CRICOS Provider No: 00300K (NT/VIC) 03286A (NSW) RTO Provider No: 0373 TEQSA Provider ID PRV12069

Data, Analytics, GenAI, Standards

FACULTY OF ARTS & SOCIETY

Associate Professor Jon Mason

jon.mason@cdu.edu.au

July 2024

CHARLES DARWIN UNIVERSITY AUSTRALIA

CHARLES DARWIN UNIVERSITY AUSTRALIA

Charles Darwin University acknowledges all First Nations people across the lands on which we live and work, and we pay our respects to Elders both past and present.

Overview – the Request

Learning and Educational Technologies Research Unit

- Research involving Learning Analytics
- Learning Analytics at my university, region or country
- Generative AI and Learning Analytics
- Standardization of Learning Analytics

… in 15 minutes

Assumptions

- Viewers of this webinar already know a lot about the theory & practice of Learning Analytics

  → *what perspective can I offer?*

- Worldwide, academic communities are confronting the challenges & opportunities provided by GenAI

  → *a paradigm shift in education is underway.*

Springer Open

Questions as data: illuminating the potential of learning analytics through questioning an emergent field

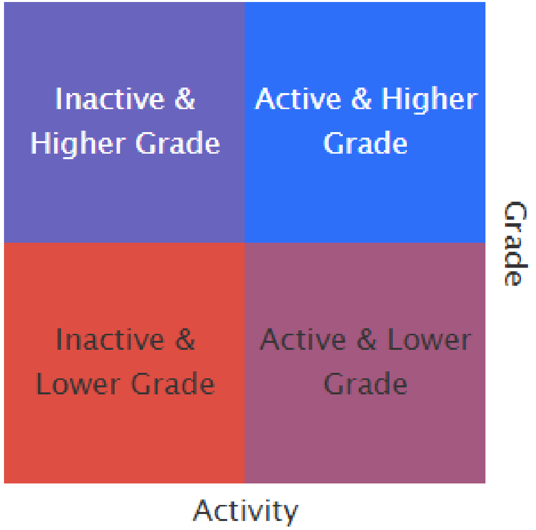
Jon Mason, Weiqin Chen & Tore Hoel

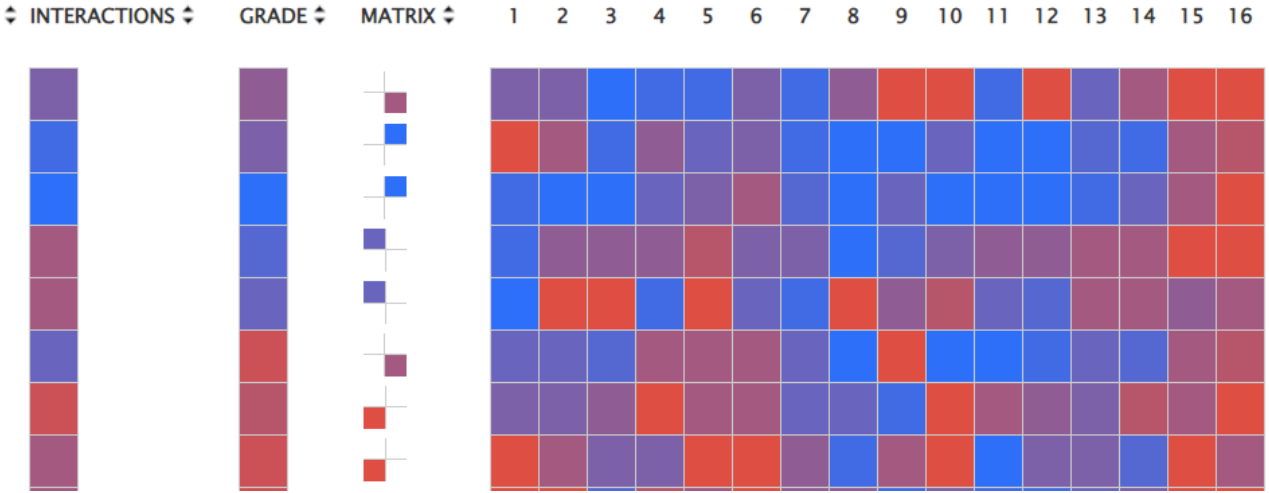
Research and Practice in Technology Enhanced Learning **11**, Article number: 12 (2016)

https://telrp.springeropen.com/articles/10.1186/s41039-016-0037-1

Engagement?

