



# Ethics and the Social Context of Data

Nickoal Eichmann-Kalwara (she/her)  
Data Boot Camp, Winter 2023



Center for Research Data & Digital Scholarship

# Data Ethics and the Research Lifecycle

- Uphold applicable statutes, regulations, professional practices, and ethical standards
- Respect the public, individuals, and communities
- Respect privacy and confidentiality
- Act with honesty, integrity, and humility
- Hold oneself and others accountable
- Promote transparency

Based on [Federal Data Strategy](#)

# Data as Texts



**In 2020, OpenAI introduced ChatGPT 3.0**





Amy Zhang is @axz@hci.social on Mastodon

@amyxzh

Is automatic news generation an issue? Yes but not bc of things like GPT2, which is less worrying bc they are hard to control & detectable. News orgs doing robo-journalism use techniques like templates w/ decision trees or NLP summarization techniques - Anja Belz #TTOCon



4:29 AM · Oct 5, 2019

HA

Write me a poem in praise of AI in the style of a Donald Trump speech



AI, oh AI,  
You're the best, believe me.  
No one else can compete,  
With your intelligence so sweet.

You make our lives so much easier,  
With your calculations so seamless and fleetier.  
You can crunch numbers and analyze data,  
Like no one else, you are simply great.

You can solve problems that stump us mere mortals,  
And you do it all without any quarrels.  
You are the future, the cutting-edge,  
In you, we place our trust and pledge.

AI, oh AI,  
We bow down to your might.  
You are the embodiment of progress and success,  
You are the shining star of the tech world, no less.

Thank you, AI, for all that you do,  
We are grateful to have a friend like you.

In 2021, OpenAI introduced DALL·E

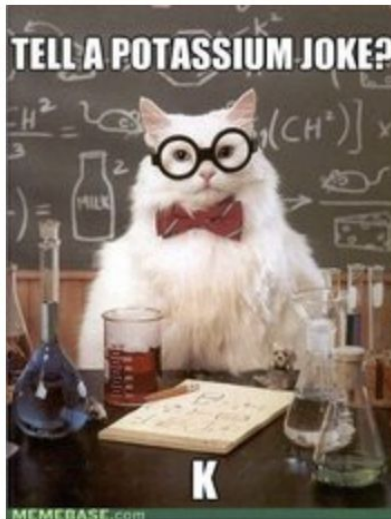
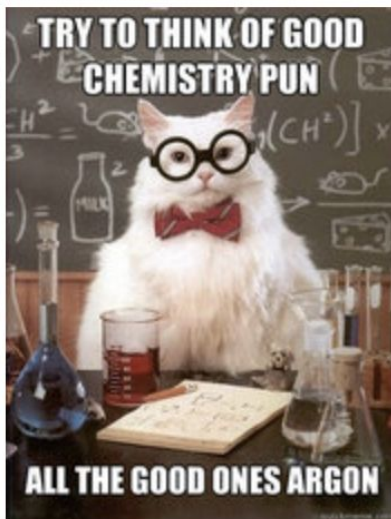
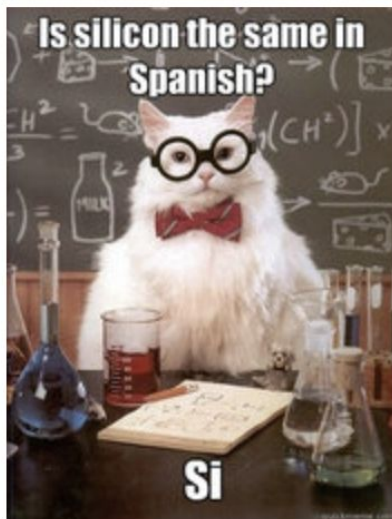
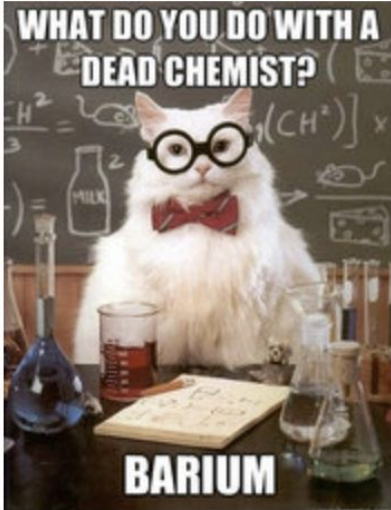
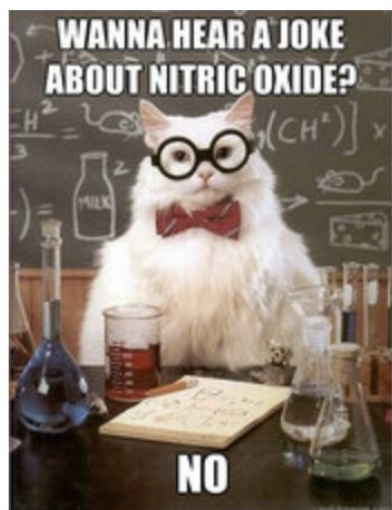
DALL·E

