CELEBRATING MEANINGFUL IMPACT 25 YEARS AND BEYOND

SMU SINGAPORE MANAGEMENT UNIVERSITY

Libraries Advancing AI Literacy in Our Communities

Samantha Seah

Librarian
User Services & Experience
SMU Libraries

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Contents

1. Perspectives on AI and AI Ethics [20 min]
2. Library as a Key Agent in AI Literacy [15 min]
3. Using and analysing AI Tools [20 min]
4. Q&A [5 min]

CELEBRATING MEANINGFUL IMPACT
25 YEARS AND BEYOND

About me

AI Literacy efforts:

- Presented "*Responsible practices for responsible libraries: The role of libraries in a world of generative AI*" at the Library Association of Singapore, Professional Sharing, May 2023
- Mentor for SMU Libraries' Hackathon 2023 advising on Alma/Primo APIs, Power Automate, library datasets, and responsible use of data
- Developed workshop "*When AI Goes Wrong*" with SMU Centre for AI and Data Governance, conducted at the start of each term since August 2023
- Presented about AI Literacy workshop "*When AI Goes Wrong*" at Chulalongkorn University's webinar series on Navigating the AI Implementation at ASEAN Libraries, April 2024
- Invited speaker at CONCERT Open Future Communication webinar series to talk about "Deconstructing AI Literacy for Librarians" in August 2024

1. Perspectives on AI and AI Ethics

Components of AI Literacy

Technical knowledge

Grasping essential concepts of machine learning, neural networks, etc. to be able to approach AI with greater confidence and understanding

Ethical Awareness

When implementing an AI tool, how sure are we that it upholds fairness, inclusivity, and transparency? What potential biases does it have? Any privacy concerns?

Critical Thinking

Leveraging IL skills to critically evaluate AI-generated information, taking into consideration the training data, the motivations of the organisations building these systems, etc.

Practical Skills

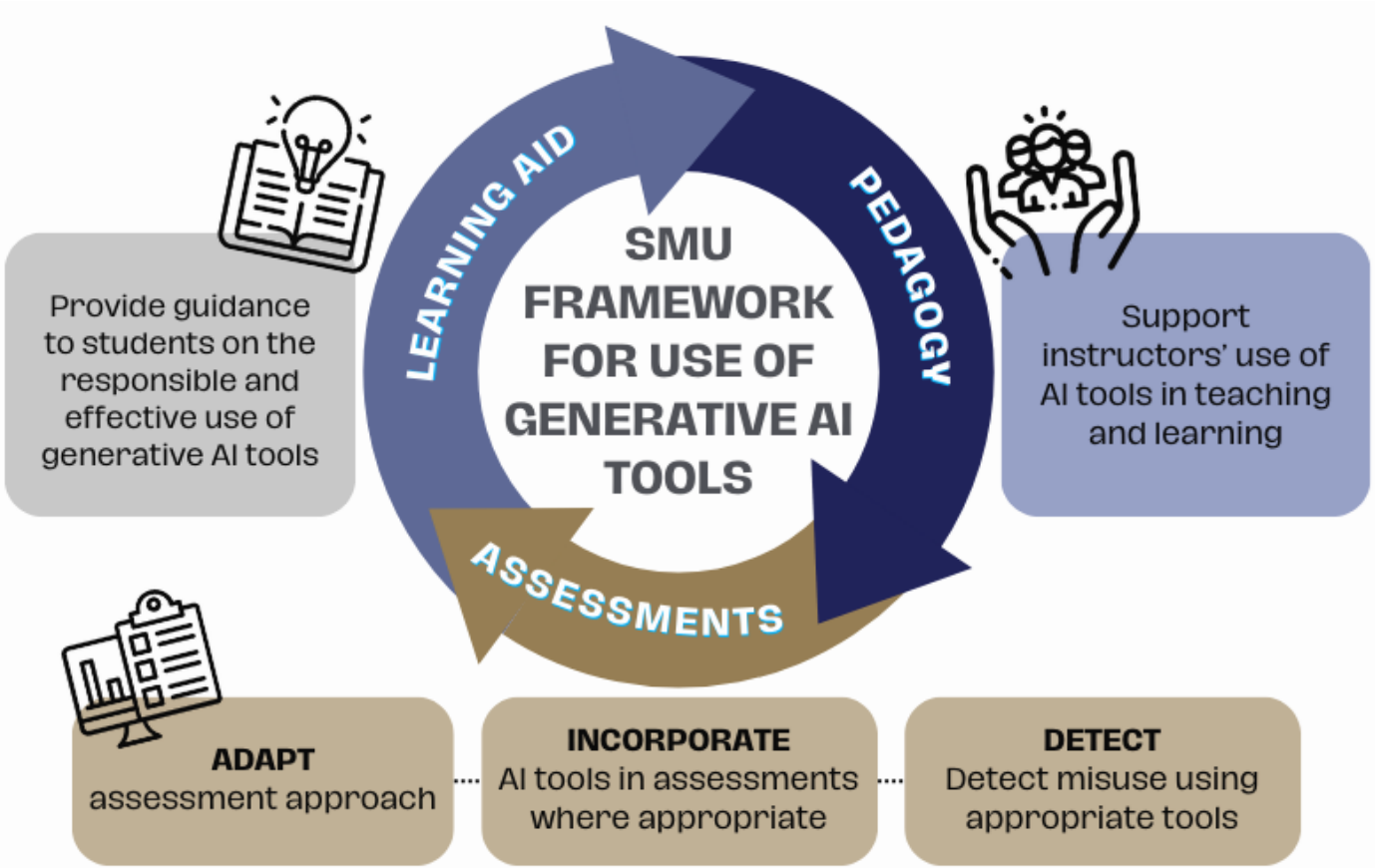
Building confidence to experiment and evaluate AI tools and using them in specific contexts

