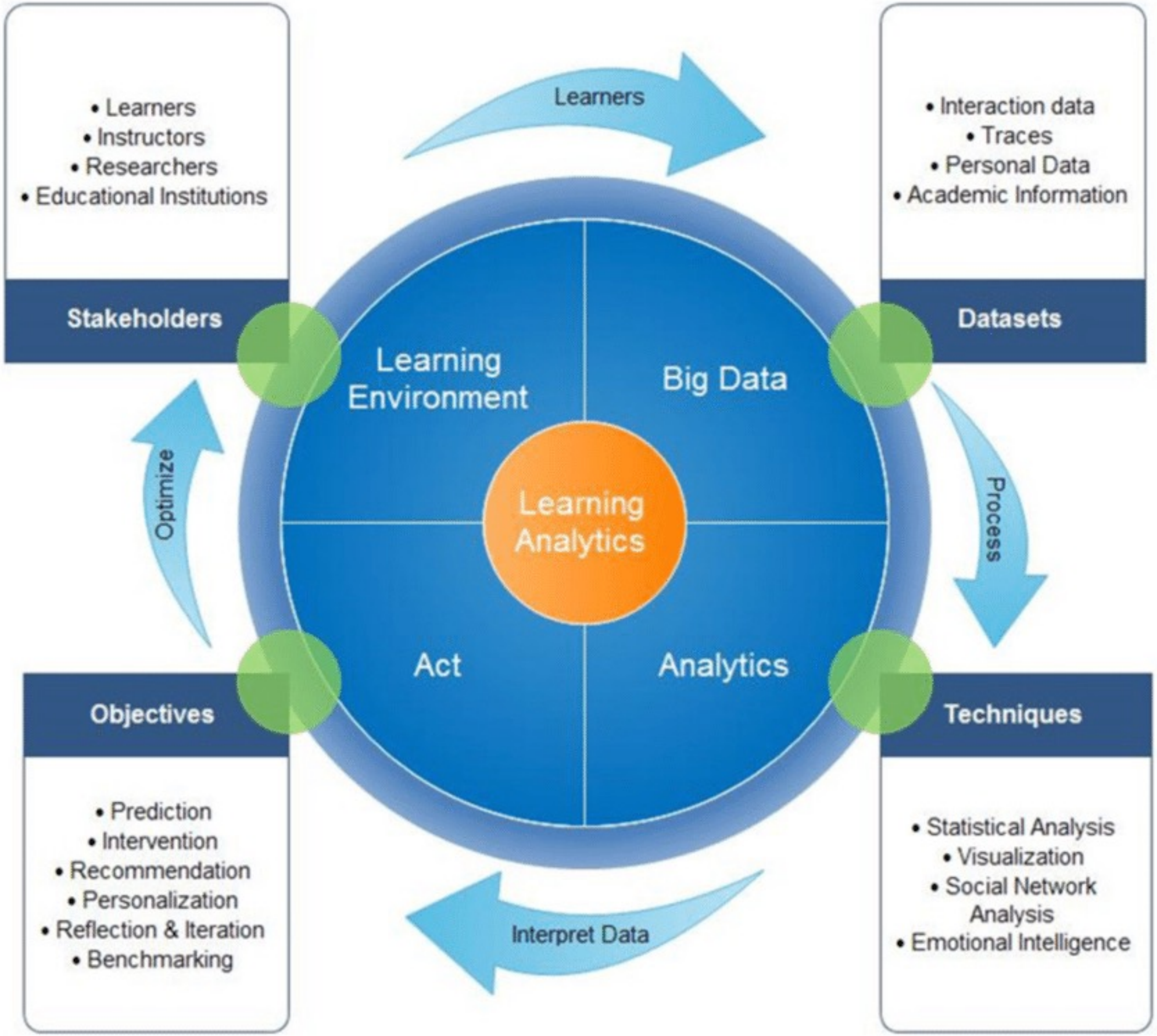
How does this work for predictive modelling of students?

Note that this type of thing can be used to predict lots of different types of things…

For example: at risk of failing a subject, at risk of dropout, emotional wellbeing, student success, language support… it depends on the university!

And it is common to limit who has access to predictions

Khalil, M., & Ebner, M. (2015, June). Learning analytics: principles and constraints. In *EdMedia+ Innovate Learning* (pp. 1789-1799). Association for the Advancement of Computing in Education (AACE).



