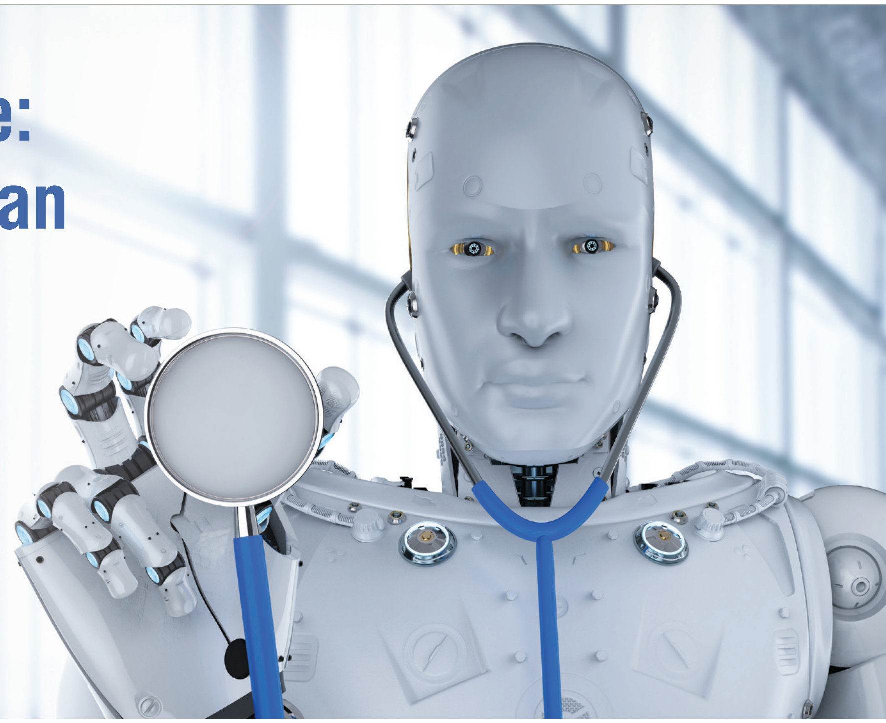
AI in healthcare: Improving human interface for patient safety



Better standards needed to make artificial intelligence user-friendly for clinicians

By Avishek Choudhury

Working in healthcare can cause major mental, physical and organizational stress, which in turn can lead to clinical hazards. Artificial intelligence may help, though recent research on such advanced technology is largely devoted to patient safety; regulatory bodies conducting such studies, such as the U.S. Food and Drug Administration (FDA), focus on improving an AI system's performance purely from a statistical viewpoint. What's missing is a need to integrate AI and cognitive ergonomics to make it more user-friendly for medical workers.

Cognitive ergonomics – the gathering of knowledge about human perception, memory and mental processes – has been neglected in the healthcare domain and the evaluation of AI from a cognitive ergonomics perspective is not well-established. Though the effect of clinicians' heavy mental load on patient safety is known, there is no framework to guide the graphical user interface (GUI) of complex AI systems and their influence on clinicians' and patients' thoughts and feelings. Here we seek to adopt a systems approach to develop and propose a conceptual framework across AI, cognitive ergonomics and patient safety.

The gap between cognitive ergonomics, AI

Cognitive ergonomics is a component of human factors and ergonomics aimed at ensuring effective interaction between technology and humans. In this article, we discuss human-system interaction from a cognitive ergonomics perspective in healthcare. In this interaction, cognitive ergonomics concentrates on mental processes such as thinking, reasoning and problem-solving, as well as psychological or behavioral interactions. In cognitive ergonomics, these aspects are studied in the context of work and other systems.

Due to the increasing complexity of healthcare, researchers are focused on the application of AI primarily in improving its ability to provide accurate diagnoses. AI systems have long held the promise of refining the prediction of diseases, such as guiding imaging for pulmonary embolisms.

A healthcare AI system has two dimensions that impact cognitive workload: Complexity of the algorithm and user interface. Many researchers have tried to simplify its underlying algorithm, yet no significant steps have been taken to improve the AI systems interface. Thus, its impact on human cognition remains uncertain.

