



# Computerized Knowledge Production: Machine Learning Models as Social Actors

Stephen Paff



# Anthropology of Machine Learning

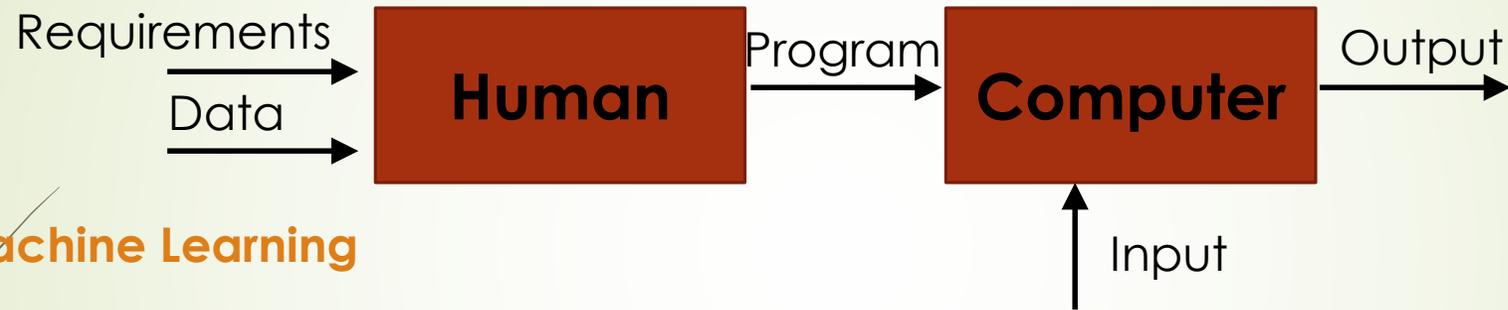
- ▶ For this presentation, I am interested in the anthropology of machine learning.
- ▶ One, major formal definition of Machine Learning:

“Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time” (Simon).
- ▶ In short, machine learning refers to:

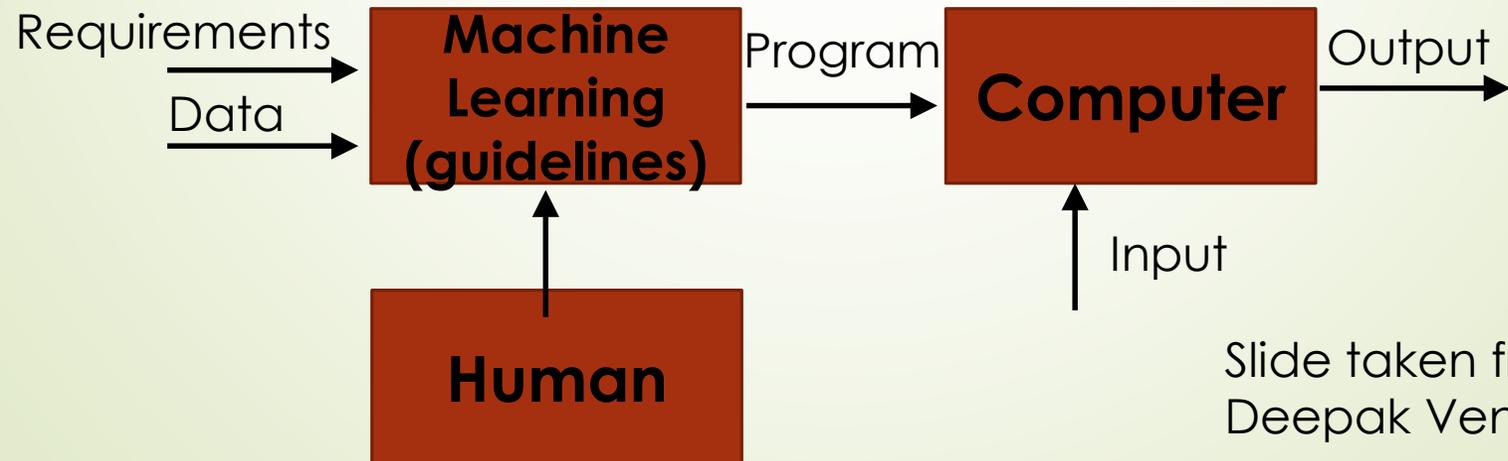
Algorithmic system which learns by developing its own way to analyze data (in the same or new context) by adapting over iterations

# Shift in Relationship in Computer Science

## Traditional Programming



## Machine Learning



Slide taken from presentation by  
Deepak Venugopal



# Key Statement:

- ▶ Data scientists study (big) data through machine learning (primarily).