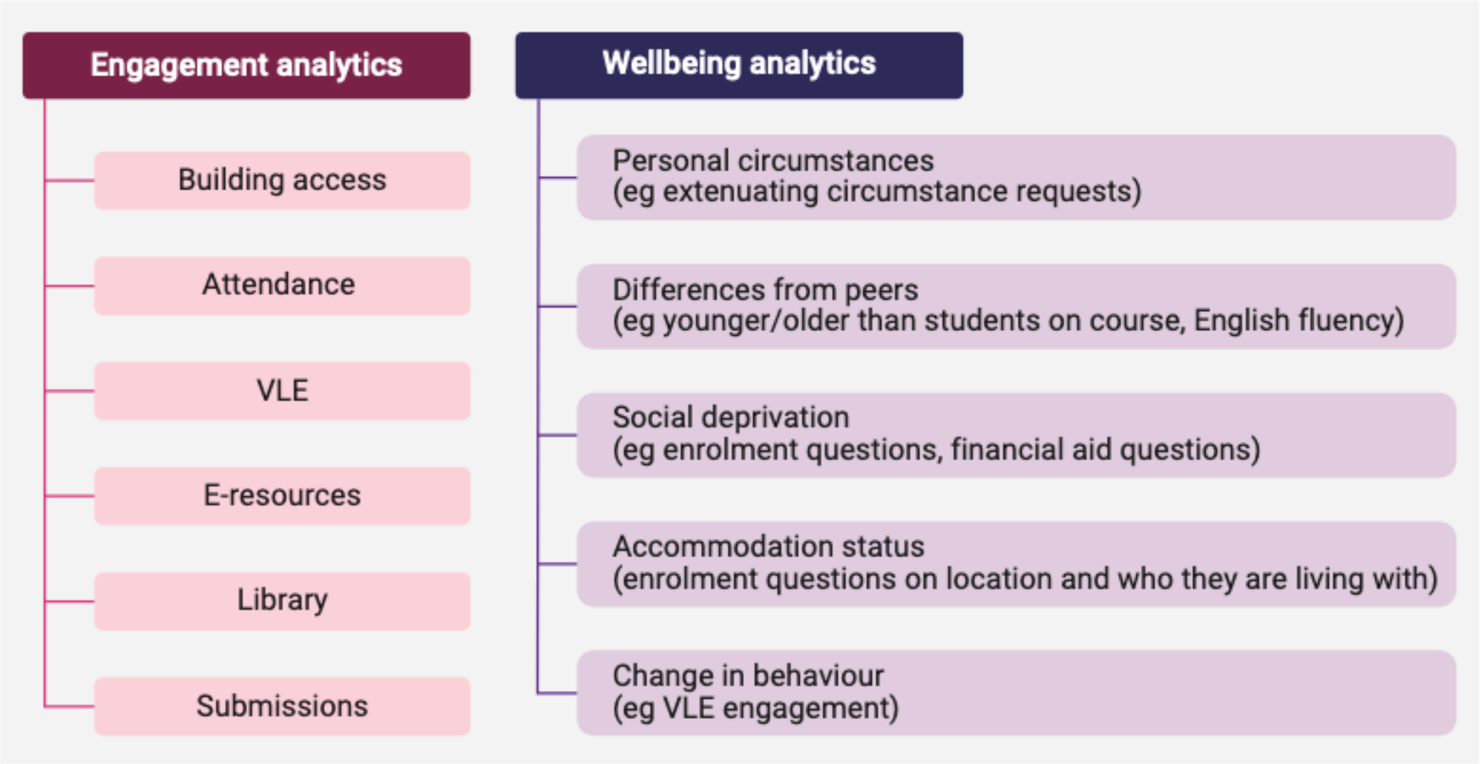
What kinds of data can these models use?

It depends!

Different types of student models use different sorts of data, and it is possible to limit what gets pulled in.

It is also possible to do things like:

