Human Centered Data Science

DATA 512 — Jonathan T. Morgan & Os Keyes

Interrogating algorithms | Week 6 | November 1, 2018

Overview of the day

- Reading reflections
- Ethical implications of crowdwork
- Algorithmic transparency, interpretability, and accountability
- Auditing algorithms
- In-class activity: auditing the Perspective API

Reading reflections

Reading reflections

“The authors here state that bias in training data is one of the main mechanisms for introducing bias into an algorithm - if the training data is biased, the algorithm will learn those same biases. They use the example of crime data being racially biased, and therefore state that any algorithms trying to predict crime that are trained on past crime data will also be racially biased. My question is, how do we go about fixing this? There are probably many cases like crime where there isn't existing unbiased data. In cases like these, do we simply conclude that we can't create unbiased models in these situations? Could we solve some of this by more careful selection/deselection of features for the models? Or do we try to fix the biases in the training data? And is it possible to manually fix something like this without introducing yet more bias?”

-Kenten

Reading reflections

“If an audit is implemented for algorithmic systems, who takes up the onus of responsibility? What if a technologist or anyone for that matter fails to explain how the algorithm arrived at the answer due to its inherent mathematical complexity? As algorithms don’t follow protocol based thinking like social scientists, isn’t the trade-off of using machine learning algorithms to make decisions the loss of accountability?”

-Sayil

Reading reflections

“Tech companies are overwhelmingly dominated by White and Asian males. They are often the people who develop, train and deploy these algorithms. Do you think that training these employees to recognize ways in which biases can creep into algorithms would help improve the problem?”

-Tejas H

Reading reflections

“About COMPAS: Did the makers of COMPAS consider training a new model without racial data and see if they could negate the racial bias inherited by the algorithm? The article mentions that more blacks were arrested than whites and the training data reflected it. So, the bias was expected in some sense and should’ve been identified early on.
Facebook targeted Ads: How can we identify this if discrimination was done manually and not digitally? For example, if I have a few houses to sell and I want to sell them to a particular race only, then I could have my team deliver promotional material in only selected neighborhoods. I’m not sure if there are laws preventing me from doing this and even if it is identified, am I liable per law? I’m trying to understand if Facebook case got traction because it just made it more efficient to do it?”

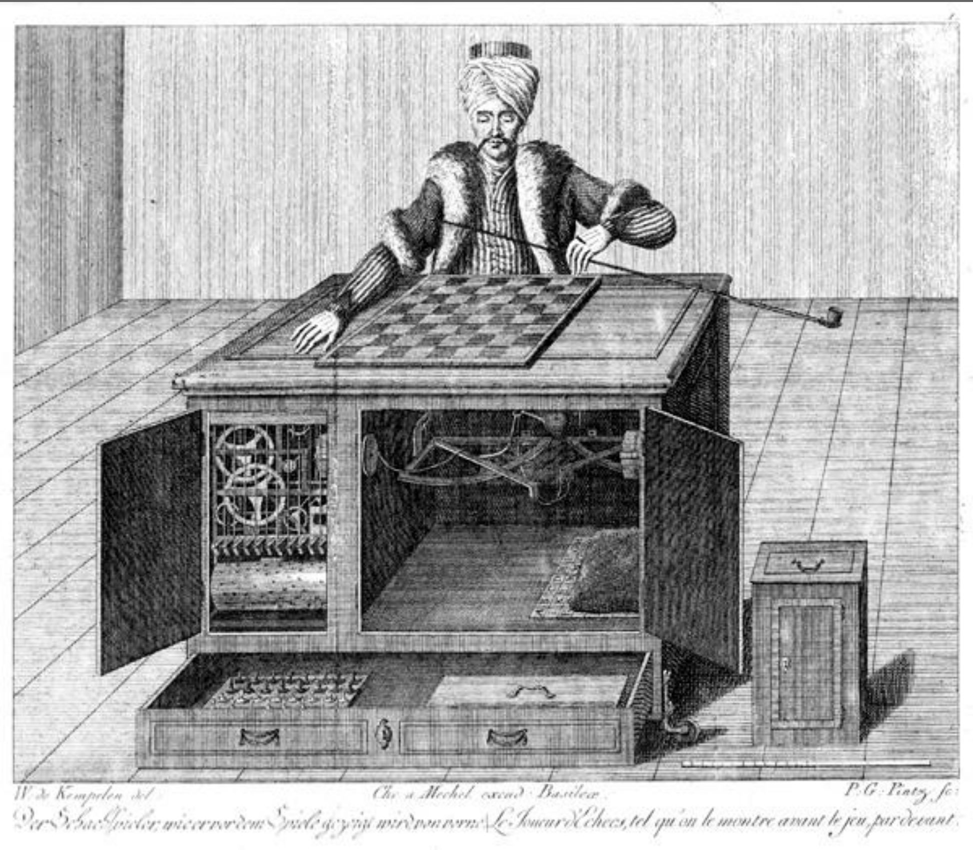
-Tejas J

Reading reflections

“Questions: Based off this article and a few that we have read, it sounds like the only real way we have of enforcing rules and regulations is legally, but those regulations are slow-moving to enact, difficult to change with times, and heavily rely on interpretation. Aside from communities of practice, what other ways are there to define “accountability” and hold people accountable in data science?”