Societal Impact

How does AI reshape the world we live in – perpetuating/disrupting existing power structures, sustainability issues, etc.

Read more here

SMU Framework for Use of Generative AI Tools



SMU's position

The University recognises the benefits of generative AI tools and is committed to integrating generative AI tools into education, while safeguarding academic integrity and rigour. At the same time, the University aims to teach students to use these tools responsibly and effectively.

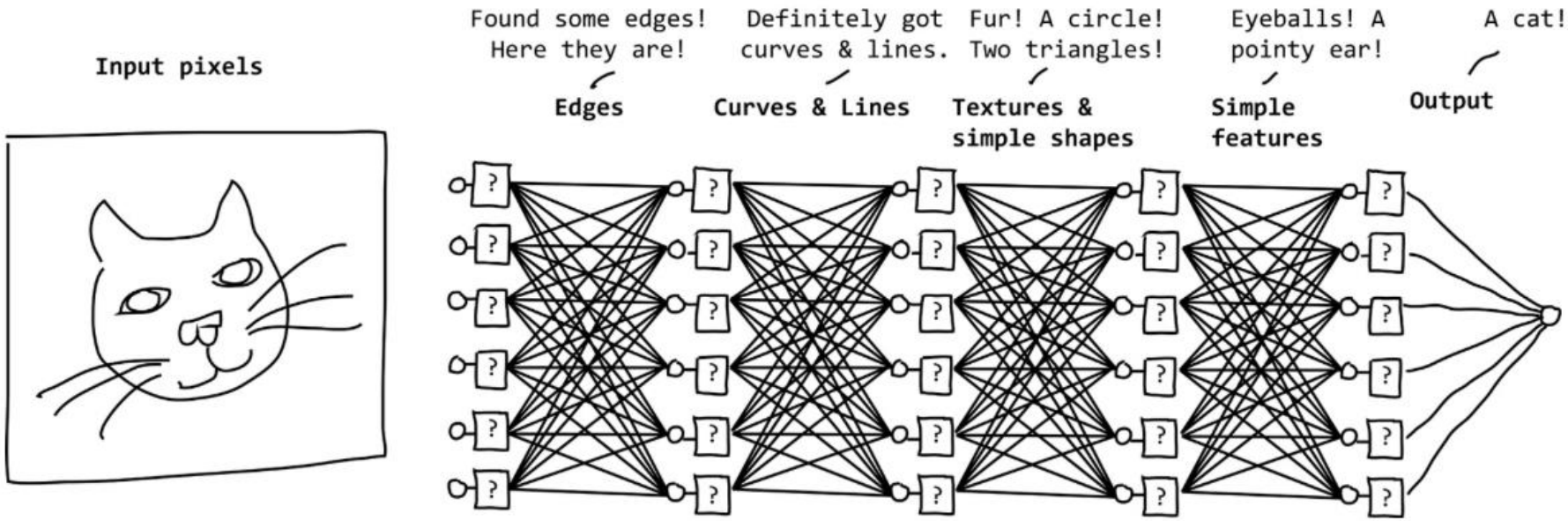
Read more here

Branches of modern AI

- Expert systems
- Predictive analytics
- Natural language processing (NLP)
- Text generation
- Machine translation
- Computer vision
- Robotics
- Speech recognition

…and more

Deep Learning



Shane (2019)

"The idea is that instead of having teams of researchers trying to find out how to find edges, you instead throw a ton of data at the algorithm and you let the data speak and have the software automatically learn from the data."

- Andrew Ng (2012),
Computer Scientist, Stanford University

Common Crawl

- Founded in 2007. Non-profit based in California, USA
- Has crawled billions of web pages since 2008 and makes data freely available via Amazon Web Services
- Currently has > 9.5 petabytes of data = 9.5×10^6 GB
- Most likely used to train the original model behind OpenAI's ChatGPT but has no affiliation with OpenAI or Microsoft

Primary goal:

- Freely make high quality web crawl data available to researchers and smaller businesses and make the playing field more equal for technology development against tech giants like Google

Language Models

Language Models:

- Machine learning model that aims to predict and generate plausible language. Autocomplete is a language model, for example.
- These models work by estimating the probability of a token or sequence of tokens occurring within a longer sequence of tokens.

Consider the following sentence:

When I hear rain on my roof, I _____ in my kitchen.

Language model determines the probabilities of different words or sequence of words that should go into the blank –

cook soup 9.4%, **warm up a kettle** 5.2%,
cower 3.6%, **nap** 2.5%, **relax** 2.2%,
etc.