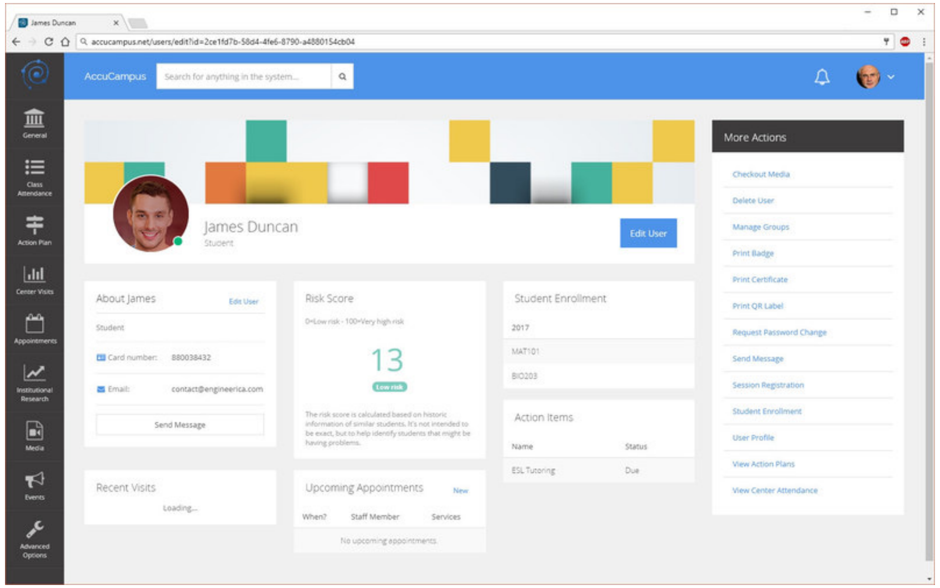
- provide a webpage that explains what models are used by a uni and what data each one uses
- require that students have a right of correction if they see incorrect data in a model about them
- let students opt out (but this can lead to bias in the models as segments of students will not be represented…)

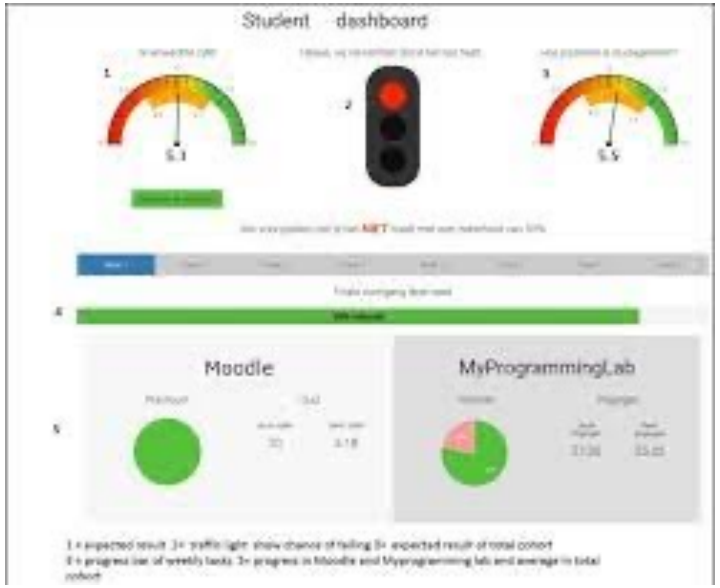


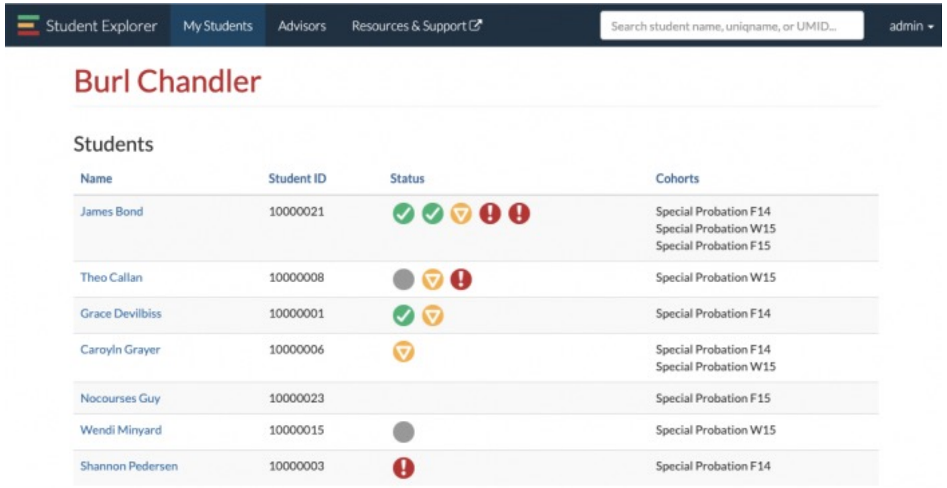
Many different types of intervention are possible!

Policy and practices differ across institutions

There is a lot of research which has worked to unpack best practice over more than 15 years







Questions?

Ethical issues

Simon Buckingham Shum
Professor of Learning Informatics, Connected Intelligence Centre

AI = Automated **Classification**

Critical Data Science and *Science & Technology Studies* remind us that **classification schemes have "politics"**

They are always defined from a perspective, for a purpose

They may silence other perspectives by erasing misfitting data

The pre-workshop stories illustrated contrasting ethics underpinning predictive models

Please reply to this with your thoughts on Story 1

Story 1: Your degree has been going fairly well, getting the grades you hoped for, and exceeding them a couple of times. However, you hit a bad patch, which just gets worse as you head for Christmas. Family stresses. Your best friend has switched to another university. To cap it all, you had to find a new rental place, and the internet is mediocre at best.

