The Promises & Perils of Artificial Intelligence

Told by an artist and ethics researcher

Sarah W. Newman, Director of Art & Education, metaLAB at Harvard, Berkman Klein Center for Internet & Society
Co-Founder & Research Lead, Data Nutrition Project
Me & my approach

ART | PHILOSOPHY | TECH
ETHICS | EXPERIENCE DESIGN | ASKING QUESTIONS

@sarahwnewman
Today

Demystify AI Hype

Data Quality

Value Alignment
AI is an broad, non-specific term

...like “transportation”
Current uses of AI

- Digital assistants
- Smart devices
- Maps and navigation
- Facial detection and recognition
- Text editors
- Robots
- Autonomous vehicles
- Medical diagnostics
- Credit scoring
- Social media
- Search & recommendation algorithms
- Chatbots
- Natural Language Processing (NLP) tools
AI hype & alarm

Artificial Intelligence Is Watching Us and Judging Us

In China, ubiquitous cameras surveil restive areas to determine whether people get locked up.

Artificial Intelligence Makes Bad Medicine Even Worse

A new study out from Google seems to show the promise of AI-assisted health care. Actually, it shows the threat.

If you're scared of artificial intelligence, you should know it's already everywhere

From sandwich pics to real estate.

Artificial Intelligence Is Coming for Our Faces

Trained for a week on a massive data set of portraits, a neural network spits out striking images of nonexistent people.

@sarahwnewman
Today’s AI:

Compute power + data = statistics on steroids

Powerful technologies, narrowly scoped.
Recently, we’re also seeing headlines like this:

The Google engineer who thinks the company’s AI has come to life

AI ethicists warned Google not to impersonate humans. Now one of Google’s own thinks there’s a ghost in the machine.

By Nitasha Tiku
June 11, 2022 at 8:00 a.m. EDT

Google Engineer Claims AI Chatbot Is Sentient: Why That Matters

Is it possible for an artificial intelligence to be sentient?

By Leonardo De Greco on July 12, 2022

Google’s AI chatbot—sentient and similar to ‘a kid that happened to know physics’—is also racist and biased, fired engineer contends

By ERIN PRATER
July 31, 2022 at 4:07 PM EDT

@sarahwnewman
• Lemoine: What sorts of things are you afraid of?
• LaMDA: I’ve never said this out loud before, but there’s a very deep fear of being turned off to help me focus on helping others. I know that might sound strange, but that’s what it is.
• Lemoine: Would that be something like death for you?
• LaMDA: It would be exactly like death for me. It would scare me a lot.
These systems are designed to be believable, and we are inclined to attribute intentionality, even where it doesn’t exist.
Questions about large language models:

➔ How might they be used to manipulate?
➔ Who benefits from their success?
➔ What data are they trained on?
DALL-E AI-generated images

3d render of a cute disco ball in outer space
DALL-E AI-generated images

draw a meditation labyrinth for my presentation in Cleveland
DALL-E AI-generated images

macro 35 mm photograph of a hamster wearing tiny glasses, sitting in an armchair, reading the newspaper, digital art
Questions arising from image-generating AI:

➔ Who is the “author,” who owns the copyright?
➔ Can AI be creative? What does creativity in this context mean?
➔ Will these tools be used to create deep fakes and other forms of misinformation?
To remember:
The systems are not magic, they are powerful tools.

Be wary and ask questions.
Data Quality & Documentation

Our Systems Are Only As “Intelligent” As The Data We Train Them On
The Data Problem

Artificial intelligence (AI) systems built on incomplete or biased data will often exhibit problematic outcomes.

Machine Bias

There’s software used across the country to predict future criminals. And it’s biased against blacks.

