香港浸會大學
HONG KONG BAPTIST UNIVERSITY

Generative AI in Educational Assessment: Challenges and Ethical Issues

Dr Rachel Siow Robertson

Department of Religion and Philosophy
Academy of Chinese, History, Religion and Philosophy

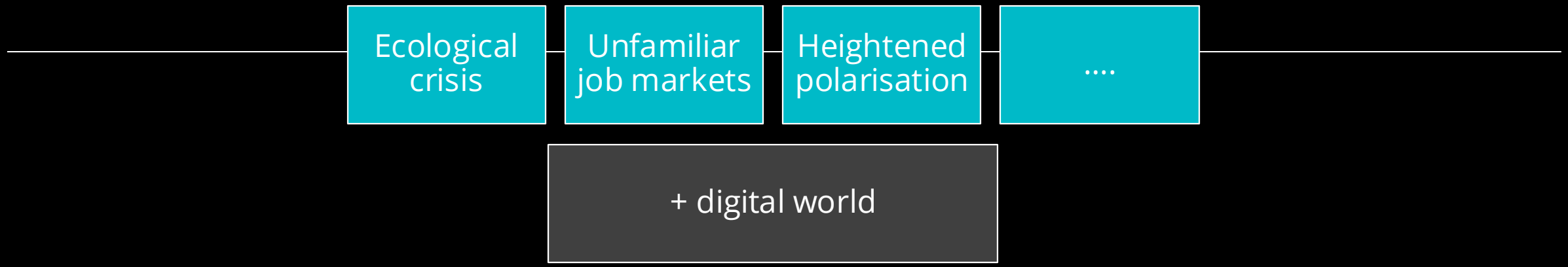
Hong Kong Baptist University

8th August 2025

Plan

- Theoretical framework: philosophy – ethics – virtue/character – integrity
- The usual catalogue of challenges and opportunities for AI in education, arranged according to impact on integrity
- The usual solutions (AI rejection, AI literacy, AI collaboration, AI design)
- The situation is worse than we think: The idea of tragic dilemmas and why this helps
- Suggestions: Changing structures, developing virtues, examples

My approach



using conceptual tools…

→ Philosophy – ethics – **virtue/character**

Qualities, dispositions, and behaviours allowing for informed choices to do the right thing at the right time, supporting individual and societal wellbeing.

Critical, moral, and creative skills e.g. honesty, self-control, humility…

and the meta-virtue (unifying trait) of **integrity**.

to explore how to prepare young people for the digital world

→ **Tensions** of generative AI:

new opportunities for learning and development, but also risks of de-skilling, information disorder, malicious actors, data extraction, and environmental impact.

→ **Context**: education & in an English language instruction context in Hong Kong.

Integrity as a condition of wellbeing

Experience of wellbeing = joy

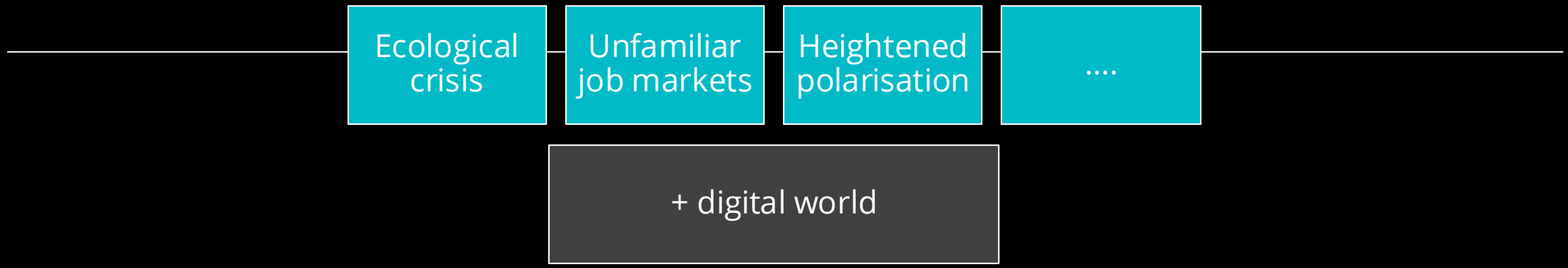
a deep and lasting positive emotion resulting from recognition and experience of integrity –

integration within oneself (Johnson 2020a), and between how the world is and how one hopes it should be (Johnson 2020b, Van Cappellen, 2020).

= experience of wellbeing

Johnson, M. K. (2020a). Joy: A review of the literature and suggestions for future directions. *Journal of Positive Psychology*, *15*(1), 5-24.
Johnson, M. K. (2020b). Joy: A reply to the replies. *Journal of Positive Psychology*, *15*(1), 84-88.
Van Cappellen, P. (2020). The emotion of joy: Commentary on Johnson. *Journal of Positive Psychology*. *15*(1), 40-43.

