# AN AI PATTERN
# Language



## M.C. Elish and Tim Hwang

# An
# AI Pattern
# Language

*M.C. Elish and Tim Hwang*

INTELLIGENCE & AUTONOMY INITIATIVE
*Data & Society*

Authors: M.C. Elish and Tim Hwang

Book design by Jeff Ytell and illustrations by Sarah Nicholls.

# CONTENTS

# 1 INTRODUCTION

Public conversations around artificial intelligence (AI) tend to focus on technologies that will develop in the far future. Most experts agree that "general artificial intelligence," the concept that a machine could exhibit all aspects of human intelligence, is decades away.[1] Our conviction is that the current thrust of research examining the implications of AI often overlooks the critically important opportunity to examine the social implications of intelligent systems in the near- to medium-term. It is important to examine long-term hypothetical possibilities, but it is equally if not more urgent to examine real trends that will be realized within this decade.

Investigating the near-term impacts of AI can be a difficult task to undertake because the research, development, and deployment of intelligent systems are nascent, fractured, and poorly understood by those outside of specific technical fields and corporate efforts. Limited frameworks, poor visibility, and the lack of technical sophistication have significant implications for media reporting, which informs public perception, as well as government policy priorities and regulatory content. In the last several

years, scholars from many different social scientific fields have turned their attention to the issues arising from big data, machine learning, and intelligent systems. Yet, these scholarly communities are often disconnected from communities of practice, including the designers, engineers, and product managers who design and deploy intelligent systems. As a result, conversations around AI in the public are often shaped more by Hollywood than empirical knowledge.

This document aims to be a bridge across this developing gap in knowledge by offering an industry pattern language through which to talk about the social implications of deploying AI and intelligent systems. The need to develop a mutually intelligible language arises from the urgency with which the implications of AI must be addressed by the currently disconnected communities who are equipped to pave the way forward.

In particular, this document highlights and draws together an existing language by articulating a series of common challenges that emerge at the junctions of engagement between human and intelligent system. We have identified a set of challenge areas, with an array of patterns that address each challenge. The patterns are intentionally abstracted in order to demonstrate common approaches across industry communities. In this way, we hope to provide insight into the different ways industry is grappling with the social impact of deploying intelligent systems.

*Document Overview*

Following this introduction, we present a taxonomy of social challenges that emerged from our interviews with practitioners within the intelligent systems industry. We took the term "intelligent systems" broadly, including sophisticated technologies ranging from what could be thought of as "big data" techniques to approaches like deep learning. We were specifically motivated to keep the definition of "intelligent systems" fluid because the term itself—in industry and in the media—is fluid and ill-defined. We did not want to artificially solve a problem with which many communities are still contending. The consequences of the slipperiness of this term will be discussed throughout the text. We also attempted to speak with a range of people who play different roles in the design and implementation of systems, from the most senior manager to the product sales person in direct contact with the customer. More about our methodology can be found below.

What we found was a set of four core challenges that emerged across all industries: (1) assuring users perceive good intentions, (2) protecting privacy, (3) establishing successful and long-term adoption, and (4) demonstrating accuracy and reliability. With the exception of data privacy, the core challenges have to do with the affective relationship to the product or system, that is, how someone feels about the technology at stake. Planning a highly automated or intelligent system necessarily involves the humans who will interact with and use it.

Significantly, our findings reveal that the ways in which practitioners are conceptualizing—in addition to addressing—social implications of intelligent systems are distinct from the way social scientific discourse is considering these implications. In a brief section following the industry challenges and patterns, we draw attention to several points of convergence and divergence in our descriptions of the core challenges by alluding to disconnected perspectives between industry practitioners and scholarly critics. While the mandates of industry professionals and academic researchers are necessarily distinct, a foundation of mutual intelligibility is imperative. With this document, we hope to contribute to this foundation.

*Project Background*

An inspirational frame (and title) for this project has been the unique collection of architectural theory by Christopher Alexander, Sara Ishikawa, and Murray Silverstein found in *A Pattern Language* (1977) and *The Timeless Way of Building* (1977). In these works, Alexander and his team developed what they viewed as "best practices" in designing a city. In the words of Alexander,

Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice.[2]

The patterns these architects envisioned are not to be strict rules, but rather structures that provide helpful support and which can—and in fact must—be embellished and developed further. The patterns embody an aspirational vision of beautiful and efficient city planning. In response to the reigning paradigms of urban renewal and centralized city planning, Alexander and his team wanted to create and share an alternative way to build a shared world.

Alexander emphasizes that the patterns are interlocking and overlapping. No pattern exists in isolation, but rather like a single puzzle piece only achieves its full potential when combined with other pieces. Alexander writes,

This is a fundamental view of the world. It says that when you build a thing you cannot merely build that thing in isolation, but must also repair the world around it, and within it, so that the larger world at that one place becomes more coherent, and more whole; and the thing which you make takes its place in the web of nature, as you make it.[3]

In *A Pattern Language*, the central problem is the built environment. While our goal here is not as grand as the city planner, we took inspiration from the values of equity and mutual responsibility, as well as the accessible form, found in *A Pattern Language*. Like those patterns, this document attempts to develop a common language of problems and potential solutions that appear in different context and at different scales of intervention.

*Methodology*

Over the course of eight months, we visited AI companies and industrial labs, attended industry and academic conferences, and conducted in-depth interviews with over thirty individuals. Our method of selecting sites and participants was



guided by an initial survey of the field of companies and thought leaders working in autonomous and intelligent technologies.[4] We then proceeded through three rounds of interviews, each round gathering new participants through a snowball method of sampling. Given this subjective and in some sense, self-selective (those who responded to our requests to meet or speak) formation of the participant pool, the people we spoke with cannot be claimed to be representative in an empirical sense. We did not explicitly seek racial, gender, or age diversity and consequently the vast majority of people we spoke with were white men under the age of fifty-five. Unfortunately, those we interviewed reflect the documented lack of diversity in STEM fields, especially the field of AI.[5] While we believe the views we present are significant and widely held, the patterns presented here are not intended to be comprehensively representative, or viewed as rules or proscriptive best practices. Rather, this document is an experiment in cataloguing and catalyzing.

