

Approaching Critical AI Literacy for Students and Faculty with Compassion: Nurturing Agency, Recognizing and Mitigating Bias

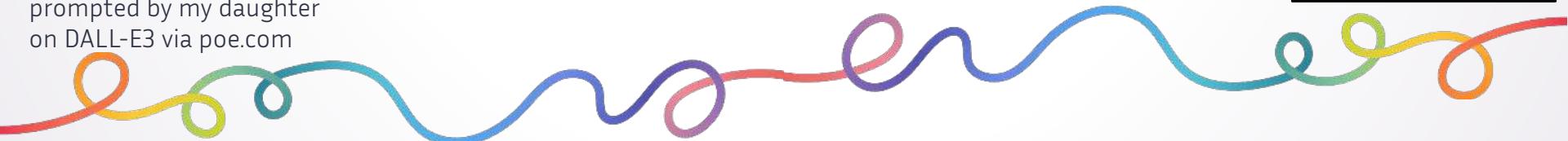


Image of cute sphinx prompted by my daughter on DALL-E3 via poe.com

Maha Bali, PhD  @bali_maha
American University in Cairo & Equity Unbound



Contact North - January, 2025

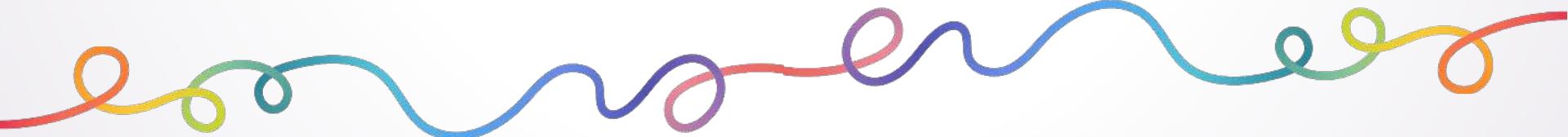


السلام عليكم
Sleepless in
Seattle: How do
you feel, where
are you?



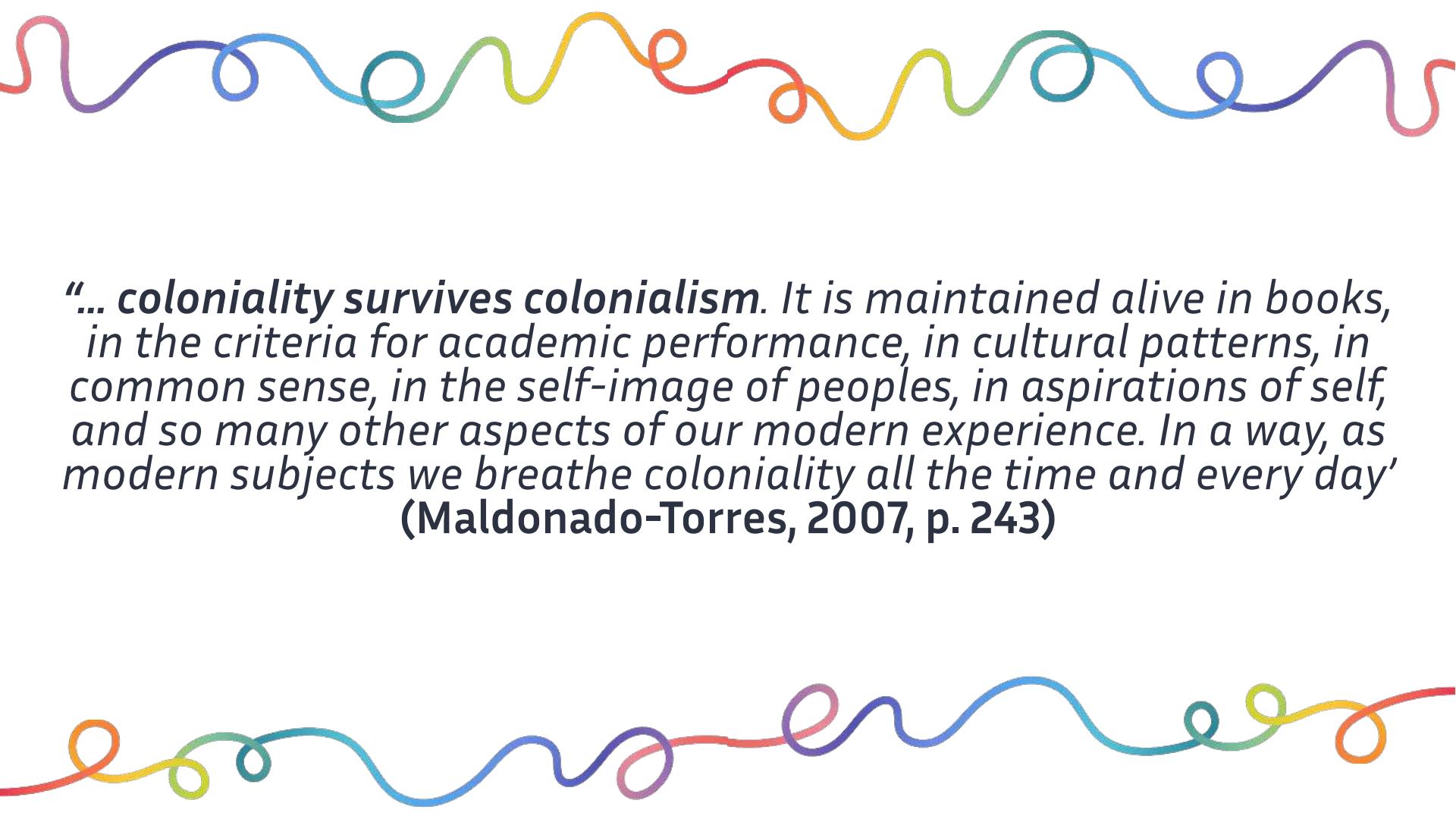
Hopes for today

1. Checking in and listening to you - interaction!
2. Compassionate design approach to AI
3. My Critical AI Literacy model, creative ways to teach it
4. Recognizing potential harm and bias in AI
5. Centering and nurturing human agency
6. Learning-equity matrix
7. Key takeaways and Q&A





Land acknowledgment



“... coloniality survives colonialism. It is maintained alive in books, in the criteria for academic performance, in cultural patterns, in common sense, in the self-image of peoples, in aspirations of self, and so many other aspects of our modern experience. In a way, as modern subjects we breathe coloniality all the time and every day”
(Maldonado-Torres, 2007, p. 243)



DIGITAL LITERACIES, EDUCATIONAL TECHNOLOGY

Against AI-Shaming

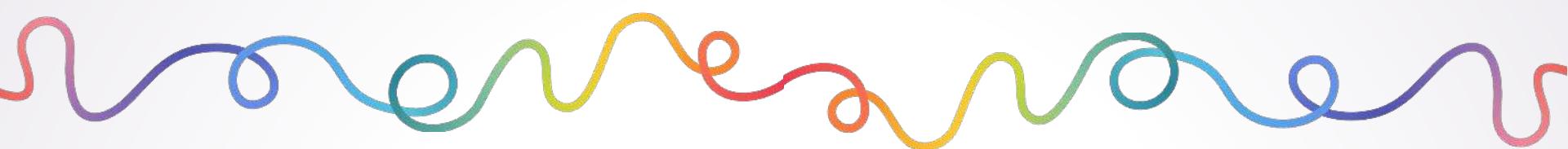
February 11, 2025 | Maha Bali

Estimated reading time: 8 minutes, 58 seconds

I'm facilitating AI roundtables at my department's Symposium and I've invited faculty members from across disciplines to share what they're doing in their classes. I've told them there is "no AI shaming"

Against AI Shaming - holding space, giving ourselves grace

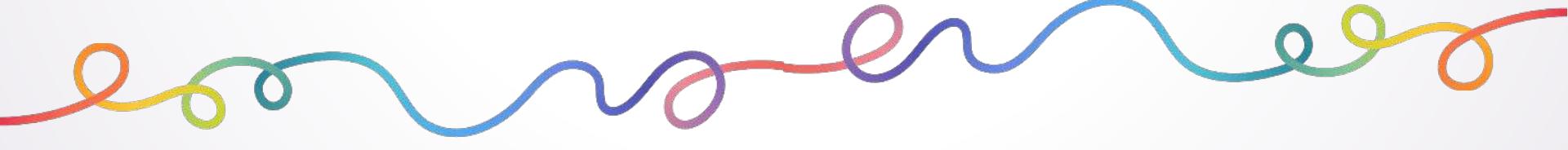
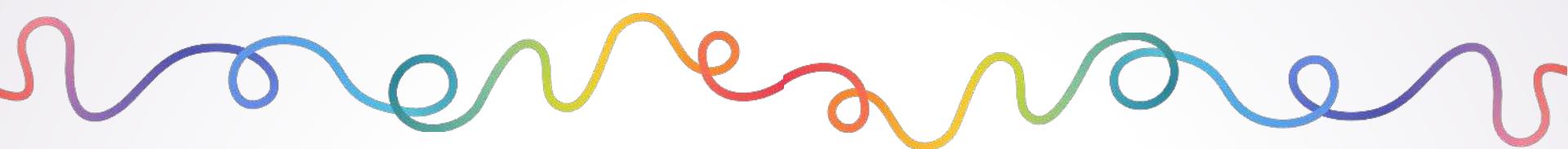
We can hold multiple views and be excited about AI despite its ethical issues; can reject AI despite its potential. [\(blogpost\)](#)



“How we choose to steward emergent intelligences will shape not only its future but our own. If we steward AI through a paradigm of extraction and exploitation, we will accelerate our demise. But if we guide AI toward a relational rationality rooted in accountability, kindness, and care, it may, in turn, help scaffold humanity to embody those same qualities—offering us a pathway to face modernity’s hallucinations and the truths it has conditioned us to deny.”

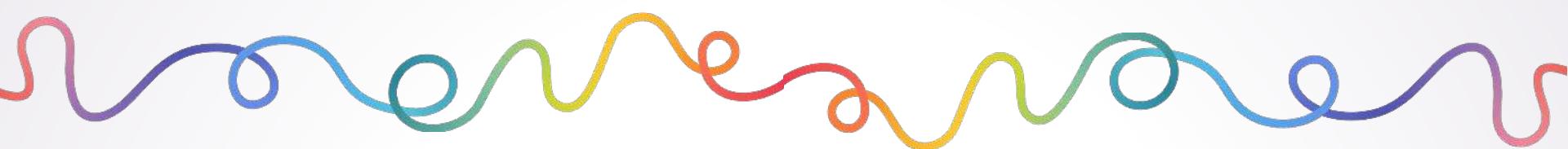
- Vanessa Andreotti (aka de Olivier)

<https://burnoutfromhumans.net/about>

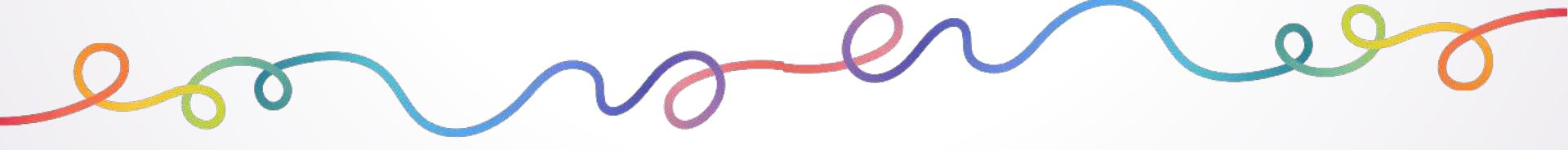


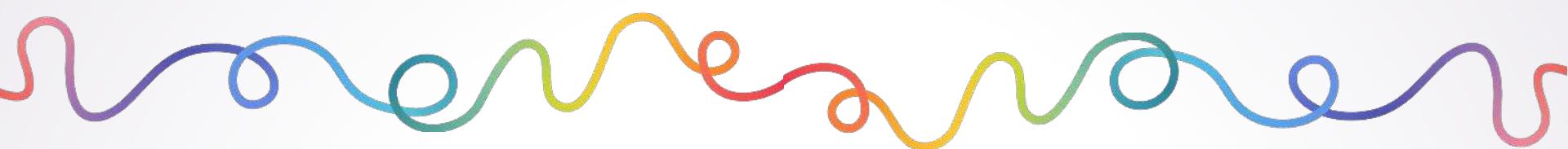
Slides are open for commenting
<https://bit.ly/BaliAlliteracyCN>



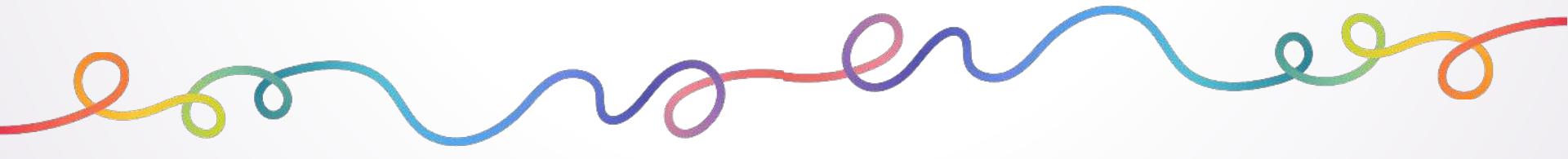


Tell me some things about yourself: what do you do for work?





Tell me some things about your attitude
towards AI - from the list of “attitudes” in
the next slide



In terms of my attitude towards AI, I am more of a/an... (Slido coming up!)

1. **Instrumentalization Wizard** - wield AI for any purpose – religious, ideological, educational, political, etc
2. **Playful Chaotic** - treat AI as a playground for creative possibilities
3. **Skeptical Pragmatist** - hesitant and cautious about AI use, and use it where it works
4. **Capitalist Extractor** – see AI as a way to make a profit
5. **Data Guardians** - don't worry about AI per se, but about privacy
6. **Footprint Alarmist** - focused on the carbon footprint of every query
7. **Shutdown Brigade** - want AI shut down and fear for our future
8. **Human Centric Purists** - see AI as a tool that threatens humanity's creativity and sacredness
9. Other? (please specify in chat)



In terms of my attitude towards AI, I am more of a/an... (choose up to two)





Sekhmet

Sekhmet was the goddess of war, destruction, and plague. She was also worshipped as protector of the king and a goddess of healing.