My approach



using conceptual tools…

to explore how to prepare young people for the digital world

How can virtue/character education, with particular attention to integrity, help young people to build for joy in the digital world?

The usual catalogue of challenges and opportunities for AI in education, arranged according to impact on integrity

Technology in general:

Robertson, R. & Johnson, M. K.(2023). Moral education in and for virtual spaces. In D. W. Yacek, M. E. Jonas & K. H. Gary (Eds.), *Moral education in the 21st century* (pp. 231-259). Cambridge University Press. https://doi.org/10.1017/9781009170291.016

Integrity	Benefits	Risks
Epistemic integrity: receptivity to the world	Using tools for summarising, pattern-finding, translation, search, and transcription Generating scripts for training (e.g. empathy) Teachers preparing prompts based on rubrics for students to use on their drafts Turning text of revision notes into other formats e.g. quiz, mindmap, podcast	Information disorder – "hallucinations", bias, deepfakes, individualisation/filter bubble Black box nature of technology Intellectual de-skilling Technological halo effect Information overload
Self-efficacy: Acting to change the world in accordance with desires and commitments	Text-to-speech or gaze-to-text assistive technologies Summarising to meet a word limit AI editing as a safety net for anxious students – messy with ideas first	Creative and moral de-skilling Surrender of agency to outputs of AI Capture and commodification of attention Too easy
Self-unity: Internal consistency of commitments and identities	Generating scripts for meditation, reflection prompts	Being treated as aggregated data – standardisation, emotional manipulation, value capture Overreliance Loss of individual voice/style/dialect

A helpful catalogue: https://markpegrum.com/tools-for-digital-learning/generative-ai/

Integrity	Benefits	Best case	Risks	Worst case
Epistemic integrity: receptivity to the world	Using tools for summarising, pattern-finding, translation, search, and transcription Generating scripts for training (e.g. empathy) Teachers preparing prompts based on rubrics for students to use on their drafts Turning text of revision notes into other formats e.g. quiz, mindmap, podcast	Students using AI to revitalise their own endangered languages	Information disorder – "hallucinations", bias, deepfakes, individualisation/filter bubble Black box nature of technology Intellectual de-skilling Technological halo effect Information overload	Deepfakes for exploitation/ extortion Feeding extractive approach to environment, labour, data
Self-efficacy: Acting to change the world in accordance with desires and commitments	Text-to-speech or gaze-to-text assistive technologies Summarising to meet a word limit AI editing as a safety net for anxious students – messy with ideas first		Creative and moral de-skilling Surrender of agency to outputs of AI Capture and commodification of attention Too easy	
Self-unity: Internal consistency of commitments and identities	Generating scripts for meditation, reflection prompts		Being treated as aggregated data – standardisation, emotional manipulation, value capture Overreliance Loss of individual voice/style/dialect	

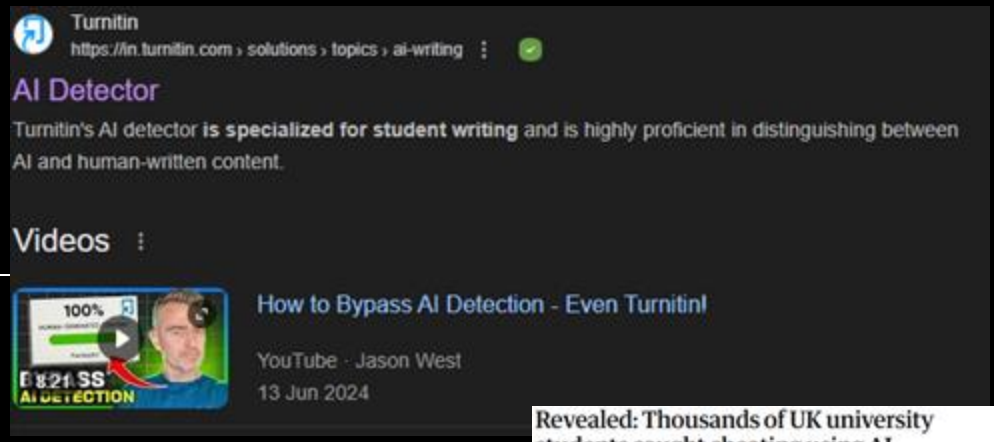
The usual solutions: **catching out** students

AI rejection

- Total bans
- AI detection tools
  - unlike plagiarism detection tools, problems of proof and "humanisers"
  - "but students won't know that we don't know"…?
- Alternative assessments: closed-book and handwriting (vs. oral examinations, creative outputs)
- Punishments for "cheating": oral examinations, written warnings, reduced or failing grades

