Measurement and Metadata

Overview

- **what is a variable**
- **what gets counted and why**
- **how this is tied to measurement**
- **differentiate between variables and metadata**
- **how they are bound to ethical dilemmas**
- **how the operationalization of variables complicates equity and justice**
- **introduction to frameworks to test for bias**

Readings
[optional]

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, *81*, 1–15.

- https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

D'Ignazio, C., & Klein, L. (2020). 4 - "What Gets Counted Counts." In *Data Feminism*. MIT Press.

- https://data-feminism.mitpress.mit.edu/pub/h1w0nbqp/release/3

Feinberg, M. (2023). Labels. In *Everyday Adventures with Unruly Data* (pp. 159–189). MIT Press.

- https://mitpress.mit.edu/9780262544405/everyday-adventures-with-unruly-data/

Jacobs, A. Z., & Wallach, H. (2021). Measurement and Fairness. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 375–385.

- https://doi.org/10.1145/3442188.3445901

Definitions	
performance model	**measures** how well an **AI/ML model** achieves its **intended task** https://www.ibm.com/think/topics/model-performance
conceptualization	the **action** or **process** of **forming** a **concept** or **idea** of something https://languages.oup.com/google-dictionary-en/
operationalization	the **process** of **transforming** abstract **concepts** into specific, **measurable**, and **observable variables** https://trainual.com/manual/operationalization
intersectional identities	how different **aspects** of a person's identity **overlap** and **interact**, creating unique **experiences** of **privilege** and **disadvantage** https://www.oregon.gov/deiconference/Documents/
confounding variables	an **external**, unmeasured **factor** that **influences** both **independent** and **dependent variables** https://amplitude.com/explore/experiment/confounding-variables
Variables	
Variable Definition	• **placeholders** for a value • **quantifiable concepts** used to represent **attributes** or **features** in the **data**
Uses of Variables	• calculate **statistics** • provide **summary** of the central **tendency** • provide **dispersion** of the **data** • formulate **hypotheses** • conduct **tests** • determine **significant relationships** in the **data** • evaluate **performance models**

Types of Variables		
	discrete	represent **characteristics** that are not inherently **quantifiable**
	nominal	• **lack** intrinsic **order**
		• **colors**, gender, **location**
	ordinal	• allow for **ranking**
		• possess **quantifiable distance** between units
		• **low** med **high**, first **second** third
	continuous	bound to a **numerical system**
	interval	• **lack** a true **zero**
		• **scaled** variable
		• **temperature**, distance **measurement**
	ratio	• has a **true zero** representing **absence** of **variable**
		• exclusively **positive** values
		• **age**, weight

Conceptualization and Operationalization	
Conceptualization	takes the idea and **gives** the **variable** an **abstract**, **conceptual**, or **theoretical** definition
Operationalization	links the conceptual **definition** to a specific set of **measurement** **techniques** and **procedures**

Example: Intelligence	
Conceptualization	• **intelligence** is doing **well** in **school**
Operationalization	• **GPA** or standardized **test scores**
Metadata	• related to **conceptualization** and **operationalization**
	• **distinct** in **form** and **function** from a **variable**
	• helps **describe**, **classify**, and **categorize** data
	• enables **users** and **systems** to **understand** and interpret **data**

	• makes it **easier** to **organize** and **locate** information
Structural Metadata	• subject **metadata**, keywords • **not** necessarily **observable** • **ex**: last **file** access **time** • inferred through **concepts** which often involve **human allocation** and/or **understanding**
Variables	• a **variable** that is **not** well **conceptualized** and **operationalized cannot measure** anything • a **tag** must conceptually **describe** the **content** to be retrieve
Categorization	• allows **researchers** to **conduct** rigorous **research** • facilitates **efficient** and **effective** data **analysis** • **improves** model **accuracy** • creates **robust databases**
Limitations	• many of the **variables** we rely on to create **automated** decision-making **tools** are **inferred variables** that rely on **proxies** that might **not** have an **established tool** for measurement • **cultural** and **social constructs** can **obscure**, under-represent, and **misrepresent data** and **demographics** • when **tools** to **operationalize** conceptual **variables** are **absent**, **existing** and often **flawed power** structures **fill** the **vacuum**
Examples of Limitations	• **gender classification** across **cultural** and **ideological** complexities • **racial classification** and **dehumanization** based on **tribalistic** genetic **memory** • clothing **designers** reinforce social **expectations** creating a second order **effect** for gender-social **norms** (**no pockets** for women = **sell more** purses)

Classifications	• **classifications** can **serve** many **purposes** • the **type** of data **stored** in a **variable** • **reductive** by design

