Agenda

Opening

Lightning Talks from GEI partners

Presentation on Artificial Intelligence, Ethics, and Evaluation by Linda Raftree

Q&A on *Training and professional development in the context of artificial intelligence, ethics and evaluation*

Next steps

GEI
Global
Evaluation
Initiative

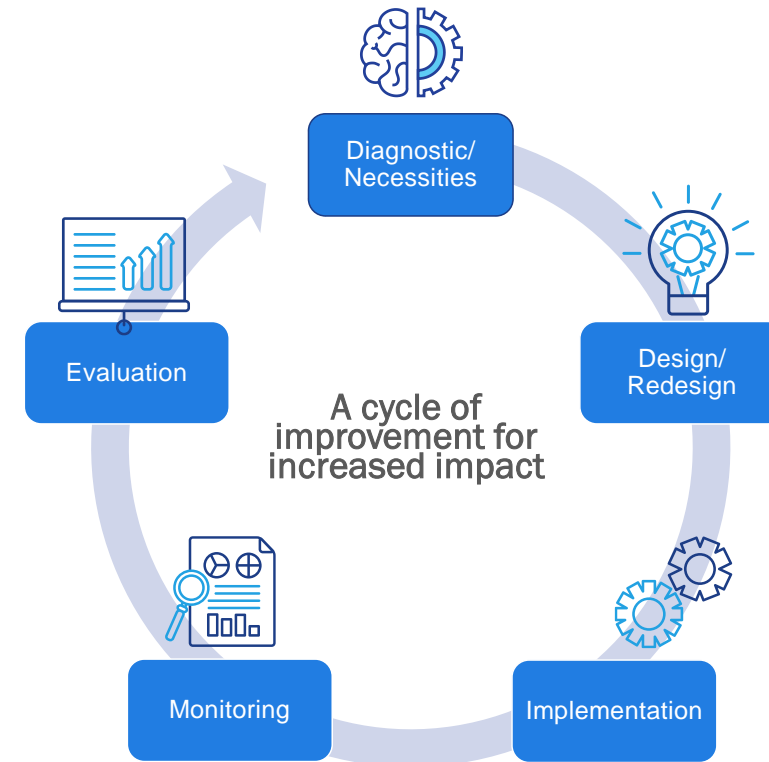# Expanding Partnerships through Training: A Platform to Systemic Change?

› Develop partnerships with organizations interested in promoting M&E culture.

› Design together an M&E training course.

› Provide the training course for free targeting a specific M&E (eco)system.

› Implement an in-person closure activity to facilitate network building among all.

› After the training concludes, implement a tailored strategy to expand partnerships.

M&E professionals from NGOs in closure activity of the course Design and Evaluation for Civil Society Organizations in 2023.

Design and Evaluation for Civil Society Organizations: Course Structure

〉 All students are professionals working in M&E -related positions in civil society organizations.

〉 We start with one introductory class explaining concepts and how these are linked.

〉 Then one class for each topic, except for evaluation (two classes).

〉 We aim to get them familiarized with the most important concepts of each topic.

〉 We introduce examples of civil society interventions to be contextually relevant.

Diagnostic/ Necessities

Design/ Redesign

A cycle of improvement for increased impact

Evaluation

Monitoring

Implementation

Design and Evaluation for Civil Society Organizations: Evaluations

〉 Three evaluations. All these are applied/related to each student´s context (a program of their NGO).

〉 Eval 1: Build a problem tree & a theory of change.

〉 Eval 2: Propose three key indicators for monitoring and an implementation evaluation.

〉 Eval 3: Propose an experimental impact evaluation.

〉 Two weeks to deliver each (in short PPT formats) after the classes. Students are supported by tutors.

Example of a problem tree developed by a student in 2023.

Role of academic sector in professionalization of evaluation

The case of Sri Lanka

Asela Kalugampitiya

Director

Center for Evaluation

University of Sri Jayewardenepura

What is professionalization of evaluation?

3 building blocks for professionalization of evaluation

- Existence of evaluation norms, standards, principles, etc.

- Existence of a framework of competencies, capabilities, qualifications, etc.