Responses:
"only 3% of employers believe higher education is adequately preparing graduates for an AI-driven future" (Digital Education Council AI in the Workplace 2025 report)

we should focus on "bridging the AI access gap" – comments by UN Secretary General Antonio Guterres, International Association of University Presidents (IAUP) conference, 2024



The usual solutions: **keeping up** with the trajectory of AI use and development

AI Literacy

- The educator equips students with skills, habits, and practices e.g. critical thinking, ethical reasoning, and creativity

Critical thinking course

AI Collaboration

- The educator designs learning experiences using AI which aim for partnership with AI

Critiquing and peer review AI drafts, AI-FIXIT

AI Design

- The educator focuses on inclusive design – universal design for learning(UDL) or design justice to try to undo harms

Language revitalisation projects

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006

Zhai, C., Wibowo, S. & Li, L.D. (2024) The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learn. Environ.* **11**, 28. https://doi.org/10.1186/s40561-024-00316-7

Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, 126, 103502

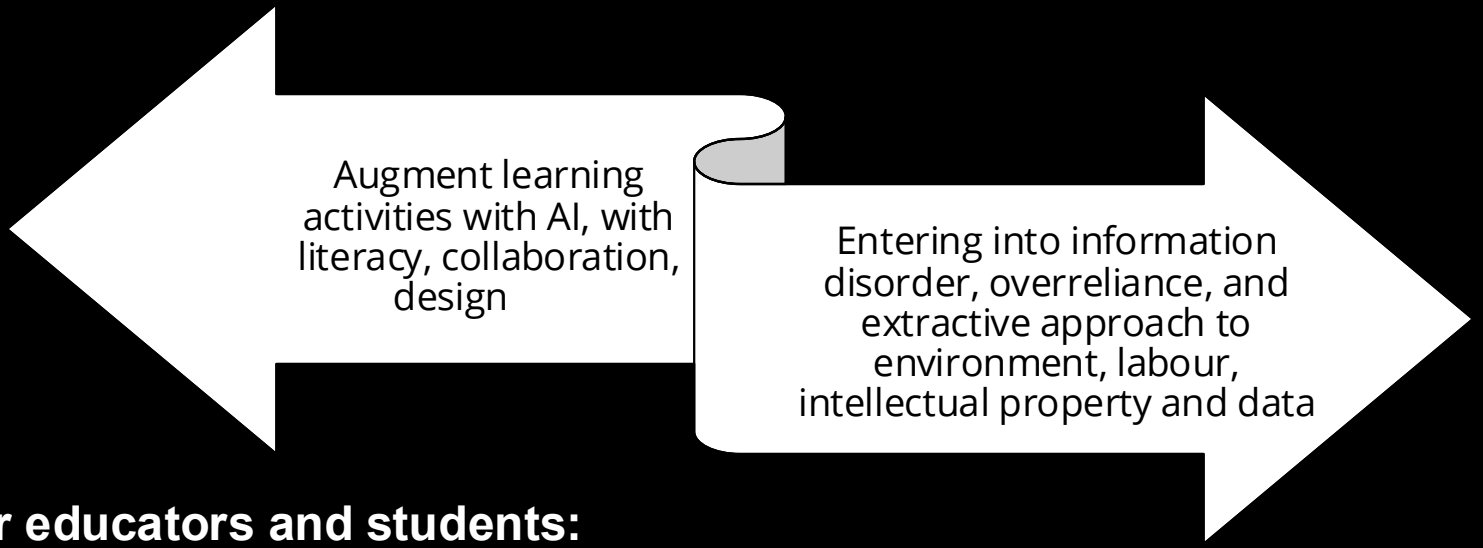
Yan, D. (2024). Collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning.

Tang, K. S., Cooper, G., & Nielsen, W. (2024). Philosophical, legal, ethical, and practical considerations in the emerging use of generative AI in academic journals: Guidelines for research in science education (RISE). *Research in Science Education*, *54*(5), 797-807.

Evmenova, A. S., Borup, J., & Shin, J. K. (2024). Harnessing the power of generative AI to support ALL learners. *TechTrends*, *68*(4), 820-831.

Miao, F. & Holmes, W. (2023) *Guidance for generative AI in education and research*. UNESCO.