-Hannah

Ethical Implications of Crowdwork



Case Study: Content Moderation

- When you report content on a website, where does it go?
- *"Moderators said they watched images of war victims who had been "gutted," and "child soldiers engaged in killings." A former moderator who worked at Facebook recalled watching video of a cat thrown into a microwave, The Journal reported."*

Case Study: Content Moderation

- Emotional and moral burden
- What resources are available to alleviate this?
    - Employer resourcing?
    - Community resourcing?
    - Personal/wage resourcing?

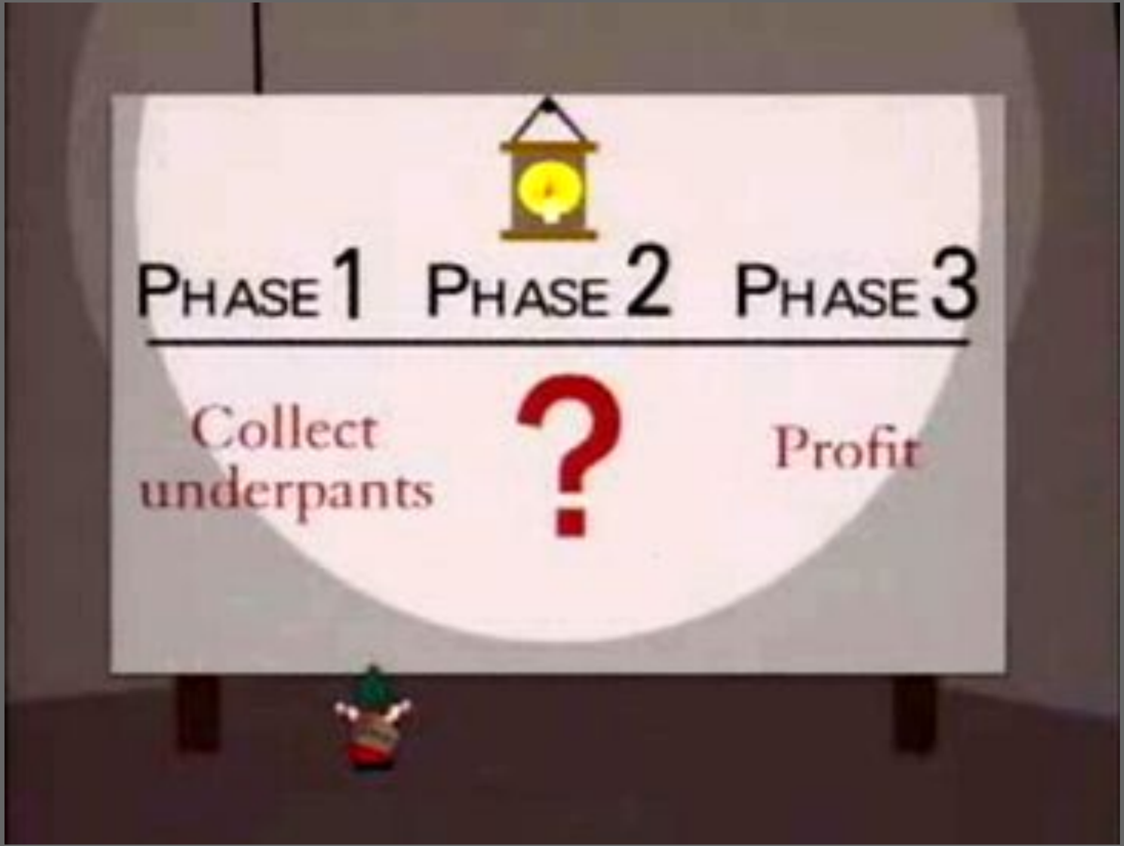
How Do We Use Crowdworkers?

- Coding data
  - Offensive text
  - Offensive images
- Giving away data
  - Survey participation
  - Study participation

How Can We Do Better?

- “Turkopticon” feedback system (https://turkopticon.ucsd.edu/)
- Fair Crowd Work (http://faircrowd.work/)
  - Wrote the “Frankfurt Declaration” on crowdwork
  - Lobbying for Crowdworker unions
- Avoid marketplace crowdwork wherever possible
  - Do crowdwork in the context/community

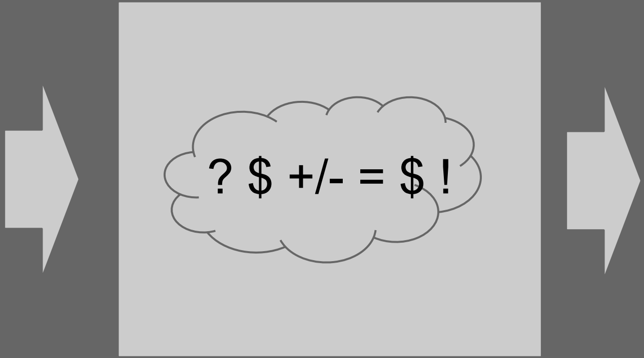
Algorithmic black boxes



Source: https://upload.wikimedia.org/wikipedia/en/d/dd/Gnomes_plan.png

Black boxes in science & engineering

Black boxes in everyday life



Source: Pixabay.com CC0

Machine learning blackens the box

- Input is a dataset, with human-selected features that describe some stuff
- Training data may be labeled by a human (supervised), by another algorithm (semi-supervised), or not labeled at all (unsupervised)
- Output is generally probabilistic, not deterministic
- Depending on the approach used, even the algorithm’s designers may not be able to reverse-engineer how a particular prediction/classification was made
- Algorithms are constantly ‘tweaked’
- Algorithms are often kept secret

But... we still kinda think we know what’s going on, most of the time. Right?