Introduction to Large Language Models by Google

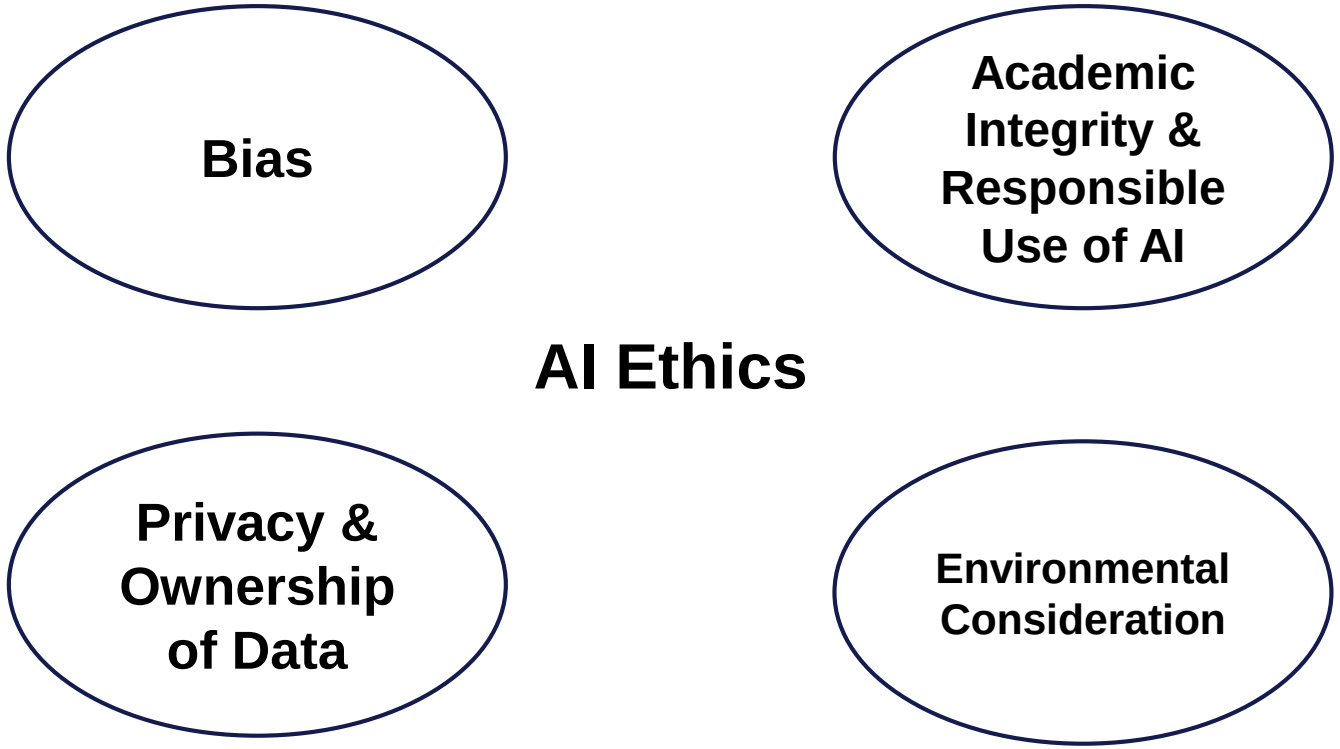
Link to video

Why AI “changes its mind”

Concerns in higher education

Impacting academic integrity, student assessment, critical thinking, and traditional teaching models, AI has ethical considerations that span higher education.

- 2025 ACRL Environmental Scan



Academic Integrity and Responsible Use of AI

How far is too far to use these AI tools in academic work/research?

If students and researchers use AI tools, are they instead losing important skills?

In allowing students to use AI tools, should they declare it? If so, how should that be done?

How students should write a thesis in the AI age

Privacy and Ownership of Data

OpenAI's ChatGPT 4 uses 13 trillion tokens (text- and code-based data) in its model

Over the past year, numerous publishers such as Wiley, HarperCollins, and Taylor & Francis have sold their content to AI companies.