Student Activity – Grade Matrix





Learner Data in LearnLine

Analytics for Learn – student demographics, access, activity, performance and more

Assessment analytics – student assessment & submission, start to finish time, date last accessed

Class Performance – Compares student unit activity related to grades

Discussion Board Analysis – information about student use, engagement, & text analysis

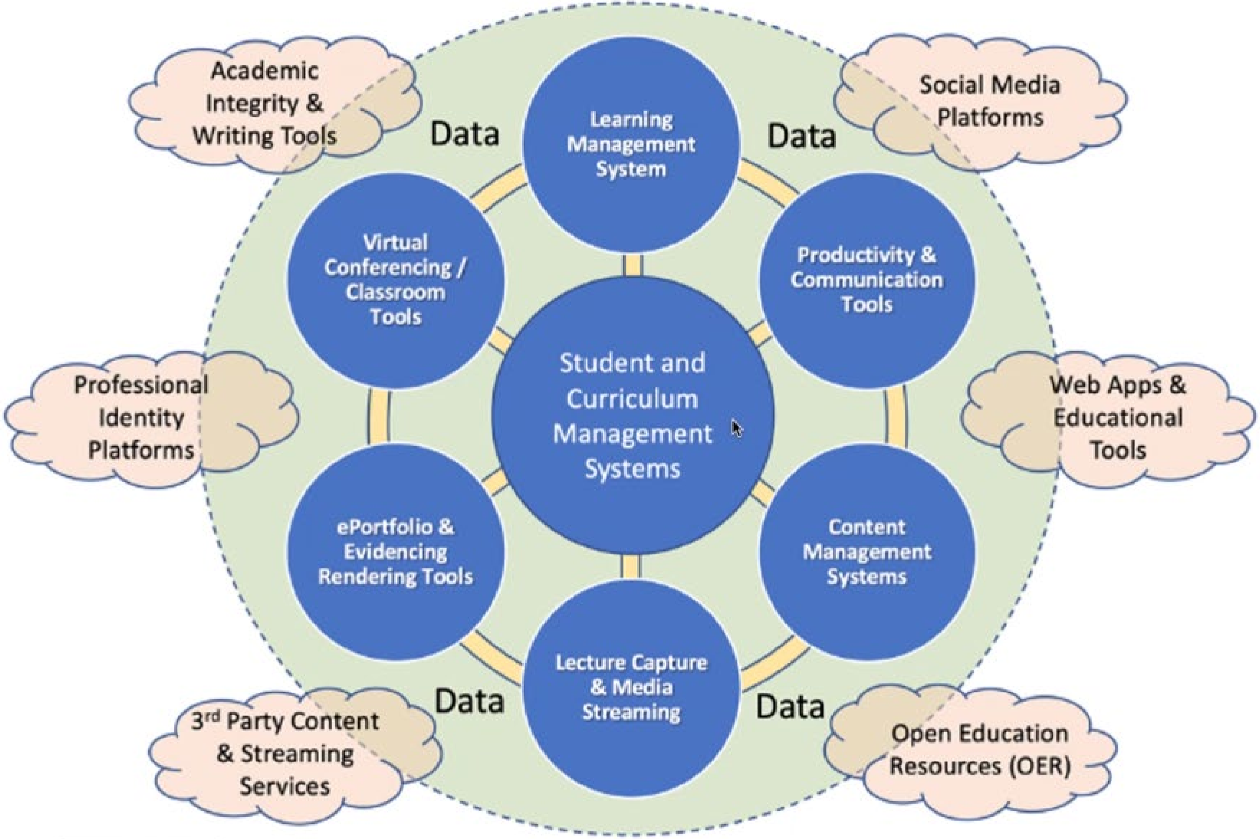
Feedback Fruits – peer feedback about student progress, comments, feedback, and reviews

Kaltura video analytics – video plays, minutes viewed, average completion rate and drop off rate, most viewed, etc

Test Question analysis – statistics on overall performance, assessment quality, individual questions

Turnitin Analytics – insights into how students engaged with assignments, submission timelines, class similarity scores, number of submissions etc

A Contemporary Technology Enhanced Learning (TEL) Ecology



michael_sankey

Overview – my Research Agenda

- Data / Data Literacy
- Analytics
- GenAI
- Standards
- Engagement
- SoTL
- Digital Frontiers
- Sense-Making

Questioning with & within the Digital Environment

Overview – my Research Agenda

- Data / Data Literacy
- Analytics
- GenAI
- Standards
- Engagement
- SoTL
- Digital Frontiers

AIED RECIFE 2024

Pre-Literate, Multi-Literate, Post-Literate: Expanding Horizons for LLMs

Jon Mason[1], Carla C. Eisemberg[1], Cat Kutay[1], Sarah Sutcliffe[1], Joanne Forrest[1], Bev Babbage [1]

https://aied2024.cesar.school/

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS

IEEE

IEEE SA STANDARDS ASSOCIATION

CHARLES DARWIN UNIVERSITY AUSTRALIA

1EDTECH

OMA3

Metaverse STANDARDS FORUM™

AI STANDARDS HUB

Advanced Distributed Learning

OER COMMONS

jisc 日本産業標準調査会 Japanese Industrial Standards Committee

OASIS
Open standards. Open source.