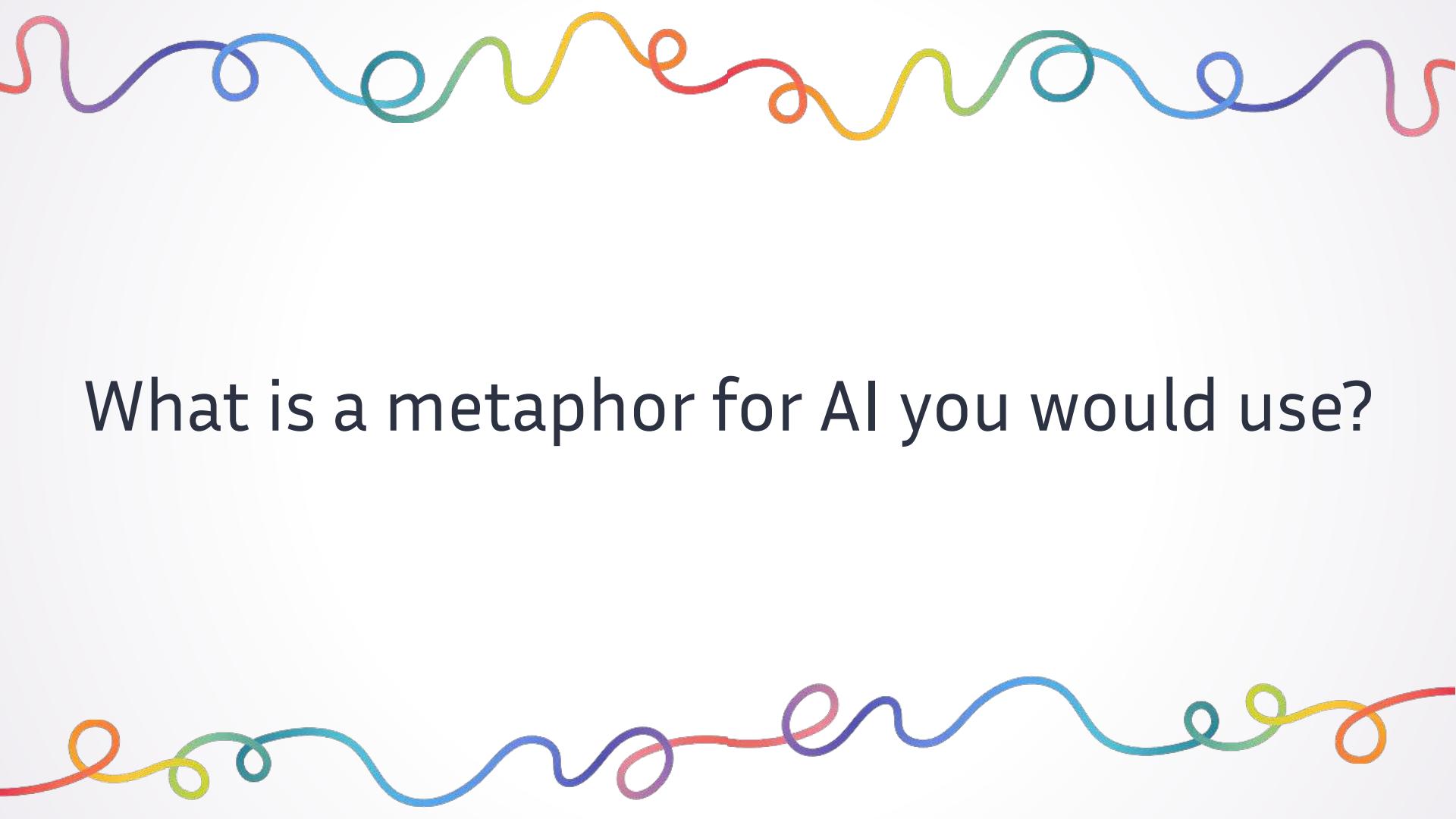
How can the same goddess be associated with war, destruction and plague, but also be worshipped as a protector of the king and a goddess of healing?

*Statue of ancient Egyptian Goddess Sekhmet, taken by me,
Grand Egyptian Museum*



And a reminder. Many companies that claim to be doing ethical AI and using AI for good have also been using AI to support war crimes at worst, and more often surveillance and exploitation of labor.

*Statue of ancient Egyptian Goddess Sekhmet, taken by me,
Grand Egyptian Museum*



What is a metaphor for AI you would use?

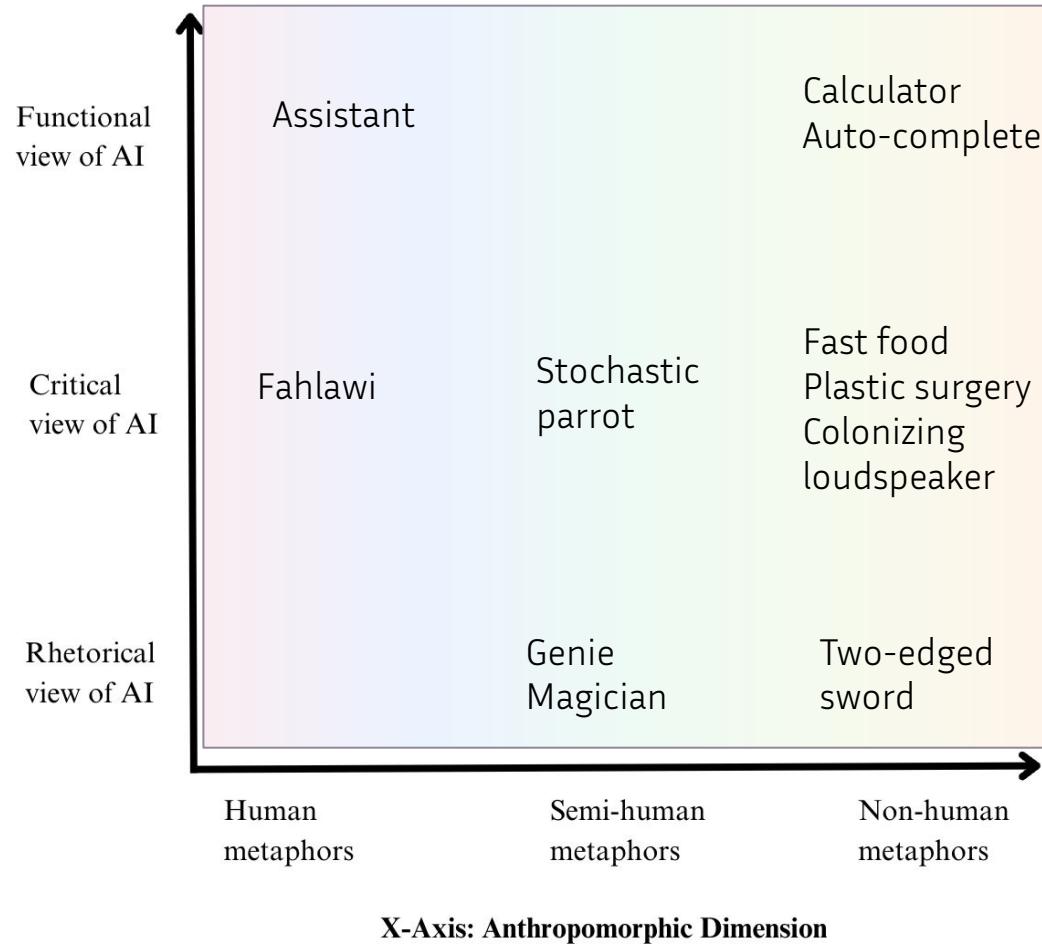
Assistant, Parrot, or Colonizing Loudspeaker? ChatGPT Metaphors for Developing Critical AI Literacies

(Gupta, Atef, Mills & Bali,
2024)

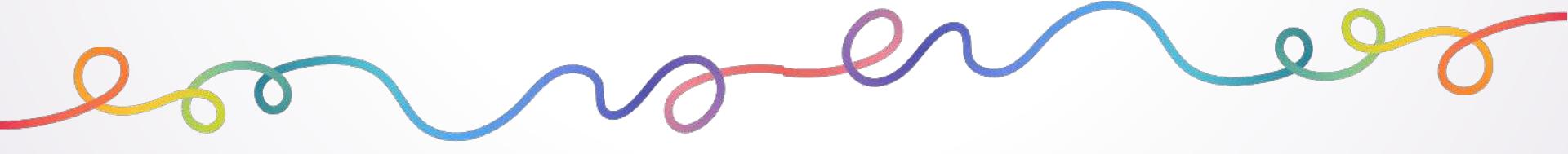
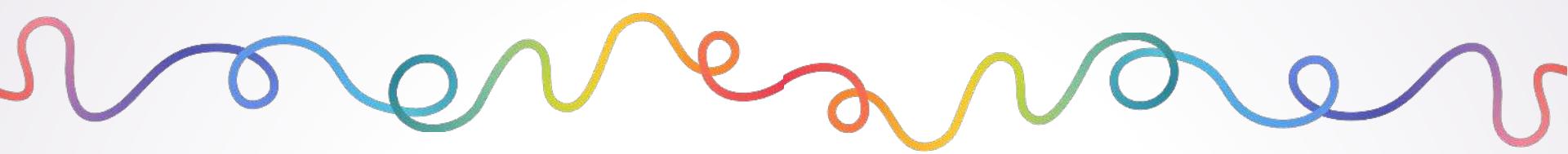
<https://openpraxis.org/articles/10.55982/openpraxis.16.1.631>



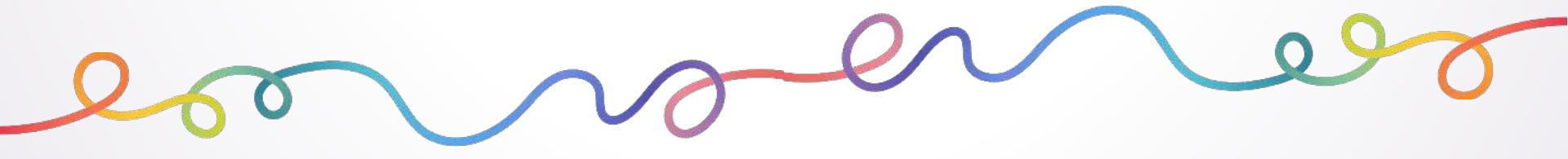
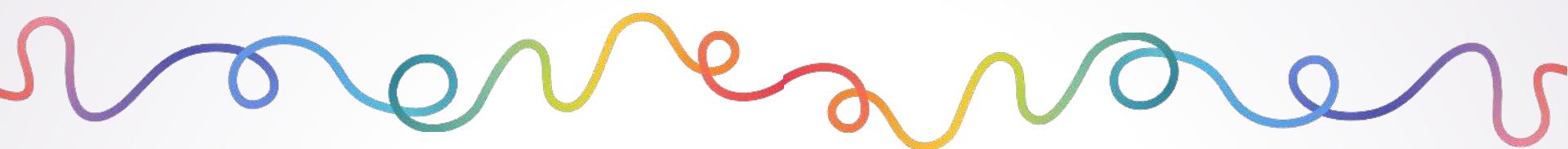
Image of Parrot, assistant and Loudspeaker generated by
Maha Bali using [DALLE-3 via poe.com](https://poe.com)



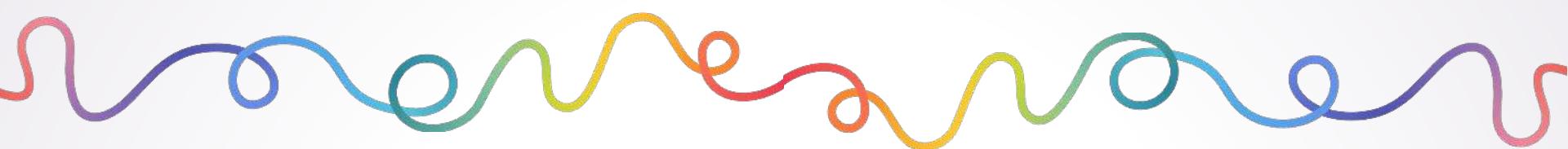
Gupta et al,
2024) -
some data
included



Responding to AI through Compassionate Learning Design

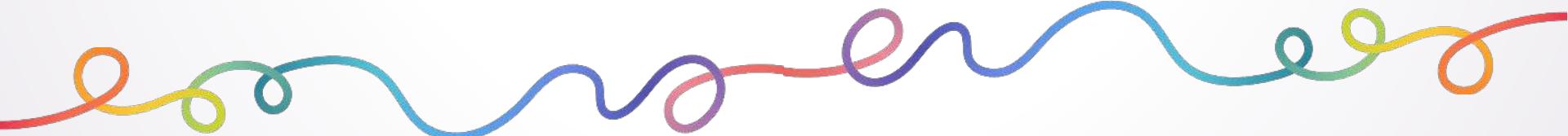


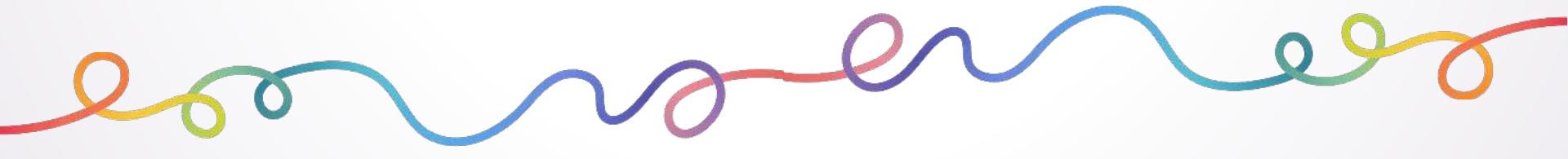
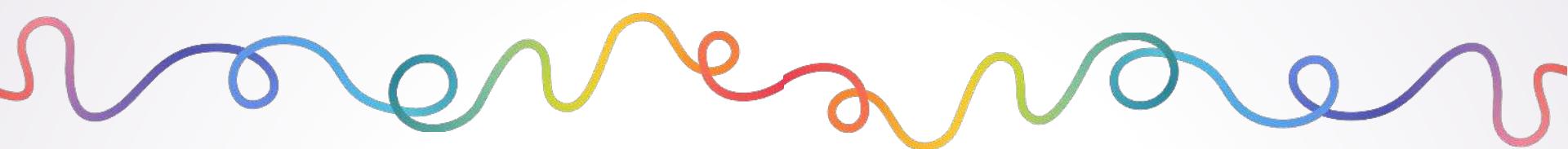
What part of your teaching/work brings you the most joy?