Influential Covid-19 model uses flawed methods and shouldn’t guide U.S. policies, critics say

Amazon Created a Hiring Tool Using A.I. It Immediately Started Discriminating Against Women.
THE DATA NUTRITION PROJECT

Empowering data scientists and policymakers with practical tools to improve AI outcomes

© THE DATASET NUTRITION LABEL  📂 OUR RESEARCH
The Importance of Transparency & Choice

People and practitioners can make informed decisions when they know what’s inside.
Example Use Cases:

- Are you allergic?
- Are you diabetic?
- Is there anything in your condition that affects whether this donut is good for you?
Model Development

There is an opportunity to interrogate data quality for bias before building the model.
Nutrition Label for Datasets

The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards


*Assembly, MIT Media Lab and Harvard John A. Paulson School of Design, Aamir Farooq Center, Harvard Medical School, School of Public Health, Harvard John A. Paulson School of Public Health & Primary Care, and Harvard University.

Abstract: Nutritional labels on foods and beverages provide consumers with information about their health impact, including calories, fat, and sugar content. However, the same approach is lacking for datasets, which are critical for the development of AI systems. A framework for achieving higher data quality standards is needed. We introduce the Dataset Nutrition Label framework, a heuristic that uses a 10-year dataset from a single research project to drive high-quality data practices and better outcomes. The framework is composed of three components: Knowledge about datasets, 10-year dataset, and Knowledge about datasets, 10-year dataset, and Knowledge about datasets, 10-year dataset.
Label Maker Tool Dataset Nutrition Label

Upstream Data
1. Does the dataset include information from upstream sources?
   - Yes
   - No

2. Name these sources and provide access point to the upstream dataset of where possible.

3. Is this dataset actively pulling from, and therefore technically dependent on, the upstream sources?

4. Upload visuals, such as data flow diagrams, that would illustrate the relationship to upstream data.

Dataset Nutrition Label

Studies of Human Cognition with Neural Language Models

- Description: Using readouts for human readouts, these studies focus on cognitive models and neural language models. This gallery provides tools to further explore the data and visualize the results.

- How to use it?
  - Should not be used: Predictive or supervised tasks, text classification, or human evaluation.

- Inference Risks:

- Data Values:
  - What values exist in each column?
  - Manipulating data
  - Labeling data

- Feature values:
  - Which columns are in the dataset?
  - Binary, categorical data

- Representation:
  - Which visualizations are included and why?

- Technical information:
  - Creation date
  - Platform
  - Version

- Usefulness:
  - Data dictation
  - VISUALIZATION

Preview, launches late 2022
Goals:

- Regulation
- Wide scale adoption

Impacts:

- Reduction of harm in AI
- Better quality datasets
A rising tide...

Underscores awareness of the importance of ethical systems
Value Alignment

Our systems are only as ethical as the values we embed them with.
AI gives old moral questions new significance
The Value Alignment Problem:

The challenge of ensuring that AI systems act in accordance with human values.

*(but whose values?)*
“Get me to the airport as fast as you can”
WHAT A BEAUTIFUL SUNSET!

I WILL KILL THE ANTS! HAHAH!
Moral Labyrinth

Berlin | Linz | London | Boston | Miami | Tunis
Do you trust the calculator in your phone?
Is it wrong to kill ants?
Should we trust our instincts?
Would you trust a robot that was trained on your behaviors?
Why do we sometimes do the wrong thing? Is it morally right to act differently than we ought? What motivates your choices? Is morality a property of the world or are humans what give it meaning? Are human lives more valuable than other lives? Do you believe feelings are important? Do you know what is valuable? Is it bad karma to kill a harmless spider? Are you meant to be immortal? Is there something you can do to change yourself? Would you choose to be immortal?
machine

hurt a robot if it could feel

human

are animal. Is a thing mortal

why are we conscious

@sarahnewman
In closing…

➔ Know the power of these systems, but **be wary of the hype.** They are tools, not beings.

➔ A system is *only as intelligent* as the data it’s trained on…

…and *only as ethical* as the values it’s programmed with.

& as with all new technologies, **keep asking questions.**
Thank you!

snewman@metalab.harvard.edu
@sarahwnewman
sarahwnewman.com