It feels like the coursework just got tougher. You know you're not staying on top of it. You're tired and low, and miss a couple of Zoom and face-to-face sessions. You got confused when the last deadline was, and only downloaded the assignment the day before it was due. Your friends Zach and Millie are doing their best to be helpful, but they're cruising through.

With 2 weeks to go before the next deadline, you get a phone call from your tutor asking how it's going, since yo don't seem to have been your usual efficient self, and would you like to have a chat? Maybe there are some thi they can help with? You're impressed! This is really helpful.

But hang on. Are they calling every one of the 800 students, or just you? Is this level of attention a bit weird? check in with Zach and Millie: turns out they didn't get phone calls, but emails encouraging them to keep u good work since they look on course for HDs. But Mimi, one of your other friends who's hit a tough patch, phoned by her tutor.

In order to provide differentiated communications like this, UTS has been tracking students' online acti model recommends which may be in need of extra support. Universities using predictive AI regard it to use every available tool to support students.

Do you agree?

Do you want to know more about how this is done?

Do you trust that UTS is doing this responsibly?

Please reply to this with your thoughts on Story 2a and 2b

Story 2. Ethical and unethical use of predictive models in two universities.

Ethical (The Open University, UK). "An example of how a teacher successfully used the predictive analytics dashboard is discussed below. This teacher was able to use the dashboard to provide timely support to a female engineering student from a Black and Minority Ethnic (BME) background with no prior higher education experience, and enable her to succeed. Prior to this, the student received 100% on the first assignment (a quiz) and 86% on her second assignment. However, in week 10, the dashboard flagged the student as unlikely to submit the third assignment. Upon further inspection by the teacher, it emerged that the student had not accessed the VLE after submitting the previous assignment three weeks earlier. When the teacher contacted the student, it became apparent that the student's lack of activity on the VLE was due to the birth of her third child. The student had not disclosed her pregnancy as she was unsure whether the university would allow her to carry on with her studies. Not only did the teacher resolve the misunderstanding, but also provided support enabling the student to get back on track.

Subsequent monitoring of the student's performance helped the teacher identify another occasion when the student had limited VLE activity and was likely to fail to submit her next assignment. Again, the teacher was able to prevent the student from giving up by identifying the problem she was facing and providing timely support. The student eventually completed the course with an average score of over 80%."

Source: Learning Analytics in Open and Distance Higher Education: The Case of th

Unethical (Mount St. Mary's University, USA). In this example, it was issued during Orientation: "This year, we are going to start th valuable tool that will help you discover more ab here at The Mount, and it is based that determine motivati like you to

Some critical questions to ask about predictive AI in education:

- Are students aware that predictive analytics are being used?
- Should it be possible to opt out?
- Are the staff who use the analytics suitably trained in how to interpret, and act on, its recommendations?
- How accurate is the model, and are its predictions helpful in time to provide support that makes a difference?
- Are there any risks following from a misclassification?
- What is the ethical responsibility of *not acting* if a predictive model flags a student?
- Does one model work for all students in all courses, or do we need to tune models to reflect the diverse ways that students are taught, and how they use online platforms?
- Could social injustices be *perpetuated* because the training data reflects historic biases that we want to overcome?
- Could social injustices be *reversed*, by closing the gap for minoritized student groups? (cf. Georgia State University https://success.gsu.edu)
- How secure is the data that has classified students, who can view this, and how long is it kept on record?

Questions?

Student Partnership in AI Workshop

Proposed pilot at UTS

Susan Gibson
Head of Data Analytics & AI
Carl Young
Data/Architect, AI Platforms

Supporting student success

UTS plan to develop a machine learning model that will predict which students are likely to leave the university prior to census or don't realise that they are enrolled and committed financially

The Data Analytics and Insights Unit plans to pilot 3 models to support student success

1. **Students likely to leave prior to census – we currently have a large amount of commencing students who leave the university prior to the census date– we want to ensure that they are provided with the appropriate support to be successful**
2. **Students sometimes enrol in courses and are unaware that they have made a financial commitment. Where students have no intention of studying at UTS we want to contact them prior to census allowing them the opportunity to withdraw**
3. **Accurate student load forecasting predictions allow faculties an opportunity to optimise resources for students prior to the beginning of semester**

Questions?

