Last time

Collaboration is hard: physical **distance matters**.

Tools can try to mitigate the effects of distance, but we are limited by the **socio-technical gap**.

Crowdsourcing gives up on tight teamwork in favor of structured contributions through open call and at massive scale.
Social Computing

Unit 3

social media

collaboration
Where we go from here

<table>
<thead>
<tr>
<th>so far</th>
<th>Ubiquitous Computing, Design, Social Computing</th>
</tr>
</thead>
<tbody>
<tr>
<td>week 5</td>
<td>Human-Centered AI</td>
</tr>
<tr>
<td>week 6</td>
<td>Cognition/Visualization</td>
</tr>
<tr>
<td>week 7</td>
<td>Software Tools/Content Creation</td>
</tr>
<tr>
<td>week 8</td>
<td>Critical Theory/Simulating People</td>
</tr>
<tr>
<td>week 9</td>
<td>Methodology</td>
</tr>
<tr>
<td>week 10</td>
<td>History</td>
</tr>
</tbody>
</table>
Human Centered AI

Unit 4

human-centered AI
working with unpredictable black boxes
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
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
Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes

Qian Yang
HCI Institute
Carnegie Mellon University
yangq2@cmu.edu

Aaron Steinfield
Robotics Institute
Carnegie Mellon University
steinfield@cs.cmu.edu

John Zimmerman
HCI Institute
Carnegie Mellon University
johnz@cs.cmu.edu

ABSTRACT
Clinical decision support tools (DSTs) promote improved healthcare outcomes by offering data-driven insights. While effective in lab settings, almost all DSTs have failed in practice. Empirical research diagnosed the poor contextual fit as the causally new form of DST. It automatically generates advisories for clinicians’ decision meetings with subtly embedded machine predictions. This design took inspiration from the notion of Unremarkable Computing, that by augmenting the users’ routines, AI can have significant importance for the users yet remain unobtrusive. Our field evaluation suggests clinicians are more likely to encounter and embrace such an DST. Drawing on their responses, we discuss the importance and intricacies of finding the right level of unremarkable- ness in DST design, and share lessons learned in prototyping critical AI systems as a situated experience.

CCS Concepts
• Human-centered computing → User centered design:

KEYWORDS

ACM Reference Format:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Copyright held by the owner/author(s).

Published by the ACM. ACM 978-1-4503-5795-2/19/05...

1 INTRODUCTION
The idea of leveraging machine intelligence in healthcare in the form of decision support tools (DSTs) has fascinated healthcare and AI researchers for decades. These tools often promise insights on patient diagnosis, treatment options, and likely prognosis. With the adoption of electronic medical records and the explosive technological advances in machine learning (ML) in recent years, now seems a perfect time for DSTs to impact healthcare practice.

Interestingly, almost all these tools have failed when migrating from research labs to clinical practice in the past 30 years [5, 6]. In a review of deployed DSTs, healthcare researchers ranked the lack of HCT considerations as the most likely reason for failure [12, 23]. This included a lack of consideration for clinicians’ workflow and the collaborative nature of clinical work. The interaction design of most clinical decision support tools instead assumes that individual clinicians will recognize when they need help, walk up record, and that they will trust the system’s output. We are collaborating with biomedical researchers on an artificial heart. The artificial heart, VAD (ventricular assist device), is an implantable electro-mechanical device used to partially replace heart function. For many end-stage heart failure patients who are not eligible for or able to receive a heart transplant, VADs offer the only chance to extend their lives. Unfortunately, many patients who received VADs die after the implant [2]. In this light, a DST that can predict the likely trajectory a patient will take post-implant, should help identify the patients who are most likely to benefit from the therapy.

We draw insight from a field study investigating the VAD decision process, searching for opportunities where ML might help [16]. The findings revealed that clinicians are unlikely to encounter or to successfully engage with a DST for help at the time and place of decision-making. For most cases, they did not find the implant decision challenging; thus, they had no desire for computational support. In addition, the highly hierarchical healthcare culture strained those physicians who make implant decisions and the...
Our trust isn’t calibrated

**Algorithm aversion**: we prefer human decision-making to AIs, even if the algorithm is better at the task [Dietvorst, Simmons, and Massey 2015]

…and especially after seeing the algorithm make an error

What if the algorithm just suggests the answer to you?

We often get influenced by the AI’s suggestion and rely on it when we shouldn’t [Buçinca, Malaya, and Gajos 2021]

But surely if the algorithm explains its reasoning?

 Doesn’t help, unless the explanation takes almost no effort to verify [Vasconcelos et al. 2023]
OVERALL ABOUT PROGRAM
CREDITS
ACTIVITIES
USAGE
CONTROL TECHNIQUES
RICE IMPLEMENTATION
INTRODUCTION
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

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]

Control

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>music box</td>
</tr>
<tr>
<td>High</td>
<td>bicycle piano</td>
</tr>
</tbody>
</table>

Automation
Agency plus automation

Generalize the user’s inputs (selecting text “Alabama”) into scripts

[Suggest alternative visualizations]
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?

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

<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>

Desired goal

Not desired goal
Mixed initiative: visually

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

\[ u(A, G) \]

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

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

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

Utility of inaction

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

\[ u(A, G) \]

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

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

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

Expected value

\[ P(G) \]

\[ 0 \]

\[ 1 \]
Mixed initiative: visually

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

\[ u(A, G) \]

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

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

Expected value

Utility of action

Utility of inaction

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

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

Expected value

\[ u(A, G) \]

Higher utility not to act

Higher utility to act

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

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

Utility of action

Utility of inaction

\[ P(G) \]

Higher utility not to act

Higher utility to act
What if we ask the user?

Asking often carries lower risk, but also lower utility

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

$u(A, G)$

$u(\text{Ask}, \neg G)$

$u(A, \neg G)$

$u(\text{Ask}, G)$

$u(\neg A, G)$

Utility of asking

Utility of action

Utility of inaction

Expected value

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

Asking often carries lower risk, but also lower utility

\begin{align*}
    u(\neg A, \neg G) \\
    u(\text{Ask}, \neg G) \\
    u(A, \neg G) \\
    u(A, G)
\end{align*}

\[ u(A, G) \]

\[ u(\text{Ask}, G) \]

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

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

Expected value

Utility of action

Utility of inaction

Utility of asking

Inaction zone

Ask zone

Act zone

\[ P(G) \]
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 might we let people train such a doorbell
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
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?

I 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 work with these tools.


References


