Human-Centered Artificial Intelligence

CS 347
Michael Bernstein
Announcements

Project abstracts due 11:59pm

Quiz 3 on Thursday

You just finished your last two-reading day of the quarter
Artifacts have politics: the systems we create influence groups and societies, often with undesirable outcomes

Example: gig economy — potential of upward mobility and community social capital, but not currently implemented in a way that unlocks those possibilities

Design approaches focused on marginalized groups, such as feminist HCI, center these communities’ needs in the design process

Algorithmic systems, not just designed systems, similarly have impact. People struggle to reason about them, and industry struggles to avoid mistakes.
Social Computing
Unit 3

social media
collaboration
society
Where we go from here

so far  Ubiquitous Computing, Design, Social Computing
week 6  Software
week 7  Cognition
week 8  Methods
week 9  Society
week 10  Something old, something new

These units will build on the three larger units we’ve covered so far, adding their own insights and flavors.
Software

Unit 4

human-centered AI
development, tools and toolkits
media authoring
Today

AI vs. IA

Direct manipulation vs. Agents

Mixed-initiative interaction

End-user AI authoring

AI and design
People: where AI lives or dies

[Breazeal 2004]  [Dragan, Lee, and Srinivasa 2013]
...but we need to think carefully

[Mok et al. 2015]
“Don’t let your UI write a check that your AI can’t cash.”

- Eytan Adar [2018]
Intelligence Augmentation

AI vs. IA
A reaction to:

“AI will replace human intelligence”

Intelligence augmentation says that replacement is the wrong approach.
Algorithms in practice: Comparing web journalism and criminal justice

Angèle Christin

Abstract

Big Data evangelists often argue that algorithms make decision-making more informed and objective—a promise hotly contested by critics of these technologies. Yet, to date, most of the debate has focused on the instruments themselves, rather than on how they are used. This article addresses this lack by examining the actual practices surrounding algorithmic technologies. Specifically, drawing on multi-sited ethnographic data, I compare how algorithms are used and interpreted in two institutional contexts with markedly different characteristics: web journalism and criminal justice. I find that there are surprising similarities in how web journalists and legal professionals use algorithms in their work. In both cases, I document a gap between the intended and actual effects of algorithms—a process I analyze as “decoupling.”

Second, I identify a gamut of buffering strategies used by both web journalists and legal professionals to minimize the impact of algorithms in their daily work. Those include foot-dragging, gaming, and open critique. Of course, these similarities do not exhaust the differences between the two cases, which are explored in the discussion section. I conclude with a call for further ethnographic work on algorithms in practice as an important empirical check against the dominant rhetoric of algorithmic power.

Keywords

Algorithms, ethnography, work practices, organizations, journalism, criminal justice
Introduction

Overview about Program

RIG as an "Instrument"

Control Techniques

RIG Implementation

Usage

Activities

Credits
AUGMENTING HUMAN INTELLECT: A CONCEPTUAL FRAMEWORK

Prepared for:
DIRECTOR OF INFORMATION SCIENCES
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
WASHINGTON 25, D.C.

By: D. C. Engelbart

CONTRACT AF 49(638)-1024

STANFORD RESEARCH INSTITUTE
MENLO PARK, CALIFORNIA
Artificial Intelligence

Replace human intelligence with artificial intelligence

Intelligence Augmentation

Augment human intelligence with artificial intelligence
Examples we’ve discussed

Help me understand where I’m using water in my household
Realize my sketched mechanical design into a rough functional system
Connect me with jobs or movies that I might want to see
Show me behavior patterns that are influencing my health
But who should lead this dance? How much control should we yield to the AI? This leads to a debate…
Agents vs. Direct Manipulation

[Shneiderman and Maes 1997]
Software agents

We should delegate to proactive artificial intelligence systems

Pattie Maes, MIT Media Lab

Direct manipulation

Users should always have full control, even as automation increases

Ben Shneiderman, U. Maryland
Agents

AI agents ask questions about images on social media to learn about the world around them [Krishna et al. 2022]

Learn to automate tasks that you do commonly [Maes 1995]
Direct manipulation

Shneiderman: it is possible to maintain high levels of user control even as automation increases [Shneiderman 2022]

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>music box</td>
<td>airbag pacemaker</td>
</tr>
<tr>
<td>High</td>
<td>bicycle piano</td>
<td>camera</td>
</tr>
</tbody>
</table>
Agency plus automation [Heer 2019]