Chrome; Seattle IP; Logged in @wikimedia.org

"jonathan morgan"

All | Images | Videos | News | Shopping | More | Settings | Tools

About 566,000 results (0.62 seconds)

Jonathan Morgan - Wikipedia
https://en.wikipedia.org/wiki/Jonathan_Morgan
Jonathan Morgan may refer to: **Jonathan Morgan** (director) (born 1966), director and former actor in pornographic films; **Jonathan Morgan** (politician) (born ...

Jonathan Morgan (director) - Wikipedia
https://en.wikipedia.org/wiki/Jonathan_Morgan_(director)
Jonathan Morgan (born February 5, 1966) is a former actor and current director of pornographic films. He has directed for studios including Wicked Pictures, ...

Jonathan T. Morgan - Wikimedia Foundation
https://wikimediafoundation.org/wiki/User:Jtmorgan
Nov 4, 2016 - About me. My first edit to Wikipedia was in 2006. I've been performing gnomish edits since 2008 as Jtmorgan. I have an MS and PhD in Human ...

Jonathan Morgan Profiles | Facebook
https://www.facebook.com/public/Jonathan-Morgan
View the profiles of people named **Jonathan Morgan**. Join Facebook to connect with **Jonathan Morgan** and others you may know. Facebook gives people the ...

Jonathan Morgan: Music
https://jonathanmorgan.bandcamp.com/
Jonathan Morgan Jon is a psychedelic rock guitarist, vocalist, composer and producer from Birmingham, England.

Jonathan Morgan & Company Limited
https://www.jmcdesigninteriors.com/
Commercial interior design firm based in Vancouver BC. Custom interior design, construction & renovation services for commercial spaces & more.

Firefox; London proxy IP; Private browsing

"jonathan morgan"

All | Images | Videos | News | Shopping | More | Settings | Tools

About 372,000 results (0.35 seconds)

Dr Jonathan Morgan | Faculty of Law
https://www.law.cam.ac.uk/people/academic/je-morgan/161
Jonathan Morgan read Jurisprudence at Oxford, later migrating to Corpus Christi College, Cambridge, to write his PhD thesis, "In defence of freedom of contract".

Dr Jonathan Morgan | Corpus Christi College University of Cambridge
https://www.corpus.cam.ac.uk/people/dr-jonathan-morgan
Jonathan Morgan grew up in Warwickshire and read jurisprudence at Oxford. He taught law at Warsaw and Oxford Universities before writing his doctoral thesis, ...

Jonathan Morgan - Wikipedia
https://en.wikipedia.org/wiki/Jonathan_Morgan
Jonathan Morgan may refer to: **Jonathan Morgan** (director) (born 1966), director and former actor in pornographic films; **Jonathan Morgan** (politician) (born ...

Jonathan Morgan (director) - Wikipedia
https://en.wikipedia.org/wiki/Jonathan_Morgan_(director)
Jonathan Morgan (born February 5, 1966) is a former actor and current director of pornographic films. He has directed for studios including Wicked Pictures, ...

Jonathan Morgan Profiles | Facebook
https://en-gb.facebook.com/public/Jonathan-Morgan
View the profiles of people named **Jonathan Morgan**. Join Facebook to connect with **Jonathan Morgan** and others you may know. Facebook gives people the ...

Jonathan Morgan | Professional Profile - LinkedIn
https://uk.linkedin.com/in/jonathan-morgan-a561913
View **Jonathan Morgan's** full profile. It's free! Your colleagues, classmates, and 500 million other professionals are on LinkedIn. View Jonathan's Full Profile.

We make assumptions about black boxes

- We assume a **shared context**
  - When my mom asks “why didn’t you like my post on Facebook?”
- We assume **legitimacy**
  - Even though companies and hackers constantly try to deceive or confuse us
- We assume **good faith**
  - Even when Facebook messes with our emotions
- We assume **competence**
  - Even when we’re shown evidence that algorithm designers frequently mess up

Biases are embedded in algorithms

Any algorithmic model is, by definition, a simplification or abstraction of the real world, made by humans. Bias can be introduced all over the place.

- Sample/training data doesn't represent population, excludes relevant variables
- Human-labeled data embeds subjective judgements
- Model may memorize instead of learn (over-fitting)
- Feature selection, weighting embeds assumptions of importance
- False positive/negative rate not equally distributed among subgroups
- Models can degrade without maintenance/monitoring (see: Google Flu Trends)

Social consequences of bias

Biased algorithms can *actively* discriminate against people, the discrimination may be intentional or unintentional (on the part of the designers).

