Mark Esposito

Professor of Business, Government & Society &
Economic Strategy Competitiveness, Harvard University

HERACADEMY
sapere_crescere_innovare

What Every Manager Should Know About Human-Centered AI:

A Manager's Introduction to Human-Centered Artificial Intelligence

Mark Esposito, Ph.D
@Exp_Mark

VERSO UN'INTELLIGENZA ARTIFICIALE ANTROPOCENTRICA

di Mark Esposito

È urgente riflettere sullo stato e sulle prospettive dell'utilizzo dell'IA in rapporto all'intelligenza e alle capacità umane. Questo saggio, rivolto a manager e dirigenti, vuole costituire una guida per orientarsi fra i moltissimi sviluppi possibili delle relazioni uomo-macchina.

Quando nel 1999 la sonda Mars Climate Orbiter, costata 125 milioni di dollari, si schiantò sulla superficie di Marte a causa di un errore di misurazione, Carl Pilcher, direttore scientifico dei progetti di esplorazione del sistema solare del Jet Propulsion Laboratory della NASA, fornì la seguente spiegazione: «Penso che il problema sia stato che i sistemi che avevamo progettato per riconoscere e correggere l'errore umano non siano stati all'altezza delle aspettative». Un problema analogo si è ripresentato nel 2013, quando un errore di calcolo nella progettazione di un nuovo sottomarino spagnolo ha generato un difetto tale per cui il sottomarino poteva immergersi ma non riemergere dall'acqua, causando il rallentamento pluriennale di un investimento da 2,2 miliardi di dollari. Entrambi gli errori nascevano da un semplice errore di calcolo e traduzione, in definitiva una questione di qualche punto decimale sbagliato. Per cui vale la pena chiedersi: e se ci fosse un sistema in grado di cogliere automaticamente tutti gli errori di questo tipo?

Com'è naturale, il quesito è già stato esplorato e i nuovi sistemi di intelligenza artificiale (IA) vengono impiegati in settori diversi per ridurre l'incidenza dell'errore umano. In tutti questi progetti è emerso un tema comune: i sistemi di IA possono essere efficaci nella correzione dei problemi una volta che sono stati identificati, ma purtroppo non lo sono altrettanto nel capire in modo autonomo cosa vada considerato il problema da risolvere al di là di ciò che è stato già isolato esplicitamente come tale. Il problema legato all'utilizzo dell'IA va oltre, dato che non si tratta semplicemente di correggere l'errore, ma di avvertire che c'è un errore che va corretto.

A tale problema di identificazione si aggiunge tutta una serie di problematiche affini che creano problemi alle aziende che cercano di costruire, comprare, impiegare e modificare le soluzioni di intelligenza artificiale. Alla base c'è la mancanza di principi e di linee guida comuni che determinino il modo in cui l'organizzazione comprende il valore dell'IA e ciò che le persone vogliono trarne. Più procedono lo sviluppo e l'adozione dell'IA, più la questione di quale valore dovrebbe apportare occupa spazio agli occhi dell'opinione pubblica, dato che spesso l'IA fa emergere quell'ansia da automazione per cui è più forte il timore che, in tutta una serie di attività e mansioni, le persone possano essere sostituite (perdendo il vantaggio competitivo rappresentato dall'apporto umano) che non la speranza in maggiori van-

Introduction

Remember the Mars Climate Orbiter incident from 1999?

Spain builds submarine 70 tons too heavy after putting a decimal in the wrong place

THE ASSOCIATED PRESS Updated: June 19, 2013

Apparently somebody in the calculations made a mistake in the very beginning and nobody paid attention to review the calculations.

HERACADEMY
sapere_crescere_innovare

Inherent Problems of AI

AI systems may be good at correcting for problems identified…but…

They are **not good at identifying what counts as a problem** that needs to be corrected **outside of what was explicitly identified** as a problem.

It's not simply correcting the mistake but **alerting that a mistake was needed to be corrected**.

There is a **lack of common principles and guidelines** shaping how organization understand the value of AI and what people want from it.

Why AI is Never Just AI

To understand how to leverage human-centered AI, managers first need to understand changes facing AI management:

1

AI technology itself is far more than the algorithm; AI is the umbrella term for a class of systems integrating talent, AI algorithm designs, data sets & data management strategies, data capturing devices, and computational power.

2

Successful AI projects are less about the tech than the **institutions and practices to which they are connected.**

3

AI systems evolve, and so do the institutions and practices once AI is used. So, the question for managers is to 'look around the corner' at what's next in value creation.

HERACADEMY
sapere_crescere_innovare

HUMAN-CENTERED AI

¨Designing and leveraging artificial intelligence for enhancing human capabilities in an effective, intelligible, and ethical manner.¨

HERACADEMY
sapere_crescere_innovare

Ensuring Ethically-Aligned AI

4 action points to consider when determining if AI solutions are ethical:

Fairness

Intelligibility & Transparency

Privacy

Autonomy

HERACADEMY
sapere_crescere_innovare

Fairness

AI solutions are often designed as aides to existing decision-making processes. AI for recruitment often serves either to reduce the time burden of reviewing initial resumes to sort for the top 5, or to assess additional data-points during an interview that might otherwise go missed by a recruiter.