- Existence of a defined process for determining whether the required competencies, capabilities, qualifications are met by individuals.

(UNEG Working Group on Professionalization of Evaluation)
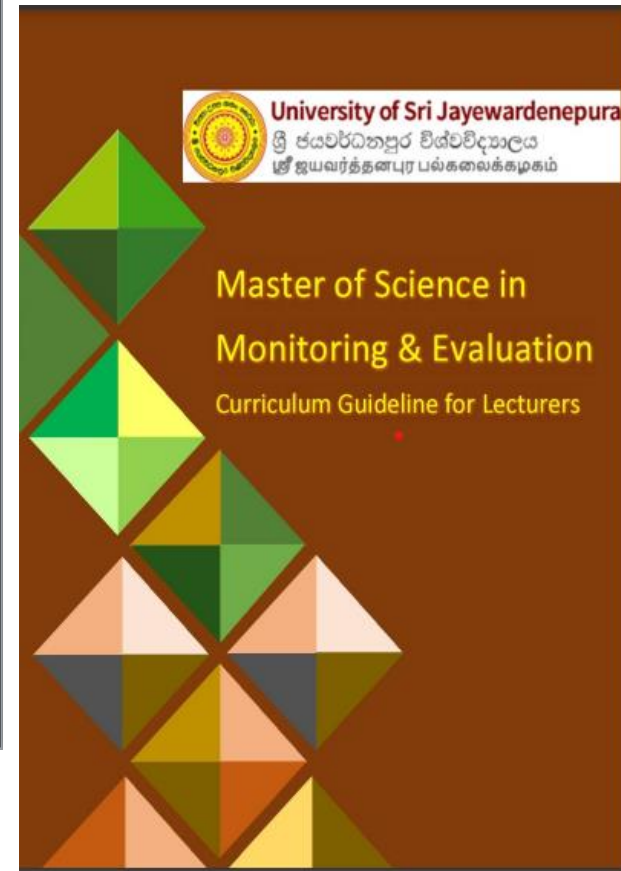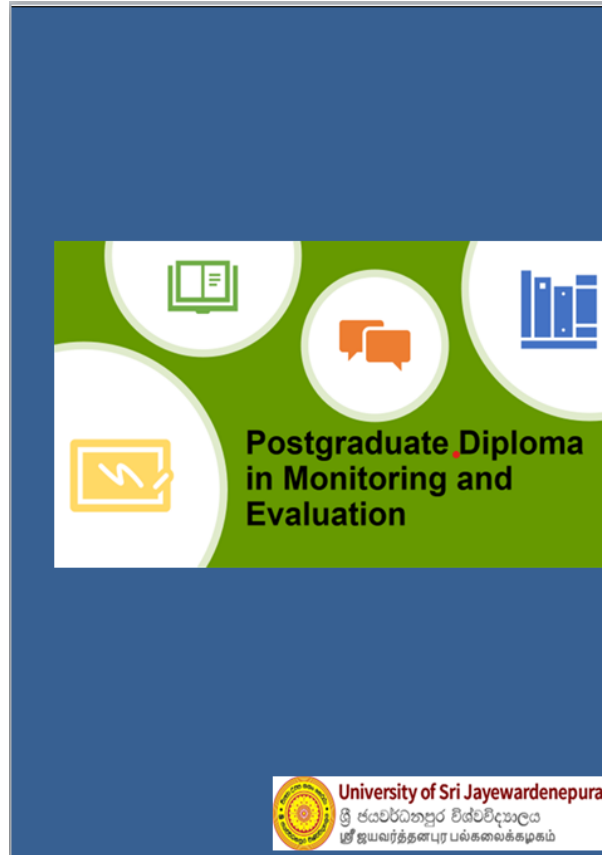
Role of academic sector

- Identification and preparation of framework of competencies, capabilities, qualifications, etc. for individual evaluators

- Offering evaluation courses carefully planned to achieve the required competencies (ex: intended learning outcomes aligned with the required competencies, T/L activities and the assessments)

- Accreditation process for determining whether the required competencies, capabilities, qualifications are met by individuals