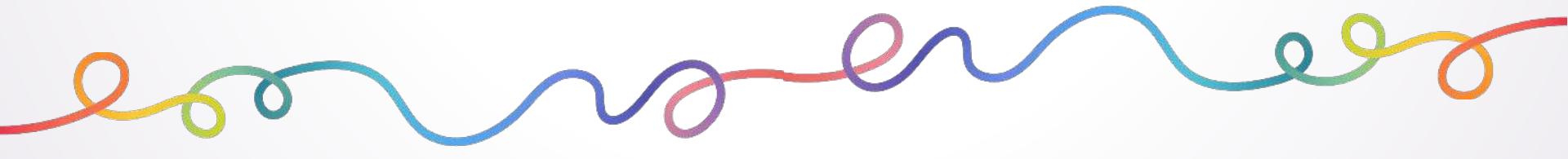
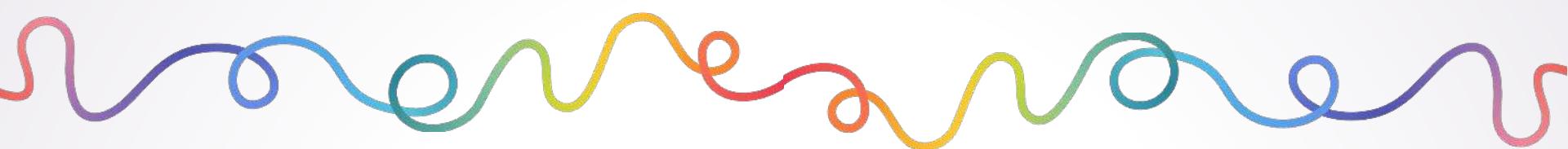
Compassion for teachers:
What is something you are grieving (due to
the existence of genAI) that you have or
haven't named out loud yet?

(inspired by this)

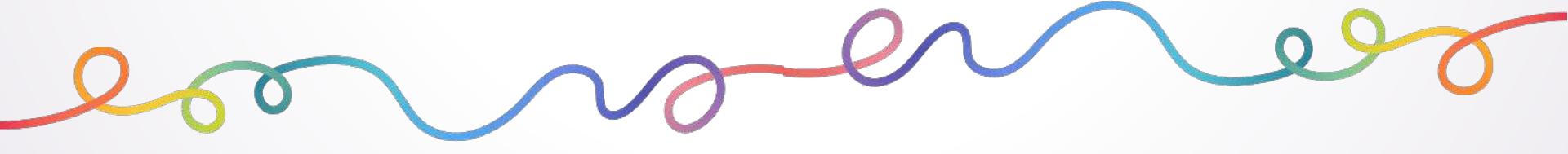
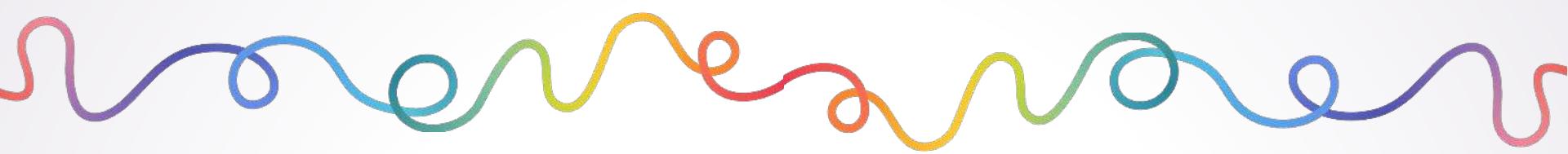




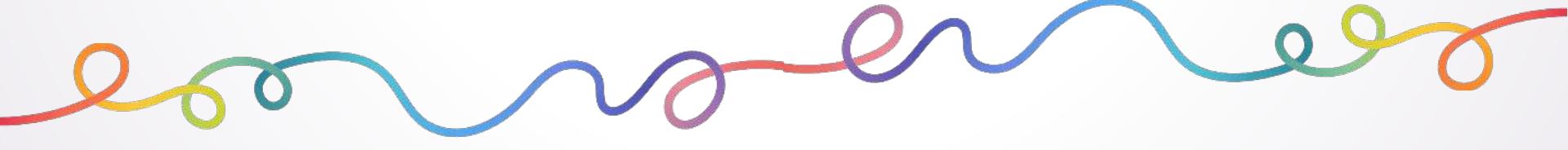
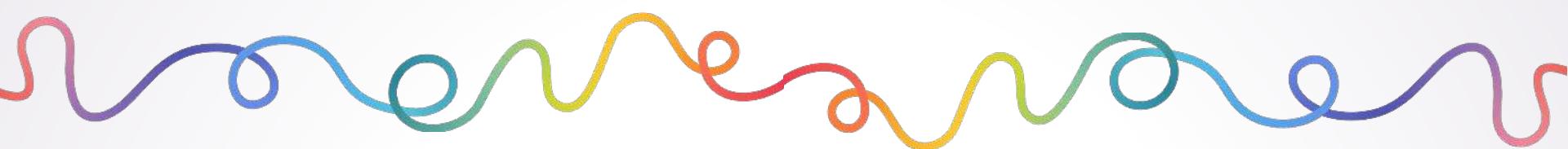
What makes students use AI as a shortcut
(or other unauthorized help)?



Understand what can lead learners to take
(unauthorized) shortcuts & build a
compassionate communal response-
[blogpost with Yasser Tamer](#)



Playful Confessions approach...



Cake analogy

(*Teaching and Generative AI: Pedagogical Possibilities and Productive Tensions*

<https://www.usu.edu/empowerteaching/publications/books/teaching-ai/>)

Cake as a metaphor for AI

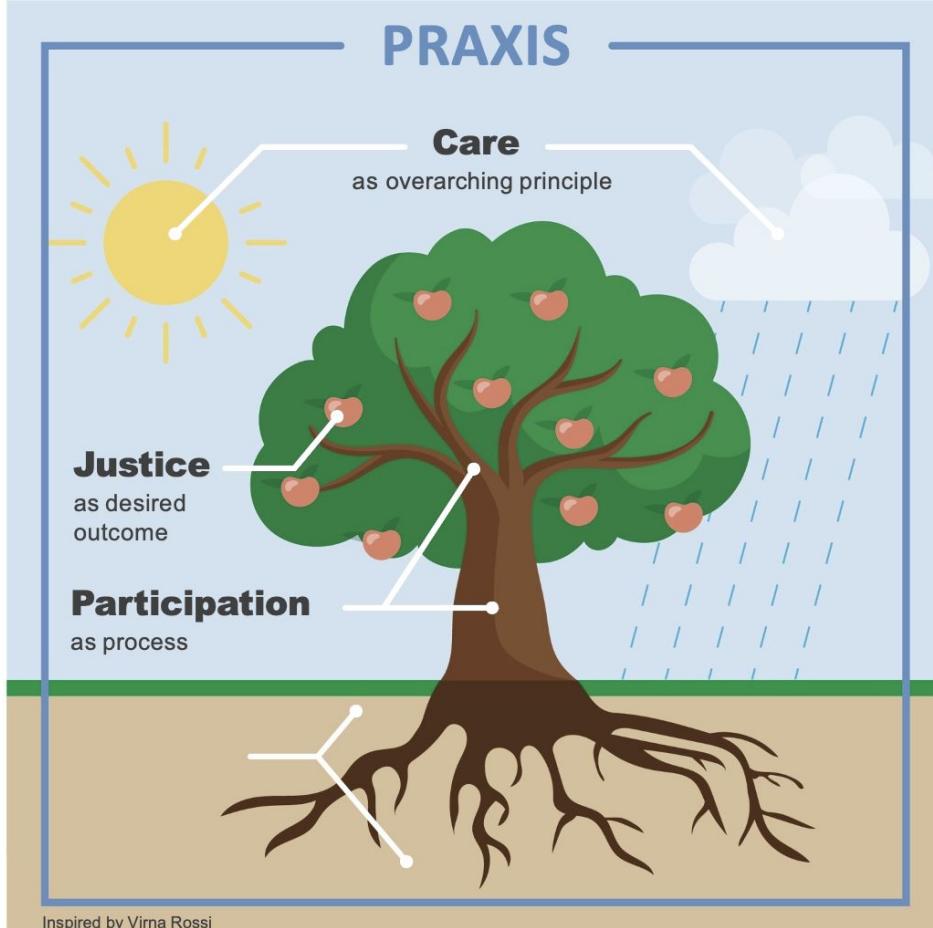


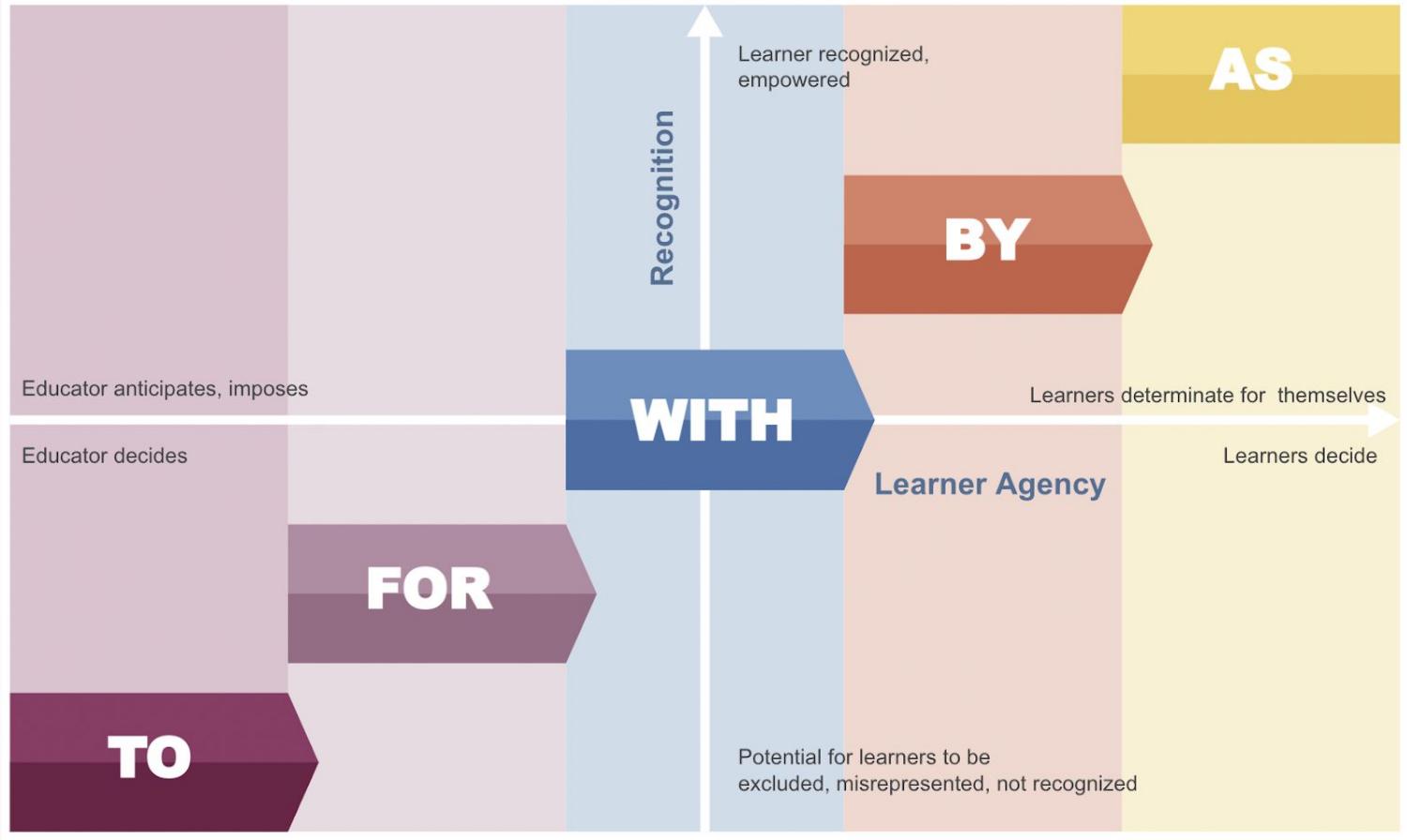
When would you
make it from
scratch, from box,
bakery or grocery?



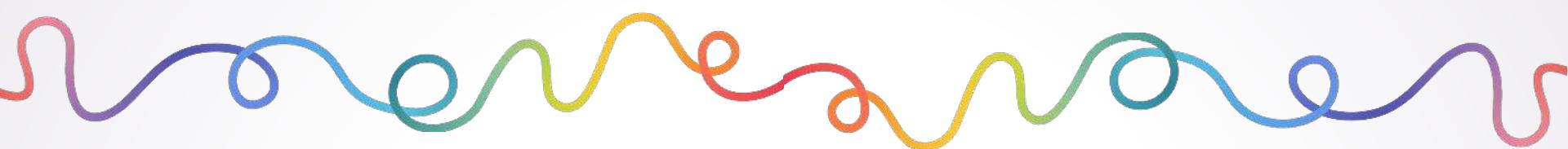
Compassionate Learning Design

Gachago, Bali & Pallitt (2022):
Compassionate learning design.
In Jhangiani, S., J. Quinn &
M. Burtis (Eds.). *Toward a Critical
Instructional Design*.
PressBooks.