UNESCO
United Nations Educational, Scientific and Cultural Organization

Sustainable Development Goals

BLOCKNET

euos
EU Observatory for ICT Standardisation

GDPR

BIG DATA EUROPE
Empowering Communities with Data Technologies

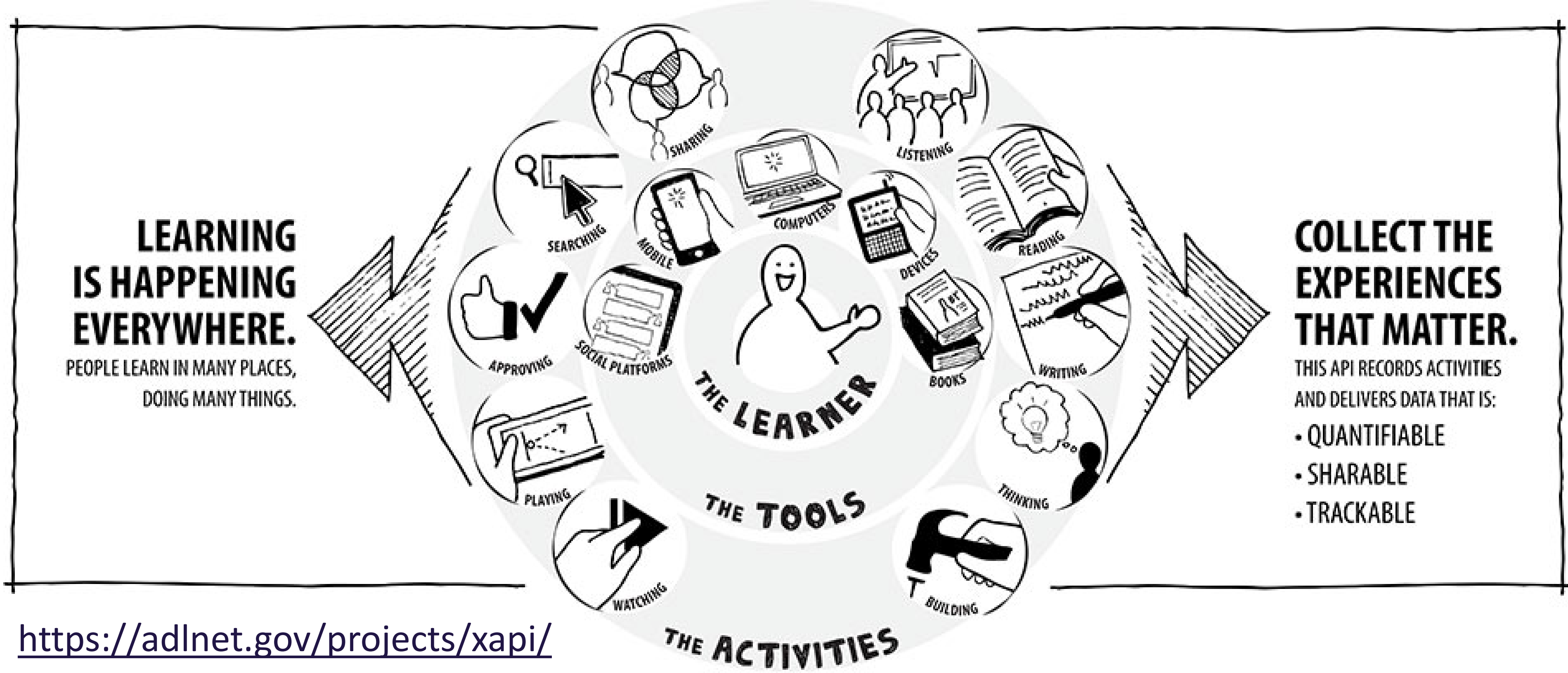
SFIA
Skills Framework for the Information Age

cc creative commons

EUBlockchain
Observatory and Forum

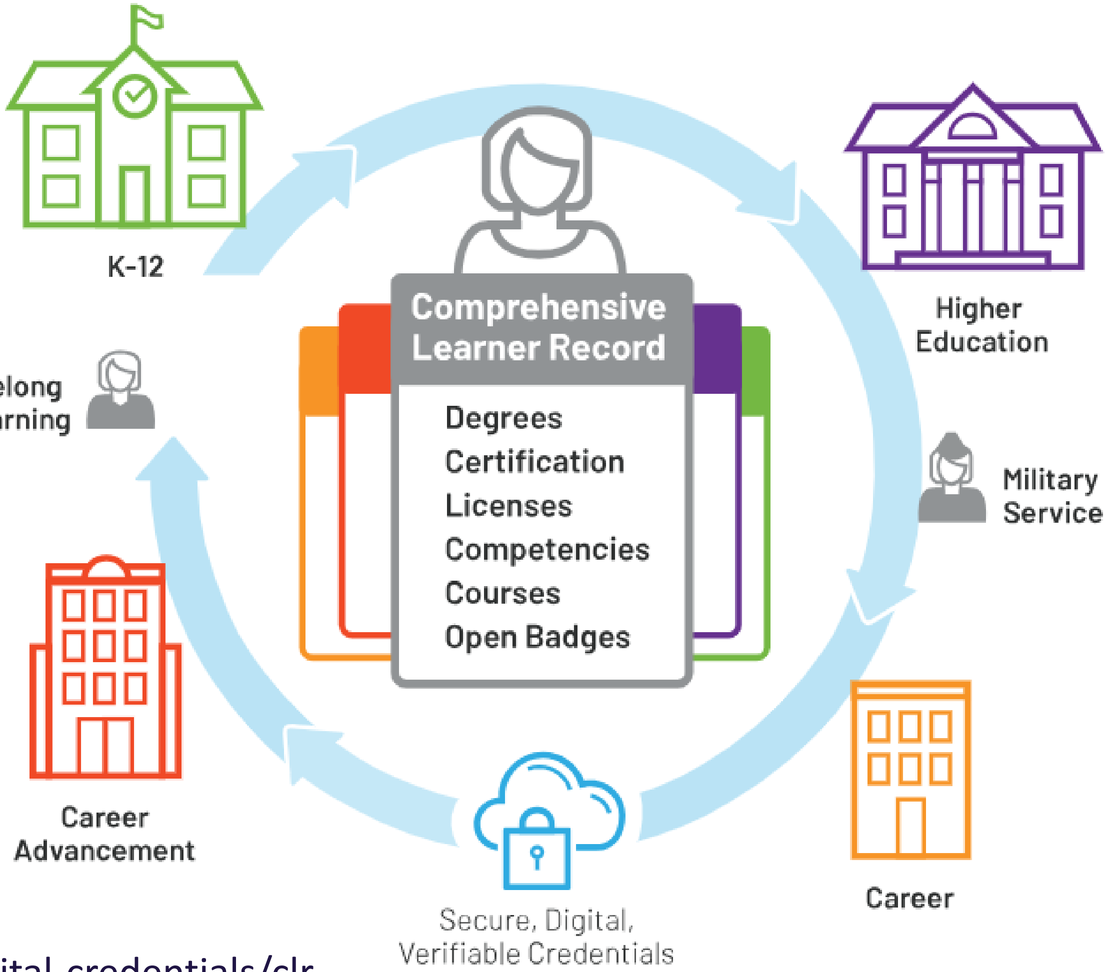
{xAPI} [IEEE approved standard (IEEE 9274.1.1-2023)](https://adlnet.gov/projects/xapi/)