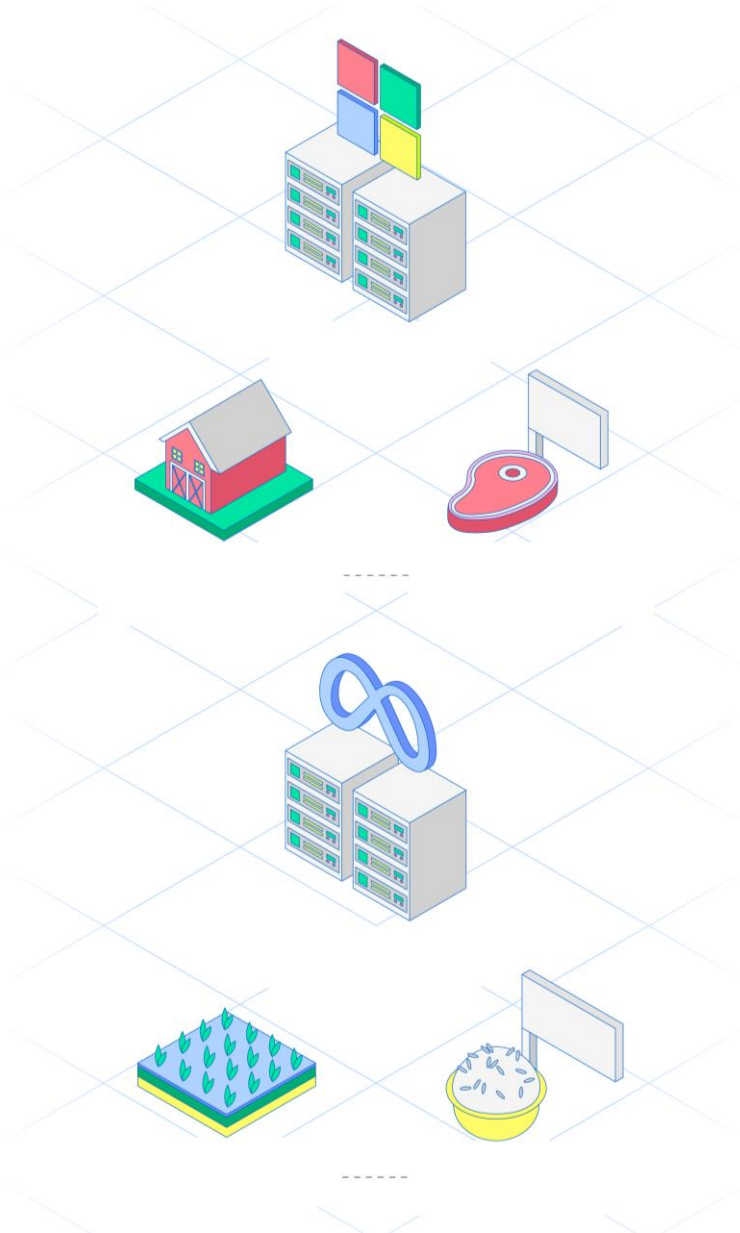
On the other side of the fence, multiple newspapers have launched lawsuits against OpenAI for harvesting content

GPT-4 Architecture, Infrastructure, Training Dataset, Costs, Vision, MoE

Environmental Consideration

- Microsoft's data centre used 700,000 litres of water training GPT-3
- ~6,992 litres of water used to produce 1 pound of beef (growing feed, cleaning farms, hydrating cows, etc.)

- Meta used 22 million litres of water training LLaMA-3
- ~4,956 litres of water used to produce 1 pound of rice



Read more:

Verma, P. & Tan, S. (2024, September 18). A bottle of water per email: the hidden environmental costs of using AI chatbots. *The Washington Post.*

Bias

Technology is never neutral. It is always imbued with the biases and experiences of its creators.

- Andrew L. Gillen (Asst. Teaching Professor in College of Engineering, Northeastern University)

What are the implications of this?

What can we do to minimise it's impact?

Opinion piece from Inside Higher Education, October 2024

Common Crawl and Generative AI

- Common Crawl provided “82% of raw tokens used to train GPT-3”
- LLM users of Common Crawl overshadow all others since the publication of OpenAI’s paper on GPT-3

Yet, Common Crawl does not represent the entire web:

- 50% of each crawl is likely to pick up URLs of previously retrieved top domains
- Crawls are not continuous, does not reflect changes between crawls
- Common Crawl respects robots.txt
- Because the infrastructure is based in the US, it skews the crawls towards English content

Using filters with Common Crawl

Four typical approaches to filtering, all automated in different ways:

1. Language filtering
2. Keywords and simple heuristics
3. AI classifiers
4. Deduplication

Examples of criticism of handling of data sets:

- If your AI classifier relies on upvotes on user-generated content (like Reddit), they cannot be representative of any population and are prone to bias
- Popular keyword lists used to filter AI risk underrepresenting vulnerable groups because it will also remove non-toxic content from LGBTQIA+ communities

Link to video

Proxies

Many concepts are difficult to define computationally

Often we come up with proxies to help us measure e.g., intelligence, fairness, “good” driver/teacher/student, health

Are these accurate measures of the concept? Are they universally accepted measures? Do they vary in definition from research to research?

Read more: *Weapons of Math Destruction* by Cathy O’Neil

Link to video

2. Library as a Key Agent in AI Literacy

Components of AI Literacy

Technical knowledge

Grasping essential concepts of machine learning, neural networks, etc. to be able to approach AI with greater confidence and understanding

Ethical Awareness

When implementing an AI tool, how sure are we that it upholds fairness, inclusivity, and transparency? What potential biases does it have? Any privacy concerns?

Critical Thinking

Leveraging IL skills to critically evaluate AI-generated information, taking into consideration the training data, the motivations of the organisations building these systems, etc.

Practical Skills

Building confidence to experiment and evaluate AI tools and using them in specific contexts

Societal Impact

How does AI reshape the world we live in – perpetuating/disrupting existing power structures, sustainability issues, etc.