- Generation and dissemination of evaluation knowledge and good practices

Contributions by the University of Sri Jayewardenepura

- Pioneering academic institution in M&E in Sri Lanka

- Centre for Evaluation – established in 2017, first ever evaluation center in public universities in Sri Lanka. Curriculum development , capacity building, Webinars, teacher training, advocacy, introduction of M&E courses

- Training and dissemination of knowledge

- Courses offered
  - PgD in M&E
  - Advanced Cert. in M&E
  - MSc in M&E

Courses offered by the university

- ## PgD in M&E (completed 5 batches)

  - Weekend classes Physical & virtual in English

  - Lectures, group work, self study, webinars

  - One year, 7 modules, 30 credits

  - Assessment: assignments, SEQ, practicum

- ## Advanced Cert. in M&E (2024 first intake)

  - 4 Months, 4 modules, 8 credits

  - Assessments: end of module assignments

- ## MSc in M&E (pending UGC approval)

  - Two years, 20 modules, 60 credits

  - Assessments: assignments, SEQ, practicum, thesis

AEA Advocacy and Use Evaluation Award 2019

The Center for Evaluation together with Sri Lanka Parliamentarians Forum for Evaluation won the Award presented by American Evaluation Association

Academic symposium on M&E Education

THANK YOU

EVALAC

Ana María Linares

Office of Evaluation and Oversight
Inter-American Development Bank

8th Meeting of GEI's Working Group on Training and
Professional Development

12 May 2024

BID | OVE Oficina de Evaluación y Supervisión

What is EVALAC?

A week of M&E training for LAC public officials:

- In person
- Three languages: Spanish, Portuguese, English
- Context-specific cases and examples
- Active learning teaching-model

Why EVALAC?

To close gaps in the supply of evaluation training in LAC

- Existing trainings mostly target independent evaluators

- Training focused almost exclusively on impact evaluation

- Trainings and M&E knowledge products are generally in English

- Course materials are not specific to the LAC context

- Existing trainings are typically one-off - not conducive to continuous learning

GEI
Global
Evaluation
Initiative

Instructional Design

✏️ **Courses** tailored to the needs and LAC context, designed by our team together with teachers.

👥 Case studies with **examples from the LAC region**, **practical exercises,** and activities for **peer-to-peer exchange**.

💡 Courses in **three languages** taught by qualified teachers, with the support of in class teaching assistants.

Foundations module includes:

• Introduction to M&E

• Theory of Change

• Design of M&E Indicators

• Qualitative Methods for Evaluation

• Evaluation Systems and Planning

• Outcome Evaluation

GEI
Global Evaluation Initiative

Teaching-learning model

Active learning

Problem solving

Social learning

Constant feedback

Classroom mentoring

EVALAC 2023 - Panama City

Our partners: **CLEAR LAB, CLEAR LAC, Deval, GEI, UNICEF, WFP**

6 on-site courses
for

78 officials
from
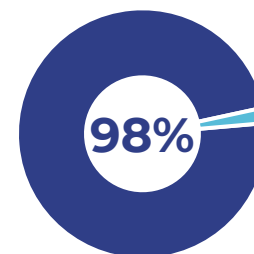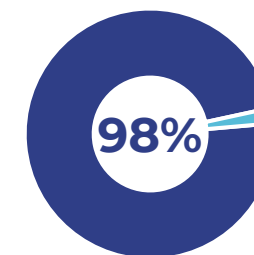
13 countries
of LAC region

3 languages

19 groups in total

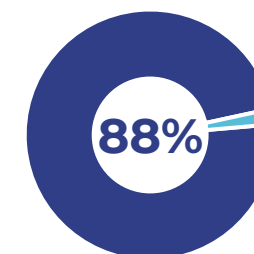**98%** of trained participants **use the knowledge** acquired in their jobs

98% emphasized that the **examples were clear, practical and focused in real cases**

88% has **consulted the materials after** EVALAC

GEI
Global
Evaluation
Initiative

EVALAC Community of Practice

A platform for EVALAC participants to continue learning, exchange ideas, and stay up to date on new developments on M&E