CHARLES DARWIN UNIVERSITY AUSTRALIA

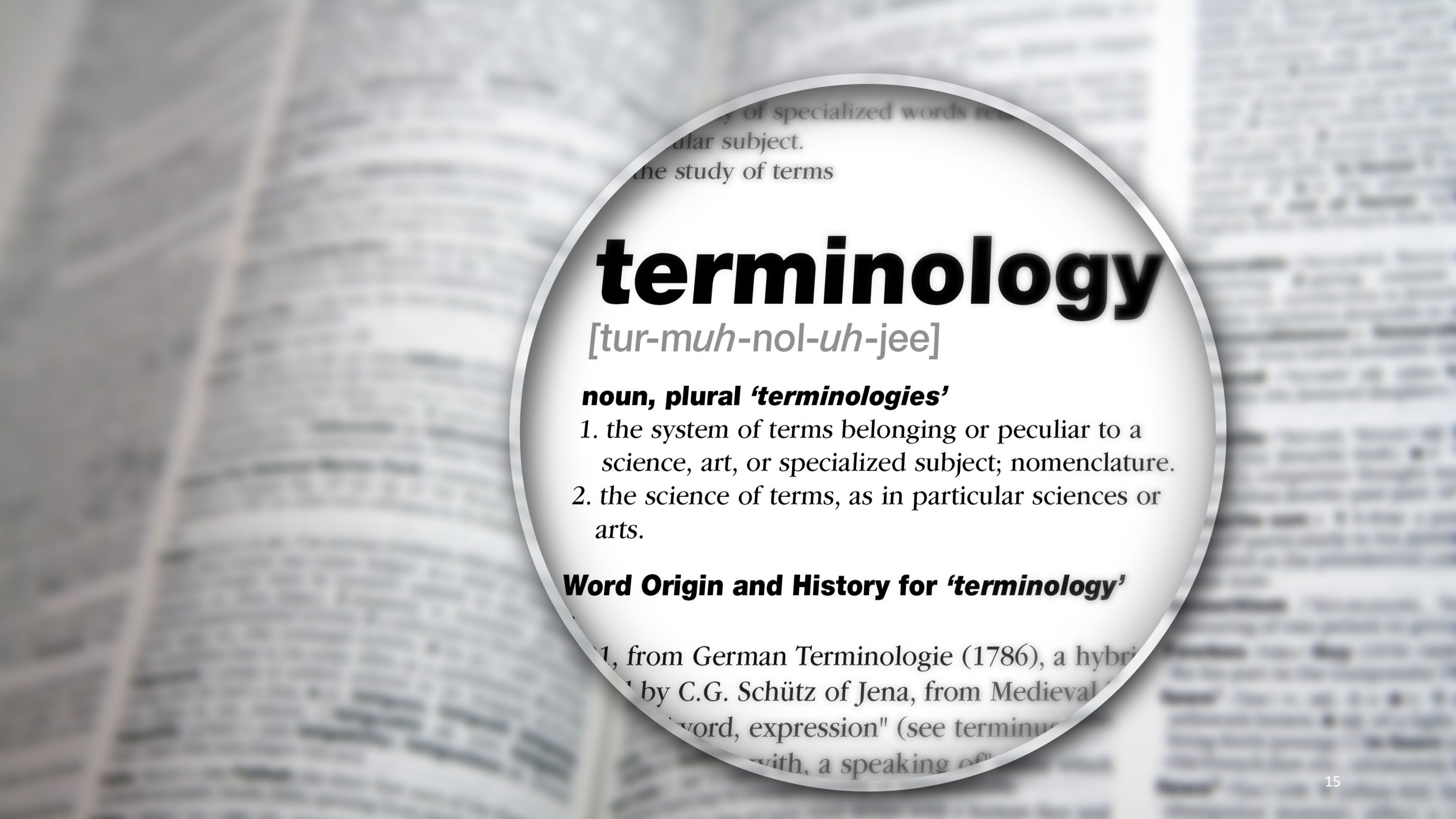


https://adlnet.gov/projects/xapi/

A customizable personal digital record for all learning experiences.