Likewise, when we raise examples from social science and academic literature, we are not claiming such examples

are comprehensive. We made our selection by identifying what we saw as useful examples of when industry and academic perspectives coincide or diverge around similar topics. Like the industry frames we highlight, our academic examples are partial in the sense that they are indexical of much larger and more complex arguments, logics, and contexts. The challenges, patterns, and examples that we elaborate are intended to be a starting point for beginning more inclusive conversations.

*Definitions: What do we talk about when we talk about AI?*

If you get your information about artificial intelligence, machine learning, or robotics from popular news sources, your information is most likely wrong. This is especially true if you glean information from headlines or the first paragraphs of a story. A *Wired* article from 2014, which was headlined, "Artificial Intelligence Is Now Telling Doctors How to Treat You," followed a familiar pattern of using one advance in technology, usually still in a testing or limited release phase, to extrapolate to all AI technology.[6] One recent intelligent system that received a great deal of popular press attention was ROSS, a legal research software assistant supported by IBM Watson that allows lawyers to ask natural language research questions, rather than having to specify keywords and search multiple databases and sources. Referring to ROSS, the third sentence of an article in *Tech Insider* online read, "Just about the only thing it can't do is fetch coffee."[7] ROSS is an example of a powerful research tool, but it is far

from replacing lawyers or fulfilling all of the aspects of legal research and work.[8] The way popular audience coverage is framed even blurs the line between fantasy and reality. A CBS Sunday Morning news segment from 2015, covering the DARPA Robotics Challenge showcase, featured an interview with the director of the science fiction film, *Ex Machina*, which features a sentient robot.[9]

Even the language that we use to talk about these systems is often misleading. There are a multitude of stories each week that suggest that robots walk and talk on their own and will soon threaten all of humanity, or that artificial intelligence will soon overtake human intelligence. In reality, technologies that will exist in the short-term bear little resemblance to popular depictions. An appreciation for the real state of the art—and not just the science fiction imagination of characters like HAL 9000 or the Terminator—allows us to appreciate the amazing ingenuity and progress that does arise.

**Artificial intelligence** can be broadly understood as a characteristic or set of capabilities exhibited by a computer that resembles intelligent behavior. Defining what constitutes intelligence is a central, though unresolved, dimension of this definition.[10] For some researchers, intelligence is based in behavior and is exhibited when a computer can sense and act appropriately in a dynamic environment. For others, intelligence is based in symbolic processing, for instance recognizing and responding appropriately to human speech. While not necessarily
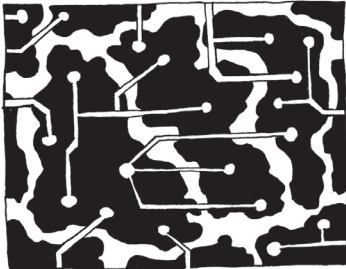
emphasized by computer scientists or engineers today, artificial intelligence has a substantial history that can be traced to antiquity and has sustained substantial transformations even in the past two decades. In light of this, we would argue that artificial intelligence, as a capacity of machines, is best thought of as a moving target. What counts as intelligence is not a static set of traits, but rather defined in relation to existing beliefs, attitudes, and technology. This is not to say that "artificial intelligence" is meaningless, but rather to call attention to its inherent instability.

Current, cutting edge work in artificial intelligence stems primarily from the field of **machine learning**. Machine learning refers to a type of computer program or algorithm that enables a computer to "learn" from a provided dataset and make appropriate predictions based on that data. Tom Mitchell's definition makes clear what "learning" means in this case: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."[11] Learning, in this case, is narrowly defined and refers essentially to the capacity for a program to recognize a defined characteristic in a dataset in relation to a defined goal and to improve the capacity to recognize this characteristic by repeated exposure to the dataset.

While artificial intelligence calls up the idea of a holistic intelligent entity, machine learning is better understood as

a specialized sub-process that can accomplish specific kinds of tasks. There are different kinds of machine learning, from supervised to semi-supervised to unsupervised. But all require a guiding human hand; an algorithm cannot begin to learn on its own. Learning involves defining datasets, appropriate variables, and metrics. Supervised machine learning algorithms are also tweaked or adjusted by programmers throughout their development.

**Deep learning** refers to a type of machine learning tasks based on models inspired by biology that are referred to as **neural networks**. Unlike the statistical approaches taken by other types of machine learning techniques, deep learning systems develop based on processes that are happening in parallel, resulting in models that are much harder to interrogate. Neural nets are often presumed to be magical because little is known about how or why they work in certain contexts. This lack of information is often transformed to suggest that the basis of neural networks is not knowable when, in fact, it is simply currently unknown.

While, neural networks were proposed as one of the primary cybernetic approaches to the newly emerging field of artificial intelligence in the 1950s, the technique was largely dismissed until the 1980s when advances in

hardware and more refined processing techniques were developed, in addition to a exponentially growing datasets. Gaining renewed interest during this period, more research and product development demonstrated the ways in which it could be effectively put to use in certain kinds of problems, such as object and speech recognition. In the past decade, more technical advances, devoted resources, and most importantly, access to immense datasets (the "fuel" for learning) has contributed to the widespread use and popularity of machine learning generally, and deep learning specifically. Deep learning techniques are increasingly embedded in popular data-driven services, from Google text and image search to the Facebook newsfeed to Amazon's recommendation systems. Many technical experts see deep learning as core to artificial intelligence, but not other branches of machine learning.

Another term that frequently comes up in discussions of intelligent technologies is **autonomy** or **autonomous systems**. An autonomous technology refers to a system that operates without human intervention. However, like AI, there is variability within the definition. What constitutes "human intervention" is key to defining autonomy and varies within different domains. For instance, the National Highway and Traffic Safety Administration (NHTSA) has developed a specific graduated scale to define autonomy in vehicles which distinguishes different kinds of human intervention.[12] Very few systems are currently completely autonomous, and all require human design and maintenance.