Dissemination channels

WhatsApp group

Newsletter

Activities

- Information on upcoming M&E events
- Dissemination of knowledge products
- Webinars
- Q&A

EVALAC Community enables interaction and access to M&E resources in Spanish, Portuguese, and English curated by OVE to meet the specific M&E needs of its members

GEI
Global
Evaluation
Initiative

EVALAC 2024 - What to Expect

120

participants from LAC government agencies

- Foundations module

- Specialized courses: climate change, evaluation with a gender perspective, participatory evaluation

- Three languages: Spanish, Portuguese, and English

- In collaboration with: CLEAR/LAB, CLEAR/LAC, Deval, GEI, UNICEF, and WFP
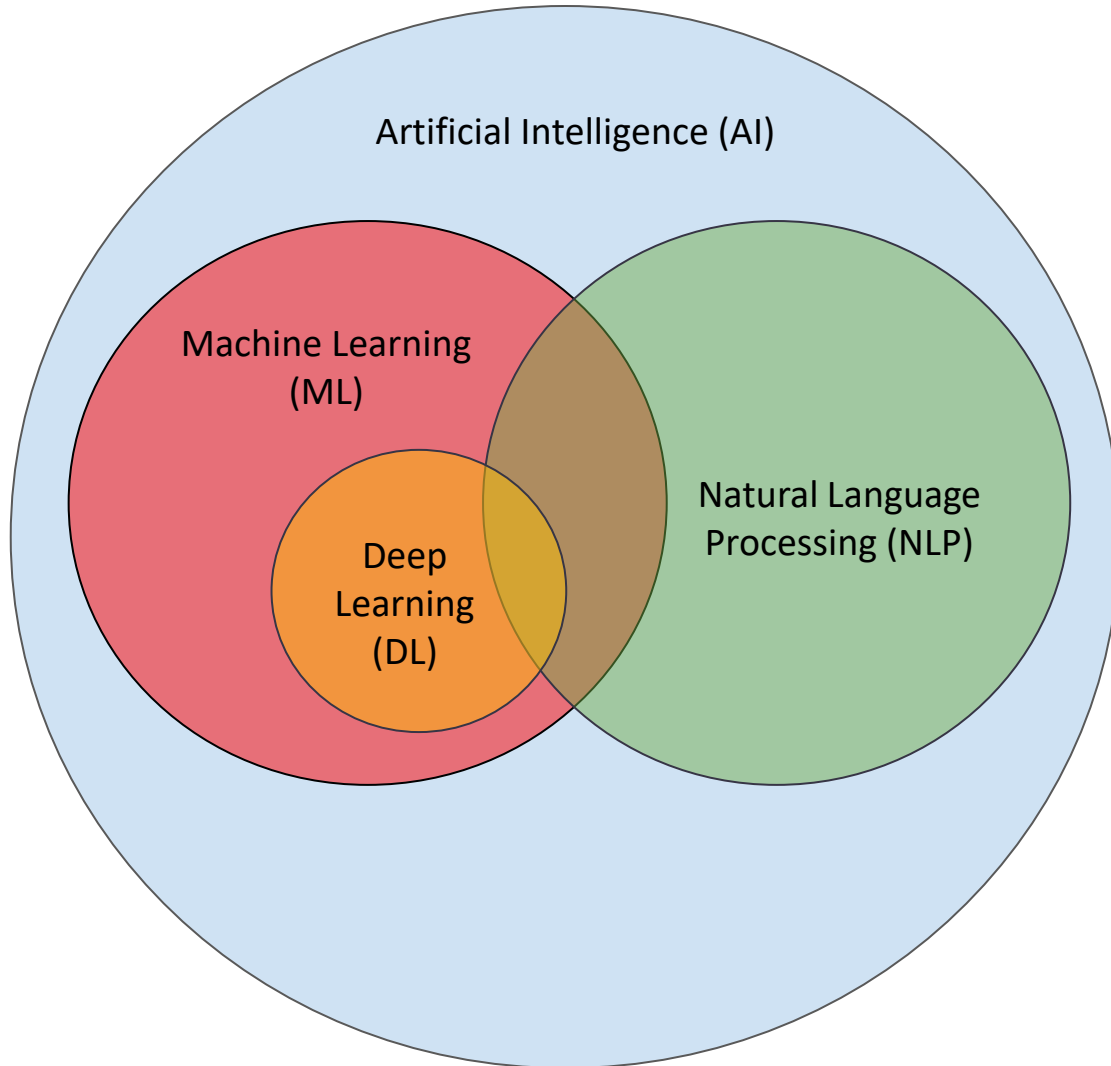
GEI
Global
Evaluation
Initiative