https://www.1edtech.org/initiatives/digital-credentials/clr


of specialized words
subject.
study of terms
terminology
[tur-muh-nol-uh-jee]
noun, plural 'terminologies'
1. the system of terms belonging or peculiar to a science, art, or specialized subject; nomenclature.
2. the science of terms, as in particular sciences or arts.
Word Origin and History for 'terminology'
, from German Terminologie (1786), a
by C.G. Schütz of Jena, from Medieval
expression" (see
with, a speaking

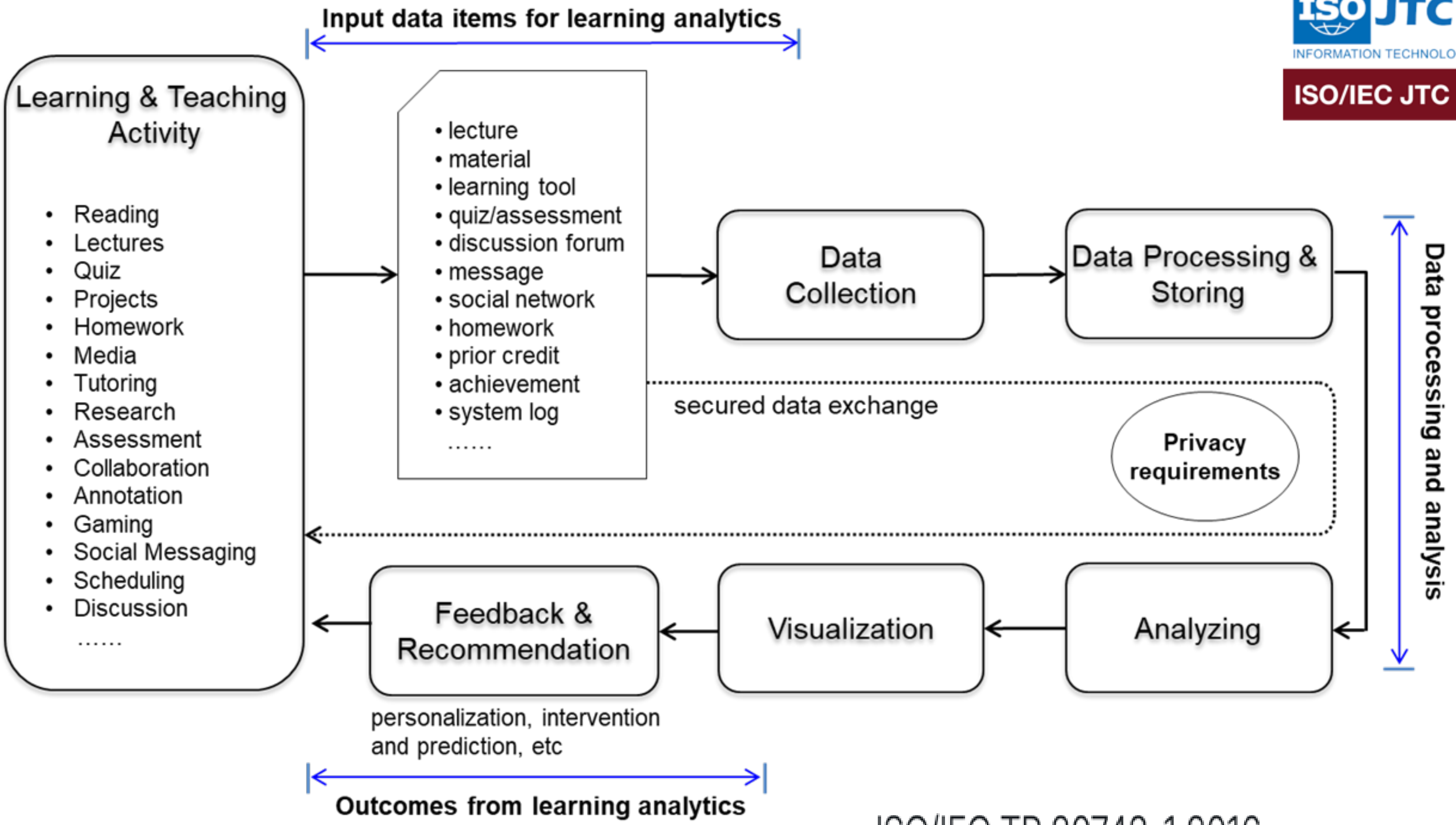
Recent Terminology …

- Artificial Communication / Cognition
- Artificial Emotional Intelligence
- Digital Ecosystem
- Conversational Agents
- Prompt Engineering
- Learning Fabric
- Learning Experience Design
- Cybersecurity Mesh
- GenAI
- Data Observability
- Hybrid Learning Spaces
- Post-Literate
- Neuroadaptive Learning
- Metaverse
- Web 3.0
- Decentralised Identity