Read more here

AI Literacy in and out of the classroom

Audience	Title	Mode	Type
Postgraduate	Literature Search with AI tools	Workshop	Curriculum Supported Learning
Undergraduate	Literature Search with AI tools	Workshop	Curriculum Supported Learning
Undergraduate	Effective Use of AI Tools in Research	Digital Learning Object	Curriculum Supported Learning
SMU Staff	Responsible Use of AI	Online Training	Staff Training Initiative
SMU Staff	Introduction to Generative AI	Workshop	Staff Training Initiative
SMU Community	When AI Goes Wrong	Bite-sized Workshop	Out of Classroom
SMU Community	Prompt Engineering & Practical AI tools	Bite-Sized Workshop	Out of Classroom

Ways to Search for Literature

Slide from curriculum supported learning workshop for UG/PG

Keyword Search

- Need to know keywords
- Search using keywords

Citation Chasing

- Needs seed paper(s)
- *Maaay* get messy

ResearchRabbit

Semantic Search

- Natural language

scite_

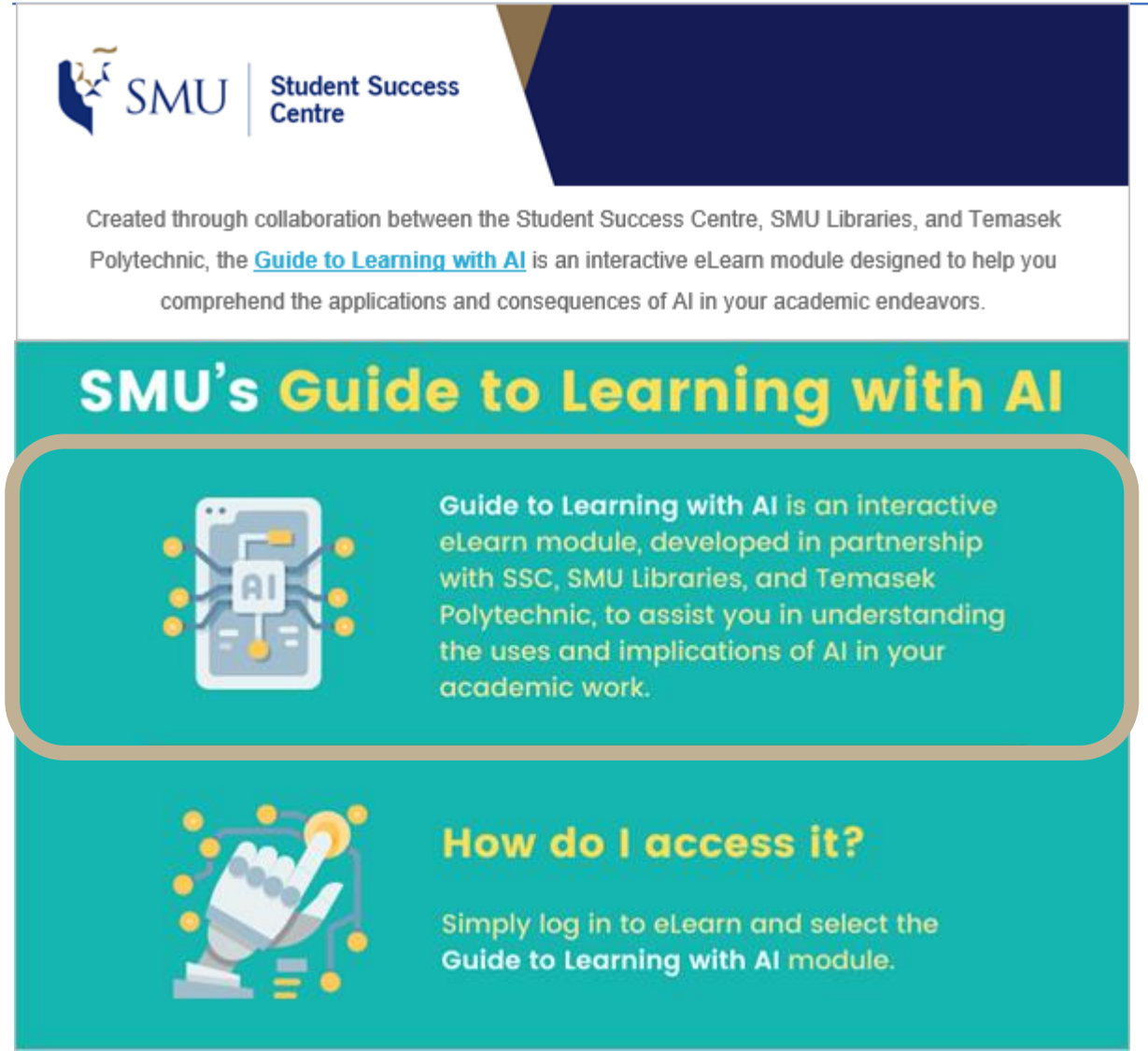
SCISPACE

Effective Use of AI Tools in Research

Screenshot from digital learning object targeted for UG

[Explore the entire DLO here]

SMU's Guide to Learning with AI



The digital learning object is now accessible SMU-wide as a 'highly recommended' content for undergraduate students.

- Partners within SMU: Student Success Centre
- Institution Partner: Temasek Polytechnic (TP)
- Target audience within SMU: 2000 UG students per year

TP Library published the Research module in POLITEMall, which is an Learning Management System used by the five polytechnics and the ITEs.