A well-designed graphical interface can help an AI user locate relevant information, then interpret and prioritize it. The significance of cognitive ergonomics and human factors in the intersection of healthcare and AI has not yet been studied and cognitive load has been assumed rather than measured. There is therefore a need to use the methods of cognitive ergonomics in healthcare artificial intelligence systems. We propose a framework to simplify a GUI of healthcare AI systems and understand its impact on clinicians' cognitive load.

Meaningful outcome and cognitive load

Those who use advanced AI, such as clinicians, might face difficulties in understanding and interpreting the outcomes of the technology. This may be due to the users' inability to form a conceptual model of the information presented on a computer screen or a device, and their lack of understanding of the system's working principle. Addressing confusion with AI technology and its interface will not only aid in safer application but can foster better satisfaction.

Nevertheless, a benchmark for examining the cognitive load generated by a healthcare AI system is not defined in the FDA's premarket clearance program. In an early-phase study of 326 hospitalized patients, the FDA approved a predictive algorithm, WAVE, which indicates vital signs' abnormalities and has led to a reduction in the average duration of patient instability. Though this was judicious under current regulatory standards, the approved system was not tested for the usability and complexity of its GUI. Is the system's interface simple or intuitive enough for an inexperienced clinician to implement and understand in a chaotic environment? It is uncertain whether WAVE has reduced clinicians' cognitive load.

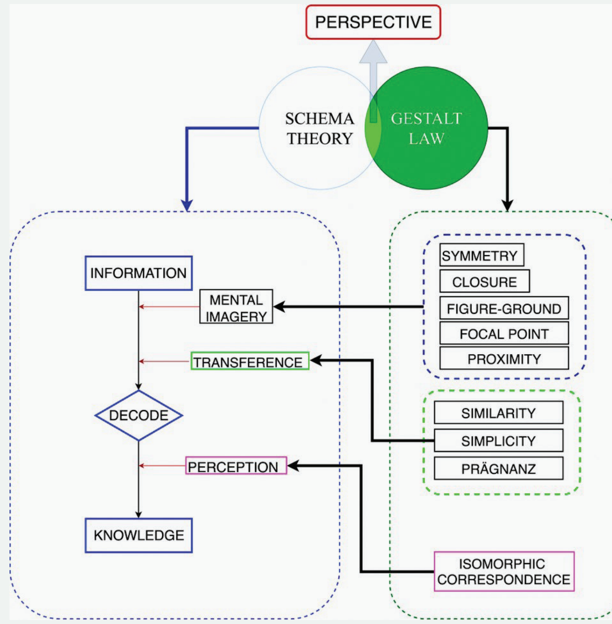
The FDA should rigorously confirm and test surrogate endpoints of new technologies to prohibit the introduction of AI systems with questionable GUI into a chaotic healthcare environment where human life is at risk. The agency must ensure that implementing an AI system not only improves diagnostic capabilities but also minimizes a clinician's time spent analyzing and interpreting complex outcomes; such necessary measures could be valuable for premarketing authorization. Unfortunately, no AI algorithms or systems that received regulatory clearance have been tested for their impact on cognitive load.

With the increase of technology in healthcare, clinicians and patients are encountering AI devices that involve complex and unfamiliar GUIs. The industry needs to conduct comprehensive and informative research to develop user-friendly AI devices and systems by considering such design needs as human memory limitation, perception and attention.

FIGURE 1

Schema and Gestalt

A look at the fundamental functioning of human perspective.



GUI design considerations

Here are aspects of GUI design that should be considered when approving new AI technology for healthcare use:

Human memory limitation. This partly involves retention theory and cognitive load theory. The human memory allows an average person to retain seven (plus or minus two) sets of information at a time. Thus, to ensure better retention, health information should be divided into smaller units that do not exceed the limits of 7 ± 2 per AI system display. Displaying fewer sets of information per screen can reduce clinicians' need to memorize data, which in turn helps minimize the cognitive load on them and their patients. In addition, using suitable colors in designing GUI is shown to enhance information retention by 50%.

Perception. This aspect employs the schema theory and Gestalt law, illustrated in Figure 1. A schema signifies the conceptual representation of how an individual decodes information and extracts contextual knowledge. For instance, when we see an image showing sun, blue sky and birds flying, we interpret it as daytime and associate a feeling of happiness. In addition, users' prior experience and knowledge plays a crucial role in enhancing their perceiving ability (transference and mental imagery).