International Standards

ISO/IEC JTC 1/SC 36

ISO/IEC TR 20748-1	Information technology for learning, education and training -- Learning analytics interoperability — Part 1: Reference model	Published (2016)
ISO/IEC TR 20748-2	Information technology for learning, education and training -- Learning analytics interoperability — Part 2: System requirements	Published (2017)
ISO/IEC TS 20748-3	Information technology for learning, education and training — Learning analytics interoperability — Part 3: Guidelines for data interoperability	Published (2020)
ISO/IEC TS 20748-4	Information technology for learning, education and training — Learning analytics interoperability — Part 4: Privacy and data protection policies	Published (2019)



ISO/IEC JTC 1/SC 36

ISO/IEC TR 20748-1:2016

Information technology for learning, education and training — Learning analytics interoperability — Part 1: Reference model

6.4 Data collection

Data collection is the process of measuring and gathering information on matters of interest from learning and teaching activities. Tracking data from learners emanate from a wide variety of platforms, e.g., when accessing learning material, using desktop computers and mobile devices, including wearable technologies and the Internet of Things. In this process, the requirements related to the data authority, control of data source, interoperability of data, and efficient flow and exchange of data are addressed.

6.4.1 Accessibility

Requirement ID	Description
R4.1.1	Accessibility requirements ***should*** be registered prior to data collection.

6.4.2 Aggregation/Integration of data

Requirement ID	Description
R4.2	To access heterogeneous learning systems or tools, aggregated profiles for the user ***should*** be supported.

6.4.3 Data interoperability

Requirement ID	Description
R4.3.1	To improve accuracy of collected data, standardized information model and controlled vocabularies ***should*** be applied to data collection API. *EXAMPLE 1: recipe for xAPI specification is an example to define activity stream type and vocabularies.* *EXAMPLE 2: IMS Caliper Metric Profile is an example to define learning activity types and vocabularies.*
R4.3.2	Data information model ***should*** cover a wide range of data types. *EXAMPLE: xAPI is an example to define statement and IMS Caliper is an example for metric profile in terms of information model.*
R4.3.3	Data information model ***should not*** be dependent on dominant products or services. *Note: xAPI and IMS Caliper are good example for open specification about information model. Institution can use both specifications to develop its profile in terms of specific*

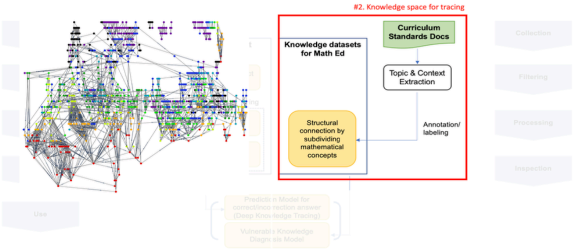
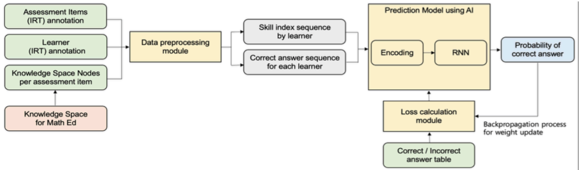
ISO/IEC CD TR 9858

Use cases on advanced learning analytics services using emerging technologies

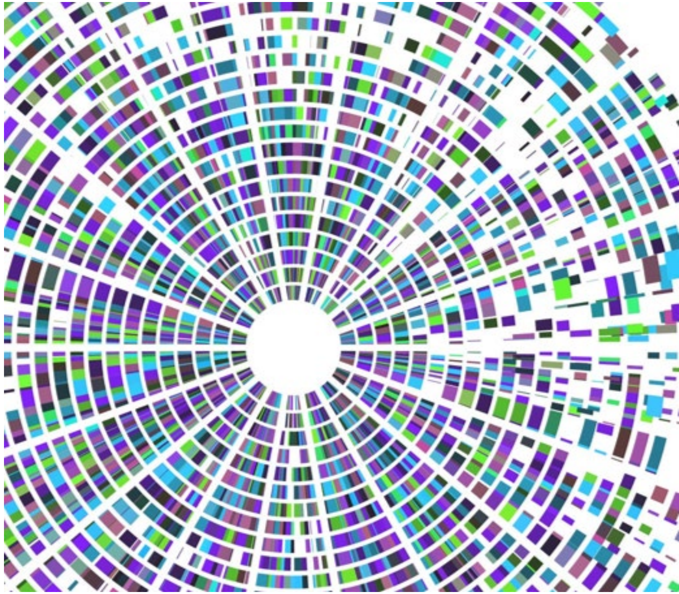
Use-Case Template

ID:	[Unique ID of this use case]
Title:	[Enter the goal of the use case - preferably as a short, active verb phrase]
Description:	[Describe the goal and context of this use case. This is usually an expanded version of what you entered in the "Title" field.]
Primary Actor:	[A person or a software/hardware system that interacts with your system to achieve the goal of this use case.]
Preconditions:	[Describe the state the system is in before the first event in this use case.]
Postconditions:	[Describe the state the system is in after all the events in this use case have taken place.]
Main Success Scenario:	[Describe the flow of events from preconditions to postconditions, when nothing goes wrong. This is the meat of the use case.]
Extensions:	[Describe all the other scenarios for this use case - including exceptions and error cases.]
Frequency of Use:	[How often will this use case be used?]
Status:	[Development status]
Owner:	[Who owns this use case, in your project team?]
Priority:	[Priority of this use case]