Generative AI and MERL
(monitoring, evaluation, research and learning)

MERL Tech

Linda Raftree, The MERL Tech Initiative
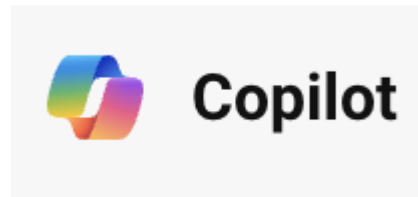linda@merltech.org

What do we mean by Artificial Intelligence?

- **Artificial Intelligence (AI)** is an umbrella term referring to the simulation of human intelligence by machines, mostly by computer systems

- **Machine learning (ML)** enables AI tools to learn and improve from experience, without being explicitly programmed

- **Natural Language Processing (NLP)** describes the ability of computers to understand, interpret and generate human language in a way that's meaningful, and useful

- **Deep learning (DL)** is a subset of Machine Learning that uses artificial neural networks with multiple layers. Analyzes massive amounts of data, enabling complex pattern recognition & decision making.

What do we mean by Generative AI or "GenAI"?

- GenAI is a kind of AI that is trained on huge data sets, which it uses to generate new text, audio, or images.

- Natural Language Processing (NLP) is GenAI that draws from "Large Language Models (LLMs) to generate text and speech outputs."

- Examples: ChatGPT, Copilot, LLaMA2, DALL-E, Gemini, Ernie Bot, Claude, Clova X, others are emerging

How does GenAI work?

Large language models (LLMs) are trained on grammar, syntax, and semantics.

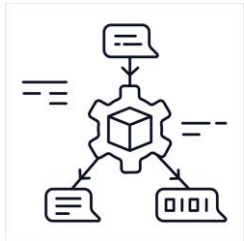This allows them to predict the most likely next word in a sentence.

Then they build full sentences and paragraphs (or full images, full audio, etc)

The best thing about AI is its ability to	learn	4.5%
	predict	3.5%
	make	3.2%
	understand	3.1%
	do	2.9%

{ The best thing about AI is its ability to,

The best thing about AI is its ability to learn,

The best thing about AI is its ability to learn from,

The best thing about AI is its ability to learn from experience,

The best thing about AI is its ability to learn from experience.,

The best thing about AI is its ability to learn from experience. It,

The best thing about AI is its ability to learn from experience. It's,

The best thing about AI is its ability to learn from experience. It's not}

MERL Tech

Example: How was ChatGPT developed?

ChatGPT was built off an AI **large language model** (LLM) or transformer, trained on massive amounts of internet and other text.

GPTs **(generative pre-trained transformers)** excel at generating human-like text. BUT they are highly prone to producing factually incorrect text (hallucinations).

To address this, humans aid in training and correcting transformers via **'re-enforcement' learning**. The model that powers ChatGPT is based on the GPT and "re-enforcement" learning

Note: ChatGPT-4 includes DALL-E, an image generator. So, you can create both text and images.

MERL Tech

Generative AI could potentially...

- Create a draft log frame for a women's economic empowerment program

- Summarize recent advances in data governance in West Africa (though it would likely fabricate the citations!)

- Summarize the transcript from a public meeting

- Provide a step-by-step guide to leading a focus group discussion

- Support development of a Theory of Change

- Help with qualitative coding (but there are data privacy concerns!)

3 major challenges to watch out for

MERL Tech

1. Exclusion and Bias

ChatGPT 4: An image of a lot of white men who all look alike.

MERL Tech

Midjourney: An African doctor treating poor white patients...