- There is interest from Joint-Poly Libraries institutions to adapt content

What is a Prompt?

It is simply an **instruction** to complete a **task**

Slide from training workshop for administrative staff

Role	Context	Task	Output
You are - an expert researcher - a workshop instructor - a project manager focusing - business consultant	[Audience] - to a 12 year old child - for a class of twenty undergraduates - **multiple stakeholders** [Scope] - within 12 months - with a budget of - within Asia	- Explain - Give me 5 examples - Give me a timeline - Give me a project plan for - Create 10 MCQs - Give me a **SWOT analysis**	- In less than 300 words - In point form - In a **table form** - With the following categories: a, b, c - With headings - In a poem / haiku - In style similar to the example provided: “Example”

ARTIFICIAL INTELLIGENCE

The AI lab waging a guerrilla war over exploitative AI

The tools Glaze and Nightshade are giving artists hope that they can fight back against AI that hoovers internet data to train. Are they enough?

By Melissa Heikkilä

November 13, 2024

DIMA KASHTALYAN

Slide from out of classroom learning for the general community

Tilt the balance

3. Using and analysing AI Tools

ROBOT

Reliability	How reliable is the information about the AI technology? If it's not produced by the party responsible for the AI, what are the author's credentials? Is there author bias? If it is produced by the party responsible for the AI, how much information are they making available? Is information only partially available due to trade secrets? How biased is the information they produce?
Objective	What is the goal or objective of the use of AI? What is the goal of sharing information about it? To inform? To convince? To find financial support?
Bias	What could create bias in the AI technology? Are there ethical issues associated with this? Are biases or ethical issues acknowledged? By the source of information? By the party responsible for the AI? By its users?
Ownership	Who is the owner or developer of the AI technology? Who is responsible for it? Is it a private company? The government? A think tank or research group? Who has access to it? Who can use it?
Type	Which subtype of AI is it? Is the technology theoretical or applied? What kind of information system does it rely on? Does it rely on human intervention?

Hervieux & Wheatley (2023)

ROBOT in practice

Let's look at a use case of AI in education

Using the ROBOT evaluation tool, what is your analysis of this technology?

May 15, 2023

Admissions Offices, Cautiously, Start Using AI

They are divided about what to do about ChatGPT, but that doesn't prevent some of them from embracing AI.

By Scott Jaschik

Taken from
Inside Higher Education

Experimenting

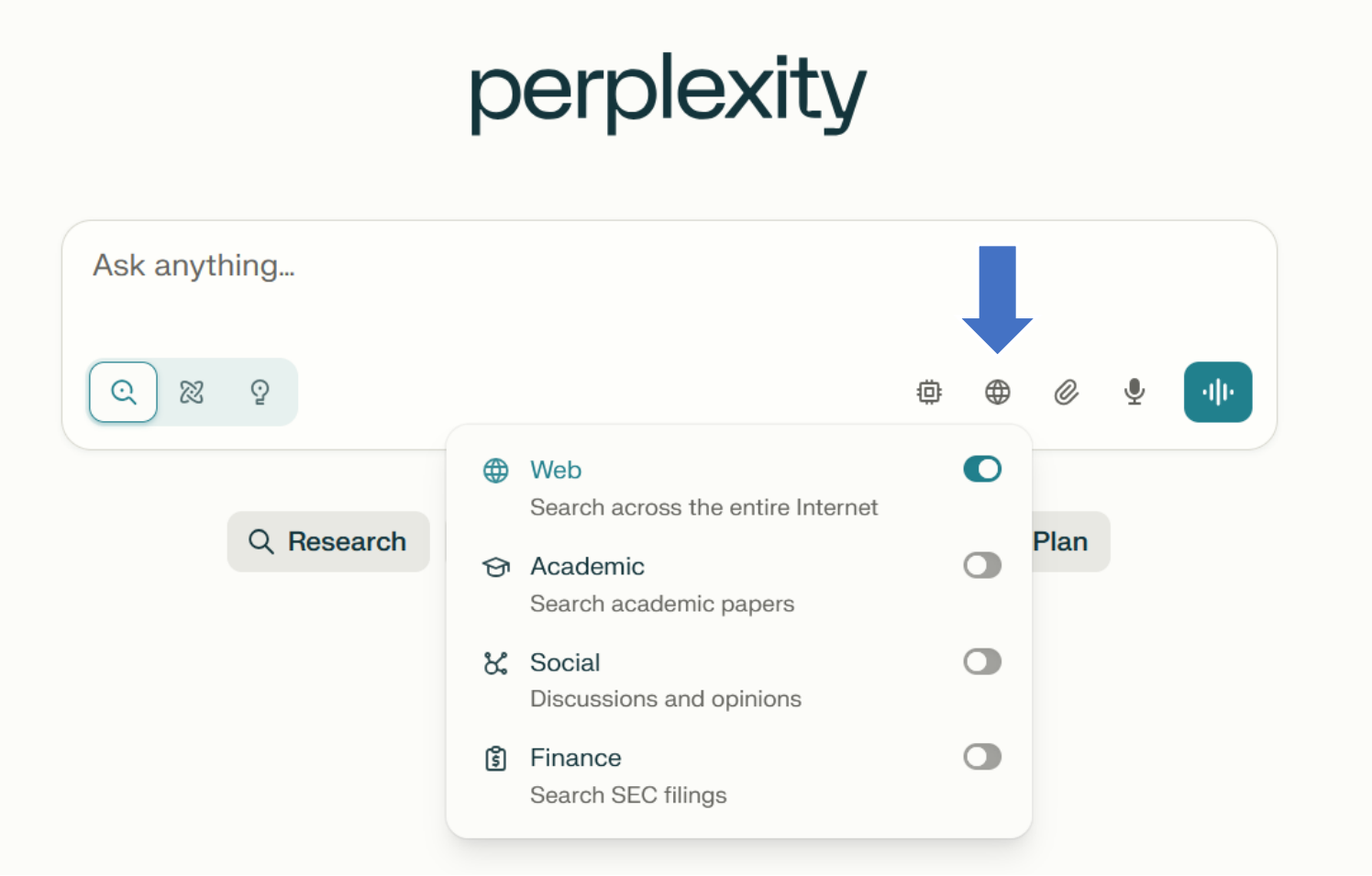
Other colleges are experimenting with AI in admissions.