Just like:

- ▶ Anthropologists study culture through ethnography (primarily).

Who	What	How
Data scientists/data science	Big data	Machine Learning
Anthropologists/Anthropology	Culture	Ethnography



# Thesis:

We, as anthropologists (and other social researchers), need:

1. Analysis: To understand what machine learning, data science, and big data look like on the ground through ethnography
2. Methodology: To work to bridge the gap between machine learning approaches and ethnography
3. Theory: To shift our understandings of who or what participates in a culture and society to include increasingly intricate computerized learning systems.

Which of these three arguments would you prefer I discuss?



# Argument 1: On the Ground Focus

Main Argument:

We need to understand what machine learning, data science, and big data look like on the ground through ethnography.



# Away from Big Data

- Most anthropological work on this topic has focused on big data.
- Big data promotes an airplane view of the topic, stifling a deep, on-the-ground, ethnographic understanding.
- Shifting focus to machine learning and data scientists resituates analysis on the ground.



# Data Myths

Two common myths surrounding these phenomena:

1. Gods: Saviors revolutionizing our world.
2. Demons: Threatening our lives (anthropologists generally here)

Both encourage a single, totalizing view, which gets in the way of understanding what they look on the ground.



# An Ocean of Data

- ▶ Nick Seaver (2015) in “Bastard Algebra” uses an elaborate ocean analogy to describe the different facets involved:
  - ▶ The Ocean Depths of Big Data
  - ▶ The Shore where most data scientists and machine learners operate
    - ▶ We need a “frog’s eye view.”
    - ▶ Where most data scientists and machine learning occur
- ▶ Main Point for my Presentation:

These systems are made up of many distinct ecosystems all of which require ethnography to understand. The big data view, which looks the whole ocean as a whole, is only one vantage point.



# Faith/Trust in Machine Learning

Ethnographically analyze how faith in machine learning drive action in the here and now

Examples:

- ▶ Thomas, Nafus, and Sherman (2018)
- ▶ Bucher (2016)

“People position algorithms in ways that make algorithms promise more than they can deliver in strictly material terms.... This is the moment of social creativity when faith in a promise delivers possibility” (Thomas 2018:4).



# Ethnography is key

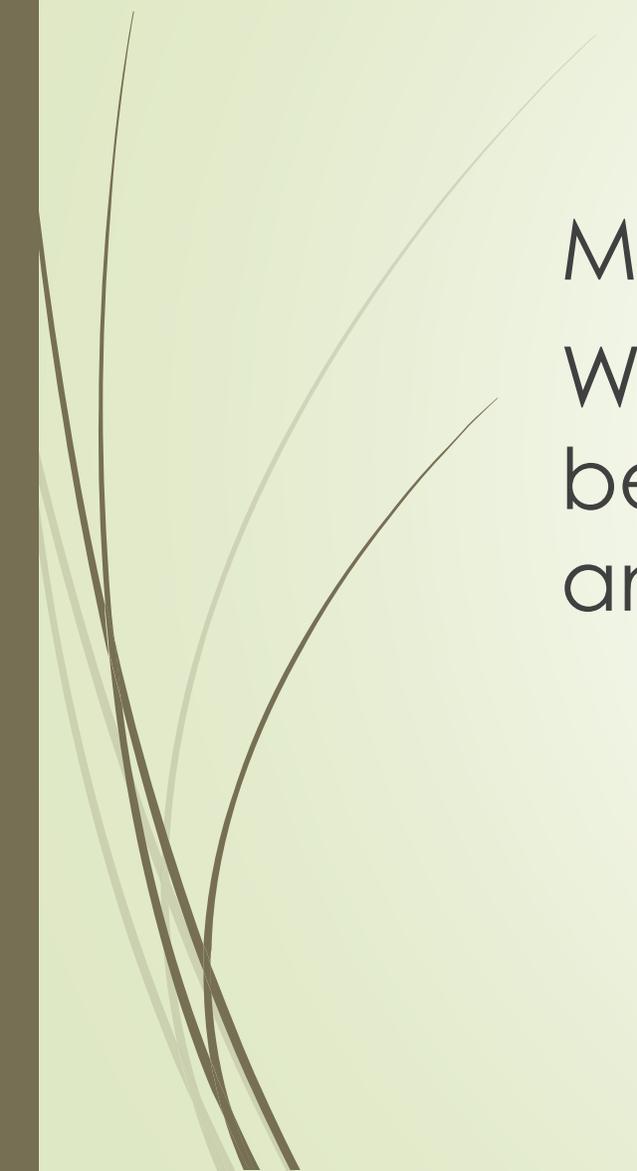
- ▶ Ethnography is still the best way to understand what these systems look like on the ground.
  - ▶ In this topic, this is best done by refocusing ethnographic research on machine learning and data scientists, instead of big data.
- 