Both artificial intelligence and autonomy in many ways refer to activities or processes that were previously carried out by humans. In this way, intelligent technologies are often technologies of **automation**. Automation refers to "a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator."[13] This broad definition positions automation, autonomy, and, we might insert, intelligence, as varying in degree rather than an all or nothing state of affairs.

Throughout this text, we use the term **machine intelligence** to refer to a broad class of intelligent technologies with aspects of artificial intelligence, machine learning, and autonomy. In our view, machine intelligence most correctly points toward the capacities and limitations of what intelligence, of all degrees, may look like embodied in a machine.

Still, we all talk about different things when we talk about AI and related terms, and this volatility in definition has significant consequences. First, time must be devoted to clarifying terms and managing expectations of technology. Understandably, people may use the term AI to generate excitement and belief in the value of a new product. However, the downside of relying on catchy buzzwords can be public misunderstandings about technology. Moreover, the laws and regulations that are crafted to address these technologies may rely on incorrect assumptions about the technology at stake.

To demonstrate the range of meanings and conceptions we encountered, below we present some of the responses we received from interviewees when we asked: What do you talk about when you talk about AI?

*Artificial intelligence tends to be both the scariest but also sort of most universal [term]. If I walked into a room and said, "We're going to talk about an artificial intelligent personal assistant," nearly everyone would have a picture in their head. The problem is everybody's picture would be different.*

– Designer, Seattle

*I think what a lot of the confusion is about is, it starts with the fact that we don't have a good definition of intelligence in humans, right? And the difference between deep learning and AI? They're completely interchangeable. I think what happened, basically, was that AI had become a dirty word, and machine learning was the great rebranding of AI.*

– Venture capitalist, New York

*AI? Well, we stopped using the term, we talked about cognitive computing, then shifted back to AI, basically came from higher ups.*

– Product manager, Chicago

*I think that artificial intelligence… my perspective is that we have very little artificial intelligence that works very well right now. We have machine learning, which I think of as being like a lower-level component.*

–Machine learning engineer, New York

*Well, the next step is this step from specialized intelligence to general intelligence, and where we're starting to build machines, which are aware of the context around them.*
　　　　　　　　– Venture capitalist, San Francisco

*People are confused when they talk about general AI, why that's not the thing that we need to think about right now. The thing that we need to think about right now that's in the here and now is the impact from systems that are quite specialized. Maybe not autonomous. They're quite specialized, but they do the thing that they're specialized for incredibly well.*
　　　　　　　　– Venture capitalist, Boston

*I forget that we still live in a world where so much of the way people think about machine learning is more informed by movies than actual practice, right? And so all the analogies and stories and metaphors come out of that world rather than what people have actually done in practice, and partly because the stuff in practice is still so nascent.*
　　　　　　　　–Software engineer, San Francisco

*Oh yes, every week, there are people who go, "Hey, do you remember the scene in Her when this happened? Why aren't we doing it like that?" "Is it more like Wall-E or is it more like Samantha?"*
　　　　　　　　– Product manager, San Francisco

*There are actually two different answers because there's how we talk about it internally, there's how we talk about what we're striving to build internally, and then there's an external*

*message. I don't mean to make that sound complicated but I think it belays a problem that we need to face in terms of creators and consumers, and how you talk to them. We talk quite a bit about how we turn machine learning into insights. We feel insights are incredibly valuable. They're context plus information plus actions.*

– Designer, Seattle

*I hate it [the term artificial intelligence] so much. There's this conception that people have, that it's the same thing as Elon Musk and Steven Hawking saying machine learning is playing with the devil. It raises interesting philosophical questions, yes, but all the things that people are worried about are science fiction. At the end of the day, the way the machine learning works is you give it a set of examples and it learns to repeat what you show it in a very constrained way. The issue that I have with artificial intelligence is that the connotation that people tend to draw from it is small robotic children. That's just never going to happen. With all of our understanding of artificial intelligence, it's not like, "Oh yes, we're right around the corner."*

– Start-up founder, San Francisco

# 2 CHALLENGES & PATTERNS FROM INDUSTRY PERSPECTIVES

How are the social implications of intelligent systems understood by those involved in their design? An important follow-up question is: when and why do engineers, product managers, designers, and investors feel it is their responsibility to address these social dimensions? While the following patterns address the first question, this section briefly examines the ways in which many engineers, product managers, designers, and investors understand these kinds of questions to be outside the scope of their job. Through our interviews, we noticed two themes in the reasons people gave: (1) the size of the company limited what they could do and (2) they had to prioritize limited resources.

Individuals in very large corporations and small start-ups both expressed that the size of their company limited their role in thinking through the social impact of their work. In large corporations, the division of the development process means that teams are highly specialized and often silo-ed. A designer at a multi-national corporation explained that "there is a whole team of lawyers I never see who deal with that stuff." The responsibility always seems to lie with another department. For smaller companies who have services that are licensed by larger companies,

the responsibility seems out of their hands. A designer working in a medical technology company expressed the sentiment with regard to consumer privacy debates that the product was a business-to-business (B2B) company, so it "wasn't really" within his control.

The director of a financial analytics startup summarized the potential implications regarding company size: "It's just so different. I guess we're both automating investment processes, but at my old firm I was one in a company of a thousand or 1,500 people. I was not anywhere near the legal and regulatory aspects of the business. I was so far removed from those other parts of the business. Doing a startup, I'm kind of responsible for everything."

A venture capitalist observed what he had seen through his experience with startups: "To be honest, like most entrepreneurs, they are just trying to keep the company going, trying to grow the company. And the larger-scale implications? Oftentimes, they won't think about it until they're quite successful or they're further along and actually have a chance to look back." A product manager at a recommendation service put this in terms of a slightly different dilemma around resource allocation. While he may want his team to go back to test or fully understand why one iteration of an algorithm worked better than another, that is time "wasted." He explained that time spent thinking about why or to what end is time that his engineers are not developing a new product. Often, the immediate costs outweigh the potential benefits.