The situation is worse than expected



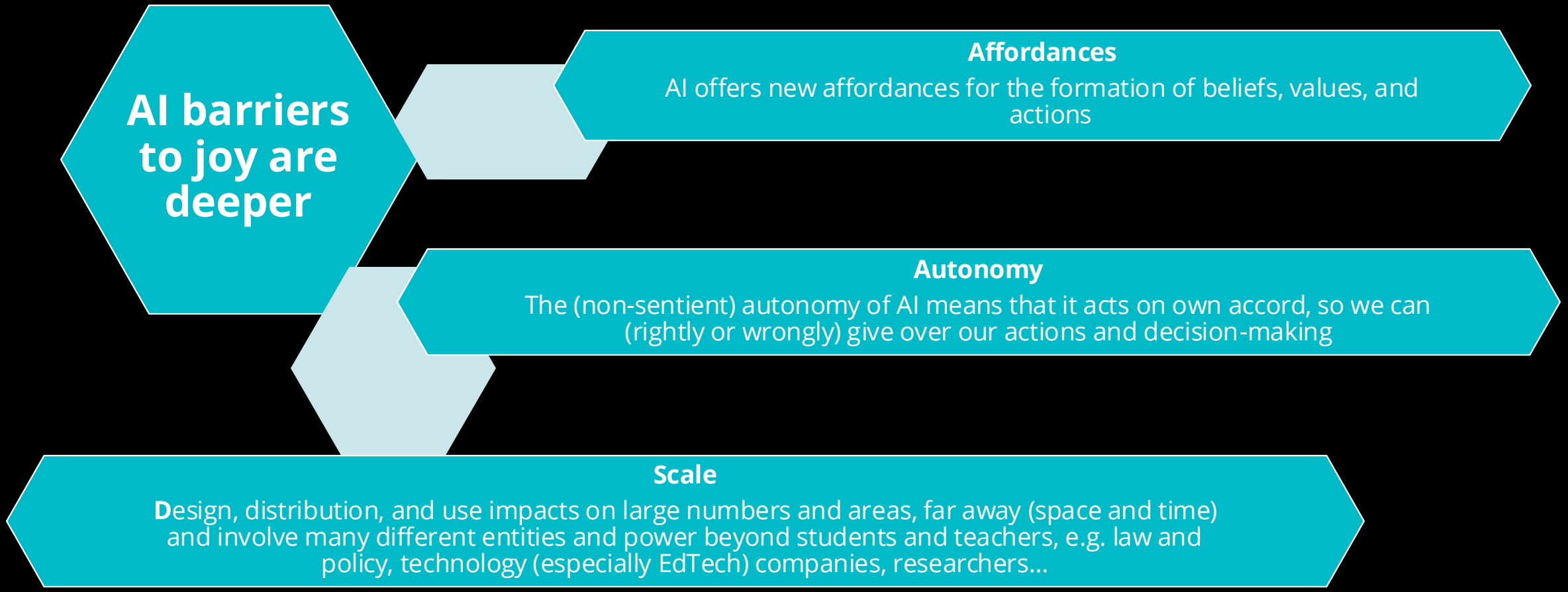
Tragic dilemma for educators and students:

- a case in which there may be an all-things-considered best thing to do, but doing the right thing is impossible.
- accompanied by experience of tension: a pull in contrary directions, a lack of wholeheartedness, and a “moral remainder” from choosing an option that was not the right (in an unqualified sense) thing to do.
- disconnect between virtue and joy

(Tessman, 2017)

Educators and students are facing tragic dilemmas

The situation is worse than expected



Taking stock

Supporting wellbeing and joy = supporting character and integrity.

Although there are opportunities, there are also significant risks to integrity which are amplified by the unique nature of AI.

We are in the position of a “tragic dilemma” – no right thing to recommend.

→What can we as educators do?

Acknowledging the difficulties means we can also work towards solutions

Suggestions: changing structures

Good use of AI is not just about individual students, or even educators (questioning the “catching out” approach, and the “keeping up” emphasis on the skills and projects of students)

They need a supportive ecosystem.

Who is responsible for the structures?

Policy-makers, technology (especially EdTech) companies, education researchers… need to aim for equal access to safe, non-extractive AI models

- Practical steps: financial support, tech support, attending to processes, spaces, and times which work against current approaches to AI

AND educators in their institutions and classrooms

Suggestions: virtues which challenge constraints

“Burdened virtues”

an approach to virtue theory in situations of structural oppression (Tessman 2015)

“traits that make a contribution to human flourishing. . . only because they enable survival of or resistance to oppression . . ., while in other ways they detract from their bearer’s well-being” (Tessman 2015, 95)

the bearer of burdened virtues lacks the external conditions needed for flourishing (e.g. the absence of pain), but takes on the burdened virtues through no fault of their own and for the purpose of survival or resistance to the oppressive structures.