Student Partnership in AI Workshop

Groupwork: AI ethics audit using the 5 principles

UTS AI Ethics Principles template: <*AI System*>

Principle		Alignment	Rating
1. Community benefit	AI should deliver the best outcome for human users, in this case, the UTS community, and provide key insights into decision making. AI must be the most appropriate solution for a service delivery or a policy problem, considered against other analysis and policy tools.	• ***Student views…***	
2. Fairness	Use of AI will include safeguards to manage data bias or data quality risks. The best use of AI will depend on data quality and relevant data as well as careful data management to ensure potential data biases are identified and appropriately managed	• ***Student views…***	
3. Privacy and security	AI will include the highest levels of assurance. The UTS community must have confidence that data is used safely and securely in a manner that is consistent with privacy, data sharing and information access requirements	• ***Student views…***	

UTS AI Ethics Principles template: <*AI System*>

Principle		Alignment	Rating
4. Transparency	Review mechanisms will ensure that the UTS community can challenge and question AI-based outcomes and will have access to an efficient and transparent review mechanism if there are questions about the use of data or AI-informed outcomes	• ***Student views…***	
5. Accountability	While AI is recognised for analysing and looking for patterns in large quantities of data, undertaking high-volume routine process work, or making recommendations based on complex information, AI-based functions and decisions must always be subject to human review and intervention. AI system owners and business owners are responsible for the management of their AI systems	• ***Student views…***	

Example only – you need to decide the rating!

Principle		Alignment	Rating
1. Community benefit	AI should deliver the best outcome for human users, in this case, the UTS community, and provide key insights into decision making. AI must be the most appropriate solution for a service delivery or a policy problem, considered against other analysis and policy tools.	• ***Student views…***	
2. Fairness	Use of AI will include safeguards to manage data bias or data quality risks. The best use of AI will depend on data quality and relevant data as well as careful data management to ensure potential data biases are identified and appropriately managed	• ***Student views…***	
3. Privacy and security	AI will include the highest levels of assurance. The UTS community must have confidence that data is used safely and securely in a manner that is consistent with privacy, data sharing and information access requirements	• ***Student views…***	

Risks template: <*AI System*>

Risk	Cause	Controls to reduce risk	Rating
			Critical
			High
			Medium
			Low

Groupwork: AI ethics audit and risk analysis

Google Docs:

https://bit.ly/uts-pai-grp1

https://bit.ly/uts-pai-grp2

https://bit.ly/uts-pai-grp3

https://bit.ly/uts-pai-grp4

https://bit.ly/uts-pai-grp5

Make notes on your Principle in Google Doc
(Group 1 = Principle 1 Etc…)

Principle		Alignment	Rating
1. Community benefit	AI should deliver the best outcome for human users, in this case, the UTS community, and provide key insights into decision making. AI must be the most appropriate solution for a service delivery or a policy problem, considered against other analysis and policy tools.	• ***Student views…***	
2. Fairness	Use of AI will include safeguards to manage data bias or data quality risks. The best use of AI will depend on data quality and relevant data as well as careful data management to ensure potential data biases are identified and appropriately managed	• ***Student views…***	
3. Privacy and security	AI will include the highest levels of assurance. The UTS community must have confidence that data is used safely and securely in a manner that is consistent with privacy, data sharing and information access requirements	• ***Student views…***	

Complete the risk analysis table based on your notes

Risk	Risk Type	Cause	Controls	Rating
				Critical
				High
				Medium
				Low

Repeat for other Principles if you have time, or after workshop

Principle		Alignment	Rating
1. Community benefit	AI should deliver the best outcome for human users, in this case, the UTS community, and provide key insights into decision making. AI must be the most appropriate solution for a service delivery or a policy problem, considered against other analysis and policy tools.	• ***Student views…***	
2. Fairness	Use of AI will include safeguards to manage data bias or data quality risks. The best use of AI will depend on data quality and relevant data as well as careful data management to ensure potential data biases are identified and appropriately managed	• ***Student views…***	
3. Privacy and security	AI will include the highest levels of assurance. The UTS community must have confidence that data is used safely and securely in a manner that is consistent with privacy, data sharing and information access requirements	• ***Student views…***	

Plenary discussion

Gregory Martin
Associate Dean for Teaching & Learning, Faculty of Social Sciences

Thank you! Next steps…

Simon Buckingham Shum
Professor of Learning Informatics, Connected Intelligence Centre

What happens now?

If you wish, continue posting ideas on Teams + Google Doc

2 June: Workshop report posted on Teams for comment

5 July: SA reps will present to the AI Operations Board

Now! Have some lunch and post on the Stickies wall

I like…

I wish…

I wonder…