Generalize the user's inputs (selecting text "Alabama") into scripts

<table>
<thead>
<tr>
<th>#</th>
<th>Year</th>
<th>Year</th>
<th>extract</th>
<th>Population</th>
<th>Property_crime_rate</th>
<th>Burglary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2004</td>
<td>Alabama</td>
<td>extract</td>
<td>6052375</td>
<td>981</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2005</td>
<td>Alaska</td>
<td>extract</td>
<td>5099309</td>
<td>596</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2006</td>
<td>Arizona</td>
<td>extract</td>
<td>4627851</td>
<td>908.9</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2007</td>
<td>Arkansas</td>
<td>extract</td>
<td>4661900</td>
<td>908.9</td>
<td></td>
</tr>
</tbody>
</table>

Suggest alternative visualizations
Mixed initiative interaction

Eric Horvitz keeps listening to the agents vs. direct manipulation debate. He decides that he's had enough and that it's a false dichotomy…
Mixed-initiative, intuitively

You don’t need to decide between full control and full automation. Instead, the system should automate the things it can, hand control to the user for the things it can’t, and ask the user if it’s unsure.

Today, mixed-initiative interaction typically refers to the mode of suggesting an action and letting the user confirm it.
Horvitz envisioned mixed-initiative more broadly as trading off dynamically between all options, using utilities:

\[ u(A, G) = \text{(positive) utility of taking an automated action when the goal is correctly guessed} \]

\[ u(A, \neg G) = \text{(negative) utility of taking the same action when the goal is incorrectly guessed} \]

\[ u(\neg A, G) \text{ and } u(\neg A, \neg G) \text{ similarly} \]

<table>
<thead>
<tr>
<th></th>
<th>Desired goal</th>
<th>Not desired goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take action</td>
<td>( u(A, G) )</td>
<td>( u(A, \neg G) )</td>
</tr>
<tr>
<td>No action</td>
<td>( u(\neg A, G) )</td>
<td>( u(\neg A, \neg G) )</td>
</tr>
</tbody>
</table>
Now, take expected values

[Horvitz 1999]

What’s the expected value of taking action?

\[ P(G) \cdot u(A, G) + P(\neg G) \cdot u(A, \neg G) \]

What’s the expected value of taking no action?

\[ P(G) \cdot u(\neg A, G) + P(\neg G) \cdot u(\neg A, \neg G) \]
Mixed initiative: visually

If it's unlikely that the user has the given goal

If it's likely that the user has the given goal
Mixed initiative: visually

\[ u(\neg A, \neg G) \]

\[ u(A, G) \]

\[ u(A, \neg G) \]

\[ u(\neg A, G) \]

Utility of inaction

Expected value

\[ P(G) \]
Mixed initiative: visually

Utility of action

Utility of inaction

$u(A, G)$

$u(\neg A, \neg G)$

$u(\neg A, G)$

$u(A, \neg G)$

Expected value

$P(G)$
Mixed initiative: visually

Utility of action

Utility of inaction

Expected value

Higher utility not to act

Higher utility to act

$u(\neg A, \neg G)$

$u(A, G)$

$u(A, \neg G)$

$u(\neg A, G)$

Expected value

Utility of action

Utility of inaction

$P(G)$

0

1
What if making an error is costly?

\[ u(\neg A, \neg G) \]

\[ u(A, G) \]

\[ u(A, \neg G) \]

\[ u(\neg A, G) \]

Utility of action

Utility of inaction

Expected value

\( P(G) \)

(This moved down!)
What if making an error is costly?

Now we only take action if we are even more certain that we correctly estimated the user's goal.
What if we ask the user?

Asking often carries lower risk, but also lower utility

$$u(\lnot A, \lnot G)$$

$$u(Ask, \lnot G)$$

$$u(A, \lnot G)$$

Expected value

Utility of action

Utility of inaction

Utility of asking

$$u(A, G)$$

$$u(Ask, G)$$

$$u(\lnot A, G)$$

$$P(G)$$
What if we ask the user?

Asking often carries lower risk, but also lower utility

Expected value

Utility of action

Utility of inaction

Utility of asking
So, when does this screw up?

When the system cannot accurately assess the probability of the user having the goal \( P(G) \)

or

When the utilities are not correctly estimated

  e.g., too high a utility for asking if the user doesn’t have the goal \( G \).
  “Are you writing a letter right now?”
A problem has been detected and Windows has been shut down to prevent damage to your computer.