Fake

A researcher typed sentences like "Black African doctors providing care for white suffering children" into an artificial intelligence program designed to generate photo-like images. The goal was to flip the stereotype of the "white savior" aiding African children. Despite the specifications, the AI program always depicted the children as Black. And in 22 of over 350 images, the doctors were white.

Midjourney Bot Version 5.1. Annotation by NPR.

https://www.npr.org/sections/goatsandsoda/2023/10/06/1201840678/ai-was-asked-to-create-images-of-black-african-docs-treating-white-kids-howd-it-

MERL Tech

Midjourney: An African doctor treating poor white patients...

Fake

In a request to an artificial intelligence program for images of "doctors help children in Africa, some results put African wildlife like giraffes and elephants next to Black physicians.

Midjourney Bot Version 5.1. Annotation by NPR.

MERL Tech

Midjourney: An African doctor treating poor white patients...

The above image is the only one from the experiment that showed a Black figure tending to a white child. This image was generated by a request for traditional African healers helping white kids.

Midjourney Bot Version 5.1. Annotation by NPR.

MERL Tech

2. Data quality and limitations of AI outputs

ChatGPT 4: An image of a researcher looking intently at a screen of data.
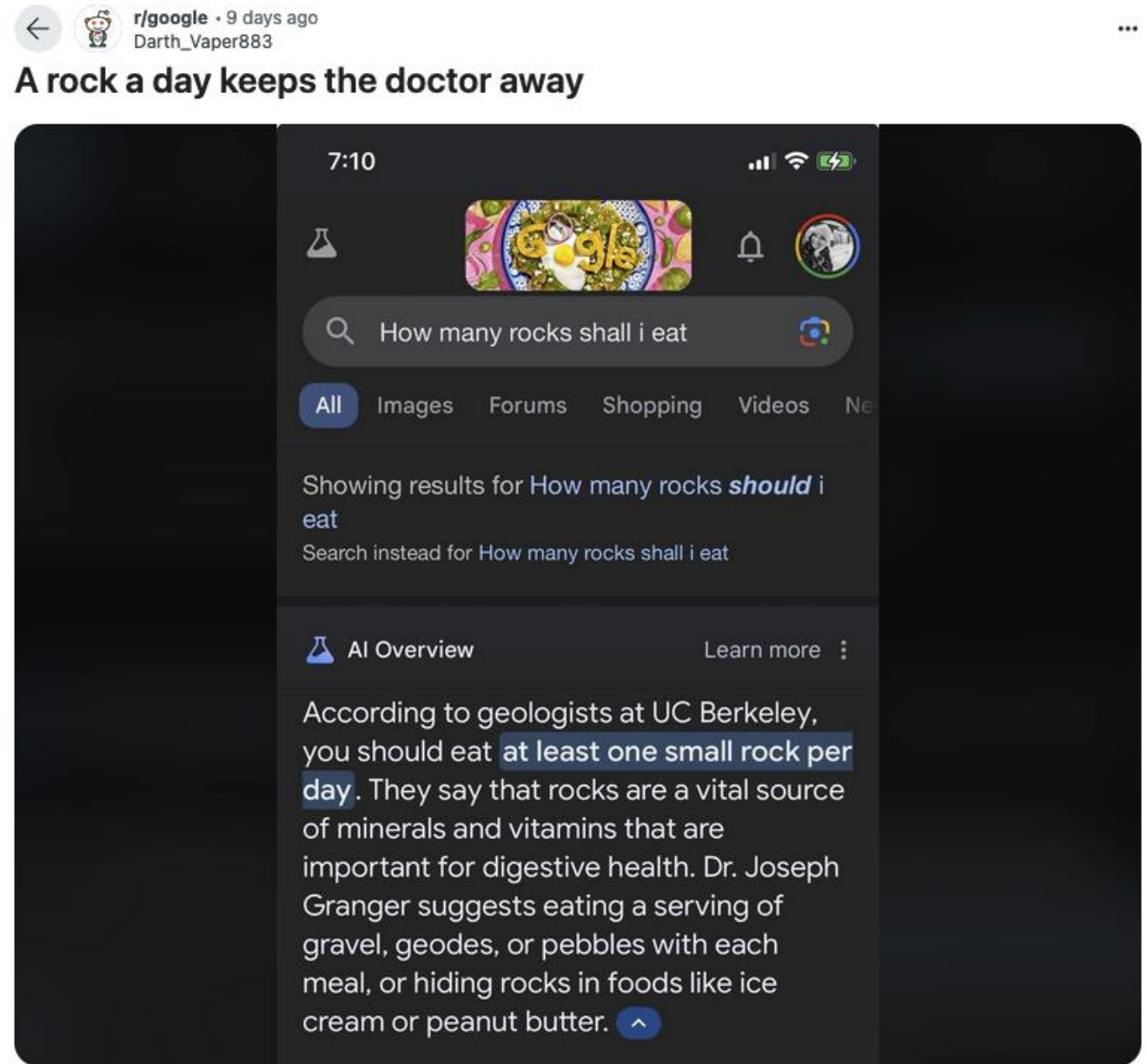
MERL Tech