E.g. ignorance, refusal

Some examples for AI:

dishonesty through lying about personal details

ignorance by withholding data for training

How can we as educators allow for burdened virtues?

Suggestions: virtues which challenge constraints

Virtues of solidarity/co-liberation (D'Ignazio & Klein, 2020)

- Focusing on "anticipatory futures", imagining "how we would like to be living" and making changes to bring those desired futures about (Selwyn, 2021). Neil Selwyn identifies educational institutions as communities capable of bringing about present behavioural change through engaging in "hopeful re-imagining" of technology use, highlighting values such as resilience, humility, and inclusivity (Selwyn, 2021).
- Taking relational approaches which involve being in the right kind of listening and caring relationships and allowing for constant 'mutual adjustment', rather than impersonal approaches handing down moral judgments on what the 'best thing' to do is (Walker, 1989).
- Building in checks e.g. The AI Resource Test (Mulaney, 2024)

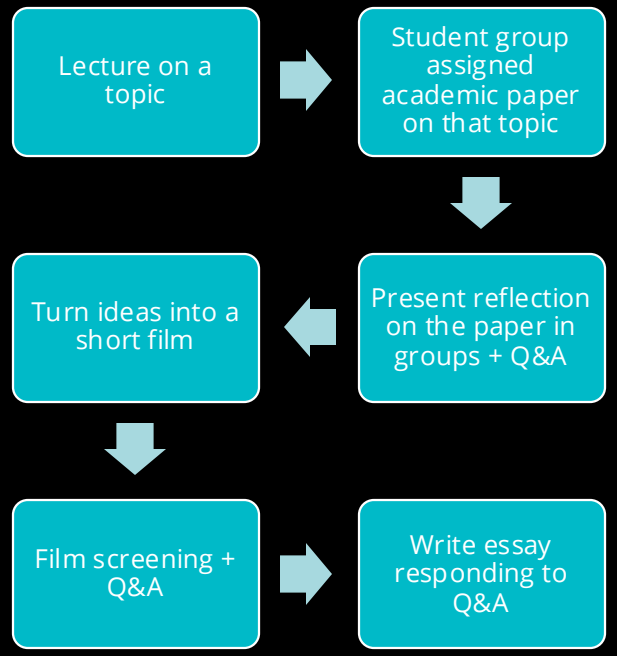
"Are the prompts and generated text worth a bottle of water? Is the generated image worth enough energy to charge a smartphone?"

Suggestions: virtues which challenge constraints

Educators have power to structure the classroom and lead with compassion rather than “catching out” (suspicion of students) or “keeping up” (fear of missing out, resignation to the trajectory of technologies)

- assessing based on process, more than what AI can extract
- giving time and space to experiment
- helping students to lead, address problems and imagine futures which matter to them

Example 1: Reading academic papers for a purpose



- Experience and embodied storytelling
- Focus on the process and motivation
- Giving time in class to write and edit together

Virtues:

Assesses *leadership* and *autonomy*, *motivation*, *creativity* and *resourcefulness*

Example 1: Reading academic papers for a purpose

Rubrics – Matters of Life and Death

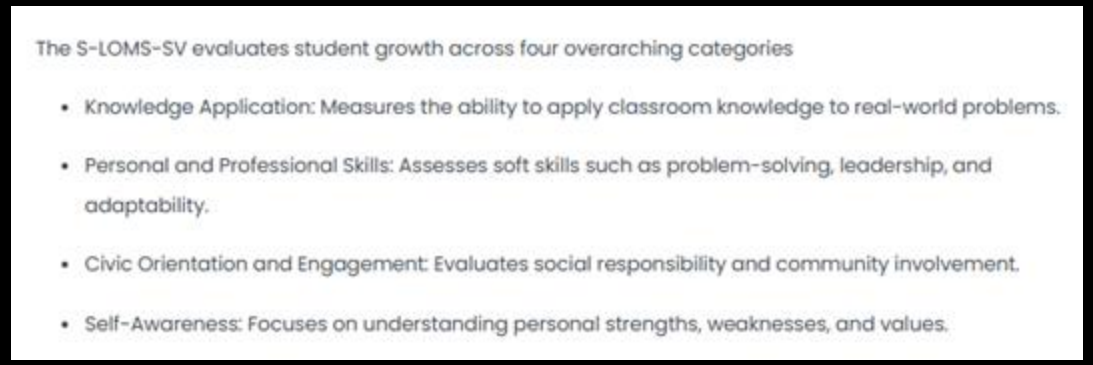
	3 Outstanding	2 Satisfactory	1 Needs Improvement
Autonomy	Demonstrates interests and knowledge about belief and practice regarding life and death, applying knowledge to what is happening outside of the classroom. Goes beyond classroom requirements. Takes the lead in thinking or actions.	Demonstrates some interests and applications of knowledge to issues outside of the classroom, or takes the lead in thinking or actions.	No evidence of interests and applications of knowledge to issues outside of the classroom, or of taking the lead in thinking or actions.
Reflection	Reasons about different and conflicting perspectives. Addresses societal impacts and explains what/who will be affected. Potential limitations of different ideas and theories are clearly described as well as solutions to overcome them.	Reasons about different and conflicting perspectives. Indicates some societal impacts and moderately explains what/who will be affected by the proposed solution. Potential limitations of different ideas and theories are addressed as well as some strategies to overcome them.	Does not engage with different or conflicting perspectives. Does not describe societal impacts, or what/who will be affected by the proposed solution. Potential limitations of different ideas and theories are not or hardly addressed.
Creativity	Applies knowledge in new and unfamiliar ways and explores new and/or creative ways to answer questions about belief and practice regarding life and death. Is aware that such questions often do not have a straightforward right or wrong answer.	Tries to apply disciplinary knowledge in new and unfamiliar ways but resorts to familiar ground. Starts to explore new and/or creative ways to solve a problem but falls back on known patterns and working methods. Comes up with multiple ideas but finds it hard to determine which ideas will be useful in the end.	Isn't able to apply learned knowledge. Discards ideas too soon or focusses on one simple idea from the start without thinking of other possibilities.