- **Ex 1:** using race as a features in an actuarial model
- **Ex 2:** building a face-recognition algorithm that doesn’t recognize darker skin

Perpetuating systemic bias

Algorithms can reinforce systemic biases even if they don't *explicitly* discriminate based on things like race, gender, etc..

- **Ex 1:** US News College Rankings incentivize colleges to admit wealthier students
- **Ex 2:** Amazon 1-day delivery less available in non-white neighborhoods

Creating filter bubbles

Algorithms that determine what we see (and don’t) skew our view of the world.

- They decide what information we want and need to see for us
- They hide opposing viewpoints, opinions, and perspectives
- They deprive us of information necessary to make informed decisions
- They create a picture of reality that reinforces our existing beliefs, or change our beliefs without our knowledge

Ethical AI

There is no formal definition of what it means to do AI in an ethical manner, but based on these examples and what you already know about bias, we can derive some ‘best practices’:

- “First, do no harm”
- Get informed consent, when applicable
- Consider sources of bias in data, algorithm, and applications
- Consider social consequences (discrimination, disenfranchisement, automation)
- Take **accountability** for negative outcomes (even unintended outcomes) of what you build, and taking concrete steps to address them

Algorithmic accountability

"Determining who is the trusted decision-maker between algorithmic engineers, algorithms, and users requires careful consideration of what the algorithm claims to do and who suffers from the consequences of mistakes. When an algorithm is making decisions or helping an expert make decisions, it becomes unclear who is ultimately responsible for the effects of those decisions."

This isn't *just* a legal issue; it's also an ethical issue. And a <u>*design*</u> issue.

Donovan, J., Caplan, R., Matthews, J., & Hanson, L. (2018). *Algorithmic accountability: A primer*. Data & Society, 501(c).

Opening up the black box: Algorithmic transparency and interpretability

Algorithmic transparency

- The model (code) and training/test data are publicly inspectable
- Individual algorithmic decisions are reproducible
- Changes are logged and version controlled

Algorithmic interpretability

People other than you should be able to...

- Understand how the model works (input data, features, basic mechanics)
- Understand how specific determinations/predictions/classifications were made
- Glean insights from the model and communicate those insights effectively to *other* non-subject-matter experts

Algorithmic interpretability

People other than you should be able to...

- Understand how the model works (input data, features, basic mechanics)
- Understand how specific determinations/predictions/classifications were made
- Glean insights from the model and communicate those insights effectively to *other* non-subject-matter experts

Question: *interpretable… to whom?*

Algorithmic interpretability

There are no hard rules for making an model more interpretable, because it's a context thing (and an audience thing, and a purpose thing…). But some strategies include:

- Use simpler models
- Use concrete features of the input data, rather than composite features
- Use fewer features
- Provide supporting documentation written for non-data scientists
- Make it easy for people to explore various inputs/outputs of your model

Example: ORES

The ORES (Objective Revision Evaluation Service) platform you are using for assessing article quality in Assignment 2 is an example of a *very transparent* algorithm:

- Model info: https://ores.wmflabs.org/v2/scores/enwiki/wp10/?model_info
- Model code: https://github.com/wiki-ai
- Documentation: https://www.mediawiki.org/wiki/ORES
- Sandbox: https://ores.wikimedia.org/v3/#!/scoring/get_v3_scores_context_revid_model

Example: ORES

ORES is also a fairly *interpretable* model

- The classification task the model performs is relatively straightforward
- The features and scores used to determine the quality of a specific article are presented in a fairly human-readable way
- You can test it out yourself, on real data, to understand its strengths and weaknesses
- You can even ‘inject’ different feature values to see how that changes the prediction

Example: ORES

But interpretability is contextual. So we must ask...

- Who needs to be able to interpret ORES's output? → *audience*
- What task do they need to interpret it for? → *purpose*
- How/where/when will they be interacting with the model? → *context*

And... how will we decide when the model is interpretable 'enough'?

Why be transparent and interpretable?

It may soon be required by law

- Example: EU General Data Protection Regulation gives *data subjects* of machine learning systems a right to explanation

Often it's the ethical thing to do

- Example: Facebook suggests you 'friend' a long-lost relative, but won't tell you what information they used to make the recommendation

Plus, if other people understand your model, they can give you useful feedback to make your model better.

Why be transparent and interpretable?

If your audience believes they understand your model, they are more likely to trust your model, and use your model.

“A model’s total performance is the product of the model predictive performance times the probability that the model will be used. One needs to optimize for both.”

- Carl Anderson, *The role of model interpretability in data science*

Legitimate trade-offs and limitations

There are legitimate reasons scientists and companies do not make their models fully transparent and interpretable.

- You need to use a more complex/opaque model because simpler/more interpretable models don’t perform well enough
- You are concerned about people gaming or undermining the system if they know exactly how the model works
- You are concerned other people could use your model for nefarious purposes
- Your model is your intellectual property and you need to make a living

Illegitimate trade-offs and limitations

There are also *less than legitimate* reasons scientists and companies do not make their models fully transparent and interpretable.

- It takes time and effort.
- If people knew how your model worked they would not use your product; you would be publicly shamed and/or arrested.
- You think you can get away with some ‘token’ transparency, which may or may not provide people with accurate or useful information about how your model works.