## Argument 2:

Main Argument:

We need to work to bridge the gap between machine learning approaches and ethnography





# Cross-Collaboration



- ▶ Anthropology and data science have much they should learn from each other and need to speak to/with each other more.
- ▶ Because this is an anthropology conference, I will focus on how what anthropologists can learn from data science.
- ▶ I will focus on three areas:
  1. How machine learning can shed new light into the quantitative/qualitative debate
  2. Creative ways to use machine learning in ethnographies
  3. Ways to work along side of machine learners in algorithm development



# Qualitative/Quantitative Debate

Common dichotomies based on qualitative/quantitative debates:

Qualitative Research	Quantitative Research
Inductive (open-ended)	Deductive (close-ended)
Flexible/Ground-Up	Prescriptive/Top-down
Subjective/Touchy-feely	Logical/mathematical

Created a false association of mathematics with deductive, top-down thinking.

Many machine learning techniques break these simple molds.



# New Approach

Many machine learning systems (particularly unsupervised learning) are:

- ▶ Both mathematical and inductive
- ▶ Like ethnography, building self-constructing approaches, not top-down
- ▶ Interpretivist research questions
- ▶ Could connect well with ethnography and could shed new light into these well-worn qualitative/quantitative battles.

(Not all are and some model more typical quantitative approaches.)



# Machine Learning

Several scholars have developed strategies for using machine learning in ethnographies:

1. Giaccardi's (2014) Thing Ethnography
2. Faßler's (2013) Inter-creativity
3. Geiger's (2011) Trace Ethnography
4. Haines's (2017) Multi-Dimensional Ethnography
5. Madsen's (2018) Transversal Collaboration