Adapted from: Valid Assessment of Learning in Undergraduate Education (VALUE) system, The American Association of Colleges and Universities

Assessment of Interdisciplinary Competencies (AIC) developed by Universiteit Utrecht

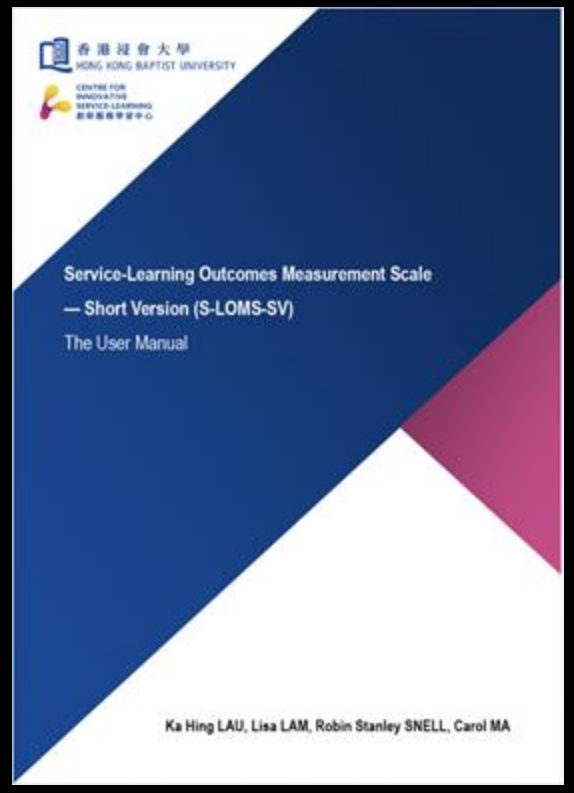
Example 2: Digital ethics as service-learning

- 8 theoretical sessions and 1 preparation session
  - using AI to prepare and translate material, and preparing prompts to demonstrate benefits and limitations of AI, although some by this point had pledged to refrain from using AI for low-grade tasks!
- 2 sessions teaching secondary school students (experience sharing, carnival games)
- Assesses: *reflection*, *civic virtue*, *ethical reasoning*, *critical thinking*