With these caveats in place, our work unearthed a set of discrete problems that practitioners did consider within their ambit. These, interestingly, were relatively independent of the size of the company and were consistent across industries. These problems were considered within the practitioner's scope because these problems were seen as key to the development and deployment process. We present them as the challenges below.

## Challenge: Assuring Users Perceive Good Intentions

Perhaps because intelligent systems are new, unknown, and variously defined, cultivating trust between a user and a system was seen as a foundational social aspect of designing and deploying intelligent systems in every sector. Transparency was the main frame through which our interviewees thought about proving their good intentions and establishing trust. What transparency means and to what degree it can be offered varied within the responses.

Perhaps because each founder or designer relies on the self-knowledge of their own good intentions or have never considered how a product they believe in might introduce harm, those we spoke with did not focus on the potential that their products might result in inequality or injustice. When we asked professionals about inequality, they tended to think about the introduction of new forms of inequality in terms of user perception or product reputation.

PATTERN 1: SHOW THE MAN BEHIND THE CURTAIN

"We want our users to know that our products are powered by people," one product manager at a music recommendation company explained. This desire reflects one perspective on how to achieve trust between human and computer: make the role of humans visible in an interaction which can sometimes seem inhuman, even if it's personalized. Speaking about a discovery feature, this product manager explained how through the design
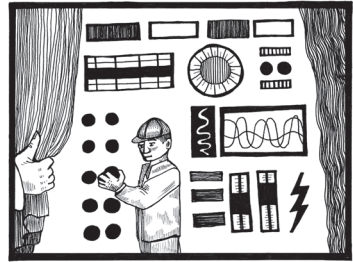


process they realized they had to make the product "not feel creepy," it had to "feel human, powered by people." He explained, "We've made it like a gift, a gift for you each week. Sometimes we get it right, but like a human, the product can have good weeks and bad weeks."

PATTERN 2: OPEN UP THE BLACK BOX

Additionally, the idea of providing transparency about how a product worked was also used as a design strategy. A lead designer at Cortana, Microsoft's personal assistant application, explained how the idea of "Cortana's notebook," was central to how they conceived of the relationship between a user and her Cortana. The notebook could be accessed by the user at any time, and "as the place where all of the inferences that we use are stored and visible to users, [the user] can go and adjust them, turn them off,

or delete them." Providing a clear and easy access point to how Cortana is building intelligence provides the user with the sense that the relationship is evolving as they, the user, want. "We found that there are a bunch of inferences we can determine that when quantified to users are just creepy," the designer explained. He continued:

"For example, we found a lot of anxiety around knowing your [the user's] home and work address. So as much as it only took us a short period of time to use GPS and other things to begin to locate where you were, actually using that data without talking to you was super creepy to our users. They thought it was neat but it was scary. We really looked at how we built this path of how do we introduce ourselves, how do we focus on setting you up to succeed with the tool, and how do we actually grow that success over a period of time so that relationship becomes more indispensable? Something we hadn't realized was that there was a lot more people wanting to be able to tune and control."

The founder of a new machine learning start-up also explained how the principle of transparency could be used as a means of establishing trust: "We were motivated by things like GitHub, open-source software, and Wikipedia, where you really see high-quality content emerge from an open refinement process where people can contribute. … Similarly, people now trust Wikipedia for getting information, but most people do not dig behind the scenes or engage in fact checking or citation checking or anything

like that. But nonetheless, people trust Wikipedia and trust open-source software because they know the process works like that and that people are behind the scenes, doing those things. So that's going to be our approach to how we get the same kind of level of trust."

PATTERN 3: DEMONSTRATE FAIR AND EQUAL TREATMENT

The idea of cultivating trust at times can look like grappling with the fairness of a system. The founder of an intelligent stock portfolio startup explained that while any kind of systematized and automated investing introduces new kinds of potential biases, this does not obviate the attention that must be paid when developing new systems, "You have fiduciary responsibility to make sure that it's [the execution of trades] is fair across accounts. In our current company, there are only two of us, and we write the code ourselves, and each of us reviews the algorithms where things like that matter." For instance, if all of a company's clients are employing the same strategy, such as, buy shares of IBM, the order in which the algorithm processes the list of clients is significant. It might seem straightforward to execute the trades down an alphabetical list, but this could mean that over time, the lowest last name would have a slight advantage over the others. This founder explained, 'You have to think, 'Is this actually

fair? Is there any ordering bias? Is there any way that this client is going to be favored, executing and submitting the trades, more than others…?' So there are plenty of details like that that you need to have in mind when writing the algorithms. … A common way to solve the problem in the algorithm is to use a random number and order clients like that. It's like pulling numbers from a hat, that there's no inherent bias in the system." Ensuring that the company would uphold its fiduciary responsibility and principles of fairness in the face of complex automation was a core systems design principle.

## Challenge: Protecting Privacy

*Privacy is still a very important part of the founding of this country [America], and so I… man, I have a lot of things to say!*

   – Machine learning product manager, Florida

Intelligence is built from data: intelligent systems gain intelligence through the acquisition and analysis of big datasets. It makes sense that the protection of that data would be of paramount importance to the practitioners of intelligent systems, both as a valuable asset of their own or as the stewards of someone else's valuable asset, such as an individual consumer's pictures or a bank providing sensitive datasets to an analytics firm working on contract. Depending on the kind of service offered and the size of the company, some people we spoke with wished that they had more control over the data used and generated.