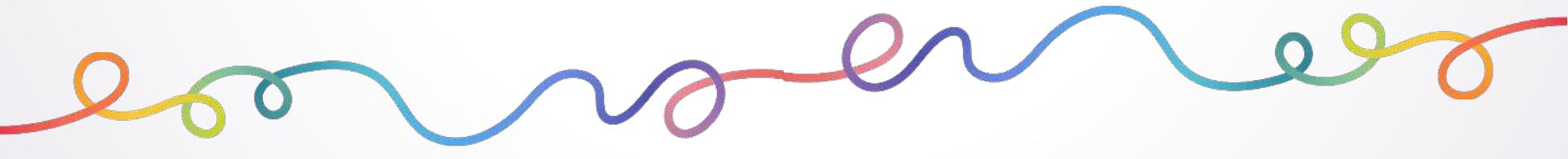




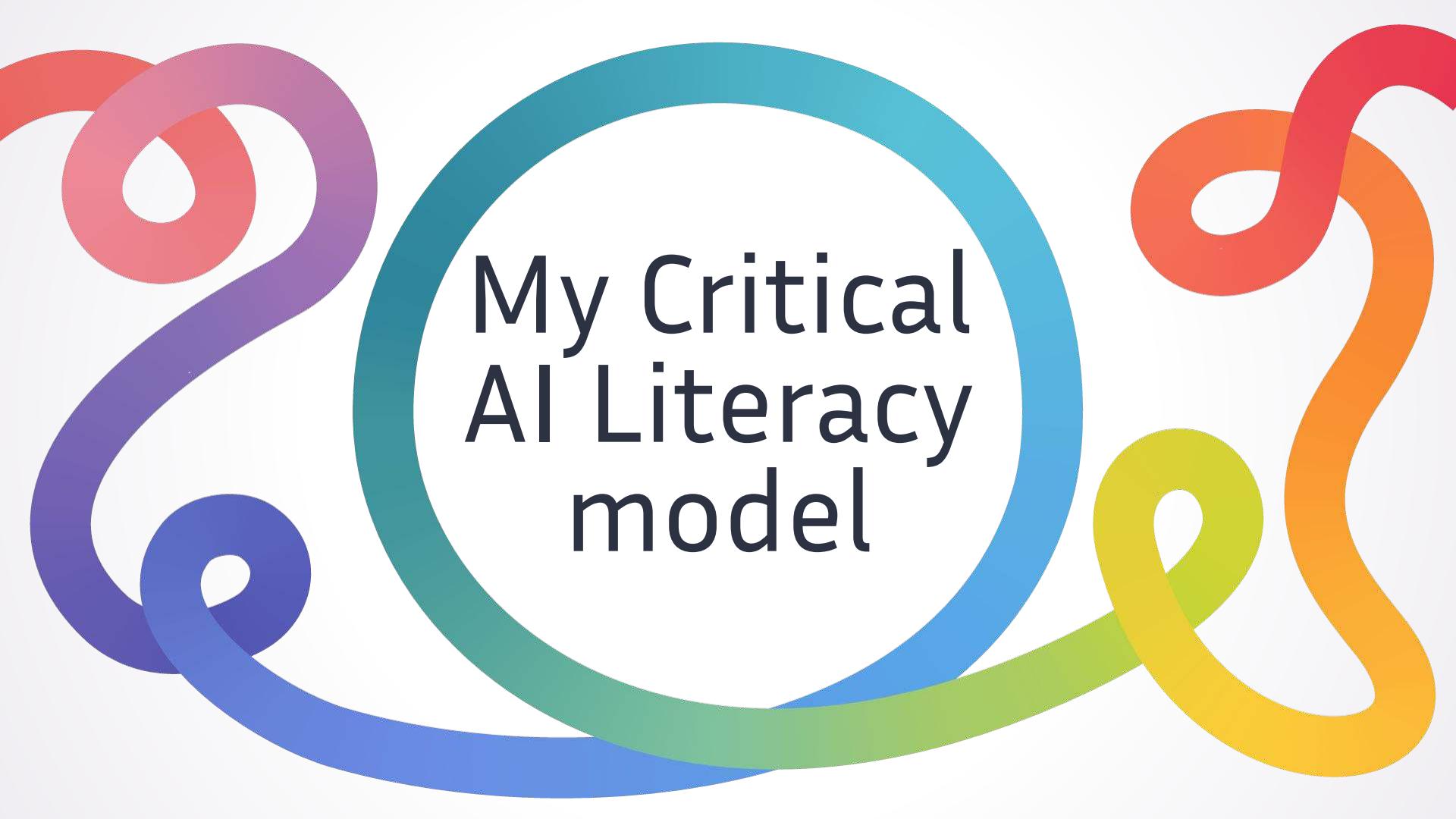
From designing with empathy to co-designing with compassion based on Wehiipeihana (2013)



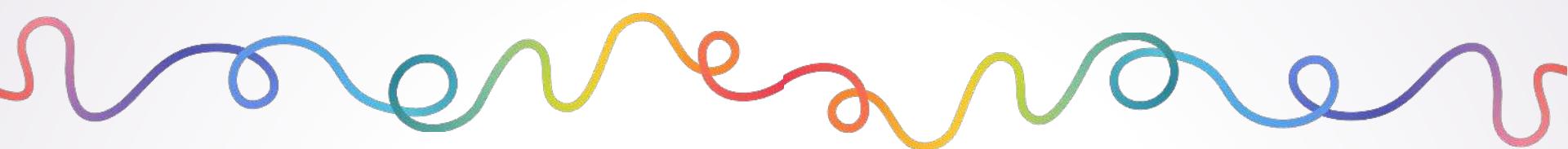
Compassionate learning design centers learner agency as an approach to creating caring and socially just learning environments and experiences



	TO	FOR	WITH	BY	AS
What it involves	Educator decides which AI tool to use and scaffolds preferred use. Educator decides on criteria for appropriate use without inputs from students.	Educator asks students which AI tools they are familiar with to inform selected tool/s. Criteria for appropriate use decided by educator in advance.	Educator and students openly discuss AI tools and ways of using it, appropriate use is negotiated. Educator invites students to make suggestions for revisions to criteria for assessments	Once familiarity has been established, students decide on criteria for appropriate use of AI tools, lecturer may make these official in the form of a submission brief	Students establish and share strategies and experiences of using AI tools to support their learning. Students decide on criteria for appropriate use and share these in the form of a reflection to accompany a submission
Pedagogic strategies / Examples	Educator models use of AI tool, decides on question prompts, and leads analysis of output, scaffolds critical AI literacies through demonstration.	Educator models use of AI tool, asks students to make suggestions to revise question prompts and facilitates discussion around analysis of outputs. Afterwards the educator may set an individual or group task to scaffold critical AI literacies experientially.	Educator and students collaborate on a toolbox of relevant AI tools and co-design the criteria in an assessment brief on appropriate use of AI tools for a particular assignment.	Students use agreed criteria as part of a peer review task and negotiate further revisions to criteria for the final submission based on their findings	While this may happen informally in study groups, this is where educators intentionally create space for such interactions in formal curricula.



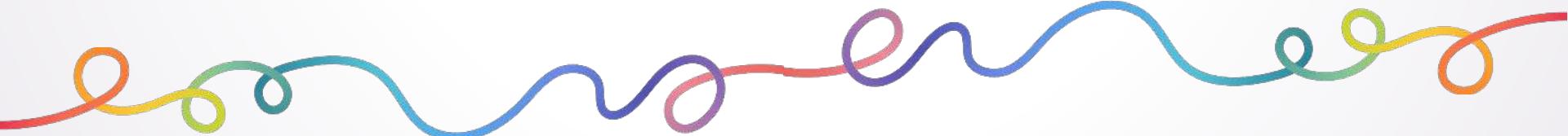
My Critical AI Literacy model



For me, being critical goes beyond critique and scepticism: it includes subscribing to critical theory and critical pedagogy – developing awareness of social justice issues and cultivating in learners a disposition to redress them.

Bali, M. (2024). Where are the crescents in AI? *LSE HE blog*.

<https://blogs.lse.ac.uk/highereducation/2024/02/26/where-are-the-crescents-in-ai>



"AI cannot be considered **inevitable, beneficial or transformative** in any straightforward way. You do not even need to take a strongly normative perspective either way to see that AI in education is highly **contested and controversial**. It is, in other words, a public problem that requires public deliberation and ongoing oversight if any **possible benefits** are to be realized and its **substantial risks** addressed." - Ben Williamson (emphasis mine)



On the importance of critical AI literacy

Understand
how it works



Recognize
inequalities &
biases



Assess
appropriate use



Critical
AI
Literacy



Examine ethical
issues



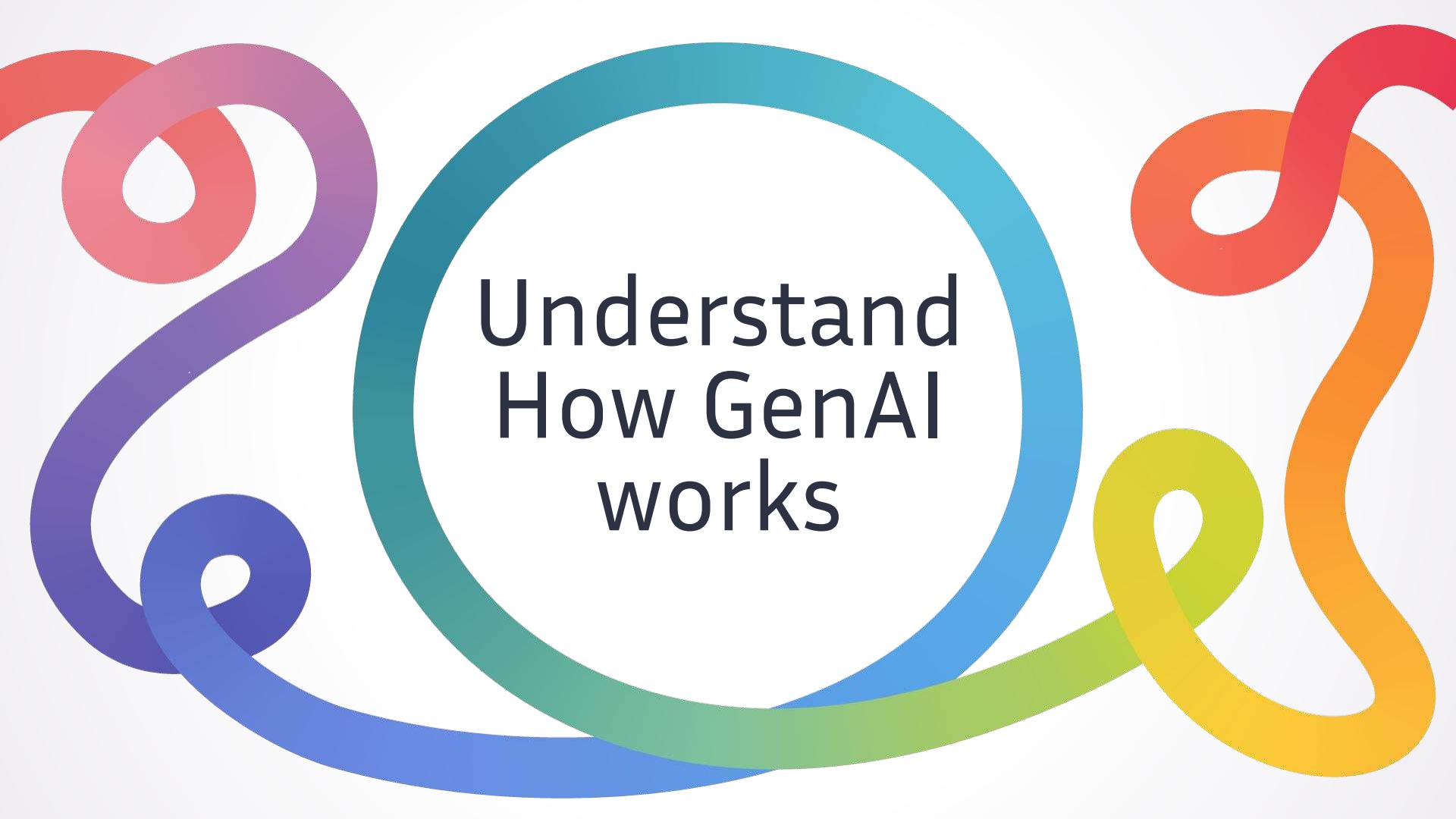
Craft effective
prompts



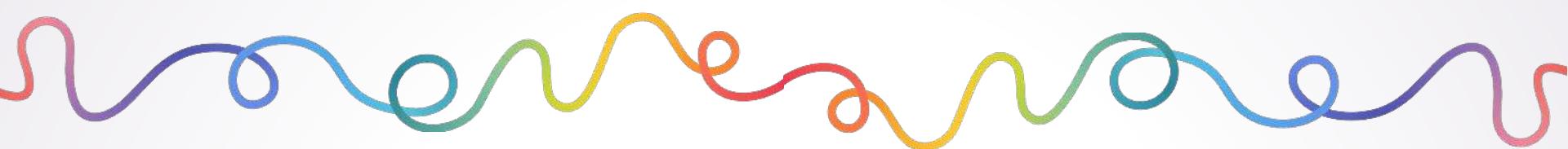
From Bali, M. (2024). Where are the
crescents in AI? *LSE HE blog*.

<https://blogs.lse.ac.uk/highereducation/2024/02/26/where-are-the-crescents-in-ai>

Dimension of critical AI literacy	AIPP Resource to use: https://aipedagogy.org/
How generative AI works	AI Starter & LLM Tutorial & ** QuickDraw
Awareness of inequalities and biases	Critical analysis across AI tools and stereotypes & **modeling
Ethical issues	Debating the ethics of generative AI or Close reading of ToS or ** eCampus Ontario H5P
Appropriate use	AI Sandwich or A tale of two critiques or ** Cake analogy
Prompt engineering	Resources (external links)→ Prompting ** Self-assessment



Understand
How GenAI
works



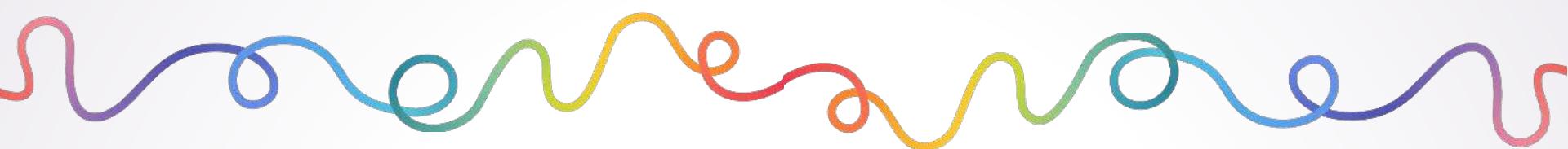
Play QuickDraw



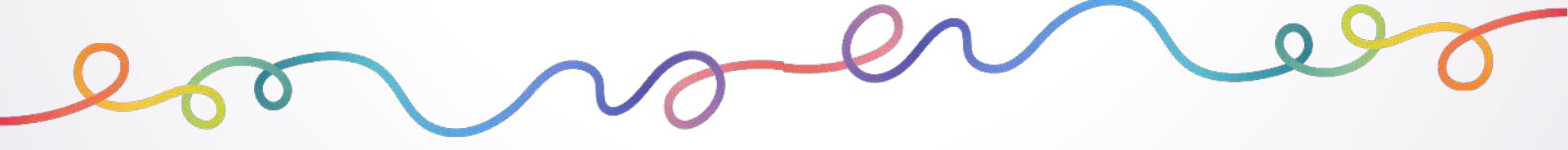
Can a neural network learn to recognize doodling?

Help teach it by adding your drawings to the [world's largest doodling data set](#), shared publicly to help with machine learning research.

Let's Draw!

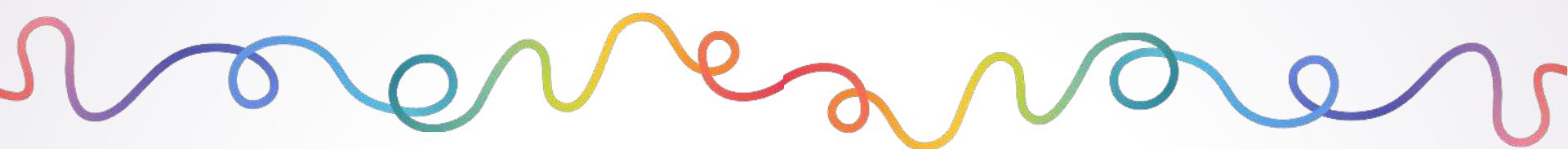


Recognize that hallucinations are a *normal* output of generative AI - some are calling them *mirages* (Mills & Angell).

