Ensuring the Safe and Secure Use of Patient Information

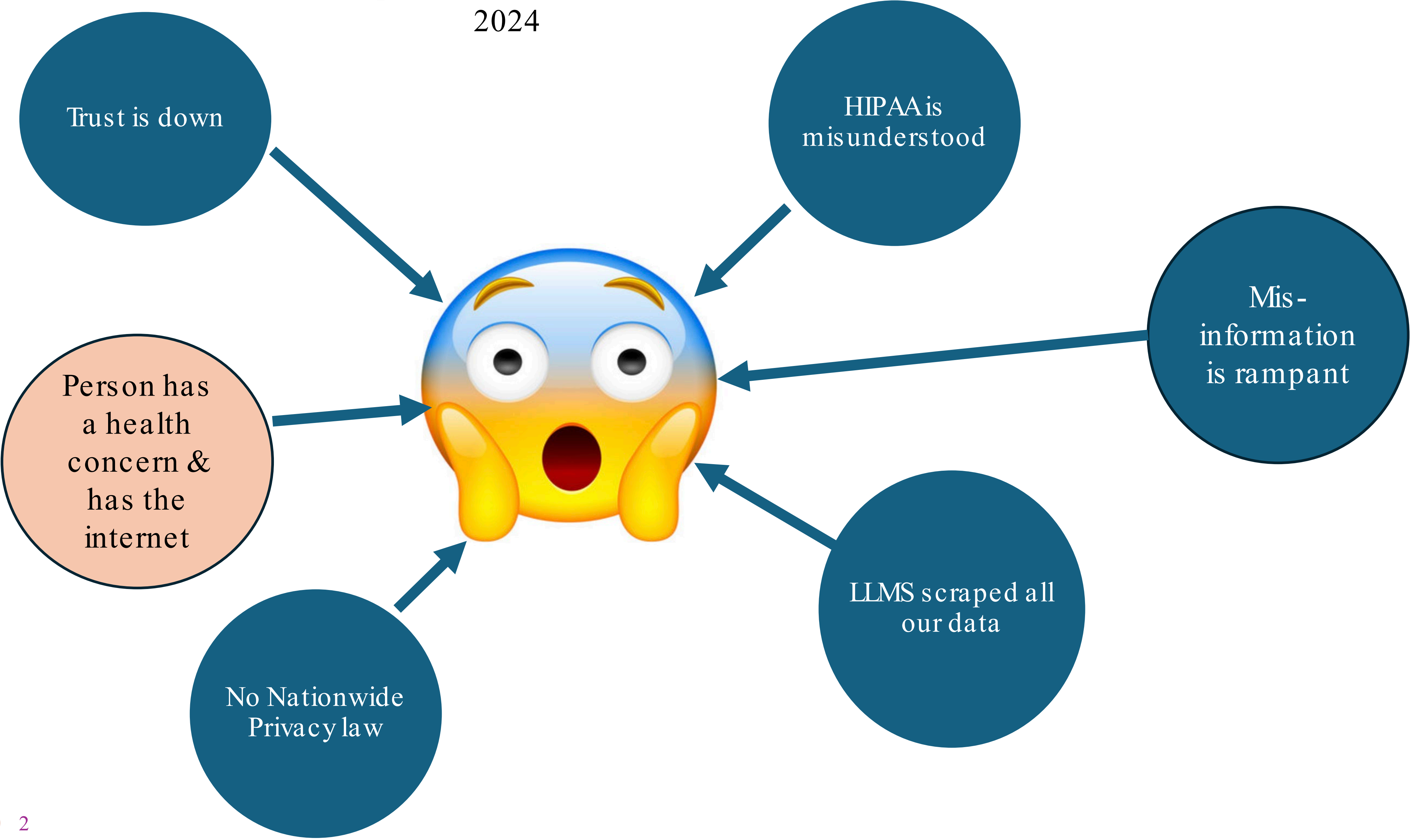
Session 1: Using Artificial Intelligence and Digital Health Technology to Support Patients from Symptom Onset to Seeking Care

part of

Diagnosis in the Era of Digital Health and Artificial Intelligence: A Workshop at the National Academies of Science, Engineering and Medicine, July 25, 2024

→ Lucia C. Savage, JD
Chief Privacy & Regulatory Officer
Omada Health, Inc.