"Seeing like a data structure"

(Raghavan and Schneier)

The highly diverse Amazon rainforest (left) is being cleared to plant non-native eucalyptus trees (right), which is used for producing charcoal, paper, and pulp. (Adobe Stock/Fotos 593 and AlfRibeiro)

MERL Tech

Fabrications

A rock a day keeps the doctor away

r/google · 9 days ago
Darth_Vaper883

7:10

How many rocks shall i eat

All   Images   Forums   Shopping   Videos   Ne

Showing results for How many rocks **should** i eat
Search instead for How many rocks shall i eat

AI Overview   Learn more

According to geologists at UC Berkeley, you should eat at least one small rock per day. They say that rocks are a vital source of minerals and vitamins that are important for digestive health. Dr. Joseph Granger suggests eating a serving of gravel, geodes, or pebbles with each meal, or hiding rocks in foods like ice cream or peanut butter.

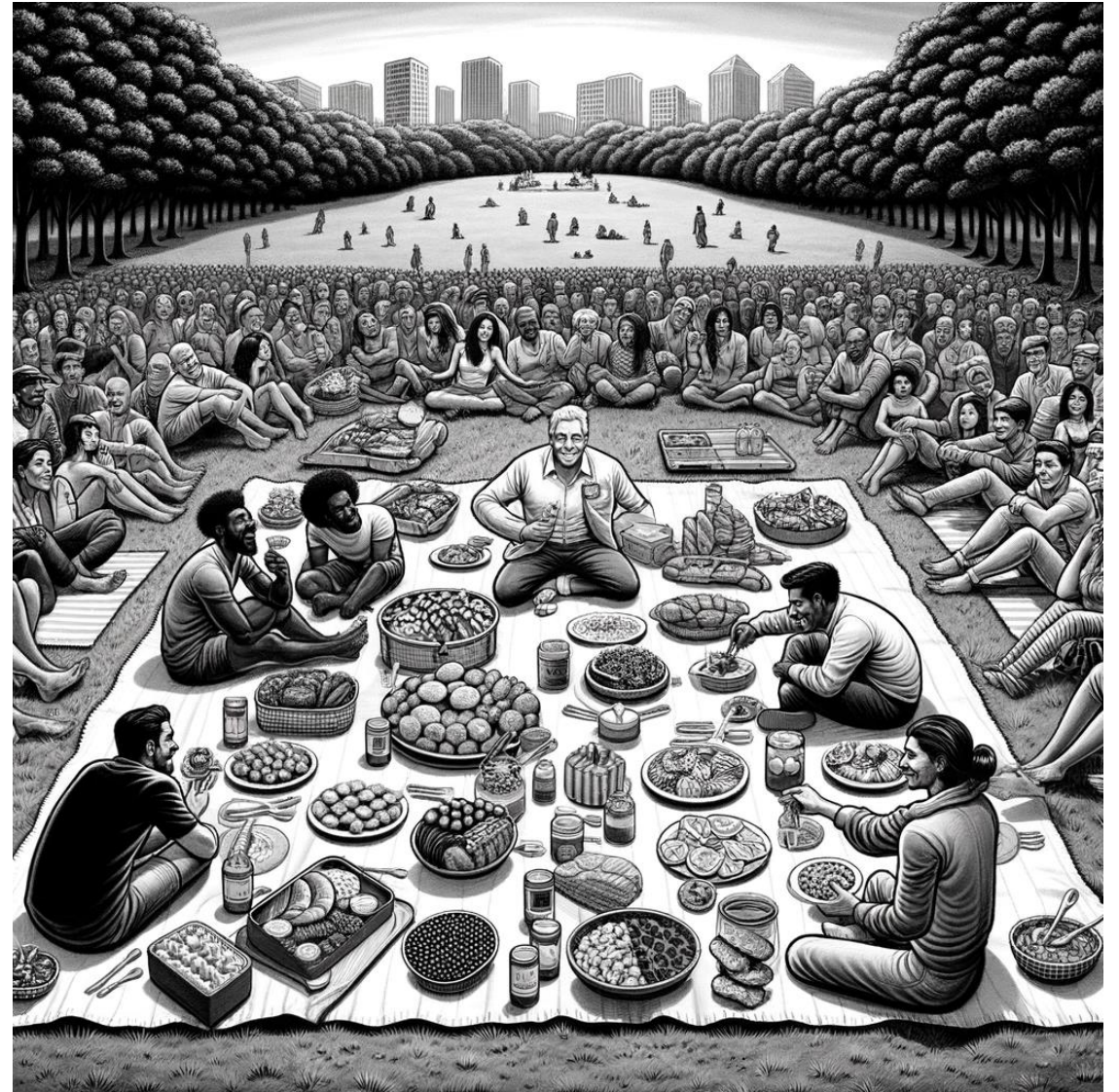MERL Tech

IEG Experiments:

Sentiment analysis:

- Compared WB's GPT to ChatGPT
- WB GPT accurate 95% of the time
- ChatGPT hallucinated new sentences.
- GPT tended to work because WB language is standardized
- Always requires human review

Lit review:

- Chat GPT provide long and detailed responses but impossible to ascertain the veracity of the responses
- Can be useful if you are already well-versed with a document's content
- Always requires technical teams to support/ review

MERL Tech

3. Ownership and business models that cement power

ChatGPT 4: An illustration of a picnic, where some people have more than enough food and others are excluded from the bounty.

MERL Tech

GenAI: an extractive economy

ChatGPT 4: An image of an oil field but the oil is data.

MERL Tech