Transference denotes the users' anticipation of the behavior of an AI system based on past experience with other computer interfaces, such as text placement and the location, appearance and functionality of buttons and icons. Mental imagery refers to the conceptual representation of how things look.

Gestalt law is explained in the table below; some law of Gestalt complements schema theory as shown in Figure 1.

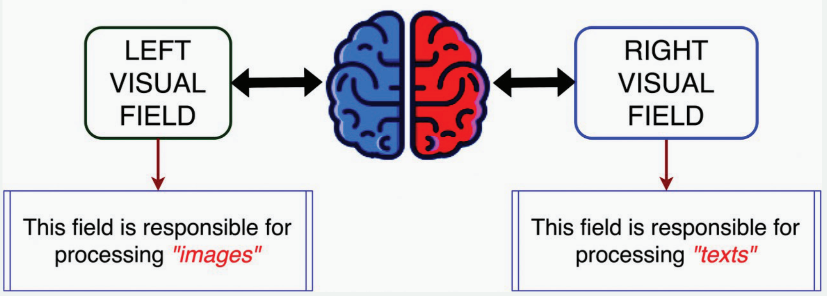
Laws of Gestalt	Meaning
Balance/symmetry	Asymmetrical images are perceived as incomplete
Continuation	Eyes typically follow a direction as shown in GUI
Closure	Shapes that are not fully closed are perceived as incomplete
Figure-ground	Different colors are perceived as different entities
Focal point	Each AI system's display should have a focal point
Isomorphic correspondence	Different users perceive a same image differently
Prägnanz	Simple GUI design are considered as good forms
Proximity	Objects that are clustered are perceived as a group
Similarity	Similar objects grouped together attract more attention
Simplicity	Users understand faster if the information is simple
Unity/harmony arrangement of the objects	Objects are perceived as one group if there is a similar arrangement of the objects

Attention. In the placement of texts and images, the left-to-right theory can be applied in designing an effective AI interface. This suggests allocating critical information on the top left corner of the screen. This is due to the responsiveness

FIGURE 2

Human visual field

Where the eyes and the brain connect to process information.



where people tend to interpret data from upper left to lower right. Additionally, the placement of texts and images should be applied considering the human visual field, which is divided into two different segments, as shown in Figure 2. Thus, the placement of texts or images according to humans' visual field minimizes the users' cognitive load.

Clinicians may experience difficulties whenever there are multiple parallel tasks overwhelming their senses. For example, an alarm noise in an emergency floor or a patient talking in an outpatient setting can be distracting if auditory information is also running on the AI system. Such multitasking should be avoided to reduce their cognitive load. Their sense of sight might be impaired when background images or texts are embedded under pieces of informative texts. Thus, AI-enabled applications should avoid such embedded content. And the need to process information simultaneously by two motor systems increases cognitive load, such as listening to a patient's distress while using a complex AI system.

Trust and meaningful use

As governing bodies decide which downstream features matter for AI systems, they should also keep in mind these systems will fail as a technology without acceptance and trust from clinicians despite having good analytical performance. Studies have shown that trust is built in a continuous manner, demanding two-way interactions between the user and the technology.

Initial trust is essential for ensuring the adoption of new technology. Trust is influenced by a user's first impression and is built based on that person's personality and institutional cues. Once trust is developed, it must be nourished to be sustained. In our context, continuous trust depends on the functioning of the AI system in reducing users' cognitive load and yielding clinically meaningful outcomes.

It has been believed that trust in technology is determined by human characteristics (personality and ability), environmental characteristics (culture, task and institutional factors) and technology characteristics (performance, process and purpose). However, the impact of meaningful use on trust has been neglected. Meaningful use, just as it is imposed on the application of EHR technology, should be imposed on AI systems to improve the quality of care.

The meaningful use of AI systems means that specific AI should be implemented and interpreted in a specific manner. Depending on the functioning of the AI algorithm, not all systems can be generalized across a healthcare system.

For instance, the WAVE platform algorithm is based on five vital signs: heart rate, respiration, oxygen saturation, temperature and blood pressure. Since such measures are common across health systems, the system could be employed by multiple diverse health systems. However, other AI-based platforms, especially those based on institution-specific EHR or image datasets, may not translate across other EHRs. Moreover, Al trained on specific datasets, such as patients from a specific institution, may not be generally applied across broader populations.