Stephen W. Harmon, executive director of the Center for 21st Century Universities at the Georgia Institute of Technology, said Georgia Tech Is "not yet using AI in admissions, but we are experimenting with it."

One project is to try to replicate admissions decisions using machine learning techniques. "We are using one of our large online master of science programs as a test case, and we are currently at about a 93 percent match with our admissions advisers' decisions," Harmon said.

What is the appeal? "The volume of applicants continues to increase at Georgia Tech, so anything we can do to improve their workflow is helpful. I don't see us relying solely on AI for admissions maybe ever, but it could become a useful tool in the process, " he said.

Hands on: Perplexity



Sample output

compare library services available in south east asia

Answer · Sources · 8 · Steps

biblioasia.nlb.gov.sg
The Singapore and Southeast Asia Collection ...

bpasjournals
[PDF] A Comparative Analysis of IIT Libraries an...

ifla
[PDF] A Force for Social Development! Ngian Lek...

National Library Board
National Reference Library - Singapore - NLB

Comparison of Library Services in Southeast Asia

Library services in Southeast Asia vary across countries but share common trends in digital transformation, outreach, and inclusivity. Below is a comparative overview of key services and features offered by national and public libraries in the region.

Digital Resources and Access

- **Singapore**: The National Library Board (NLB) offers extensive digital collections, including the Singapore and Southeast Asia Collection, which features digitized manuscripts, maps, and archival materials in multiple languages. NLB is also part of international digital library consortia and collaborates with overseas institutions for resource sharing and digitization projects [1].
- **Malaysia**: The National Library of Malaysia has launched the Ubiquitous Library Portal, aiming to provide digital access to library materials for all residents, including those in rural areas. The Sabah State Library offers a large e-resources collection accessible through various online portals and mobile apps [3] [5].