The Matrix of Domination Patricia Hill Collins: Black Feminist Thought	**structural**	• encompasses the **laws** and social **structures** that perpetuate **oppression**
	disciplinary	• bureaucratic **processes** and **hierarchies** that are used to **uphold inequity** even after laws are **changed**
	hegemonic	• the consolidated **ideas** about who has **power** and who does **not** are often **passed** through **media**, **art**, and **group think**
	interpersonal	• latent systemic **inequities** in daily **interactions** **reinforce** power **balances**

Models, Audits, and Accounts

Measurement Modeling	• ensure that the model is performing as designed • ensure model variables are being operationalized
Audit Studies	• gauging the **consistency** of measured **processes** • **routing** part of **civic infrastructure**
Audit Algorithms	• investigation of algorithms to uncover inherent or situationally derived biases and potential discrimination • designed to repeatedly and systematically validate algorithmic inputs and ensure verifiability and objectivity of outputs • determine if intersectional identities impact results

Considerations for Construct Validity

Construct Validity	• ensuring that measurements are accurately represented • ensuring the intended construct will consider the impact of these measurements

	• ensuring the measurements align with known hypotheses • assessing consequences of their usage • inherently nuanced and relies on critical reasoning to establish a degree of validity
Face Validity	• gauge plausibility of measurements
Content Validity	• ensure that operationalization fully captures the construct of what is being measured
Convergent Validity	• assesses whether the measurements align with those other established models • identify potential overlaps and discriminatory patterns
Discriminant Validity	• determine if operationalization could capture other constructs • determine if there are confounding variables
Predictive Validity	• highlights how proxies may relate to various observable properties not always factored into models
Hypothesis/ Consequential Validity	• explore how the world is shaped by the measurements • determine if measurements inadvertently perpetuate inequality • considers broader implications
Consequential Validity	• investigates the consequences of using the measurements obtained from the model

Live Session Preparation	*Week 3: 30 Oct 2025*

***Consider** an **object** that you **use every day**. **Who** do you **think** this **object** was **designed** for? **How** do you **know**? Who **can't use** this **object**? **What** might **inclusive design** look like?*

I spend most of my day at this computer for one reason or another, so I am choosing as my object my keyboard. To limit the scope, I am going to narrow this specifically to the QWERTY keyboard.

The QWERTY keyboard was invented in 1870 by Christopher Latham Sholes. Sholes, a Pennsylvania native, moved to the newly established Wisconsin Territory in 1837 to take on training as a printer's apprentice at his elder brother's newspaper. He quickly advanced in the realm of periodical publication to run several newspapers in different frontier towns to eventually receive an appointment to federal office by President Lincoln. This shift in vocation granted Sholes with unprecedented free time to pursue his interests as an engineer and inventor. Adapting a page numbering device invented by his friend John Pratt, Sholes invented the typewriter receiving a patent in June of 1968. He continued to work on and make improvements to his device throughout the rest of his life. In response to recurring jamming issues with adjacent frequently-used keys in his current design, he began to formulate what would eventually, in 1870, become the QWERTY keyboard. Sholes sold his patent in 1873 to the Remington Arms Company, a company uniquely suited to mass-produce his design, and the well-recognized result was the iconic Remington Typewriter.

The typewriter was not-surprisingly initially intended for commercial printing. As literacy rates rose through the 19th century, commercial publication moved from a smattering of fringe hobbyist, into a network of powerful economic and cultural leaders. As the familiar pattern of cultural adoption of new technologies, the ability to create printed documentation spread through the business and eventually the private sectors. As the use of typewriters accelerated to become staple in the workplace by 1910 in the United States, typing was a skill initially dominated by women. Though, by 1915, typing became a core

skill taught in most American high schools with widespread use peaking in the mid-20th century.

Some of the limitations of the keyboard are realized when adapting for physically handicapped users. The QWERTY keyboard, in particular, reveals its limitations with its Anglo-Latin-centric layout and character-set. Much effort has been made to adapt the QWERTY keyboard to cater to users with physical, visual, or cognitive disabilities. 'Large Keycap' keyboards featuring larger, spaced keys, single-handed layouts, ergonomic designs, high-contrast designs, and on-screen keyboards that can be controlled by mouse click, touch, and even neural interface, are all efforts that have successfully brought this technology to those previously excluded. One design idea that could make the keyboard more universal and globally accessible would be a dynamic keyboard that stores and renders various layouts. This could be achieved through a capacitive touch screen with a customizable layout, or a physical-key-based design with kindle-paper-like key faces that facilitate preprogrammed and customizable layouts.

References

https://www.britannica.com/biography/Christopher-Latham-Sholes

https://www.opticjam.com/the-history-of-publishing-from-ancient-tablets-to-digital-media

https://www.youtube.com/watch?v=X8yZy2-ex8E&t=134

https://www.wati.org/wp-content/uploads/2017/10/Ch4-ComputerAccess.pdf

Live Session Notes	Class Meeting: 21 Jan 2026
*	
*	
*	
*	