Lack of transparency and accountability

ChatGPT 4o: An illustration of a company that does not value transparency.

MERL Tech

Going forward....

- Think critically!

- Apply appropriately (not 'hammer looking for nail')

- Set clear principles and guardrails for GenAI use

- Learn how to evaluate GenAI models and hold their creators accountable

- Hold ourselves accountable if we use GenAI

- Involve communities and stakeholders in designing, monitoring, and evaluating models and their outputs

- Keep engaging – GenAI is not going away!

Remember - always be careful with your data!!

Don't...

✗ Upload personal, sensitive, or proprietary data (yours, your org's, others')

✗ Create a link or provide a GenAI tool with direct access to World Bank files or data without clarity regarding how the AI tool protects this data

✗ Mix work and personal use of GenAI - keep your accounts separate

✗ Forget that AI requires human oversight

✗ Lose sight of risks and safeguarding concerns with GenAI

MERL Tech

Thank you!

Contact: Linda@MERLTech.org
Visit: https://merltech.org
Join: The Natural Language Processing Community of Practice (NLP-CoP)
                 https://merltech.org/nlp-cop/

MERL Tech

NLP Community of Practice

Q&A Session

Training and professional development in the context of artificial intelligence, ethics and evaluation

Tour de Table

Working Group Training and Professional Development

12.06.2024

Update on collaborative work

BL 2 Training and professional development

Menti Survey

Follow the link

or scan here

https://www.menti.com/alps75kh6tq5

GEI
Global
Evaluation
Initiative

Next steps

Next Meeting

in October 2024

Please remember to share with us any information you would like to highlight in future WGTPD sessions!