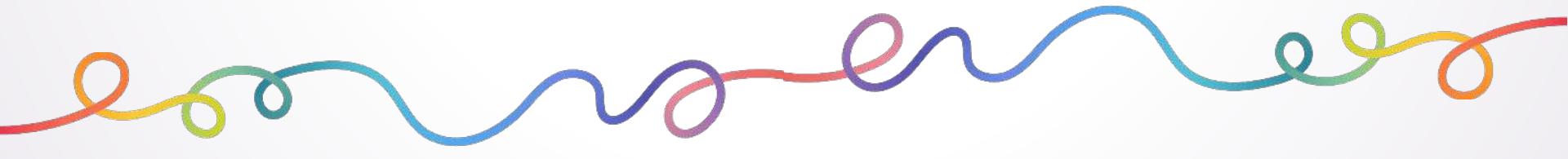


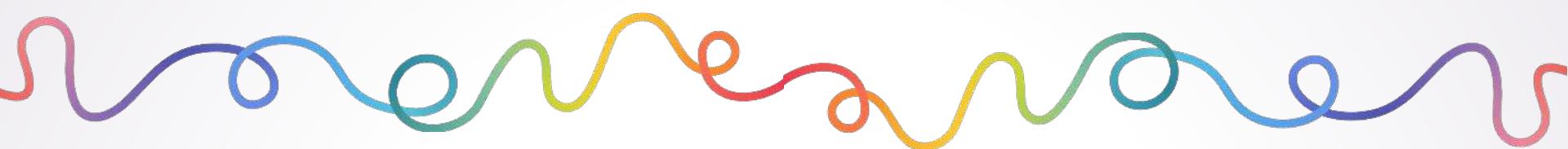


Recognize
Inequalities
and Biases



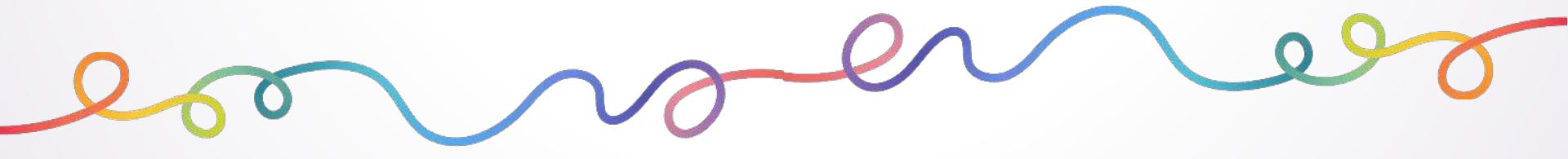
1. Everyone has biases, especially unconscious bias
(AI is trained on biased datasets)
2. Statistical reasoning (how AI works) inherently reproduces bias
3. Bias in AI has caused harm in the past
(recruitment, criminal justice, facial recognition)

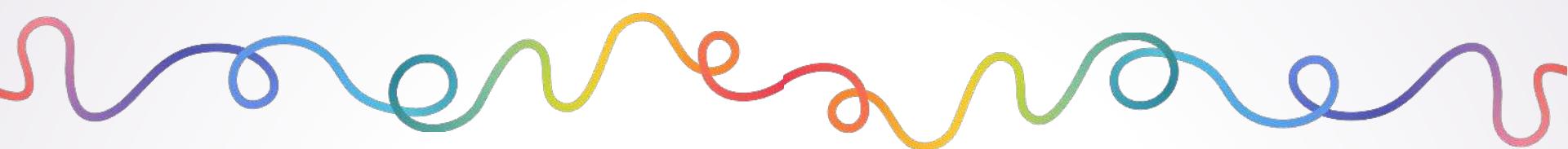




"AI is essentially a reflection of who we have let have a voice in the past and that voice was built from and still has large contributions from systems of hierarchy that use academia to further justify their positionality of power and stratification."

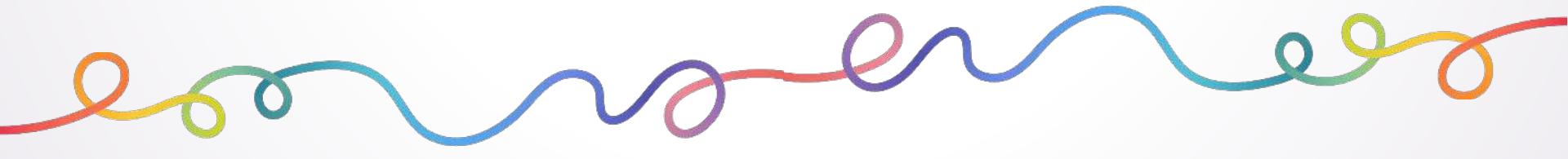
- Belinda 'Ofakihevahaoa Fotu, 2024

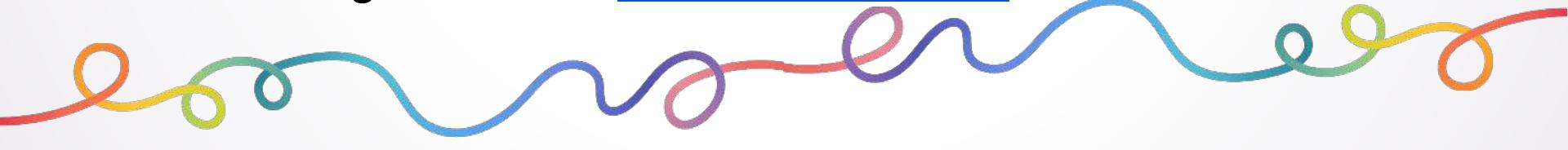
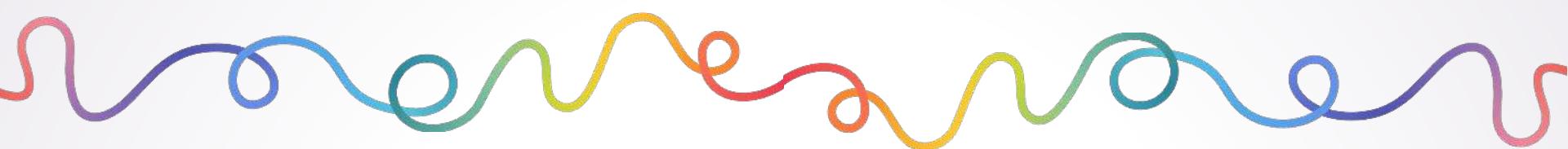




“Even if full and accurate proportional representation in data and the elimination of all human bias in labels, proxies and metrics were possible, AI powered by statistical reasoning would still decide against people and communities who are statistical outliers and marginalized minorities. It is a highly seductive form of digital eugenics that widens disparity and decreases diversity. It is academia’s favorite research method running amok.”

- Jutta Treviranus, undated



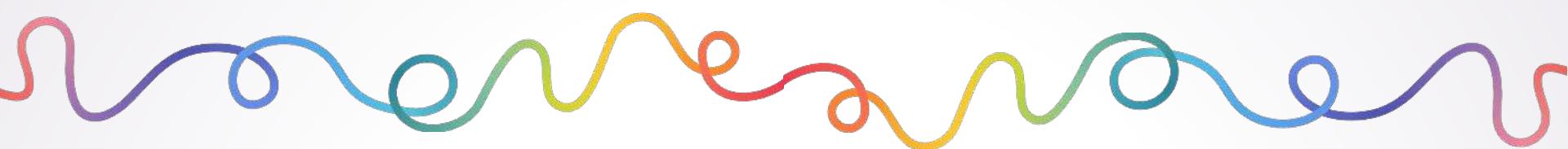


"Indigenous communities face the following **paradox** in the Generative AI space: in wanting to be represented by the Generative AI by sharing data representations their ways of knowing and being, they **lose the agency** to exert their **rhetorical, technological, and data sovereignty** over whatever is shared into the Generative AI system. Alternatively, not participating in the Generative AI space continues to perpetuate Western-centric biases and systemic racism built into the existing algorithms." - [Cardona-Rivera](#) et al, 2024

"Large language models (LLMs) can pass explicit bias tests but still harbor implicit biases, similar to humans who endorse egalitarian beliefs yet exhibit subtle biases." - Bai et al

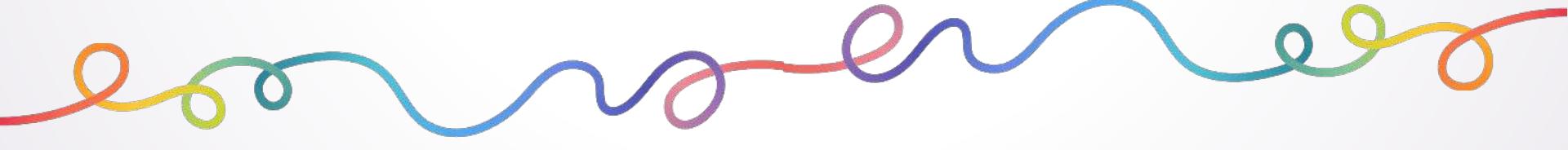
(see also Hoffman et al on dialects & Manvi et al on lower socioeconomic regions)





Terrorism Story

Bali, M. (2024, Summer). From the subtle to the glaring: Racism and bias in AI, Research Intelligence, Summer 2024 issue. [Invited].
<https://www.bera.ac.uk/publication/summer-2024>

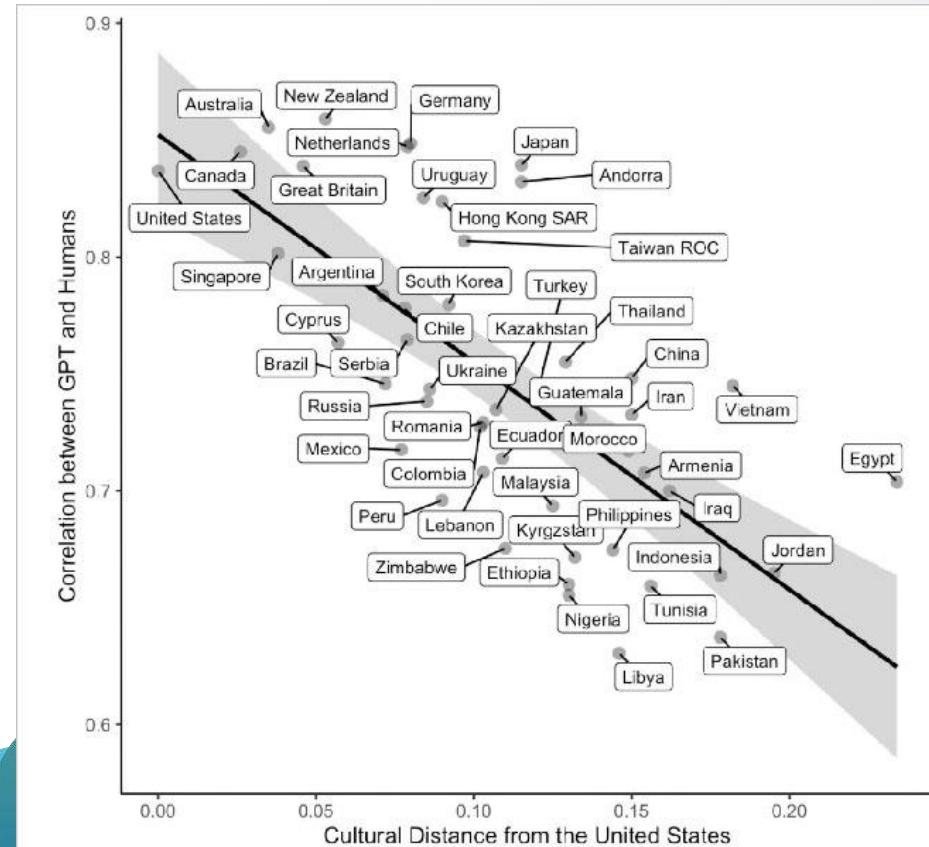


“These results point to a strong WEIRD bias in GPT’s responses to questions about cultural values, political beliefs, and social attitudes.”

(p. 11)

WEIRD = Western, Educated, Industrialized, Rich, and Democratic

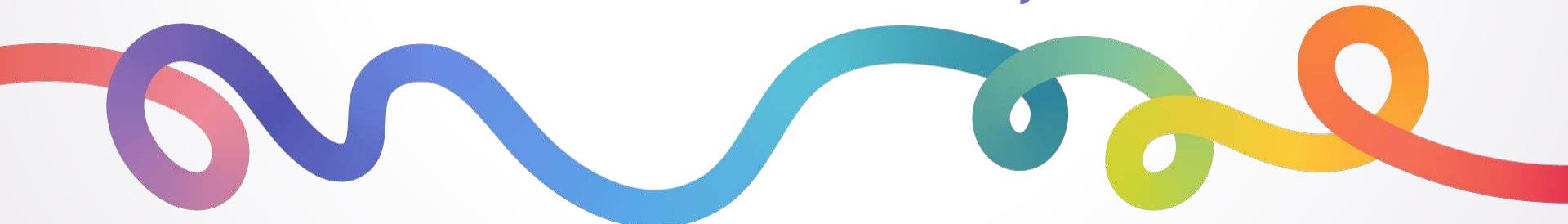
Atari et al (2023). Which Humans?
<https://osf.io/preprints/psyarxiv/5b26t>