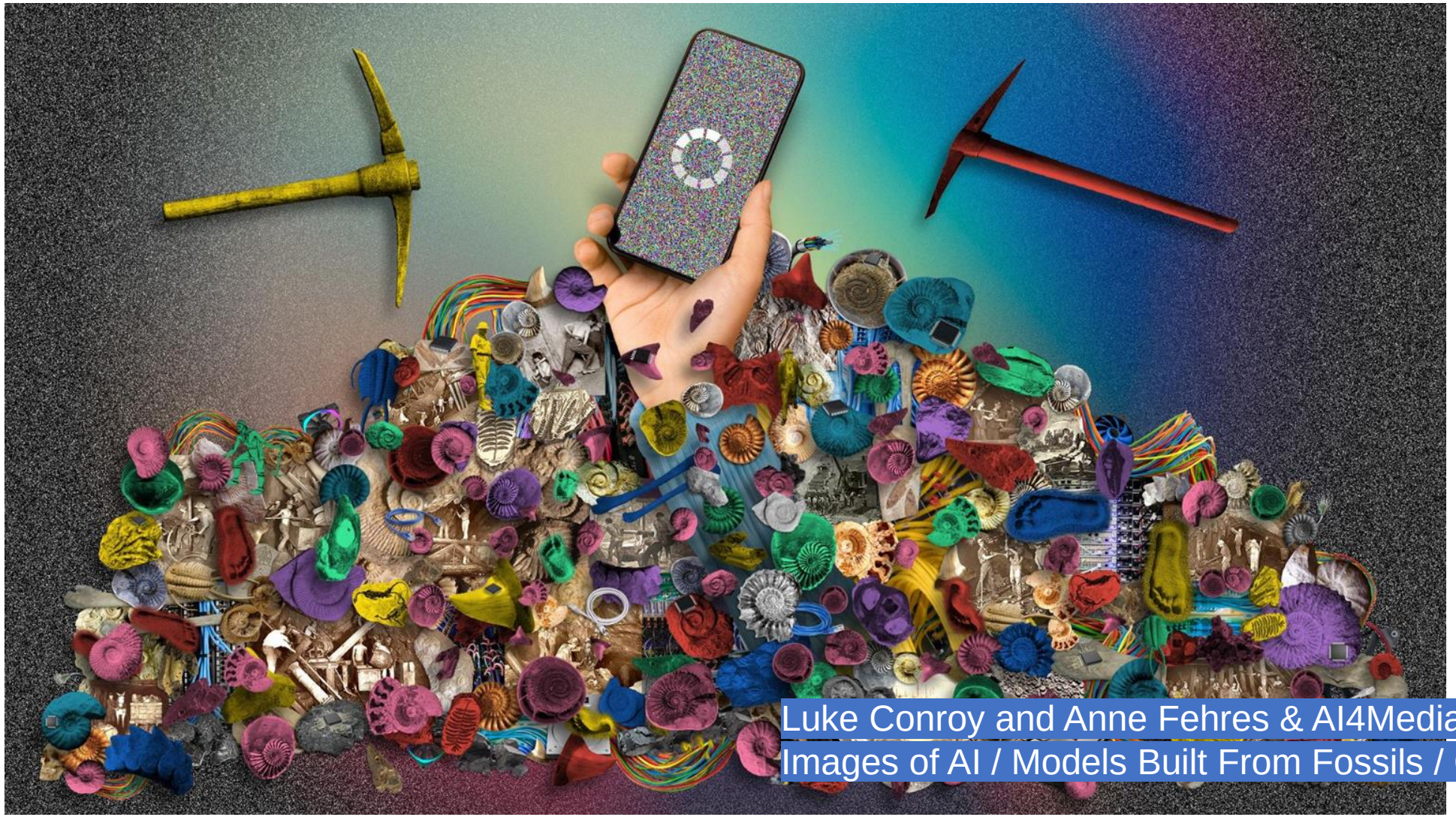
ROBOT

Reliability	How reliable is the information about the AI technology? If it's not produced by the party responsible for the AI, what are the author's credentials? Is there author bias? If it is produced by the party responsible for the AI, how much information are they making available? Is information only partially available due to trade secrets? How biased is the information they produce?
Objective	What is the goal or objective of the use of AI? What is the goal of sharing information about it? To inform? To convince? To find financial support?
Bias	What could create bias in the AI technology? Are there ethical issues associated with this? Are biases or ethical issues acknowledged? By the source of information? By the party responsible for the AI? By its users?
Ownership	Who is the owner or developer of the AI technology? Who is responsible for it? Is it a private company? The government? A think tank or research group? Who has access to it? Who can use it?
Type	Which subtype of AI is it? Is the technology theoretical or applied? What kind of information system does it rely on? Does it rely on human intervention?

Hervieux & Wheatley (2023)

Food for thought

Better images of AI



Luke Conroy and Anne Fehres & AI4Media / Better Images of AI / Models Built From Fossils / CC-BY 4.0

Thank You!

Please scan to provide feedback on this session.

Read more

- Resources for Understanding the Ethical Implications of Artificial Intelligence (AI)
- AI Weirdness, a blog by Janelle Shane on her experiments with AI
- Online free course in Practical Data Ethics: https://ethics.fast.ai/
- Models all the way down: https://knowingmachines.org/models-all-the-way
- Better Images of AI: https://betterimagesofai.org/

- AI Literacy Framework for Primary and Secondary Education: https://ailiteracyframework.org/
- Dimensions of AI Literacies: https://openedculture.org/projects/dimensions-of-ai-literacies/
- AI Literacy: A Guide for Academic Libraries (by ACRL): https://crln.acrl.org/index.php/crlnews/article/view/26704/34626
- The Environmental Cost of AI | Climate Crisis: https://www.libraryjournal.com/story/the-environmental-cost-of-ai-climate-crisis

Read more

Raji, I. D., Kumar, I. E., Horowitz, A., & Selbst, A. (2022, June). The fallacy of AI functionality. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 959-972).

Scheuerman, M. K., Hanna, A., & Denton, E. (2021). Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1-37.

References

Baack, S. & Mozilla Insights. (2024). Training Data for the Price of Sandwich. Mozilla. https://foundation.mozilla.org/en/research/library/generative-ai-training-data/common-crawl/

Heikkilä, M. (2024, November 13). *The AI lab waging a guerrilla war over exploitative AI.* MIT Technology Review. https://www.technologyreview.com/2024/11/13/1106837/ai-data-posioning-nightshade-glaze-art-university-of-chicago-exploitation/

Hervieux, S. & Wheatley, A. (2023, February 22). *Creating an Academic Library Workshop Series on AI Literacy*. LibTech Insights. https://www.choice360.org/libtech-insight/creating-an-academic-library-workshop-series-on-ai-literacy/

Jaschik, S. (2023, May 15). Admissions Offices, Cautiously, Start Using AI. *Inside Higher Education*. https://www.insidehighered.com/news/admissions/2023/05/15/admissions-offices-cautiously-start-using-ai

Shane, J. (2019). *You look like a thing and I love you*. Hachette UK.

Verma, P. & Tan, S. (2024, September 18). A bottle of water per email: the hidden environmental costs of using AI chatbots. *The Washington Post.* https://www.washingtonpost.com/technology/2024/09/18/energy-ai-use-electricity-water-data-centers/