

ECSE 557 Week 3 Student Guidelines

Tutorial 3: Values and AI systems

Background

Last tutorial we looked at three dominant ethical theories and applied them to resolve ethical dilemmas. We will turn our attention to value ethics in this tutorial and learn about how understanding values can inform design of products/processes. During the tutorials, the teacher assistant will go through an application of Value Sensitive Design (VSD) for an AI application and then you will be asked to apply VSD on another case study. The takeaways from this tutorial can be used to complete your first assignment and components of your project milestones.

Overview

The tutorial will start with TA's demonstration of value sensitive design process with a focus on value tension. Following the introduction, you will work in groups of 3 or 4 in a breakout room to independently apply the concept of value tension within a given scenario. You will complete Tasks 1 - 3 in these groups and report your discussion to the larger class when you return to the main Zoom room at the end of the class.

What to submit

You do not need to submit anything for this tutorial; however, you can use the ideas covered in this tutorial to inform your work in Assignment 1.

Task 1: Stakeholders and their values (5 min)

- Take some time to review the case study and identify the direct and indirect stakeholders.
- Identify the values that are likely most important for the direct and indirect stakeholders.

Task 2: Technology and embedded values (5 min)

- Describe how the technology is interacting/impacting with the direct and indirect stakeholders
- Identify possible values embedded in design of the HireVue platform.
 - What are the values embedded in how data is collected?
 - What are the values embedded in the choice of mode?
 - What are the values embedded in the output and the user interface?

Task 3: Value tensions (10 min)

- Compare technology embedded norms against the stakeholder values:
 - Do you see any obvious value tensions?
 - Elaborate on the nature of these tensions and their impact.
- Compare stakeholder norms with respect to each other:
 - Do you see any obvious value tensions?
 - If so, elaborate on what are the potential tensions and how do they manifest themselves in the case study? How would they impact technology adoption and impact?
- After completing tasks 1-3 you will be invited back to the main room where you will be asked to report back on your discussion.

Scenario

The following case study is excerpts from [The Washington Post article on the HireVue interview system](#). To meet the time-constraints of the tutorial we have only picked parts of the article; however, feel free to review the article on your own time for the full details. Since this article HireVue has [conducted algorithmic audit](#) of their products and has [dropped facial monitoring](#) based on the results.

An artificial intelligence hiring system has become a powerful gatekeeper for some of America's most prominent employers, reshaping how companies assess their workforce — and how prospective employees prove their worth. Designed by the recruiting-technology firm HireVue, the system uses candidates' computer or cellphone cameras to analyze their facial movements, word choice and speaking voice before ranking them against other applicants based on an automatically generated "employability" score.

AI researchers argue that the system will assume a critical role in helping decide a person's career. But they doubt it even knows what it's looking for: Just what does the perfect employee look and sound like, anyway? "It's a profoundly disturbing development that we have proprietary technology that claims to differentiate between a productive worker and a worker who isn't fit, based on their facial movements, their tone of voice, their mannerisms," said Meredith Whittaker, a co-founder of the [AI Now Institute](#), a research center in New York. "It's pseudoscience. It's a license to discriminate," she added. "And the people whose lives and opportunities are literally being shaped by these systems don't have any chance to weigh in."

Loren Larsen, HireVue's chief technology officer, said that such criticism is uninformed and that "most AI researchers have a limited understanding" of the psychology behind how workers think and behave. Larsen compared algorithms' ability to boost hiring outcomes with medicine's improvement of health outcomes and said the science backed him up. The system, he argued, is still more objective than the flawed metrics used by human recruiters, whose thinking he called the "ultimate black box." "People are rejected all the time based on how they look, their shoes, how they tucked in their shirts and how 'hot' they are," he told The Washington Post. "Algorithms eliminate most of that in a way that hasn't been possible before." The AI, he said, doesn't explain its decisions or give candidates their assessment scores, which he called "not relevant." But it is "not logical," he said, to assume some people might be unfairly eliminated by the automated judge. "When 1,000 people apply for one job," he said, "999 people are going to get rejected, whether a company uses AI or not."

HireVue said its system dissects the tiniest details of candidates' responses — their facial expressions, their eye contact and perceived "enthusiasm" — and compiles reports companies can use in deciding whom to hire or disregard. Job candidates aren't told their score or what little things they got wrong, and they can't ask the machine what they could do better. Human hiring managers can use other factors, beyond the HireVue score, to decide which candidates pass the first-round test.

The system, HireVue said, employs superhuman precision and impartiality to zero in on an ideal employee, picking up on telltale clues a recruiter might miss. Major employers with lots of high-volume, entry-level openings are increasingly turning to such automated systems to help find candidates, assess résumés and streamline hiring.

ECSE 557 Tutorial 3: Values and AI systems

Nathan Mondragon, HireVue's chief industrial-organizational psychologist, told The Post the standard 30-minute HireVue assessment includes half a dozen questions but can yield up to 500,000 data points, all of which become ingredients in the person's calculated score. The employer decides the written questions, which HireVue's system then shows the candidate while recording and analyzing their responses. The AI assesses how a person's face moves to determine, for instance, how excited someone seems about a certain work task or how they would behave around angry customers. Those "Facial Action Units," Mondragon said, can make up 29 percent of a person's score; the words they say and the "audio features" of their voice, like their tone, make up the rest.

To train the system on what to look for and tailor the test to a specific job, the employer's current workers filling the same job — "the entire spectrum, from high to low achievers" — sit through the AI assessment, Larsen said. Their responses, Larsen said, are then matched with a "benchmark of success" from those workers' past job performance, like how well they had met their sales quotas and how quickly they had resolved customer calls. The best candidates, in other words, end up looking and sounding like the employees who had done well before the prospective hires had even applied.

After a new candidate takes the HireVue test, the system generates a report card on their "competencies and behaviors," including their "willingness to learn," "conscientiousness & responsibility" and "personal stability," the latter of which is defined by how well they can cope with "irritable customers or co-workers." Those computer-estimated personality traits are then used to group candidates into high, medium and low tiers based on their "likelihood of success." Employers can still pursue candidates ranked in the bottom tier, but several interviewed by The Post said they mostly focused on the ones the computer system liked best.

HireVue offers only the most limited peek into its interview algorithms, both to protect its trade secrets and because the company doesn't always know how the system decides on who gets labeled a "future top performer." The company has given only vague explanations when defining which words or behaviors offer the best results. For a call center job, the company says, "supportive" words might be encouraged, while "aggressive" ones might sink one's score. HireVue said its board of expert advisers regularly reviews its algorithmic approach, but the company declined to make the system available for an independent audit. The company, Larsen said, is "exploring the use of an independent auditor right now, to see how that could work."

HireVue launched its AI assessment service in 2014 as an add-on to its video-interview software, which more than 700 companies have used for nearly 12 million interviews worldwide. The Utah-based company won't disclose its revenue, the cost for employers or a full list of clients.