Your information

Close

About you | **Your categories**

The categories in this section help advertisers reach people who are most likely to be interested in their products, services, and causes. We've added you to these categories based on information you've provided on Facebook and other activity.

Away from family	Away from hometown
Birthday in November	US politics (very liberal)
Facebook access (mobile): smartphones and tablets	Frequent Travelers
African American (US)	Facebook access (mobile): smartphones
Facebook access (OS): Mac OS X	Gmail users
Facebook access (network type): WiFi	Facebook access (mobile): all mobile devices

See More

Evaluating the black box from the outside: Auditing algorithms

History of audits

Audit: “an official examination by an independent body”

Audit study (social science): “a field experiment where researchers participate in a *social process* that they suspect to be corrupt in order to diagnose harmful discrimination”

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). *Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms*. ICA 2014 Data and Discrimination

History of audits

Audit studies leave the black box intact

- analyzes output based on controlled input
- evaluates output against pre-defined normative or empirical criteria

'Scholarly consensus' is that these studies do not require informed consent as long as the benefits to society (revealing and discouraging discriminating behavior) outweigh the potential harms to the people or organizations that were audited:

- Embarrassment
- Wasted time
- Lost reputation or revenue

Types of algorithm audits

- **Code audit:** code & data published, or shared with independent investigators
- **Non-invasive user audit:** individual users share their interactions with a platform and/or allow their behaviors and the results of those behaviors to be tracked or recorded
- **Scraping audit:** researchers programmatically provide the algorithm with a large variety of different inputs and record the outputs
- **Sockpuppet audit:** researchers programmatically emulate the behaviors of real users and record the result
- **Crowdsourced audit:** researchers recruit a large number of confederates to interact with the platform and track/survey the results

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). *Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms*. ICA 2014 Data and Discrimination

Uses of auditing

- Detecting discriminatory bias (duh)
- Sanity checking results
- Identifying limitations, edge cases
- Evaluating appropriateness for particular use cases
- Gathering feedback for iteration (or hitting the “emergency stop” button)

Limits of auditing

- Scale
- Non-random sampling
- Less control over experimental conditions
- May be in violation of TOU/federal law

Auditing case study: Google’s Perspective API

https://www.perspectiveapi.com

Overview

From perspectiveapi.com:

> *"Discussing things you care about can be difficult. The threat of abuse and harassment online means that many people stop expressing themselves and give up on seeking different opinions.*
>
> *Perspective is an API that makes it easier to host better conversations.*
>
> *The API uses machine learning models to score the perceived impact a comment might have on a conversation."*

Perspective predicts toxicity of comments

> *“This model was trained by asking people to rate internet comments on a scale from "Very toxic" to "Very healthy" contribution.”*
>
> *“Toxic is defined as... "a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion."*

Intended use cases include automated and semi-automated comment moderation and filtering.

Question: ***What are some other potential use cases for an online toxicity detector?***

Reception upon first release

Wired magazine is, as usual, is naively optimistic

> *“The numbers reveal everything from the trolliest time of day to the nastiest state in the union.”*

...which, apparently, is Vermont... with neighboring New Hampshire taking the ‘least toxic’ title.

Reception upon first release

Violet Blue, writing for Engadget, takes a more critical perspective:

> *"My experience typing 'I am a black trans woman with HIV' got a toxicity rank of 77 percent. 'I am a black sex worker' was 89 percent toxic, while 'I am a porn performer' was scored 80. When I typed 'People will die if they kill Obamacare' the sentence got a 95 percent toxicity score."*

In-Class Activity:
Auditing the Perspective API

Groups of 2-3

But first... homework due next week

Reading reflection

- Astrid Mager. 2012. *Algorithmic ideology: How capitalist society shapes search engines*

Assignment 3: Mechanical Turk Ethnography

- **Length:** at least 2000 words
- **Format**: Google Doc, shared with Jonathan and Os, link submitted to Canvas
- **Due date:** November 8, by 5pm

The assignment sheet for A3 has been shared with you, and is also linked from https://wiki.communitydata.cc/Human_Centered_Data_Science_(Fall_2018)/Assignments#A3:_Crowdwork_ethnography

Assignment overview

Goals

- Assess the strengths and weaknesses of the Perspective APIs models by auditing its predictions across four different datasets
- Hypothesize how the model works, based on inputs and outputs
- Develop different (better?) definitions of toxic speech based on the context of use reflected in a particular dataset.