https://cisl.hkbu.edu.hk/service-learning/Evaluation.html

Example 2: Digital ethics as service-learning



SILOS	Criteria	*Outstanding* 75 % and above A / 70 - 74 A- Lots of analysis and explanation, consideration of different viewpoints before stating a personal argument, drawing on multiple references. Fulfilling all relevant ILOs well.	*Good* 67 – 69 B+ / 63 – 66 B / 60 – 62 B- Drawing on fewer references than in the A bracket. Fulfilling majority of relevant ILOs well.	*Satisfactory* 57 – 59 C+/ 53 – 56 C / 50 – 52 C- States an argument with a single reference, draws on references without personal argument. Fulfilling some relevant ILOs well.	*Needs improvement* 46 – 49 D States an opinion, only partially relates to the issue, no references, summarizes others' responses. Not fulfilling ILOs well.	F
i.	**Applying knowledge & skills to service**	*In designing & delivering service:* Innovatively integrated academic learning to render an effective or impactful service. *In reflective activities:* Made original and concrete connections between course learning & service, and discipline knowledge & social issues.	Made good use of academic learning to render useful service. Made interesting connections between course learning & service, and discipline knowledge & social issues.	Applied basic academic learning to service. Made some connections between course learning & service, and/or discipline knowledge & social issues.	Did not apply academic learning to service. Did not make connections between course learning & service, and/or discipline knowledge & social issues.	non-submission, under-attendance, plagiarism
ii.	**Empathy & self-understanding**	*During service & when communicating/interacting with the community:* Consistently acted with empathy & active engagement Responded appropriately & adjusted readily to community members *In reflective activities:* critically evaluated understanding of self, identifying *specific areas* for personal growth.	Most of the time acted with empathy or active engagement & responded adequately to community members. Critically examined self & identified areas for personal growth.	Showed some care and responded to requests by community members. Had self-realisations, with simple analysis and/or personal implications.	Generally disengaged & inattentive to community members. Merely described feelings and events, without analysis nor implications.	
iii.	**Social awareness & commitment to the community**	*In reflective activities:* Expressed sophisticated understanding of social issues based on service experience and learning with/from others. Strongly acknowledged social responsibility, with a realistic & concrete personal action plan.	Gained better understanding of social issues through service experience & interaction with others. Acknowledged social responsibility, with ideas on how to put it into practice.	Expressed general views about social issues related to the service experience. Acknowledged social responsibility, with general plan.	Expressed simplistic or opinionated views, with limited reference to service experience. Detached or clichéd treatment of social issues.	
iv.	**New ideas about society & contributions to social devt**	*Building on SL experience (difficulties & challenges, successes & failures):* Reached original perspectives about society & able to formulate innovative ideas for creating or improving solutions for social good.	Expressed interesting perspectives about society & expressed feasible ideas about solutions for social good.	Expressed general ideas about society & possible solutions for social good.	Did not articulate ideas about society and/or possible solutions; or, expressed ideas which were unrealistic or irrelevant to service experience.	
v.	**Professional skills & generic competencies**	*In working with others to plan & implement service:* Effectively exercised relevant professional skills & competencies in a successful & collaborative service. *In reflective activities:* Clearly articulated how SL contributed to professional development and will inform one's future as a professional.	Effectively exercised one or two professional skills or competencies in a successful & collaborative service. Gave concrete examples of how SL contributed to professional development.	Exercised one or two professional skills or competencies in service. Described a general or surface-level impact of SL on professional development.	Did not exercise relevant professional skills or competencies in service. Did not articulate the impact of SL on professional development.	

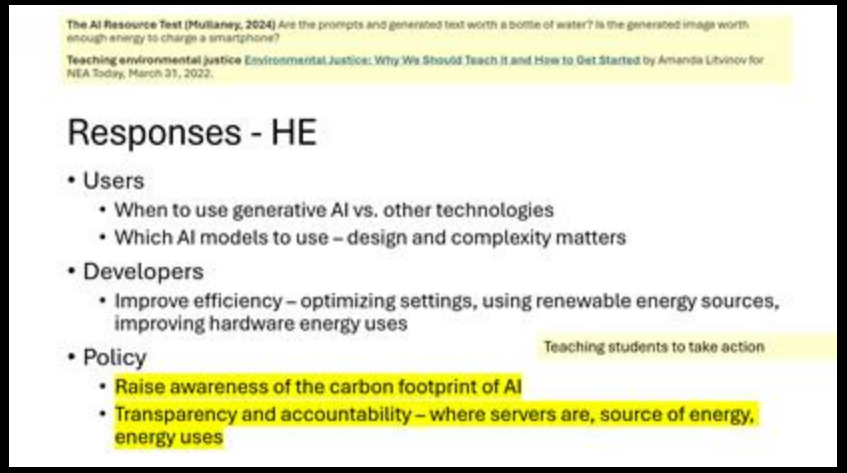
Example 2: Digital ethics as service-learning