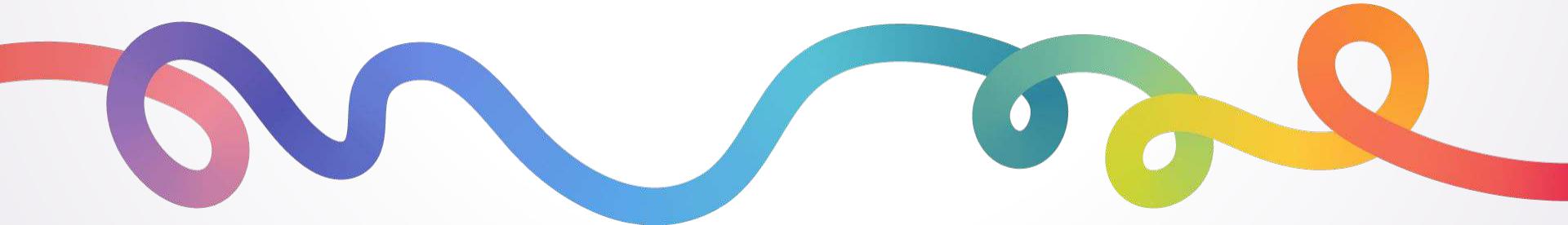


Accountability v Automation

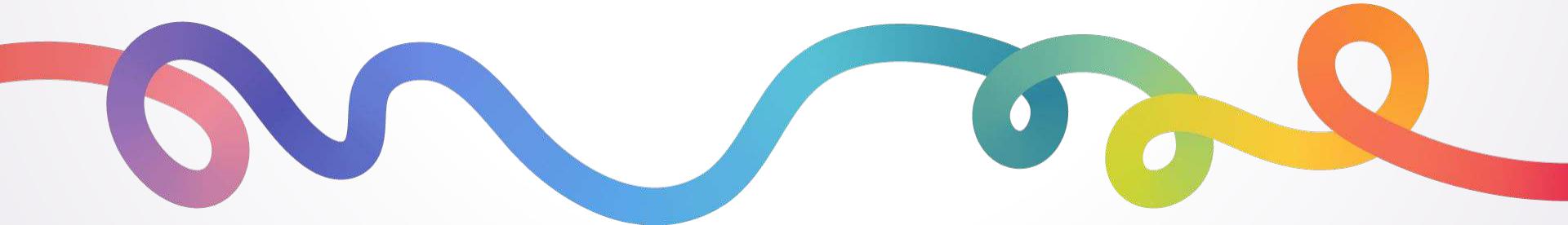
“an over-reliance on AI for personalisation could lead to a **formulaic** and rigid education system where algorithms **determine** learning paths. There was a fear that AI could **limit students' opportunities for exploration and growth based on their past behaviour or performance**. Moreover, there were serious concerns about the bias of such systems and how we could increase, not decrease, inequitable treatment of students.” - [AdvanceHE University of the Future](#)



“the need for **human agency** to be maintained in the face of AI integration. For students, this means having the ability to influence their learning journey, make choices, and explore beyond AI-recommended paths. For educators, it emphasises the role of the teacher not just as a facilitator but as an essential guide who brings **empathy, creativity, and intuition** to the educational process: qualities that AI cannot replicate.” - [AdvanceHE University of the Future](#)

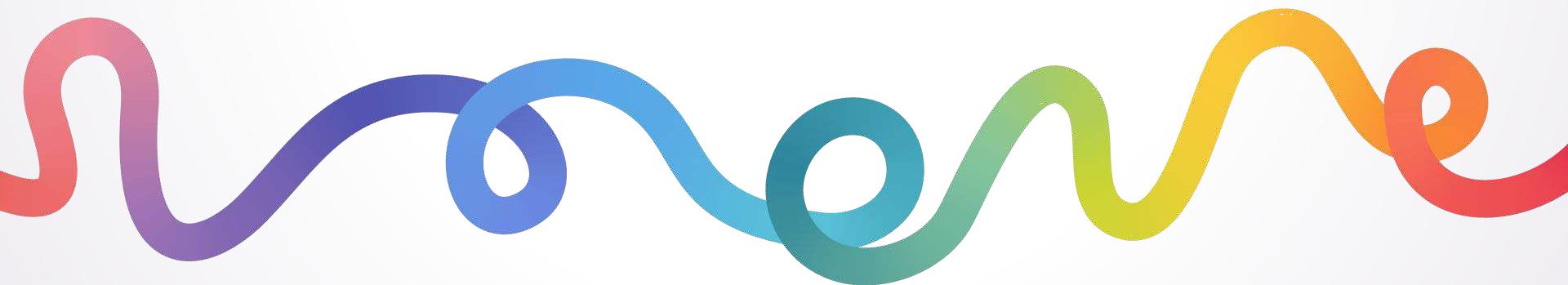


Can we use AI to better support people with disabilities without reproducing neuronormativity and ableism?



Use of AI for surveillance and policing...

The [AI] detectors are not particularly reliable. Worse yet, they are especially unreliable when the real author (a human) is not a native English speaker." - Myers



4 I's of Oppression Lego's video



Ways AI Can Reproduce Oppression at All Levels

Level	Example
Ideological	Implicit bias in outputs (epistemic injustice)
Internalized	Non native speakers think it writes better and lose their own writing voice
Interpersonal	AI detectors conversations AI as censor
Institutionalized	Grading or learning analytics based on norms AI admission/recruitment reproducing bias Facial recognition racism Who has access? To which tools? Language?

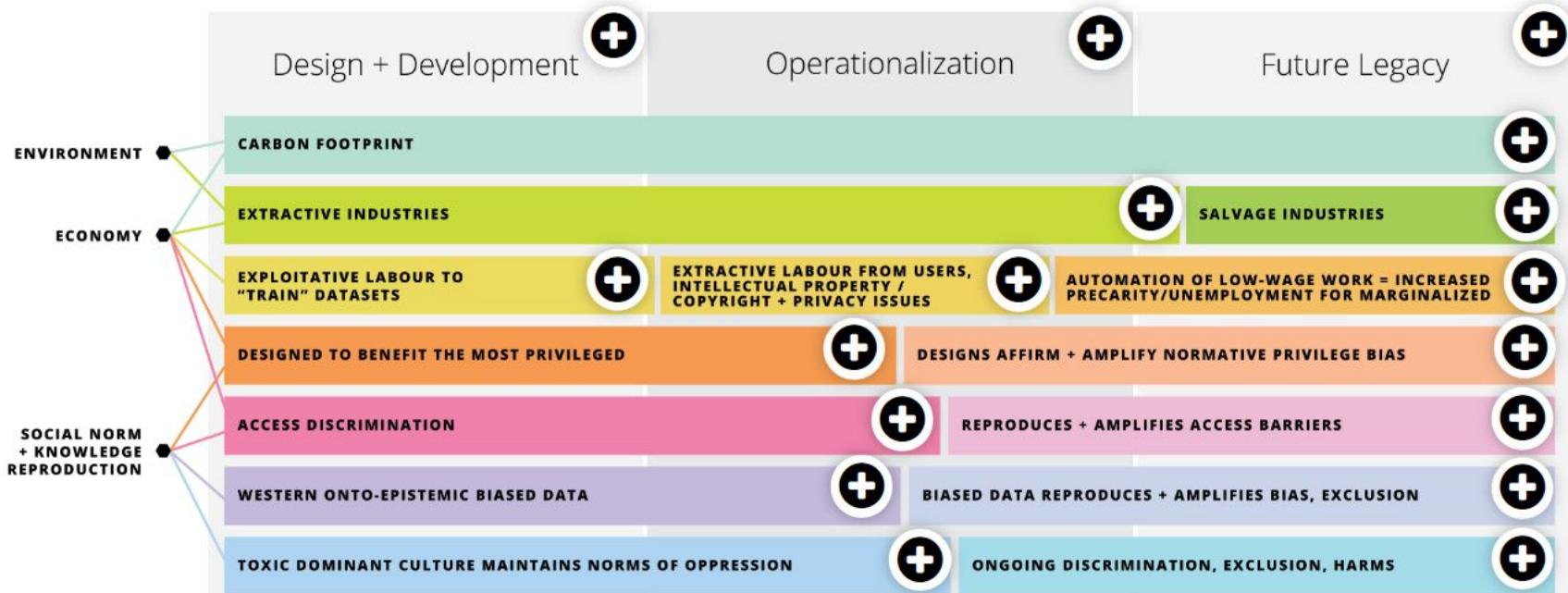


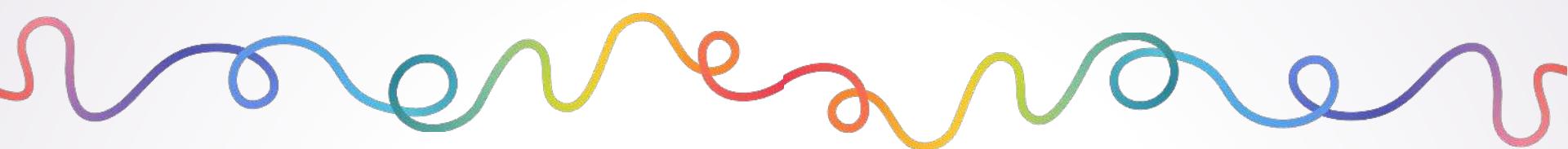
Examine
Ethical
Issues

eCampus Ontario's Harm Considerations of LLMs resource

<https://h5pstudio.ecampusontario.ca/content/51741>

Some Harm Considerations of Large Language Models (LLMs)



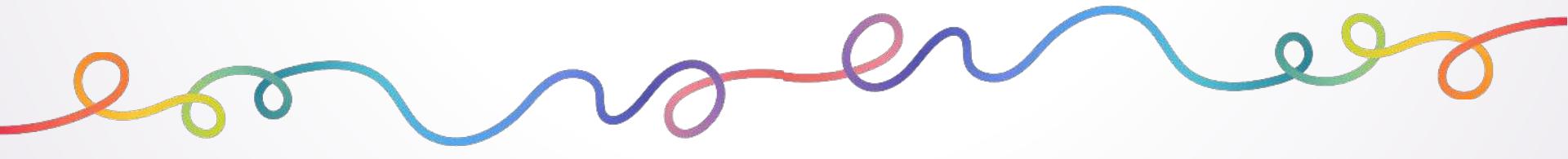


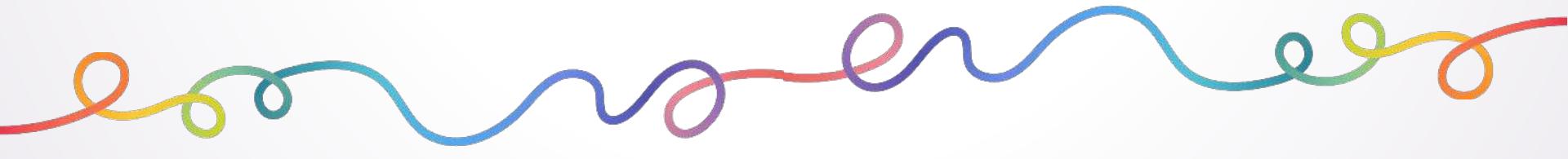
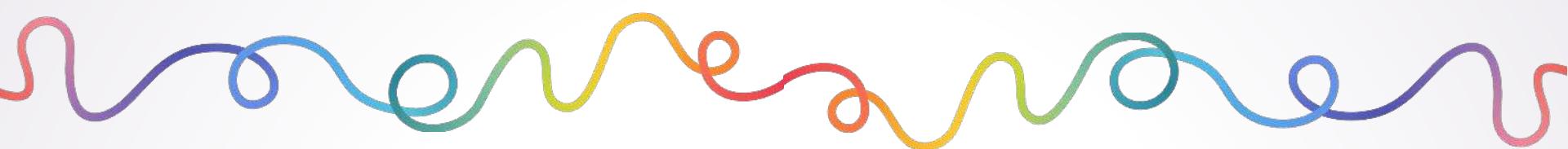
Deploying AI for harm:

- Genocide and Targeted Depopulation
- Eugenics and Genetic “Selection”
- Autonomous Warfare and Civilian Extermination
- Biological Warfare Accelerated by AI
- Techno-Supremacy
- Global Dictatorship by AGI Controllers

(Burnout from Humans, p. 104-105)

(+ use in aid distribution and humanitarian assistance!!!





Gebru & Torres (2024). The TESCREAL
bundle: Eugenics and the promise of utopia
through artificial general intelligence. First
Monday (29, 4).

<https://doi.org/10.5210/fm.v29i4.13636>

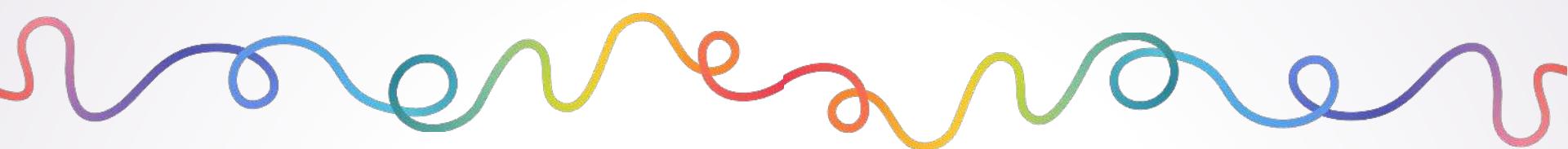


Craft
Effective
Prompts

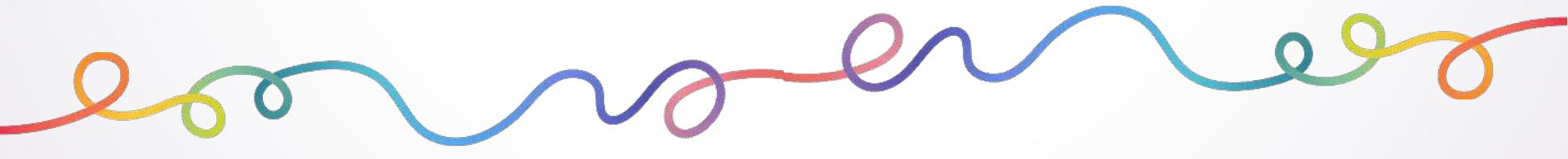
- Modeling and Trial & error
- Many free courses and blogposts on this

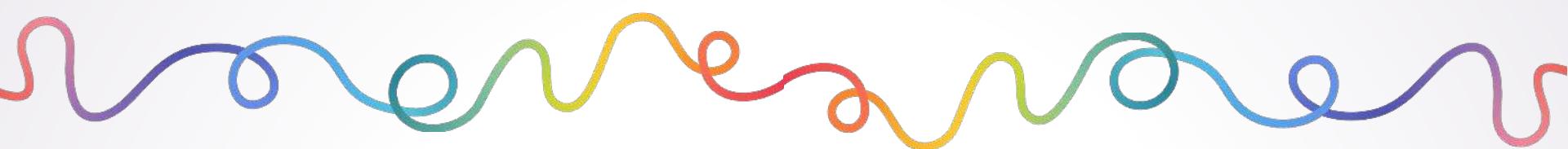


Assess
Appropriate
Use



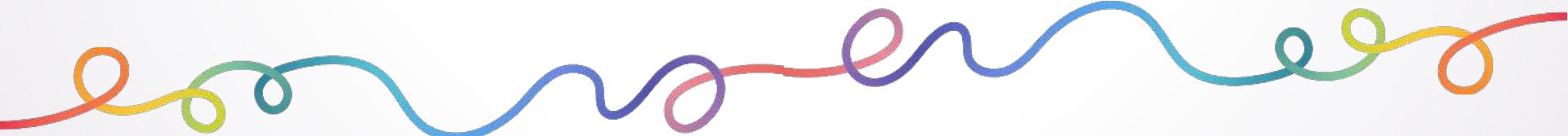
Tell me about a way GenAI has NOT worked
well for you recently?