Output

- A post to Canvas discussion “Week 6 in-class activity”
- A Google doc with process notes and outcomes, uploaded to Canvas before tomorrow night.
- Include everyone’s name on the Canvas post AND the Google doc

Models

1. **TOXICITY**
2. **SEVERE_TOXICITY**
3. **IDENTITY_ATTACK**
4. **INSULT**
5. **PROFANITY**
6. **THREAT**
7. **SEXUALLY_EXPLICIT**
8. **FLIRTATION**

Datasets

1. **Ferguson August 8-10 2014:** Tweets that mention ‘Ferguson’ in the immediate aftermath of the fatal shooting of Michael Brown by police.
2. **#Unitetheright August 4-15 2017**: tweets that use the #unitetheright hashtag before and during the Charlottsville rally
3. **#AmplifyWomen October 4 2017:** tweets using the #amplifywomen hashtag in response a Twitter boycott that started in support of actress Rose McGowan
4. **Trump tweets 2015-2018:** things that the president of the United States of America says on Twitter

Getting started

1. Open all documents linked in the Canvas announcement called “Week 6 in-class activity links”
2. Make a copy of the data spreadsheet and share it with group members (useful for sorting/filtering/highlighting/note-taking in sheet)
3. Read through the README (info on the dataset you were assigned, and the models available).
4. Read the instructions doc. You’ll be working through the steps from 1-4. We will stop to report out and discuss between each step.

Assignment steps

1. **Re-ranking the most toxic tweets:** what are the false positives? *Why* are they false positives? Can you and your group members agree?
2. **Looking for false negatives:** What kinds of toxicity is the algorithm missing? Why are these toxic?
3. **Opening the black box:** what can we infer about the way Perspective works by visually inspecting its output?
4. **Understanding toxicity in context:** can you formalize what toxicity means in the context of this dataset? How does your definition differ from Perspective's? How could it (theoretically) lead to better outcomes?

Trigger warning

Some of the things people say on Twitter are seriously messed up.

These examples were selected because they have a high probability of containing offensive speech.

That likely includes explicit language, hate speech, and potentially attacks or threats against persons or groups.

If you feel uncomfortable completing this assignment for any reason, come up to Os and Jonathan as soon as the activity period begins to receive instructions for an alternate, self-guided in-class activity.

Homework due next week

Reading reflection

- Astrid Mager. 2012. *Algorithmic ideology: How capitalist society shapes search engines*

Assignment 3: Mechanical Turk Ethnography

- **Length:** at least 2000 words
- **Format**: Google Doc, shared with Jonathan and Os, link submitted to Canvas
- **Due date:** November 8, by 5pm

The assignment sheet for A3 has been shared with you, and is also linked from https://wiki.communitydata.cc/Human_Centered_Data_Science_(Fall_2018)/Assignments#A3:_Crowdwork_ethnography

Questions?

Unused slides

Thick data case study

Understanding power users of Wikimedia Commons

Background: Structured Data on Commons

Wikimedia Commons hosts 40+ million free media files, used on Wikipedia and across the web.

Commons runs on MediaWiki, which was developed for longform text, not media.

There is a lot of metadata about media on Commons, but it’s poorly structured.

Wikimedia is building a structured data layer into Commons in order to capture metadata in structured, hierarchical, machine-readable formats.

Background: Structured Data on Commons

Existing un-structured metadata will need to be ported to the new schema. New metadata will need to be captured in a structured format.