A computer scientist and founder of a small Boston biometrics company expressed his frustration: "We enable data privacy as much as we can, but responsibility is with the vendor, like Samsung." Another computer scientist who had recently founded a predictive analytics start-up also explained how he had begun to think more about differing privacy laws between the United States and EU, "Should it be opt in versus opt out? Should I, as the company, have the ability to make that choice on behalf of the user?"
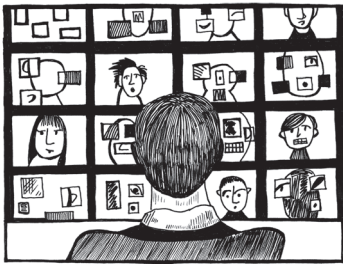
### PATTERN 4: DATA SECURITY IS THE FOUNDATION

One sales manager of a predictive analytics firm explained that explicitly addressing security and privacy was an important part of the pitch: "We use it as a sales tactic to say that the marketplace is concerned about this, and we've got a very strong response to it, and we should be the trusted advisor to ease those concerns so that they can move forward."

For this manager, and most of the others with whom we spoke, data security and data privacy are intertwined concepts. Data security is what allows data privacy. And privacy, when we spoke with interviewees and asked about the social aspects of product adoption, data privacy—for good and bad—was nearly always the first issue to be raised. There was a widespread sense that data privacy needed to be addressed in systems design because it was a primary concern for users and the general public.

PATTERN 5: ESTABLISH A CATCH AND RELEASE DATA PATTERN

The founder of a Miami facial recognition software company we interviewed explained that he has had to deal with privacy issues on both a practical and theoretical level: "We have people who come to us all the time and say, "Oh, I love what you guys are doing, but I'm also so scared, like is this the end of my privacy? Will people know what I'm doing? I feel like Minority Report, right, like I walk into a mall and all the ads change just for me." And we actually try to assure them that no one, at least none of our customers that are coming to us, are interested at all in creeping anybody out. There is very little money to be made in creeping people out. In fact, there's only money to be lost."

This founder explained how the company has chosen to mitigate privacy concerns: "So number one, we have a multi-standard environment, for example Walmart and Target can't share information with each other. I mean, maybe they can on their side, if that's what they want to do. But from our perspective, their data is completely separate. Number two, we take in a video stream, we identify certain points on the face, and from those points we know if you're male or female or your age, and then we take that image and we throw it away on device. For instance, for a camera

at a Walmart in, say, Arkansas, the image of your face comes into that camera on the device (we just use small Android-like hard drives on these devices). We process the video stream either in the device or in the store, depending on the configuration. From that, we get the demographic information out. That goes to a file or a porting API. Just the 'Female, 42, at this location' goes up, but the image itself gets deleted and kind of thrown away right there on the spot. Actually, when I say, 'delete,' it actually just is passed through and it's just void and deleted, and it's never even saved to begin with. It's been processed real-time. So that's how we, from a design perspective, keep those things [privacy violations] from happening."

PATTERN 6: TAILOR EXPECTATIONS TO CONTEXT

"Thinking globally, internationally, is a necessity when designing systems," an autonomous vehicles researcher pointed out to us. Sitting in Silicon Valley, the car company she worked for was headquartered on the other side of the world and the car would eventually be shipped to dozens of countries. The reality of mass-produced international products, which operate in specific, local contexts complicates the conceptualizing of design problems. One area where this was reflected to us was in the cultural specificity of the notion of privacy. "We do a lot of research and we do a lot of thinking about this topic," the founder of a facial recognition software company explained, "It has been interesting to us that in emerging markets, privacy is not always as important to people as it is in the United

States." Moreover, the context of use changes how privacy needs to be taken into account. Because privacy can mean different things about different kinds of information, and between different individuals or institutions, privacy needs to be considered as a relation, not a fixed attribute.

PATTERN 7: BE PATIENT.

Many of those with whom we spoke expressed a kind of "wait and see" attitude with regard to privacy concerns. "There's no absolute notion of privacy," explained an investor in financial technologies based in San Francisco. "And my bet," he continued, "is that over time, I mean over the long span of time, we will just all be much more willing to give away information to sets of services in a way that will probably make our relationship to the government or to companies unrecognizable, compared to today."

PATTERN 8: IGNORE THE ANXIETY AROUND PRIVACY: IT'S A RED HERRING

In fact, in the fundamental relativity of privacy, a number of interviewees held the view that focusing on data privacy is misguided. The founder of a leading machine learning company based in Silicon Valley explained that privacy is simply "a historical construct" and is mistaken as important today. He used the example of reactions to Gmail to illustrate his point: "When Gmail launched, 'privacy' was a big issue. The Google engine could read all your mail! But it is a machine reading it, not a human. And it's being

read for purely the purpose of ad revenue. And now no one cares." He found that focusing on the notion of privacy held back innovation. He concluded, "Yes, there's a sense of impatience on my part. We will adapt to new technology and we will evolve."

Another venture capitalist based in New York had reached a similar conclusion. He emphasized that the current focus on privacy holds back not just innovation, but more importantly, the means to build a better society in an age of digital technology: "I believe all the advocacy for privacy that's currently taking place is a terrible, horrible, bad idea. I think that the only logical construction of society going forward is one of transparency and post-privacy. The reason I believe that is because I believe that democracy can only work in an environment of mutual trust, where we figure out how to construct a government by and for the people. I know that we distrust our government in the US, but to double down on that distrust in the way Apple is doing and others are doing right now, I think will lead us to the very government we fear. It will lead us to a totalitarian government, and it will lead us to computational devices that are locked down. And do we want a world of locked-down computing devices, or do we want a world of the free flow of information, even if that free flow means that a lot more is known about each individual?"

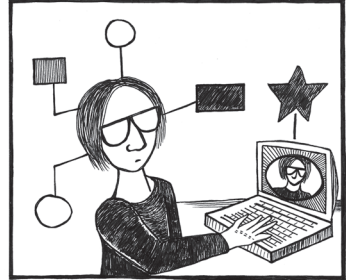## Challenge: Establishing Successful and Long-term Adoption

Intelligent systems present a new kind of design challenge in the sense that the machines, products, or services being developed are not human and yet are meant to approximate humans in some way. Design principles for interacting with "intelligent" machines continue to develop, but too often focus on interface design instead of the holistic system experience. One primary dimension of this interaction is the experience of control and long-term management. How and to what extent should the users of or collaborators with intelligent systems be in control? Many we interviewed understood this as a social challenge because it related directly to the adoption of a product or service, which is understood to be a function of more than technical capabilities.