Increasing AI interoperability may necessitate developers to deliver more specified data to confirm that predictive algorithms will achieve reliable, replicable and valid results. Indeed, regulators should focus on balancing the clarity of predictive models without impeding the proprietary interests and intellectual property of algorithm developers.

Healthcare AI systems are just getting evaluated and being made available for clinical use, so the influence of the existing regulatory framework on patient outcomes is yet to be determined. It is also uncertain what impact the 21st Century Cures Act passed to relax regulatory standards for low-risk health technology will have on the value and quality of predictive algorithms. The FDA's Digital Health Innovation Action Plan, issued in 2017, launched a precertification program to analyze clinical outcomes of AI-based algorithms. Such efforts should be acclaimed but improved based on our recommended norms.

Some developers may disparage the overregulation and standardization of a vaguely understood field. Indeed, a pledge to regulate healthcare AI systems will emerge over time and impose financial costs to stakeholders. Policymakers should also be sensitive to the stability between regulation and innovation in this evolving field. ❖

Avishek Choudhury is a Ph.D. student and research assistant in the School of Systems and Enterprises at Stevens Institute of Technology in Hoboken, New Jersey. He received his master's degree in industrial and systems engineering from Texas Tech University. His research interests include healthcare systems, artificial intelligence, clinical decision-making, patient safety and network science. He is an IISE member.

5 examples of how technology is changing healthcare

Artificial intelligence, machine learning, wearable sensors and virtual reality have taken root in various aspects of industrial and systems engineering, nowhere more influential than in healthcare. Here is a sample of the new technological processes in the works:

Seeing through a robot's eyes helps those with motor impairments live independently

Photo by Henry Clever/Phillip Grice, Georgia Tech

Henry Evans, a California man who helped Georgia Tech researchers with improvements to a web-based interface, uses a robot to shave himself. The team worked to develop robotic body surrogates to help people with profound motor deficits interact with the world.

An interface system using augmented reality technology could help individuals with profound motor impairments operate a robot to feed themselves and perform routine personal care tasks such as scratching an itch and applying skin lotion. The interface displays a "robot's eye view" of surroundings to help users interact with the world through the machine.

The system, described in the journal *PLOS ONE*, could help make robots useful to people by using standard assistive computer access technologies such as eye trackers and head trackers.

"Our results suggest that people with profound motor deficits can improve their quality of life using robotic body surrogates," said Phillip Grice, a recent Georgia Tech Ph.D. graduate and first author of the paper.

Researchers used a PR2 mobile manipulator manufactured by Willow Garage, a wheeled robot with two arms and a "head" and the ability to manipulate objects such as water bottles, washcloths, hairbrushes and even an electric shaver.

The study made the PR2 available across the internet to 15 participants with severe motor impairments. They learned to control the robot remotely using their own assistive equipment to operate a mouse cursor to perform a personal care task. Eighty percent of the participants were able to manipulate the robot to pick up a water bottle and bring it to the mouth of a mannequin.

In a second study, researchers provided the PR2 and interface system to Henry Evans, a California man with limited control of his body who tested the robot in his home for seven days. He not only completed tasks, but also devised novel uses combining the operation of both robot arms at the same time – one to control a washcloth and the other to use a brush.

"The system was very liberating to me, in that it enabled me to independently manipulate my environment for the first time since my stroke," Evans said.

Source: Georgia Tech

AI can detect 26 skin conditions as accurately as dermatologists, better than primary care doctors

Skin conditions are among the most common kind of ailment, resulting in an estimated 25% of all patient treatments worldwide and up to 37% of clinical patients having at least one skin complaint.

With dermatologists facing enormous caseloads and general practitioners tending to be less accurate in identifying skin conditions, researchers at Google investigated an AI system capable of spotting the most common dermatological disorders seen in primary care. Their paper, "A Deep Learning System for Differential Diagnosis of Skin Diseases," reports accuracy across 26 skin conditions when presented with images and metadata about a patient's case, which they say is on par with U.S. board-certified dermatologists.

Google software engineer Yuan Liu and Google Health technical program manager Dr. Peggy Bui explained that dermatologists don't give just one diagnosis for any skin condition but instead generate a ranked list of possible diagnoses narrowed by lab tests, imaging, procedures and consultations. The system does the same by processing inputs that include one or more clinical images of the skin abnormality and up to 45 types of metadata.