People and data collection 2024



Trust is down

HIPAA is misunderstood

Person has a health concern & has the internet

Mis-information is rampant

No Nationwide Privacy law

LLMS scraped all our data

People and Data Collection 2024

- Trust in big tech specifically, and business generally, remains at an all time low, per [Edelman Trust Barometer]
- People have figured out why the ads in their social media are so personalized and
  - While it can be incredibly helpful to have the internet ”understand” you,
  - It also can be creepy
- The rules on data collection are confusing, even for professionals. Imagine what its like for an ordinary patient
  - HIPAA has great branding, but few understand it
  - HIPAA does not apply to general internet browsing or social media
  - The new state-based consumer laws add patchworks to an existing patchwork
  - We do not have a nationwide consumer privacy law
- The current large data models, from Google, OpenAI, and Anthropic, for example, all got their data from general internet activity, but most people probably haven’t realized that yet.

Five Familiar Principles for Using Data from Patients to Develop and Deploy AI

1. Do No Harm
   - Even HIPAA is based on this concept in part
   - We know how to do this in the healthcare sector
   - Data collection and use in other sectors does not have this embedded concept
2. Do Not re-Identify data that is not identifiable
   - Do you ever need to actually know who the specific person is, to develop AI for diagnostics?
   - Of course, you need to diagnose a person, potentially with that AI output.
3. Trust the Individual and their community to know what's best for them
   - Stop patronizing people.
4. Be honest and transparent about your data collection and curation.
   - If Individuals “opt out,” that’s their choice.
   - If opt outs make your dataset biased, then find another transparent, honest way to build out a fair and unbiased dataset.
5. There actually are rules. Follow them.

Appendix:

Supplemental materials from NHLBI Workshop on Artificial Intelligence and Hypertension, March 29-31, 2023, published at:

Daichi Shimbo, MD, et al, **Transforming Hypertension Diagnosis and Management in The Era of Artificial Intelligence: A 2023 National Heart, Lung, and Blood Institute (NHLBI) Workshop Report, Hypertension, July 16, 2024, https://doi.org/10.1161/HYPERTENSIONAHA.124.22095**

Principle #1.

People altruistically support medical research. They want to help advance knowledge. They “get it” that their data and experiences could help with the next breakthrough. But your use of data from real people is a privilege. Don’t abuse it by letting bad things happen to people who volunteer data for research. The concept of “do no harm” and holding people’s health care information secret is thousands of years old for a reason. Look to this principle as your foundation-stone.

Do No Harm

Principle #2

The data comes from actual people. AND
The data used to train AI tools should be HIPAA de-identified and anonymized.
These are specific legal and statistical standards. Stick to them.
You are legally precluded from re-identifying de-identified data, and probably contractually prohibited as well. So,

DON'T RE-IDENTIFY THE DATA WHEN YOU COMBINE DATA SETS. In fact, watch out for this when building your data sets

Principle #3

Trust the Individual and their community to know what's best for them; stop patronizing people.

For decades, the health care system and research institutions have not adequately addressed the needs of medically underserved populations, or have failed to understand the role of whole communities (as opposed to individuals) in co-creating research data. Or, White/dominant communities have thought they understood what was "harm" (see principle #1), but through a biased lens. As a result, the very people and communities whose data we need donated to ensure fair data sets are even more skeptical that donation won't cause harm or be exploitative. Don't assume you know; ask the community.

Principle #4.

Be <u>honest and transparent</u> about your data collection and curation. If Individuals "opt out," that's their choice. If opt-outs make your dataset biased, then find another transparent, honest way to build out a fair and unbiased dataset.

What can be very upsetting is when you mislead (fool, cheat, bamboozle) people about what you are collecting from or about them, and how you are using it. Examples abound from Cambridge Analytica (unsupervised social science survey collected data resulted in a political profiling large "anonymous" data set), to screen scraping of people's images, to letting pharmaceutical companies "see" what is going on in on-line patient support groups. Did any of you Facebook users know that when you upload a photo, you enable Facebook to use that photo however it sees fit?

Principle #5.

There actually are rules. Follow them.

This is not “the wild west.” It may be that what is subject to FDA regulation when it is AI is under development but for how to fairly, privately and securely use data from individuals, rules and standards exist in the U.S. and internationally. For starters, you’ve got The Belmont Report (1977), ethics, GDPR, HIPAA, state and nationwide privacy laws in most other developed nations.

Fair, private and secure use of data donated by individuals, or harvested from them lawfully, is possible, **but the rules are multi-layered, complex, and can be fact specific. Fair data use is not a DIY project.** Find a good lawyer and partner with them. They want you to succeed. Please don’t look it up on Reddit or Git Hub or Twitter, your research colleagues or with a chatGPT search.