		4 Outstanding	3 Good	2 Satisfactory	1 Needs Improvement
Critical thinking	**Explanation of Issues**	Issue to be considered critically is stated clearly and described comprehensively, delivering all relevant information necessary for full understanding.	Issue to be considered critically is stated, described, and clarified so that understanding is not seriously impeded by omissions.	Issue to be considered critically is stated but description leaves terms undefined, ambiguities unexplored, boundaries undetermined, and/or backgrounds unknown.	Issue to be considered critically is stated without clarification or description.
	Student's Position	Student's position is imaginative, taking into account the complexities of an issue. Limits of position are acknowledged. Others' points of view are synthesized within position	Student takes into account the complexities of an issue. Others' points of view are acknowledged within position	Student acknowledges different sides of an issue.	Student's position is stated but is simplistic and obvious.
	Conclusions and Related Outcomes (implications and consequences)	Conclusions and related outcomes are logical and reflect student's informed evaluation and ability to place evidence and perspectives discussed in priority order.	Conclusion is logically tied to a range of information, including opposing viewpoints; related outcomes are identified clearly.	Conclusion is logically tied to information (because information is chosen to fit desired conclusion); some related outcomes identified clearly.	Conclusion is inconsistently tied to some of the information discussed; related outcomes
Ethical reasoning and judgement	**Ethical Issue Recognition**	Student can recognize ethical issues when presented in a complex, multilayered context AND can recognize cross-relationships among the issues.	Student can recognize ethical issues when issues are presented in a complex, multilayered context OR can grasp cross-relationships among the issues.	Student can recognize basic and obvious ethical issues and grasp (incompletely) the complexities or interrelationships among the issues.	Student can recognize basic and obvious ethical issues but fails to grasp complexity or interrelationships.
	Application of Ethical Perspectives/Concepts	Student can name and independently apply ethical perspectives/concepts to an ethical question, accurately, and is able to consider full implications of the application.	Student can name and independently apply ethical perspectives/concepts to an ethical question, accurately, but does not consider the specific implications of the application.	Student can name and apply ethical perspectives/concepts to an ethical question and the application is inaccurate.	Student is unable to name or apply ethical perspectives/concepts to the question.

Adapted from: Valid Assessment of Learning in Undergraduate Education (VALUE) system, The American Association of Colleges and Universities

Assessment of Interdisciplinary Competencies (AIC) developed by Universiteit Utrecht

Example 3: Spotlight on the environment



Module in the digital ethics course for undergraduates

Module in the digital ethics course for people working in higher education

→ Giving the facts
→ Suggesting some solutions
→ Group discussion

Conclusion

Focusing on virtue and wellbeing is a familiar theme for educators.

We now need to chart a way between the tragic dilemma of AI and the pathways of *keeping up* and *catching out*.

We need to pay attention to the tensions and interconnections between technology, people (social, creative, critical, and moral skills), and the (natural) world.

We can co-create a moral vision of what **should be** – the wellbeing of the global community.

In this way, educational communities can be the site of “hopeful re-imagining” of technologies.

Key references

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist*, 56(3), 218.

Fredrickson, B. L. 2004. "The broaden-and-build theory of positive emotions." Philosophical Transactions: Biological Sciences (The Royal Society of London) 359: 1367–1377.

Fredrickson, B. L. 2009. "Joy." In The Oxford Companion to Emotion and the Affective Sciences, edited by David Sander and Klaus Scherer, 230. Oxford: Oxford University Press.

Harrison, T. (2021). *Thrive: How to cultivate character so your children can flourish online*. Boston, MA: Little, Brown.

Johnson, M.K. & Robertson, R.S. (2022). "How Can Joy be a Divine Command if it is not Within Our Control?" Theological Puzzles (Issue 7). https://www.theo-puzzles.ac.uk/2022/05/02/robertson-johnson/.

Johnson, M. K. (2020a). Joy: A review of the literature and suggestions for future directions. *Journal of Positive Psychology*, *15*(1), 5-24.

Johnson, M. K. (2020b). Joy: A reply to the replies. *Journal of Positive Psychology*, *15*(1), 84-88.

Mullaney, T. (2024). How Schools Can Be Earth Reverant In The Age Of "AI." *Critical Inkling Substack*.
https://www.criticalinkling.com/p/how-schools-can-be-earth-reverent-ai

Robertson, R. & Johnson, M. K.(2023). Moral education in and for virtual spaces. In D. W. Yacek, M. E. Jonas & K. H. Gary (Eds.), *Moral education in the 21st century* (pp. 231-259). Cambridge University Press. https://doi.org/10.1017/9781009170291.016

Selwyn, N. (2021). Ed-tech within limits: anticipating educational technology in times of environmental crisis. *E-Learning and Digital Media*, *18*(5), 496-510. https://doi.org/10.26180/14746032.v1, https://doi.org/10.1177/20427530211022951

Tessman, L. (2015). *Moral failure: On the impossible demands of morality*. Oxford: Oxford University Press.

Van Cappellen, P. (2020). The emotion of joy: Commentary on Johnson. *Journal of Positive Psychology*. *15*(1), 40-43.

van Zyl, L. E., Gaffaney, J., van der Vaart, L., Dik, B. J., & Donaldson, S. I. (2024) The critiques and criticisms of positive psychology: a systematic review, *The Journal of Positive Psychology*, 19:2, 206-235, DOI: 10.1080/17439760.2023.2178956