PATTERN 9: ALWAYS ASK: WHO IS BEING MADE THE HERO?

One human factors engineer working at logistics software startup described how he thinks that "everybody wants to be the hero of their own story." The central question for him and his team, as he saw it, is "how do you actually under the hood have autonomy going on, and yet make the user feel like they're in control? If you have an autonomous, or semi-autonomous technology coming in, and they're not the hero, they are going to resist that." He and his team have iteratively designed their software alongside a group of farmers who would potentially use the software. This kind of user-centered design generates a

product experience that effectively takes into account the people and contexts in which a product is be used.

Similarly, a leading human-computer interaction designer explained how he framed the problem of integrating new intelligent systems into existing workflows by beginning with his own experience. "At the beginning of my career, I worked in an IT department, and we were all about making people's jobs simpler. But I realized that was a bad idea. No one wants that, no one wants their job to be simpler because that means their job isn't necessary. But everyone wants to be better at their job, because being better is about adding value to the company." For this designer, framing the assistance that technology brings as enabling a worker more—rather than as making a job simpler—was a means to build better intelligent systems that people would be glad to work with. Others we interviewed also expressed the sentiment that intelligent systems could "make people more effective in an increasingly difficult world," as one engineer at a computer vision company put it.

PATTERN 10: PLAN FOR THE ROLE OF HUMAN RESOURCES

Still, others we spoke with who were closer to the actual use of these systems confronted unanticipated challenges no matter how the systems were framed as useful. One

analyst working for the city of San Francisco's expressed her frustration that "while everyone is talking about big data, we need, we use, *little data*. It's not just what you can get, it's what you can understand and get other people to understand."

An unexpected challenge raised by several interviewees was correctly preparing for the process of introducing a new technology into an organization. Many initially underestimated the human resources that would be required to facilitate new technical resources. A product sales person working within IBM explained a scenario he had seen: "There's a visionary within the organization, and they're like, 'Absolutely. This makes sense. This is the direction we need to go.' They buy the software, and then reality sets in." He explained that sometimes a company doesn't realize it will need people and internal skill to integrate the system. He continued, "That's where I see more of the bumps and the hurdles. It's when that vision doesn't marry up with the reality. I think some folks fall into the category of thinking it's like, 'I buy a solution. It's like the iPhone, and within a minute I could start making phone calls,' when in reality there is a methodology and there's a process that takes time, expertise, and it's iterative. There's an immediacy that everyone wants in this day and age, where oftentimes the implementation can be more time consuming, resource involved, and more challenging than initially anticipated."

## Challenge: Demonstrating Accuracy and Reliability

The perception of accuracy and reliability can also be a challenge to establish because of perceived technological complexity. As the founder of a machine learning company explained, "There's so much mystique related to machine learning. There's a combination of the customers wanting to understand it as much as possible but at the same time, assuming they can't get into the details. Because it's unknown, there's this powerful desire to understand but at the same time, because it's complicated, even if you start explaining it to people, typically people will just kind of zone out after the first little portion."

PATTERN 11: EXPLAIN THE CONDITIONS OF ACCURACY

"The tensions with clients are typically around accuracy," explained an engineer at an intelligent sentiment analysis company. Some clients and users want to understand how this kind of analysis is better than what they may be currently using.  The engineer continued, "People just always ask, 'I want to know how accurate it is. Give me the accuracy.' Well, in our case, 'accuracy' is actually a really poor metric to know how good this API is because you have 111 categories. We could say, 'Great, we've got forty-five percent accuracy.' They're like, 'Forty-five percent? That's terrible. I'm looking for something close to 100.' But we're like, 'No wait, okay, hold on and think about it for a second. There

are 111 categories. Many of them are very similar, like fitness and dieting for instance, very similar.' They're like, 'Look, give me a number that I can say is accuracy.' We usually come up with proxy numbers where we can give you the best sense that corresponds to what you believe is accuracy."

PATTERN 12:  PROVE SUCCESS BY SHOWING FAILURE

Alternately, a few of those we interviewed used lapses in accuracy, as much as demonstrations of accuracy, as a way to establish trust (Pattern 2). A product manager for a recommendation system said that she and her team would ask themselves, "How do we also communicate that the system is fallible? How do we let you [the user] know when we think we're right, when we think we're close, when we think we're wrong, and how do we ultimately get to a place where the user actually decides? I [the user] decide that yes, that restaurant was a great suggestion or no, it's terrible, I absolutely don't want to eat Thai tonight."

PATTERN 13: ESTABLISH A BASELINE

Another way those we interviewed thought about accuracy was to think about accuracy not as an absolute, but in comparison to humans. A recent grad who had founded a machine learning start-up told us, "The way that I think about it [accuracy] is against the human gold standard. The human gold standard is shockingly good and shockingly bad in different cases. This is actually a problem that I faced very tangibly in my past work in that we have a lot
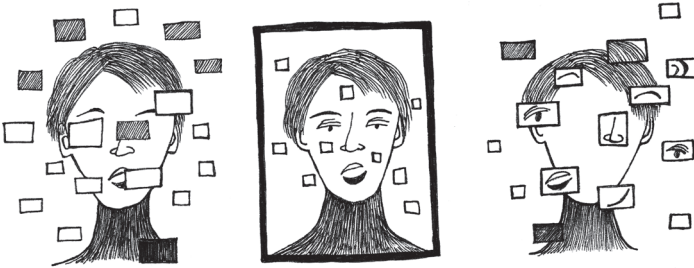
of assumptions that the human gold standard is very good but it turns out it's really bad. For example, the 'how old are you' type software, if you take an average person and show them a picture of a face and ask them how old is the person in this picture, it's really difficult. They are on average extraordinarily bad at it, though they may still be good at recognizing the face and things like that. But it becomes this interesting philosophical debate which is: is the goal to be most empirically correct or is it to match what a person thinks?"