jisc 日本産業標準調査会
Japanese Industrial Standards Committee

Use Case ID	UC 2-01
Title	Diagnosis testing for Math based on deep learning (AI) and knowledge map of Math curriculum standard
Contributor (name)	Yong-Sang Cho (Korean expert, i-Scream Edu)
Source (name or url)	i-Scream Home-Learn™ AI scanning service for Math diagnosis testing *Note: url is www.home-learn.co.kr, but this page is just available in Korean. If you need further information regarding this use case, please contact to the contributor (zzosang@i-screamedu.co.kr)* Related articles • Deep Knowledge Tracing (Chris Piech, et al, 2015) • How Technology Can Change Assessment (UNESCO Policy Brief, 2012) Related previous use case(s) • Changing assessment through the learning analytics (ISO/IEC TR 20748-1)
Main stakeholders	• K1 to K-9 Students (who need to take diagnosis testing for Math) • Teachers (who want to know intervention for specific students)
Description (Overview of data collected analyzed, visualization example, etc.)	According to use case 'Changing assessment through the learning analytics' in ISO/IEC TR 20748-1, *"Learning analytics is useful for monitoring how students are going about learning and solving problems. This can be achieved by embedding learning assessments within the learning experience and analyzing process data in log files that capture every click and keystroke. It is known in the gaming industry as "stealth assessment," where tracking performance data is part of the game. This approach can reduce test anxiety because the lines between learning and assessment are blurred.* *It is important to note that embedded assessments do not need to be hidden assessments. In fact, there are examples where providing students with the results of embedded assessments can drive greater learning and engagement. For example, the popular online game World of Warcraft continually assesses player progress and presents feedback to the player in the form of a heads-up display that appears on the game screen (it can be compared to a dashboard). The information is highly motivating and shows where they should focus their attention and learning efforts, so they can do better and open up new levels within the game*". Like games, diagnosis testing, in which the assessment system (connected to item bank) predicts the probability of correct answers to the next item according to the learner's level of understanding, and determines the appropriate item in real time, can be particularly useful in mathematics subjects. Based on the correct/incorrect answers of the sample group of learners who took each assessment conducted in the subject unit of the math curriculum, the (a) comprehension level and (b) true score of each student as well as the (c) difficulty, (d) discrimination and (c) pseudo-guessing of assessment items were expressed as features of datasets. To track the weak knowledge of learners, based on the math curriculum standard, subject matter experts of math education constructed the 'mathematics knowledge relationship among elementary, middle, and high school, by mapping to achievement statements and each level of topics / units.  **Figure 3** Knowledge space tracing part on the AI model for math Ed
Technologies / Technical Approaches and Applications	**Prediction and recommendation technologies** Prediction technology on math education is mainly used to recommend the next assessment item by calculating the probability that a learner will answer the correct answer to unsolved questions. The prediction AI model usually uses an RNN (recurrent neural network) based deep learning model. Recently, many research are underway to improve prediction accuracy by utilizing as known as the transformer model. This use case shows the accuracy is little bit over 85%, and more than 90% accuracy is recommended for reliable service. In detail, four types of datasets – (a) assessment item annotation (IRT values regarding assessment item), (b) learner level annotation (IRT values regarding assessment), (c) node on the knowledge space for each assessment item, (d) correct and incorrect answer table of learners) - are used for AI training. To improve accuracy for the prediction probability through the data processing module both skill index sequence by learner and correct answer sequence for each learner are generated from the input datasets (a) and (b). The training process of the AI predictive model consists of three steps. First, predict the probability of correct answers to assessment items the learner has not yet solved through the AI prediction model for correct answers for each item. Then, the loss is calculated by comparing the predicted value with the actual result of the correct and incorrect answer table, and then the AI training process is performed in the direction of minimizing the loss function.  Figure 4 AI training process of prediction function regarding correct answer for math In this use case, the process of tracing the learner's weak knowledge consists of an iterative process. If an assessment item with a low probability of correct answer via the AI prediction model is actually found as an true result, the AI diagnostic model

(some) **Questions Arising**



- The data might tell a story – but is storytelling a science?
- What data might be missing?
- Can we trust the data?
- How should consent be managed?
- Who ‘owns’ the data?
- Are educational institutions acting ethically in their collection & use of student data?
- Who determines the boundaries of personal & institutional data?
- What are the ethical considerations?
- What kind of data governance model best suits learning analytics?
- How can we balance compliance within a local jurisdiction & an international context?
- Self-directed learning involves a blend of formal & informal learning – how can we meaningfully capture data from both domains?