To test the system's accuracy, researchers compiled diagnoses from three U.S. board-certified dermatologists and the AI system's ranked list of skin conditions achieved 71% and 93% top-1 and top-3 accuracies, respectively. When the system was compared with clinicians, the team reported its top three predictions demonstrated a top-3 diagnostic accuracy of 90%, or comparable to dermatologists (75%) and "substantially higher" than primary care physicians (60%) and nurse practitioners (55%).

Source: Google

AI is learning to read mammograms

Researchers in the United States and Britain have found that artificial intelligence can help doctors do a better job of finding breast cancer on mammograms.

According to an article in the journal *Nature*, the new system for reading mammograms is not yet available for widespread use. Google paid for the study and worked with researchers from Northwestern University in Chicago and two British medical centers, Cancer Research Imperial Centre and Royal Surrey County Hospital.

The system performed better than radiologists in diagnosing known cases from images. Reviewing scans of 3,000 women in the U.S., the system produced a 9.4 percent reduction in false negatives, when cancer is missed, and a 5.7 percent reduction in false positives, where there is no cancer. A study of 25,000 mammograms from Britain showed that AI reduced false negatives by 2.7 percent and false positives by 1.2 percent.

About 33 million screening mammograms are performed each year in the United States. The test misses about 20 percent of breast cancers, according to the American Cancer Society, and false positives are common.

"There are many radiologists who are reading mammograms who make mistakes, some well outside the acceptable margins of normal human error," said Dr. Constance Lehman, director of breast imaging at the Massachusetts General Hospital in Boston.

To train computers to read the mammograms, the authors used scans from about 76,000 women in Britain and 15,000 in the United States whose diagnoses were already known.

Sources: New York Times, www.nature.com

Virtual reality could boost flu shot rates

Researchers at the University of Georgia and the Oak Ridge Associated Universities in Oak Ridge, Tennessee, have conducted a study using virtual reality simulation to show how flu spreads and its impact on others as a way to encourage more people to get vaccinated. It is mostly aimed as a communication tool at the "flu vaccine avoidant" 18- to 49-year-old adults.

The research, "Using Immersive Virtual Reality to Improve the Beliefs and Intentions of Influenza Vaccine Avoidant 18- to 49-year-olds," was published by the journal Vaccine.

The Centers for Disease Control and Prevention reports only 26.9% of 18- to 49-year-olds in the U.S. received a recommended annual influenza vaccination during the 2017-18 flu season. The low current acceptance of flu vaccination makes it important to identify more persuasive ways to educate them.

The 171 participants in the study were divided into four groups, each exposed to a different form of persuasion: A virtual reality experience, a video based on the VR platform, an e-pamphlet with text and images from the video, and a control group provided only with the CDC recommendations.

The virtual reality group wore headsets that enabled them to experience events and controllers enabling them to actively participate in the story. It resulted in participants showing greater confidence that a vaccination would protect others, more positive beliefs about flu vaccine and increased intention to get a vaccination.

Source: University of Georgia

Photo by Sarah Freeman, University of Georgia

University of Georgia faculty members tested methods of delivering effective flu vaccination messages through print, video and virtual reality.

Wireless sensors on patients' skin tracks health

Stanford engineers have developed wireless sensors that can detect physiological signals emanating from the skin and beam wireless readings to a receiver clipped onto clothing.

The sensors can be applied like Band-Aids on various areas. To demonstrate the technology, researchers stuck sensors on the wrist and abdomen of one test subject to monitor pulse and respiration by detecting how the skin stretched and contracted. Sensors on the elbows and knees tracked arm and leg motions by gauging tightening or relaxation of the skin when muscles flexed.

Zhenan Bao, the chemical engineering professor whose lab described the system in an Aug. 15 article in *Nature Electronics*, thinks this wearable technology, called BodyNet, will first be used in medical settings such as monitoring patients with sleep disorders or heart conditions. The goal is to create an array of wireless sensors on the skin that work with smart clothing to accurately track a wide variety of health indicators.

"We think one day it will be possible to create a full-body skin-sensor array to collect physiological data without interfering with a person's normal behavior," said Bao.

Sources: Stanford University, Nature Electronics