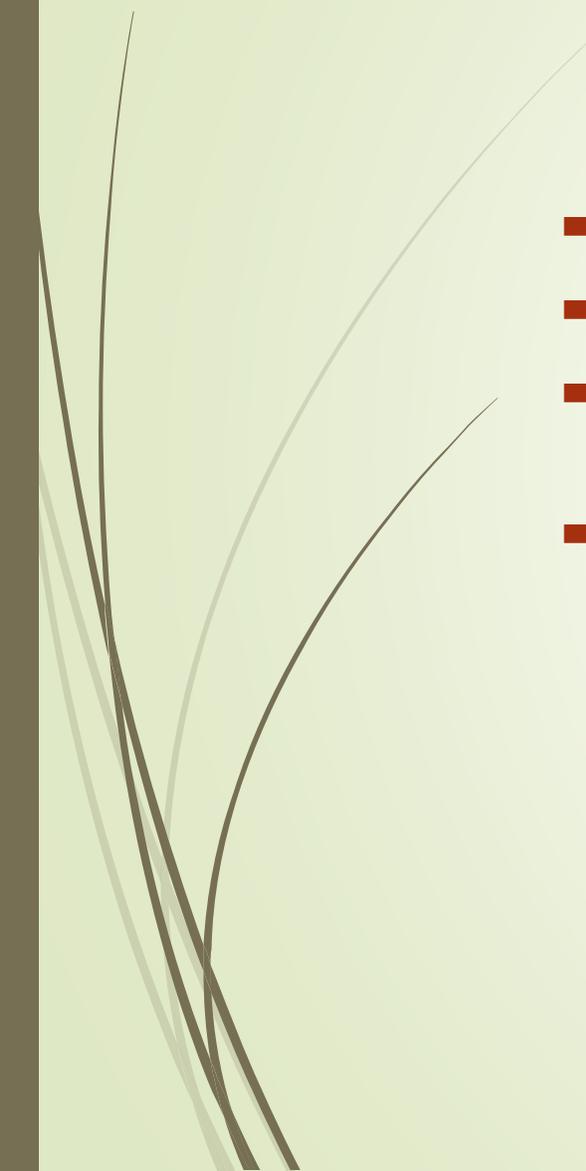


# Ethnographies of Machine Learning

- ▶ Ethnographies of machine learning systems to help develop and (re)design their development.
  - ▶ E.g. Eslami (2106) analyzed folk theories of the Facebook Newsfeed
- ▶ Provides a great venue for addressing ethical and design issues on the ground



# New Conversation

- ▶ Anthropologists have been rightfully fearful of quantitative research.
  - ▶ Our animosity may cause us to ignore the important subgroups within
  - ▶ Machine learning provides a significant perspectival shift within quantitative research
  - ▶ We should use to provide new conversations
- 



## Argument 3:

### Main Argument:

Machine learning algorithms provide one major context in which to broaden our understanding of what has agency and what participates in a culture.

### Because:

They develop their own “knowledge” outside of the immediate domain of the human creators.



# Broader Theoretical Debate

- ▶ There has been a wider anthropological debate about what has agency.
- ▶ E.g. Latour (1993) and Keane (2006), the narrative of modernity hinges on regulating agency only to humans, specifically away from the material world.
- ▶ But do non-human entities have agency?
- ▶ If so, can they participate in a culture?



# Agency

My quick definition of agency is:

The ability to influence (aka be in relationship to) other entities





# Non-Human Agency

- ▶ Two broad examples of non-human agents:
  1. Material things/entities (e.g. Latour)
  2. Social processes/structures (e.g. Foucault)
- ▶ But machine learning algorithms straddle both.
- ▶ E.g. machine learning can be considered a technology both as a computer technology and technology of power.



# Computer Learning

- ▶ Machine learning algorithms do learn, generating unique insights as they iterate through data in ways not directly anticipated by humans
  - ▶ The “knowledge” they learn is not always understandable by humans.
  - ▶ They communicate with humans and act/react along with humans.
  - ▶ As such, do these systems participate in the cultures in which they are created/used?
- 



# Machine Learners

- ▶ Adrian Mackenzie (2017) conducts an archaeology of machine learning
- ▶ For him, the primary agent is a human-computer relationship, created by machine learning processes
  - ▶ Out of which new human and computer subjectivities, identities, and relationships emerge
  - ▶ His concept of Machine Learners (the primary entity in this relationship) refers both to human coders and machine learning algorithms



# Computerized Agents



- ▶ Machine learning systems seem to have agency and participate in cultures.
- ▶ We need to reconceptualize our understanding of both to understand how a computer program could have agency.
- ▶ They also blur implied distinctions between objects and institutions. These distinctions of things need to be redefined.



# Conclusion:

Anthropologists should explore machine learning anew in order to revitalize their understanding of the three part phenomena of machine learning, data science, and big data

Thesis:

We, as anthropologists and other social analysts, need to:

1. To understand what all three look like on the ground fully through ethnography, especially the first two: data scientists, machine learning, and big data.
2. To bridge the gap between machine learning analyses and ethnography
3. To shift our understandings of who or what participates in a culture and society to include increasingly intricate computerized learning systems.



# Bibliography:

For my Bibliography (in Google Drive), go to the following link:

<https://bit.ly/2uBMlrD>