THE BASIC STEPS OF A PIA

Source: https://privacy.org.nz/publications/guidance-resources/privacy-impact-assessment-toolkit/

- Physics
- Chemistry

- Natural sciences
- Social sciences
- Life sciences

- Applied Sciences
  - Technology
  - Engineering
  - Industrial

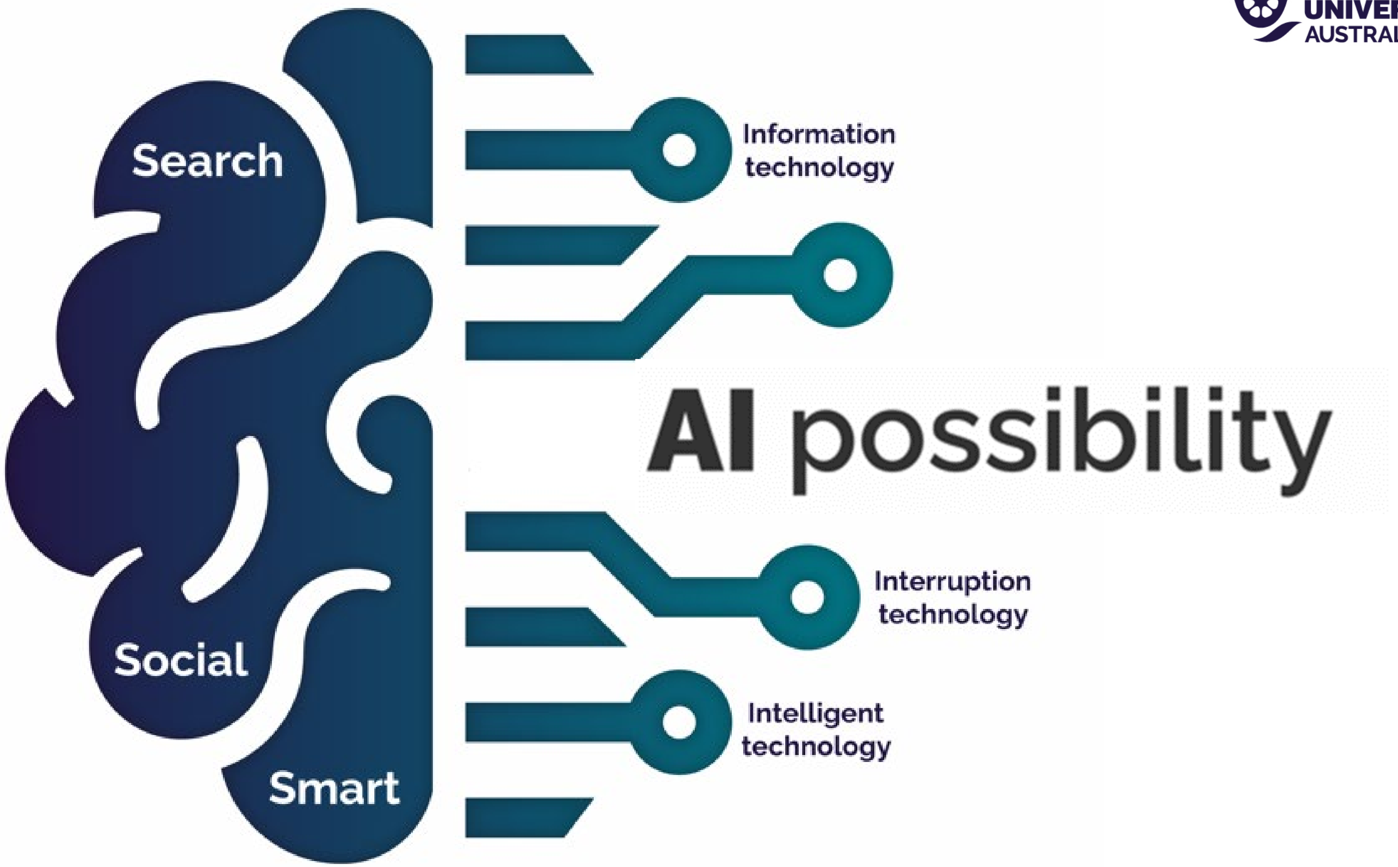
What are the disciplinary trajectories of Data Science & Artificial Intelligence?

- Information & Computing Sciences
- Engineering
- Social Sciences
- …

Interdisciplinarity

Data Governance

- The origins and destinations of data flows are complex
- Humans are also agents (‘tools’) of data production – data that is monetized by Big Tech
- ‘Ownership’ of personal data is contestable & this is why privacy laws like GDPR exist
- A data governance framework requires specification of rules, roles, & responsibilities … and policies to ensure privacy, security, & data quality are managed safely, ethically & responsibly.



https://thesociologicalreview.org/magazine/june-2023/artificial-intelligence/the-search-continues/

Wrapping up …

- Digital ‘ecosystems’ increasingly complex
- The digital environment is more than a collection of tools
- Data does not ‘speak for itself’
  - instruments that produce it require scrutiny
- Text production is now trivial
  - & so is data production – raising validity concerns around existing LA systems
- A paradigm shift is underway in Education
  - Dialogic Web – moving beyond ‘search’
  - Post-literate
  - Human-Computer Collaboration
  - Co-production of knowledge
  - New conceptual frameworks needed

jon.mason@cdu.edu.au