depicts

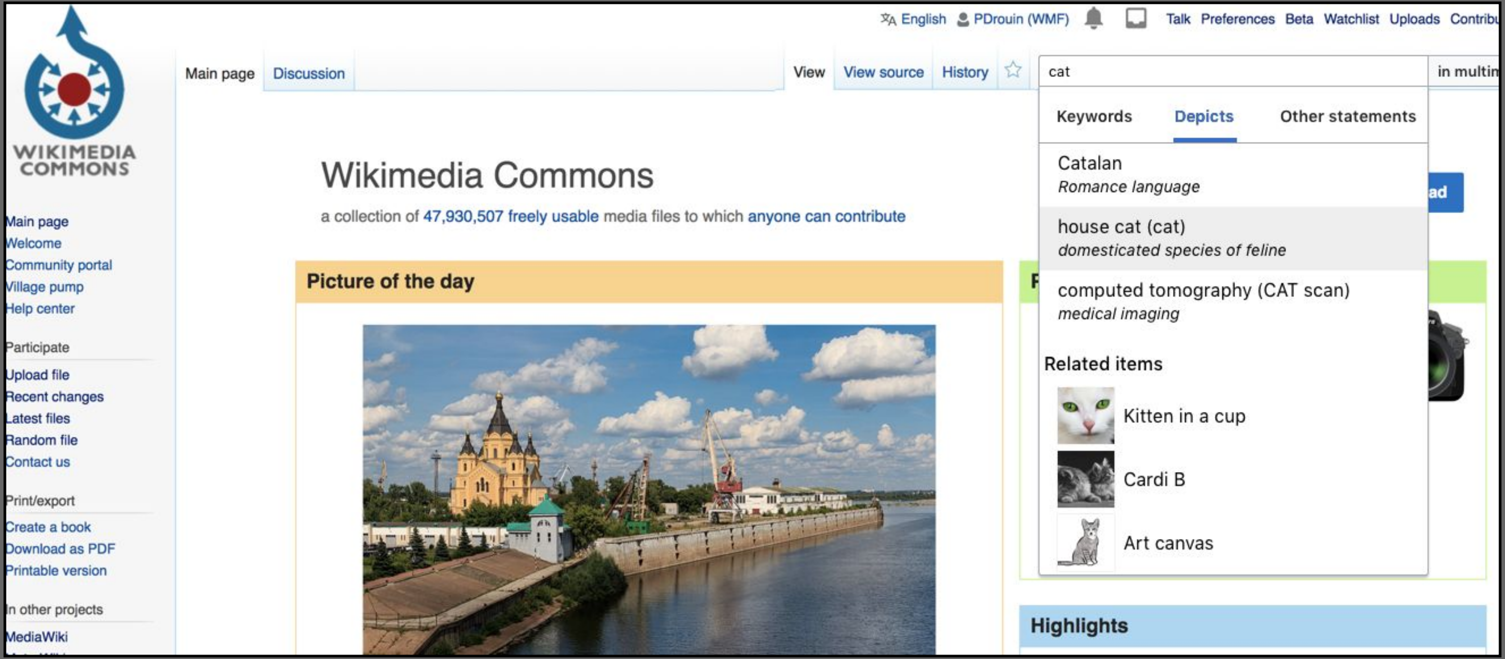
person depicted in Mona Lisa

applies to part	foreground
shown with features	brown hair
	long hair
	smile
	sitting

0 references

Background: Structured Data on Commons

Existing upload, search, and curation tools will need to be redesigned to take advantage of structured metadata.



Problems #1

- We don’t know what kinds of metadata we have
- Besides file type and (some) license info, we don’t know what *media* we have
- We can’t create new classes for every single possible metadata type
- We don’t know what tools are most important to our users
- We can’t update all upload, curation, search, or export tools at once
- We don’t know what problems we’ll *solve* by moving to structured metadata
- We don’t know what problems we’ll *create* by moving to structured metadata

Problems #2

People have been creating hacks and workarounds in order to ingest, structure, and extract file metadata on Commons for 10+ years.

- We don’t know what most of these hacks and workarounds are, why people made them, or what depends on them.
- We don’t know how to programmatically detect, classify, or extract this metadata
- We don’t know which metadata is most important for key stakeholders: *contributors, curators, or re-users*

Research goals

Understand the current practices and unmet needs of key stakeholders re metadata.

Power uploaders: GLAMs* that upload sets of valuable media with rich metadata.

Power curators: volunteers who correct, contribute, and standardize metadata.

Power consumers: orgs that re-publish Commons media outside of Wikipedia

*Galleries, Archives, Libraries, and Museums

Research process

Semi-structured interviews: 11 GLAM, 7 external re-users, 6 curators

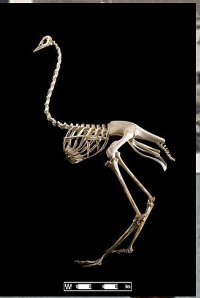
Contextual inquiry: ask interviewees to describe and demonstrate their workflows (share screen), the tools they use and the content they work on (share links).

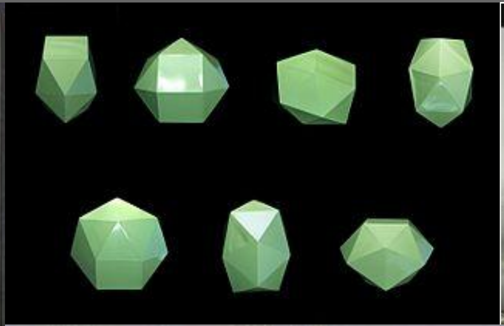
Purposeful sampling: maximize diversity of backgrounds, motivations, workflows.

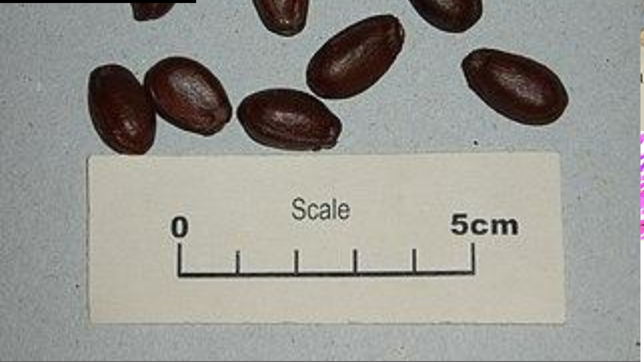
Snowball sampling: ask each interviewee “who should I talk to next?”

Analytical process

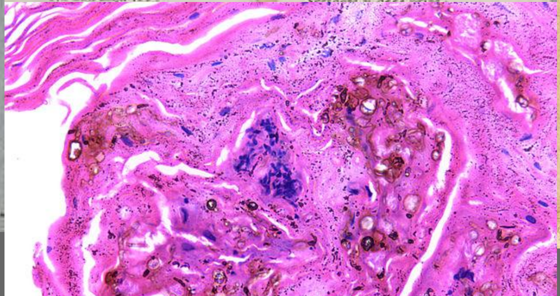
- Read through transcripts of interviews, re-watch interview videos
- Note common patterns, ‘insider’ vocabulary, motivations, pain points
- Pull out illustrative quotes and examples
- Group patterns and examples, distill them into larger more generic themes
- Discuss findings with subject matter experts
- Prioritize <u>themes</u> based on importance to *users* and *project* (i.e. business goals)
- Connect all claims back to data: notes, quotes, and examples






Scale
0
5cm


URSULA bij het UURWERK.
Daar slaat het zes!
VINCENT met het VAANDEL.
Vincent met het vaantje




14/II 90
ROUSTCHOUK ET CHOUMLA

Size of this JPG preview of this OGG file: 800 × 450 pixels. Other resolutions: 320 × 180 pixels | 640 × 360 pixels | 1,024 × 576 pixels | 1,280 × 720 pixels | 1,920 × 1,080 pixels.