Unsurprisingly, all of the people we interviewed expressed the opinion that their software programs are generally better and more accurate than humans in the same context. As one venture capitalist in San Francisco put it, "Machines are not biased. They don't think all Asians look alike, you know? They don't have these preconceptions that some of us have. We're like, 'Who cares whether it's an Indian or Japanese face?' For them, it's just the data of the face, and so once they got better, they've lost all of those biases. They don't wake up and they're grumpy, or half bleary-eyed from a hangover from the night before. They... There

could be biases in the training data, absolutely. But my point is once you've cracked the nut, those are things that you can study and correct, et cetera, and you can detect them much more precisely than you could detect them in a human." As the founder of a very successful machine learning company put it, "We are more flawed than our algorithms."

## 3 DIFFERENT LANGUAGES, DIFFERENT PERSPECTIVES

In this paper, we identified a set of core challenges and patterns that emerged from our research and discussions with industry actors. As stated in the introduction, while we believe these patterns are representative of currently dominant industry perspectives, these patterns do not comprehensively cover the range of ways in which industry is approaching intelligent systems. Moreover, as AI has become the latest buzzword and occasioned an influx of capital, the industry will be expanding rapidly in the next several years and will necessarily evolve in composition and shift perspectives.
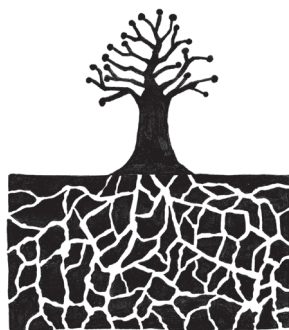
The patterns that emerge from our conversations with industry practitioners have much to tell us about the current state of developing and deploying intelligent systems. But much can also be learned from how these same issues are framed by scholars and critics— and, notably, the differences that exist between these perspectives. In particular, it becomes clear how the

different contexts within which experts must deal with technological problems inform the very ways in which "a problem" is framed and constituted. Like a language, these differing perspectives produce different ways of thinking and talking, ways that may be opaque to others outside those environments. This is, perhaps, an obvious point but one that bears repeating in the context of the endeavor to address the social implications of intelligent systems, which requires collaboration between industry practitioners, policy makers, academic researchers, and other experts. When these diverse communities come to the same table, are they talking about the same things?

For instance, confronting the privacy implications of intelligent systems has been a central concern for academic and industry practitioners, but with differing commitments. While the language of industry uses terms like "data collection" and "data security" to describe the kinds of techniques that can impinge upon individual privacy, academic discussions tend to use terms such as "surveillance" which invoke the asymmetrical dynamics of control and power. While the industry patterns demonstrate an emphasis on finding levers to control the local effects of an intelligent system, social scientific scholars have been more attuned to systemic effects and invoke the language of governance.[1] While both communities care about issues of privacy, what these issues are and why they matter is distinct.

Scholars are quick to highlight the economic interests of corporate production as a factor that fundamentally biases the technological systems, while industry practitioners see market forces as a functional check on their practice. Yet, in the same vein, industry practitioners recognize that they are designing for particular users or markets, often prompting critics to challenge the equity of the produced tools. Economic factors – and attitudes towards capitalism – often create a division between researchers and practitioners that has little to do with technology, except to the degree that technology mirrors and magnifies existing differences in attitude about socio-technical systems.

Each challenge as it emerges from industry is essentially—and understandably—about developing a productive relationship between a user or client and a product or company. If we were to think about what such parallel challenges might be in the context of social science, these challenges would likely be about articulating a social value that exceeds any product-user relationship. Because most researchers are focused on broad social implications and systemic factors, their approach to analysis is often at odds with the individual-consumer-centric perspective of the practitioners we encountered. The dilemma we see is not that communities differ in focus but rather that this difference can create an unbreachable barrier to understanding and collaboration.

# 4   CONCLUSION

Intelligent systems are in the early stages of being developed and deployed. While they may change nearly every aspect of life in the far future, in the near term we can begin to analyze and plan for social impacts we are already beginning to see. To date, much of this nascent analysis and planning has been siloed not only across different disciplines but also between industry, academia, and government. This disconnected environment results in conversations not being heard by communities who need to hear them, and perspectives not being fully understood when they arise from an unfamiliar position. This pattern language has attempted to bring multiple conversations and communities together, in an experimental form. Distilling the range of perspectives into patterns begins the work, we hope, of developing a common language and mutual intelligibility from which to address the social implications of deploying AI and intelligent systems.

In the past year, the specter of superintelligence and the singularity has begun to recede. Media coverage, as well as scholarly conferences and output, have begun to focus on more pressing concerns, many of which are addressed in this document.[1] Still, the need for a nuanced

understanding of the social implications of intelligent systems remains. Moreover, recent appeals to developing codes of ethics and developing "ethical technology" hold out promise, but far from automatic success. The social implications of intelligent systems deployment are highly contextual and must ultimately be examined as such.

One user experience designer explained to us that he wished more people thought about the profound power that designers of intelligent systems wield. He told us about a situation in which he was confronted with this consequential power, "There was an idea [to potentially work on and develop] that if someone yelled at their phone or cursed at it, could we make it run faster? I very quickly made a comment that I don't want to teach my children that yelling at someone makes them perform better. I want to use this technology to help build inter-human social models." The choices we make about how to develop technology ripple beyond any one product or discrete interaction.

As stated in the introduction, this document has been an experiment in cataloging and catalyzing ideas. The pattern language should be a living document, evolving as new technologies and new norms emerge. All of the patterns could be expanded and refined, and as more truly intelligent systems are deployed, a set of best practices could begin to be more comprehensively developed.