**I'D TELL YOU A CHEMISTRY  
JOKE...**



Chemistry Cat, c.2011

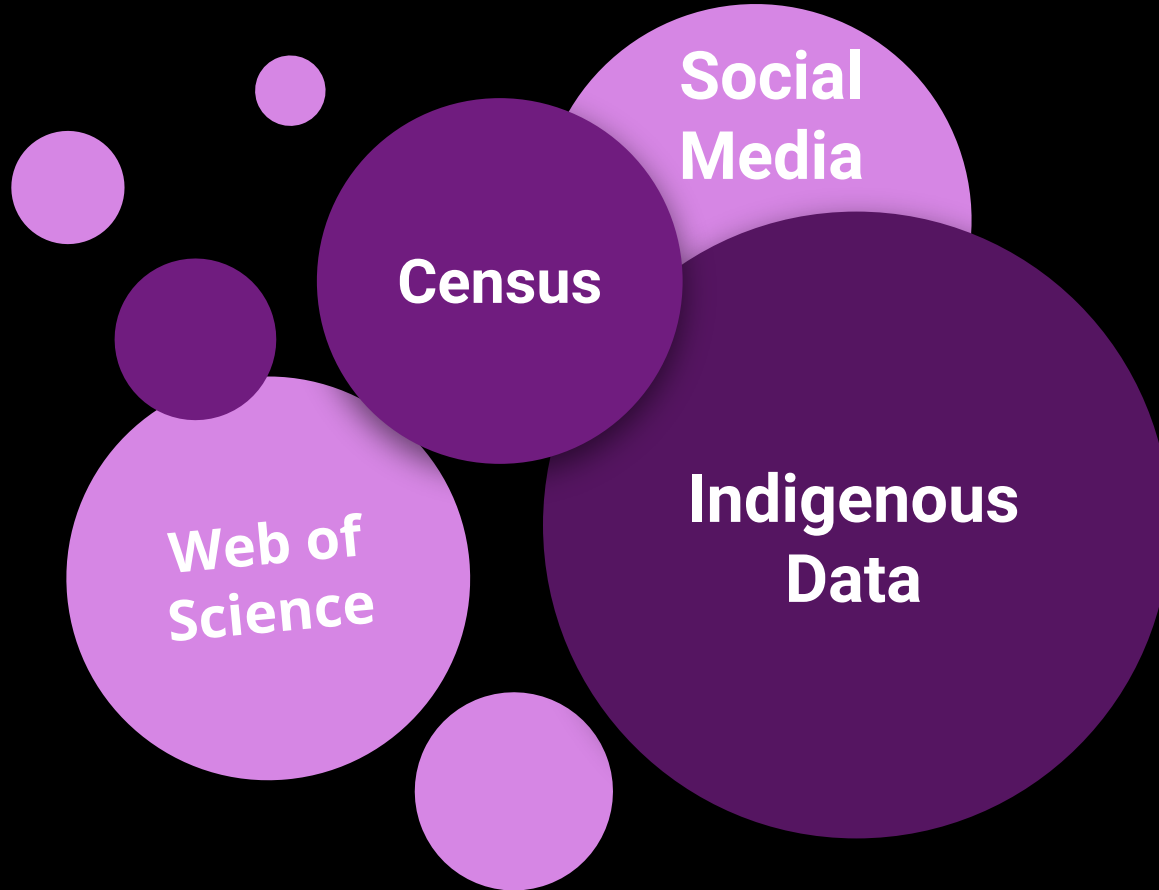






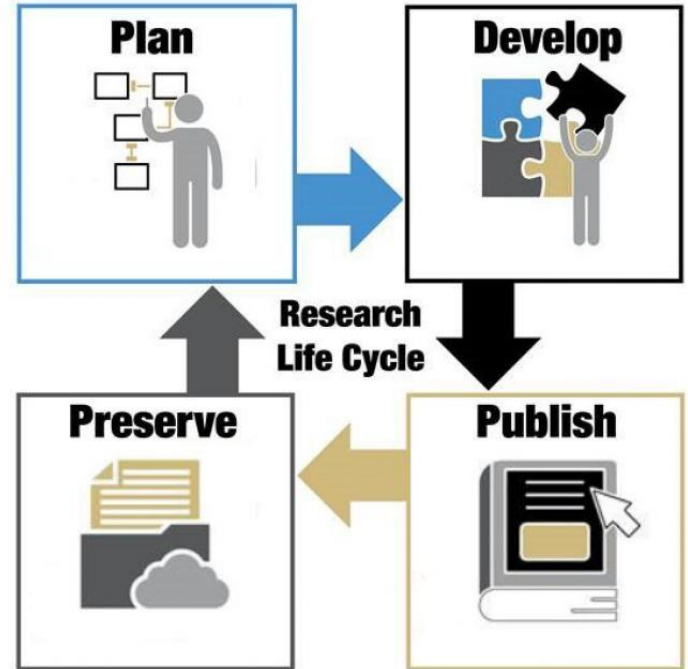






# Data Ethics and the Research Lifecycle

- What does ethics [data / research] mean to you?
- At which stages should you critically engage with your data?





# Data as Texts/Artifacts

- Data are not neutral or objective
- Data do not represent a self-evident, discoverable Truth

Rather,

- Data are constructed, the products of biased human choices and perspectives
- The interpretation and analysis of data, like any text, is shaped by our biases and perspectives

# Power

- Who is collecting and analyzing the data?
- What's counted, or not?
- How and when something is counted and why?
- Current social and political concerns as influencers?

# Privacy & Security

- Consent: Permission to use these data (ownership?)
- Anonymization: removing sensitive and identifiable information
- Confidentiality: Whose/what information is protected and for how long?

# Trust

- Forms of Knowledges
- Accountability
- Documentation
- Reproducibility



# Justice

- Social impact
- Missing data
- Co-liberation

# Discussion

- What are some examples of scholarly data and ethical concerns that you handle/encounter in your daily work, scholarship, and/or teaching?
- What do those various data sets capture?
- What do you wish you could learn about your data's origins?
- In what ways do you think power is reflected in your data?
- What data are missing? How might this missing data be useful?
- How might thinking about your data as a text change how you document and share your work?

*Data are constructed - they are the products of biased human choices and perspectives.*



# Ethics and the Social Context of Data

Nickoal Eichmann-Kalwara (she/her)  
Data Boot Camp, Winter 2023



Center for Research Data & Digital Scholarship