Help to unburden recruiters and candidates of biases for the process to be more meritocratic?

VS.

Affirm prior biases, assumptions, stereotypes, or conventions.

HERACADEMY
sapere_crescere_innovare

Intelligibility & Transparency

Solutions demand a means of deciphering AI outputs. This problem goes by different names - explainable AI (XAI), transparent AI, and/or intelligibility.

This problem is while inputs are understood, the outputs often cannot be understood in relation to the reasoning or rules the AI developed to produce those outputs.

A neural network was trained on picture of cats and dogs to identify each.

BUT

We cannot understand the specific weighting and rules it created to generate such identification patterns.

Privacy

AI generates rules from specifications - these specifications can be engineered by humans, as in the case of AlphaGo or, as in almost all cases, they are derived from data. The higher quality the data, the better the outputs can be expected to perform.

In many cases the data and means of collecting it are pursued without the consent of those impacted by the data.

Designing facial recognition systems for governments to identify, apprehend, and catch criminals.

VS.

Governments using these systems and wrongfully accusing and apprehending innocent citizens based on data collected without consent.

HERACADEMY
sapere_crescere_innovare

Autonomy

All AI solutions make a claim on the problems of a firm and the benefits of using human labor by being created as a mechanism which can enable human skills, replace them, or serve as a hybrid of the two.

Designing a chat-bot to take over redundant questions for consumer information.

VS.

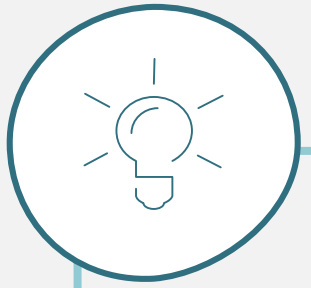
Designing a chat-bot to take over the entire customer service process.

HERACADEMY
sapere_crescere_innovare

The Human in Human-Centered AI

Human-centricity should serve as an essential design reminder to consider:

The needs and problems of ***real*** people in ***real*** contexts

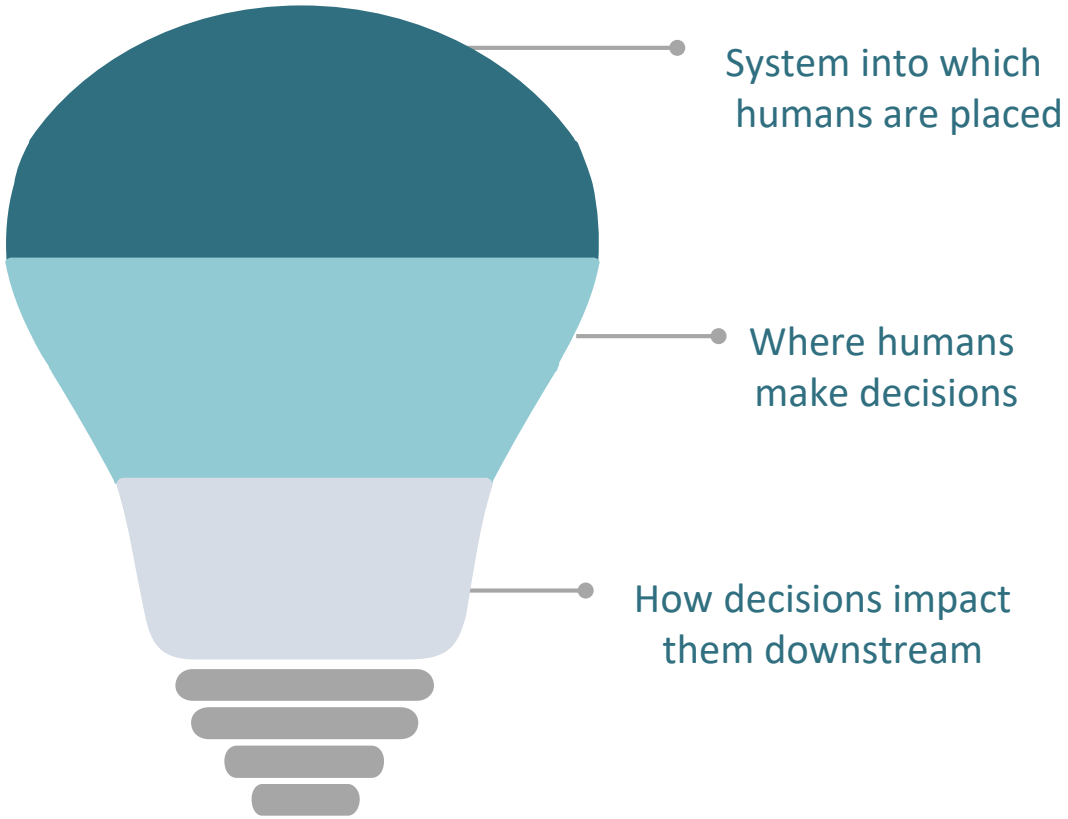


How AI will impact how people ***actually behave*** rather than how a company ***wants them to behave***.