In *A Pattern Language*, Alexander and his team aimed to create modular examples that would empower individuals and communities to take control of the built environment

and build a world that reflected and embedded certain ideal values. *An AI Pattern Language* is not prescriptive in this sense, but does aim to be empowering. The patterns in this document call attention to the sometimes nascent, sometimes explicit ways in which humans makes choices about the development and deployment of these systems. Intelligent systems are not out of control nor are they magical nor are they objective perfection. They are compounds of specific human choices, and every decision matters.

# 5 ACKNOWLEDGMENTS

The Intelligence & Autonomy initiative aims to reframe debates around the rise of machine intelligence. By expanding the analysis of social and policy issues arising from intelligent systems, the project aims to expose the recurring problems which confront designers and policymakers in implementing these systems across domains as varied as healthcare, policing, capital markets, social media, and transportation. I&A is also dedicated to expanding the frame through which we understand the rise of intelligent systems through historical research, placing recent innovations in the context of former waves of automation and technological development. By doing so, I&A aims to produce more effective policy-making and also temper some of the claims being made about these technologies.

Data & Society is a research institute in New York City that is focused on social and cultural issues arising from data-centric technological development. Data & Society is committed to identifying issues at the intersection of technology and society, providing research that can ground public debates, and building a network of researchers and practitioners who can offer insight and direction.

## INTRODUCTION

1. John Markoff. 2015. "Software is smart enough for SAT, but still far from intelligent." *New York Times* Sept 20. http://www.nytimes.com/2015/09/21/technology/personaltech/software-is-smart-enough-for-sat-but-still-far-from-intelligent.html (accessed 5/15/2016). For more in depth and historical reviews see, Rupert Goodwins. 2015. "Demystifying artificial intelligence: No, the Singularity is not just around the corner." *ArsTechnia UK* Dec 21 http://arstechnica.co.uk/information-technology/2015/12/demystifying-artificial-intelligence-no-the-singularity-is-not-just-around-the-corner/ (accessed 6/2/2016); Edward Moore Geist. "Is artificial intelligence really an existential threat to humanity?" 2015. *Bulletin of the Atomic Scientists* Aug 9. http://thebulletin.org/artificial-intelligence-really-existential-threat-humanity8577 (accessed 6/2/2016).

2. Christopher Alexander, Sara Ishikawa, and Murray Silverstein. 1977. *A pattern language: Towns, buildings, construction.* New York: Oxford University Press: x.

3. Ibid. xiii

4. We were guided by the overviews of others working in the field, including: Venture Scanner. 2015. "The State of Artificial Intelligence in Six Visuals." *Venture Scanner Blog* Aug 12. https://medium.com/@VentureScanner/the-state-of-artificial-intelligence-in-six-visuals-8bc6e9bf8f32#.6i73j1agq (accessed 5/15/2016); Dorian Pyle and Cristina San Jose. 2015. "An executive's guide to machine learning." *McKinsey Quarterly* June. http://www.mckinsey.com/industries/high-tech/our-insights/an-executives-guide-to-machine-learning (accessed 5/15/2016).

5. Jane Margolis and Allan Fisher. 2003. *Unlocking the Clubhouse: Women in Computing.* Cambridge, MA: MIT Press.

6. Daniela Hernandez. 2014. "Artificial Intelligence Is Now Telling Doctors How to Treat You." *Kaiser Health News* and *Wired.* June 2. http://www.wired.com/2014/06/ai-healthcare/ (accessed 4/22/2016).

7.  Chris Weller. 2016. "The world's first artificially intelligent lawyer was just hired at a law firm." *Tech Insider.* May 16. http://www.techinsider.io/the-worlds-first-artificially-intelligent-lawyer-gets-hired-2016-5 (accessed 6/20/2016).

8.  For a detailed empirical analysis of what tasks intelligent software can perform in legal firms see Frank Levy and Dana Remus. 2015. "Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law." Working Paper. December 30. Available at SSRN: http://ssrn.com/abstract=2701092 (accessed 6/20/2016).

9.  CBS Sunday Morning. "The future of robots and artificial intelligence." *CBS*. Video. June 14. http://www.cbsnews.com/videos/the-future-of-robots-and-artificial-intelligence/ (accessed 6/20/2016).

10.  As computer scientists Stuart Russell and Peter Norvig observe in their widely used AI textbook, the history of artificial intelligence has not produced a clear definition of AI but rather can be seen as variously emphasizing four possible goals: "systems that think like humans, systems that act like humans, systems that think rationally, systems that act rationally." Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, NJ: Prentice Hall, 1995: 27.

11.  Tom Mitchell. 1997. *Machine Learning*. McGraw Hill: 2.

12.  NHTSA Press Release. 2013. U.S. Department of Transportation Release Policy on Automated Vehicle Development. May 13. See also Robin Murphy and James Shields. 2012. *DOD DSB Task Force Report: The Role of Autonomy in DoD Systems*. Defense Science Board, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics. Washington, DC.

13.  Parasuraman et al. 2000. "A Model for Types and Levels of Human Interaction with Automation." *IEEE Transactions on Systems, Man and Cybernetics* 30(3).

## DIFFERENT LANGUAGES, DIFFERENT PERSPECTIVES

1. For instance, scholars such as Danielle Citron, Frank Pasquale, Kate Crawford and Jason Shultz have argued for new forms of "technological due process," among other mechanisms of accountability. Danielle Keats Citron. 2007. "Technological Due Process." *Washington University Law Review* 85: 1249-1313; Kate Crawford and Jason Shultz. 2014. "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms." *Boston College Law Review* 55(1): 93-128.

## CONCLUSION

1. In May of 2016, The White House announced a series of workshops and an interagency working group to prepare for the future of artificial intelligence. The workshop and public symposium in New York City, AI Now, in particular addressed important social and economic impacts in the near-term of AI. Videos and summaries of the event can be found on the AI Now website, https://artificialintelligencenow.com.