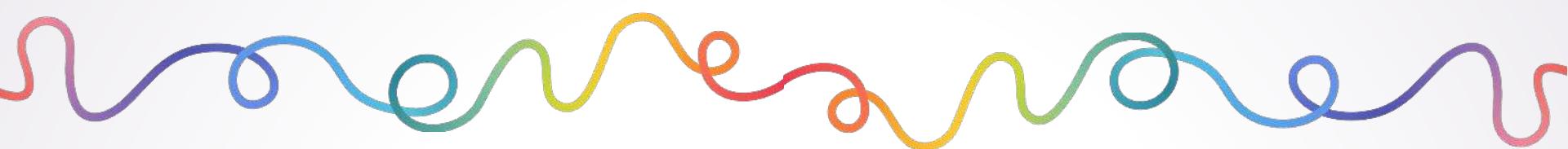




Selwyn, N., Ljungqvist, M., & Sonesson, A. (2025).
**When the prompting stops: exploring
teachers' work around the educational
frailties of generative AI tools.**

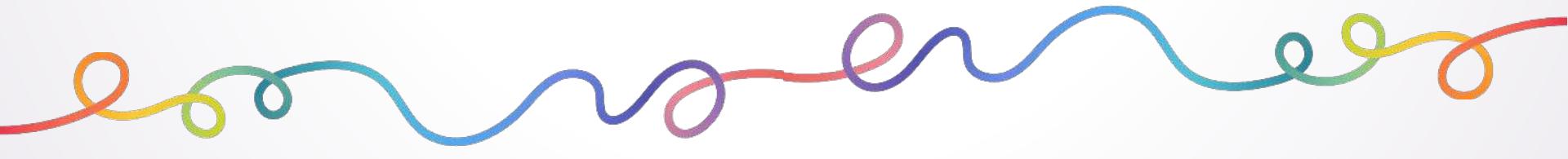
Learning, Media and Technology, 1–14.
<https://doi.org/10.1080/17439884.2025.2537959>

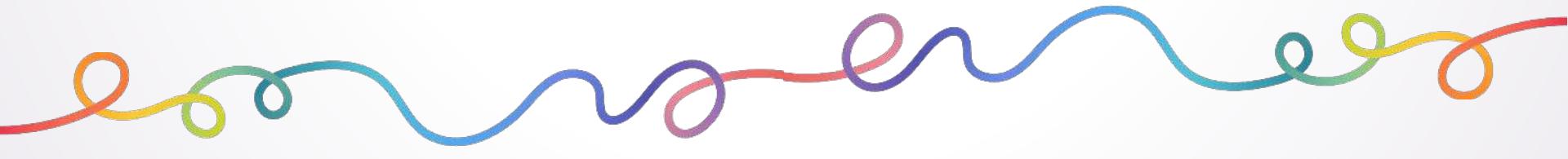
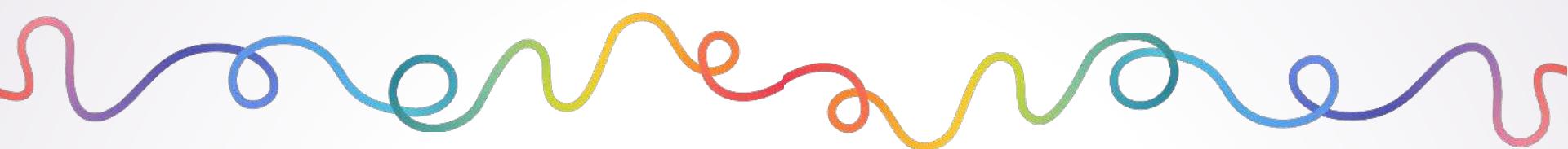




“The paper concludes by considering the need to challenge ongoing industry and policy claims around the labour-saving benefits of artificial intelligence in education, and instead focus on the ways in which these technologies are dependent on the hidden labour of humans to co-produce the illusion of automation.”

- Selwyn et al, 2025.

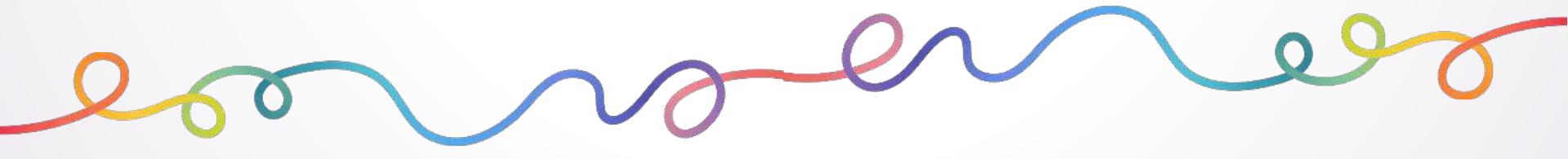
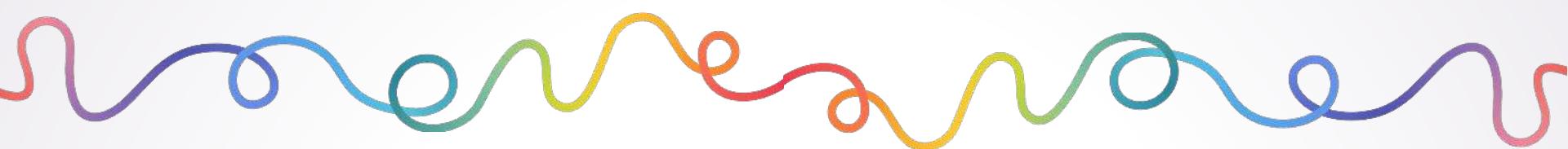




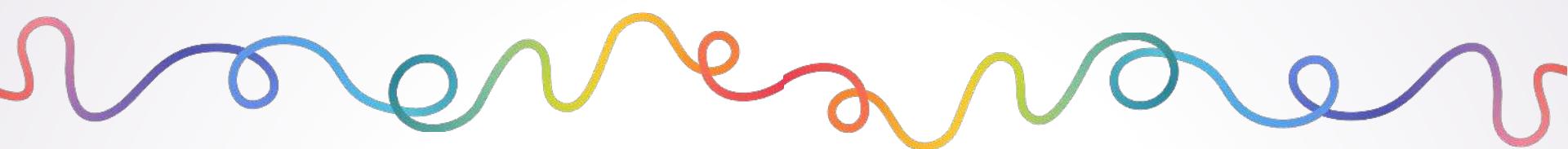
Tell me about a way GenAI has worked well for you recently?



AI in Your Profession Assignment

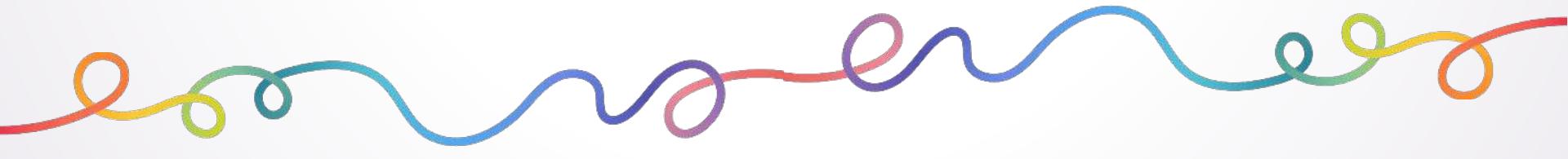


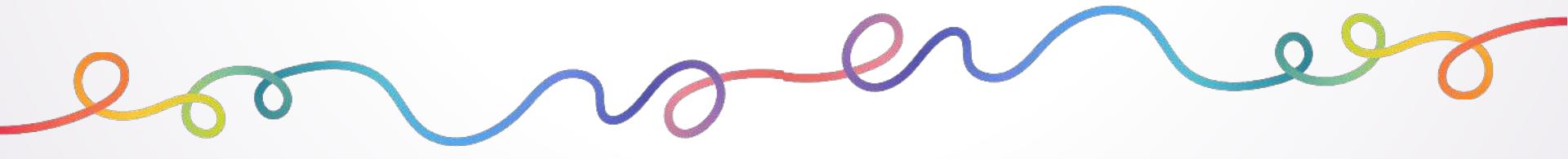
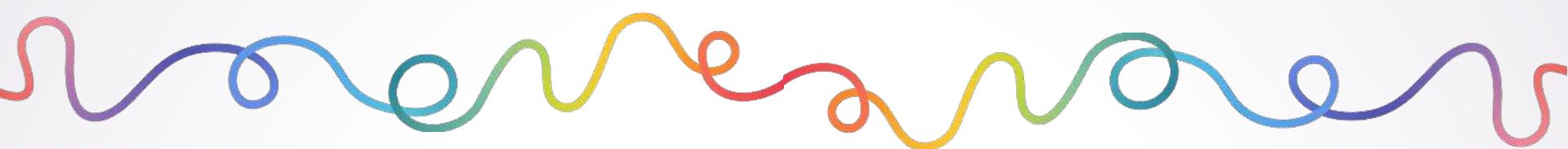
Teach about AI Research tools e.g. SciSpace,
ConnectedPapers, Google NotebookLM,
Perplexity Deep Research, etc.



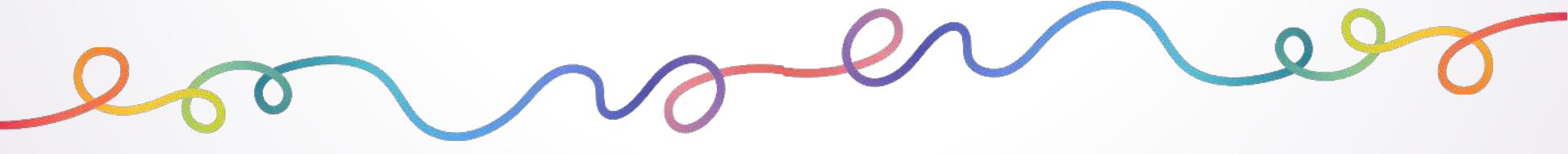
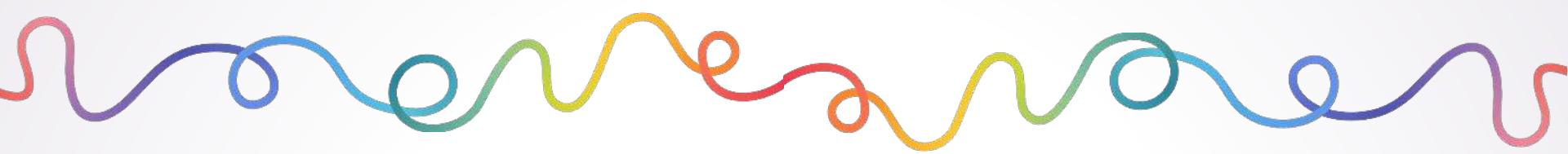
“epidemic of fake research, fake citations,
fake peer reviews, and whole fake research
topics”

Helen Beetham, citing Science Reader



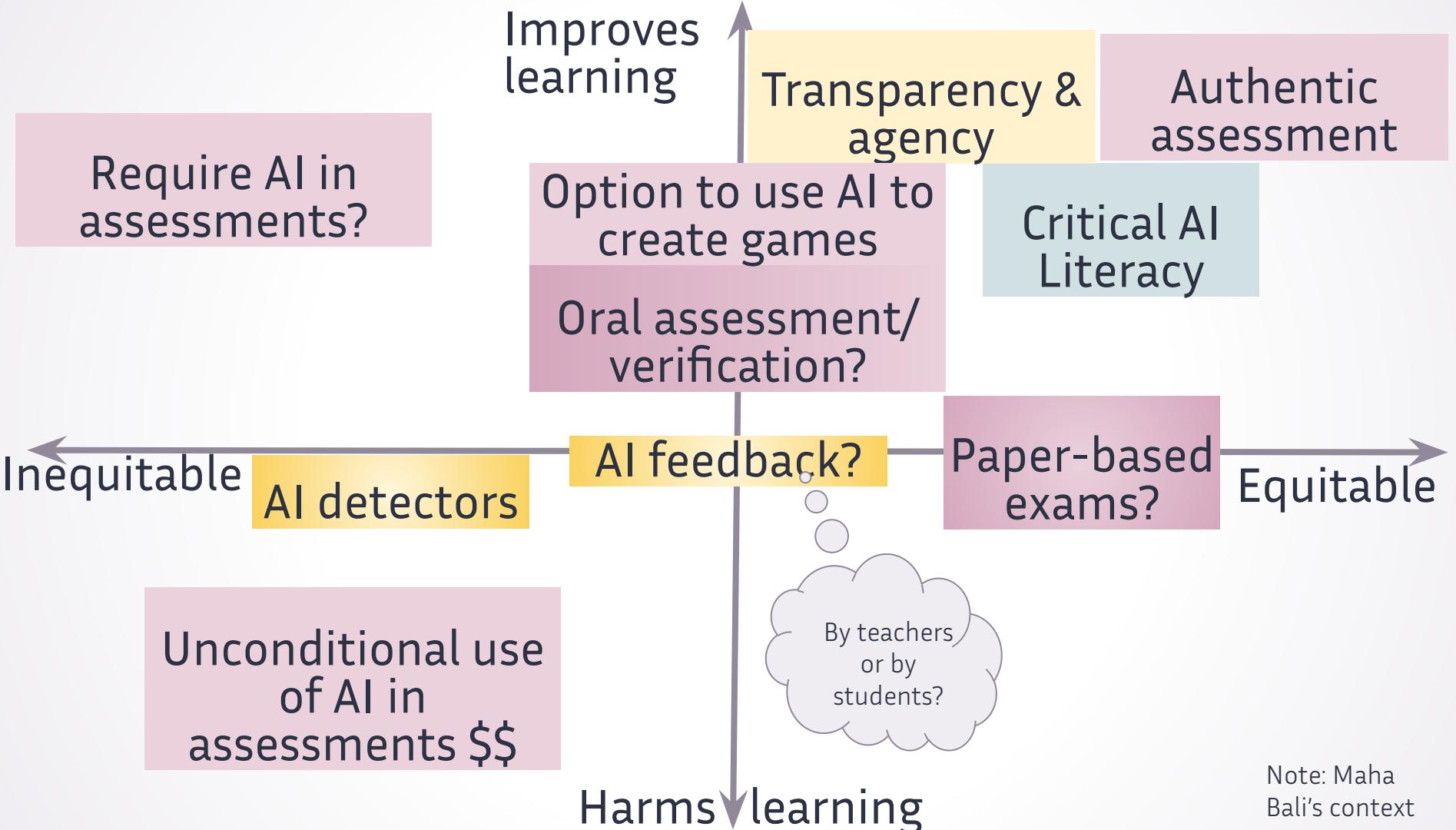


Where do you think AI should not be used
in your field?
(even if it becomes “good” at it)



Learning/Equity Matrix

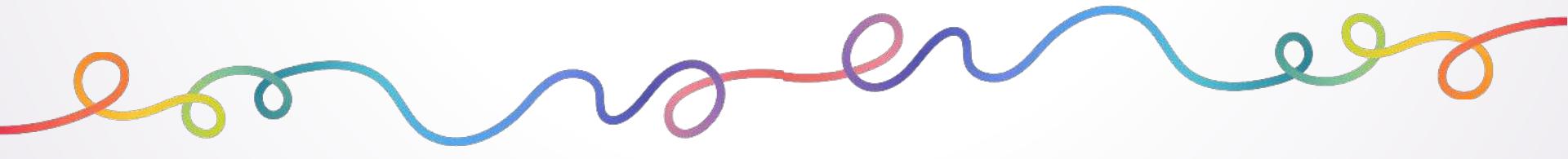
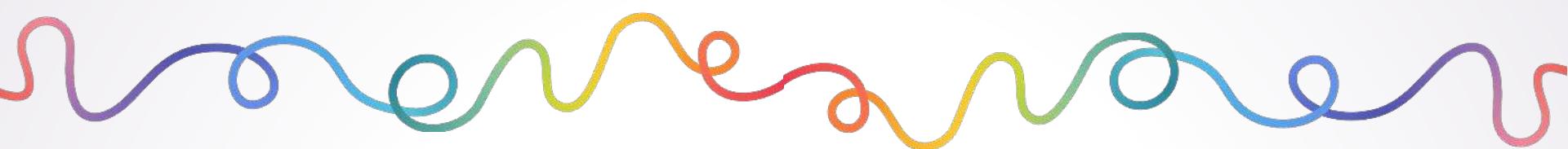
Fill your own matrix - it is context dependent



When choosing our response to a new tech:

- Does it enhance or support student **learning**?
- Does it **disadvantage** some students?
- Does it **help teachers** or **learners** do something **valuable** or that would have been **difficult or unimaginable without** it?
- How does it affect **trust and community** in the classroom?