HERACADEMY
sapere_crescere_innovare

The Best Human-Centered AI

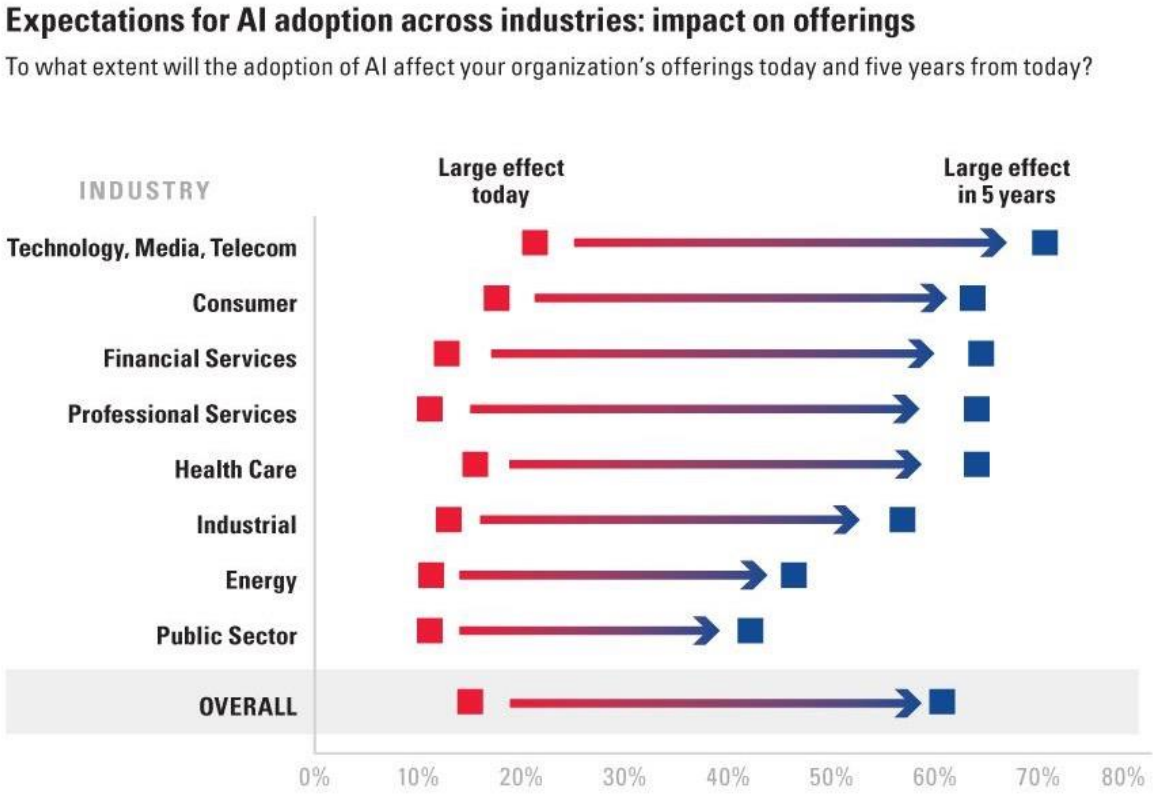
The best human centered design understands:



HERACADEMY
sapere_crescere_innovare

What Managers Should Consider

1. What is the problem or need to which AI will be applied and what is the desired outcome?
2. What kinds of decisions are implied by the problem and who will be making those decisions?
3. What are the environments in which decisions will be made and how can that environment change?



HERACADEMY
sapere_crescere_innovare

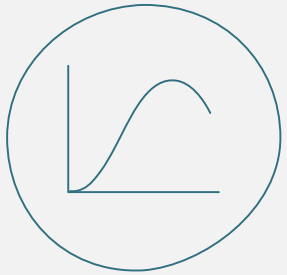
For Building & Buying Human-Centered AI

The following are some ideas worth exploring as managers attempt the integration of the symbio-intelligence that the human centered AI can foster:

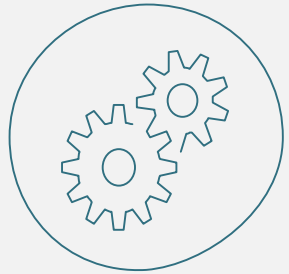
Ensure cognitive diversity in design and implementation teams to avoid blind spots

Ensure the management and external firm staff have a common understanding of the problem, needs, and value the AI can supply

Clarify what AI is supposed to do and how people will interact with it in the context of existing processes.



Establish clear processes for ‘changeability management’ to control and frame any updates



Establish clear processes for assessing mental models across user chain

HERACADEMY
sapere_crescere_innovare

Conclusion

¨The amount of collective thinking ahead will become a required competitive necessity to acquire and fulfil, if we want the next decade to be the one where humans understand that ***the real power of AI is not for the sake of technology, but for the improvement of social constructs and paradigms****.¨*

HERACADEMY
sapere_crescere_innovare