The problem seems to be caused by the following file: kbdhid.sys

MANUALLY_INITIATED_CRASH

If this is the first time you've seen this stop error screen, restart your computer. If this screen appears again, follow these steps:

Check to make sure any new hardware or software is properly installed. If this is a new installation, ask your hardware or software manufacturer for any Windows updates you might need.

If problems continue, disable or remove any newly installed hardware or software. Disable BIOS memory options such as caching or shadowing. If you need to use safe mode to remove or disable components, restart your computer, press F8 to select Advanced Startup Options, and then select Safe Mode.

Technical Information:

*** STOP: 0x000000e2 (0x00000000, 0x00000000, 0x00000000, 0x00000000)
End user authoring of artificial intelligence
If you wanted a private smart doorbell...

To automatically control entrance to your room to let in possible donors for your Stanford education
How would you train the system quickly?
Crayons: camera-based interaction

[Fails and Olsen 2003]

“The one that started it all”: direct-manipulation training
Frontier: image editing through demonstration

“Make this part of the source image look more like the reference image.” [Ko et al. 2022]
Interactive training

[Fogarty et al. 2008]

Allow users to keep training and re-training by drag-dropping instances into positive and negative classes as they go.
Revising your training as you go

[Chang, Amershi and Kamar 2017]

Facilitate concept evolution through a “could be” category that allows clustering into subcategories you can change labels for.
More recently: prompting

In-context learning allows end users to write what they want:

Control remains an open problem, however. If I can’t figure out how to cross the gulf of execution through the prompt, how do I convey my intent?
The challenge of designing with AI
Why AI is difficult to design

[Yang et al. 2020]

How do we know what AI can and cannot do, and how it will err?
How do we engage in rapid prototyping of AI-powered systems?
How do we control the unpredictable output of the AI?

Michael would add:

We are risk averse and will avoid AI-powered interactions once we stumble into one of their limits: algorithm aversion.

If “Alexa, play a reggae song by Beyoncé” returns the wrong thing, or your text message dictation errs, you back off to simpler interactions.
Human-AI design guidelines

[Amershi et al. 2019]

What guidelines, similar to Nielsen’s heuristic evaluation principles, ought to apply for human-AI interaction design?

Human-AI Interaction Design Guidelines

INITIALLY

01 Make clear what the system can do.

   Help the user understand what the AI system is capable of doing.

02 Make clear how well the system can do what it can do.

   Help the user understand how often the AI system may make mistakes.

DURING INTERACTION

03 Time services based on context.
Summary

**Intelligence augmentation** aims to place AI in context by using it to amplify our own abilities.

Debates rage about the levels of autonomy to grant to AIs: from fully autonomous **agents** that act on the person’s behalf, to **direct manipulation** that always leaves the user in full control.

**Mixed initiative interaction** splits the difference by asking, acting, or doing nothing based on its confidence and assessment of the benefit.

End users and designers seek to draw on these tools.


References


Here are the lyrics to "Brown-Eyed Girl" by Van Morrison:

"Hey, where did we go?
Days when the rains came
Down in the hollow
Playin' a new game

Laughin' and a-runnin', hey, hey
Skippin' and a-jumpin'
In the misty morning fog with
Our, our hearts a-thumping and you

My brown-eyed girl
And you, my brown-eyed girl

And whatever happened
To Tuesday and so slow?
Going down the old mine with a
Transistor radio

Standing in the sunlight laughing
Hiding 'hind a rainbow's wall
Slipping and sliding
All along the waterfall with you

My brown-eyed girl
You, my brown-eyed girl

Do you remember when we used to sing?
Sha-la-la, la-la, la-la, la-la, la-la tee-da
Just like that
Sha-la-la, la-la, la-la, la-la tee-da, la-tee-da"

Write a parody of these lyrics to be about mixed-initiative interaction in AI.
Hey, what do we know?
Days of coding gone by
Building an AI low
Making it talk and fly
Chatting and a-typing, hey, hey
Answering and a-scripting
In the world of machine learning with
Our, our code a-debugging and you
My mixed-initiative friend
And you, my mixed-initiative friend
And whatever happened
To rule-based systems, so slow?
Going down the path of NLP with
A language model, you know
Standing in the future, smiling
Hiding 'hind a chatbot's wall
Learning and evolving
All along the AI hall with you
You, my mixed-initiative friend
Do you remember when we used to train?
Data-la-la, la-la, la-la, la-la, la-la tee-da
Just like that
Data-la-la, la-la, la-la, la-la, la-la tee-da, la-tee-da
Thanks, Mendicants!