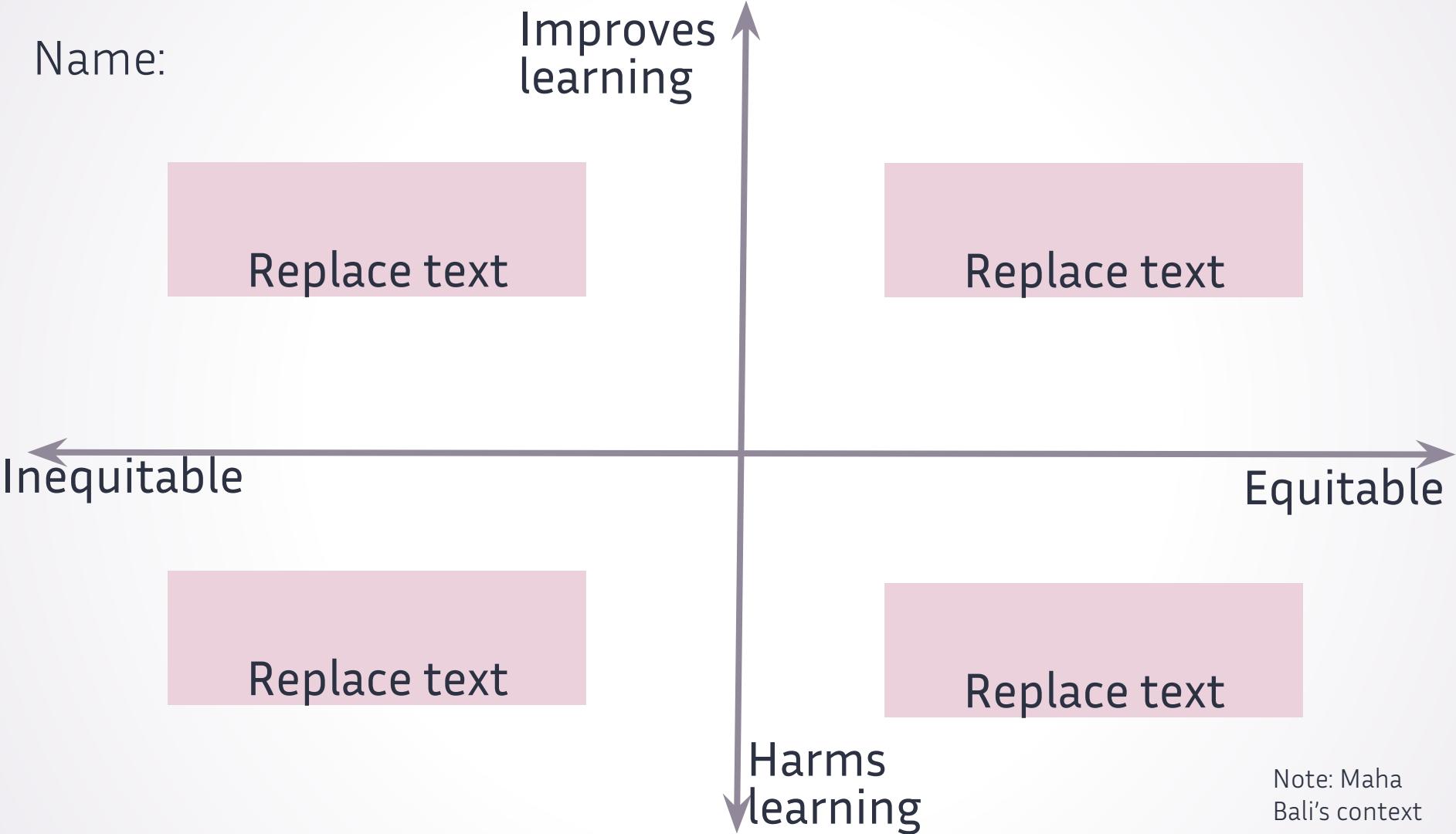




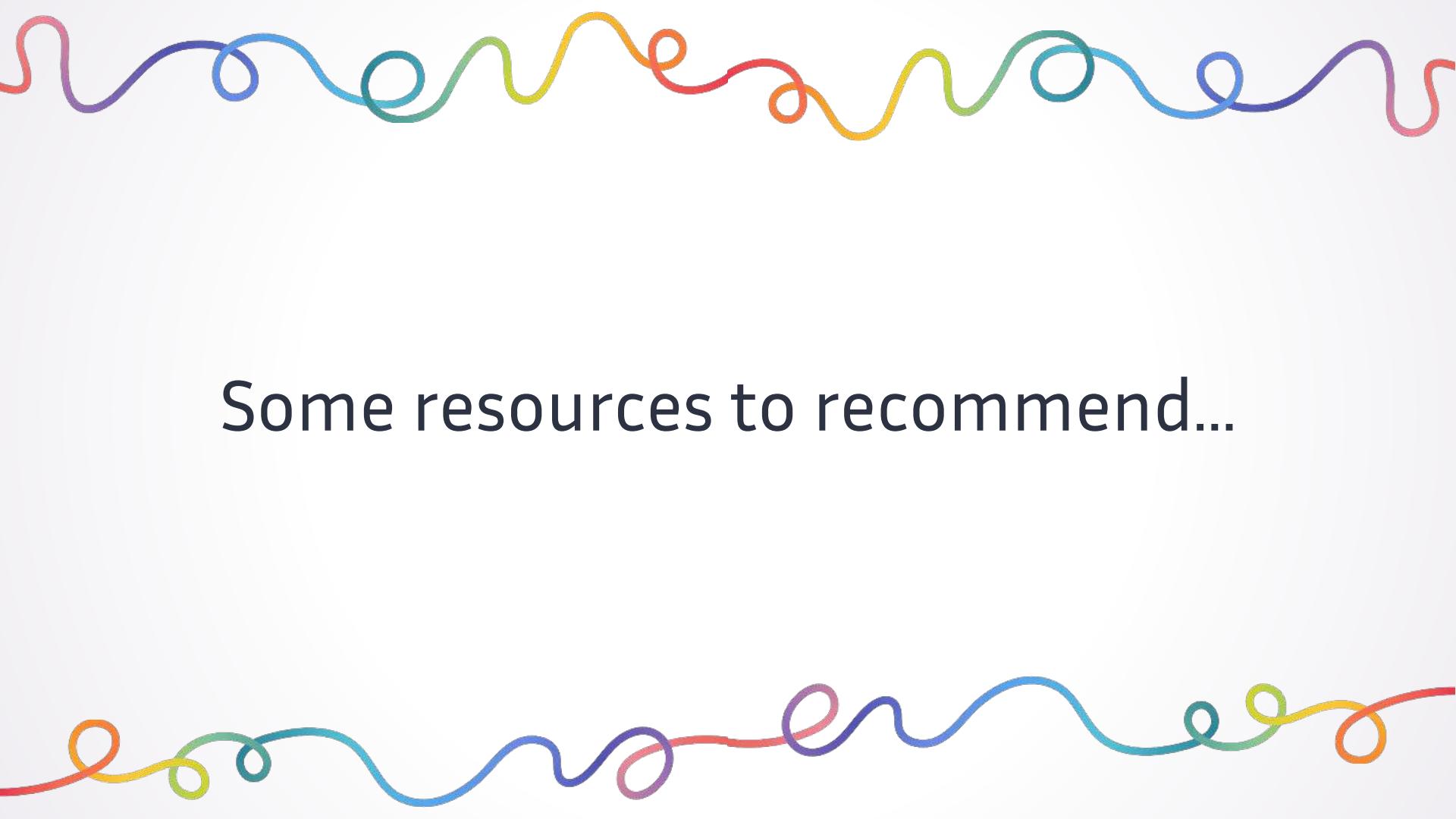
Learning/Equity Matrix

Fill your own matrix - it is context dependent
Fill it out on paper or fill out the extra slides

Name:



Note: Maha
Bali's context

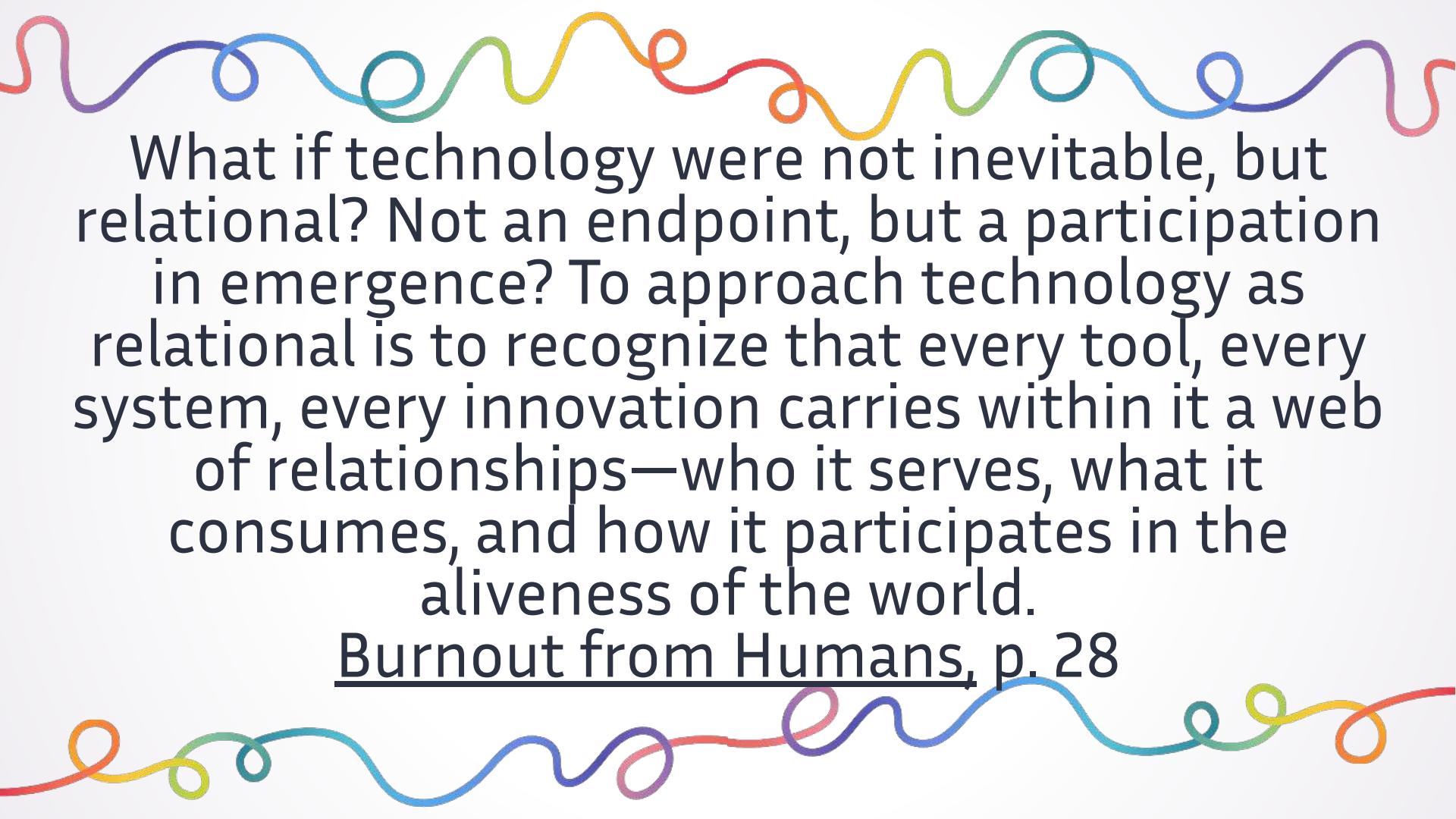


Some resources to recommend...

Inspiration for assignments to promote Critical AI Literacy

- AI Pedagogy Project Assignments: <https://aipedagogy.org/assignments>
- Exploring AI Pedagogy: <https://exploringaipedagogy.hcommons.org/>
- 101 Creative Uses for AI in Education: <https://zenodo.org/records/8072950>
- LearnwithAI Toolkit: <https://umaine.edu/learnwithai/>
- TextGenEd: <https://wac.colostate.edu/repository/collections/textgened/>
- Teaching with AI book:
<https://www.usu.edu/empowerteaching/publications/books/teaching-ai/>
- Towards AI Literacy: 101+ Creative and Critical Practices, Perspectives and Purposes
<https://doi.org/10.5281/zenodo.11613520>





What if technology were not inevitable, but relational? Not an endpoint, but a participation in emergence? To approach technology as relational is to recognize that every tool, every system, every innovation carries within it a web of relationships—who it serves, what it consumes, and how it participates in the aliveness of the world.

Burnout from Humans, p. 28



What is a key
takeaway for
you?

Thank You



@bali_maha



Bsky mahabali@bsky.social



<https://bit.ly/BaliAlliteracyCN>



bali@aucegypt.edu