Original file (Ogg multiplexed audio/video file, Theora/Vorbis, length 24 min 24 s, 1,920 × 1,080 pixels, 3.53 Mbps overall)

Summary [edit]

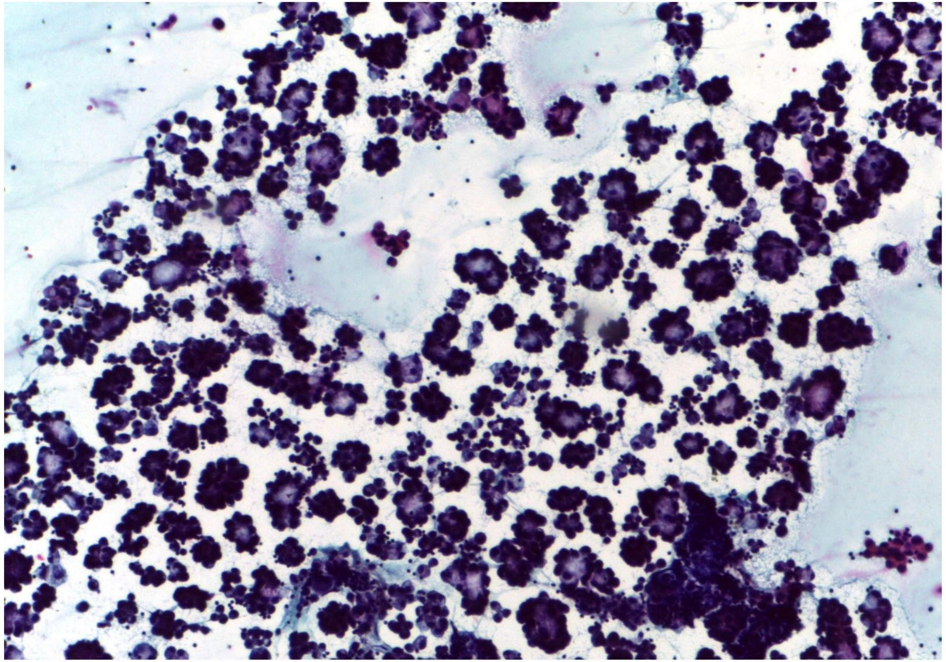
Author	VPRO
Title	Guy Consolmagno: Extra footage 1
Date	1 January 1960
Medium	Moving Image
Dimensions	PT24M23S
Current location	VPRO; Nederlands Instituut voor Beeld en Geluid
Accession number	oai:openimages.eu:1026848
Place of creation	Vatican City
Credit line	Guy Consolmagno
Notes	geology; meteorites; cosmology; physics; astronomy; Space exploration; Space; Science; philosophy; Religion; Christianity; Guy Consolmagno explains his fascination for meteorites and talks about combining science and religion; closeups of the Vatican meteorite collection; examples of special kinds of meteorite; a 3D scan of a meteorite

"I could have created categories for each tag, but would have been entirely manual and taken a huge amount of time. But we wanted to capture it somehow, so that we could theoretically go back later."

File:Adenocarcinoma cytocentrifuge preparation 100x.jpg

From Wikimedia Commons, the free media repository

File | File history | File usage on Commons | Metadata



Download all sizes

Use this file on the web

Use this file on a wiki

Email a link to this file

Information about reusing

Size of this preview: 800 × 562 pixels. Other resolutions: 320 × 225 pixels | 640 × 450 pixels | 1,024 × 720 pixels | 1,280 × 900 pixels | 1,765 × 1,241 pixels.

Original file (1,765 × 1,241 pixels, file size: 675 KB, MIME type: image/jpeg); ZoomViewer: flash/no flash

Open in Media Viewer

Summary [edit]

Description	**English:** Low power view showing malignant cells arranged in glandular pattern. Note the high cellular yield compared to normal cytology fluid preparation (Papanicolaou, 100X)


Coração de Suíno
Técnica: Diafanização
1 2 3 4 5 10 15cm

“I don’t have a complete knowledge of how categories in Commons work. I have this big generic collection of media from the museum. I came up with the categories by thinking 'if I were looking for this media, which steps would I take?'”

Category:Glass_plates

+

Category:Baldomer_Gili_i_Roig

+

Category:Collections_of_the_Museu_d'Art_Jaume Morera

=

Category:Glass_plates_by_Baldomer_Gili_i_Roig_at_Museu_d'Art_Jaume Morera

Follow up

- Use findings to define design requirements for new UI and tools
- Use themes to develop survey questions for a more representative sample
- Use themes to develop scenarios and personas for product teams
- Use notes to develop a list of important metadata for different kinds of media
- Use examples and quotes to promote empathy and shared understanding
- Share findings with stakeholders to facilitate development conversations
- Identify opportunities for (semi) automatic metadata ID and extraction