Designing Interactive Intelligent Systems for Human Learning, Creativity, and Sensemaking

Hariharan Subramonyam
University of Michigan
Ann Arbor, MI, USA
harihars@umich.edu

ABSTRACT

The focus of intelligent systems is on "making things easy" through automation. However, for many cognitive tasks—such as learning, creativity, or sensemaking—there is such a thing as too easy or too automated. Current human-AI design principles, as well as general usability guidelines, prioritize automation, and efficient task execution over human effort. However, this type of advice may not be suitable for designing systems that need to balance automation with other cognitive goals. In these cases, designers lack the necessary tools that will allow them to consider the trade-offs between automation, AI assistance, and human-effort. My dissertation looks at using models from cognitive psychology to inform the design of intelligent systems. The first system, Florum, looks at *automation after human-effort* as a strategy to facilitate learning from science text. The second system, TakeToons, explores *automation as a complementary strategy to human-effort* to support creative animation tasks. A third set, SmartCues and Affinity Lens use *AI as a last-mile optimization* strategy for human sensemaking tasks. Based on these systems, I am looking to develop a design framework that (1) classifies threats across different levels of design including automation, user interface, expectations from AI, and cognition and (2) offers ways to validate design decisions.

ACM Classification Keywords

Human-centered computing Interaction design; Computing methodologies Cognitive science; Computing methodologies Intelligent agents

Author Keywords

Human-AI; Interaction Design; Cognitive models;

INTRODUCTION

Intelligent applications allow non-experts to create animated stories, journalists to quickly churn out news articles, and students to solve complex equations by simply taking a picture.

UIST'19 Adjunct, October 20–23, 2019, New Orlens, LA, USA

DOI: http://dx.doi.org/10.1145/3332167.3356878

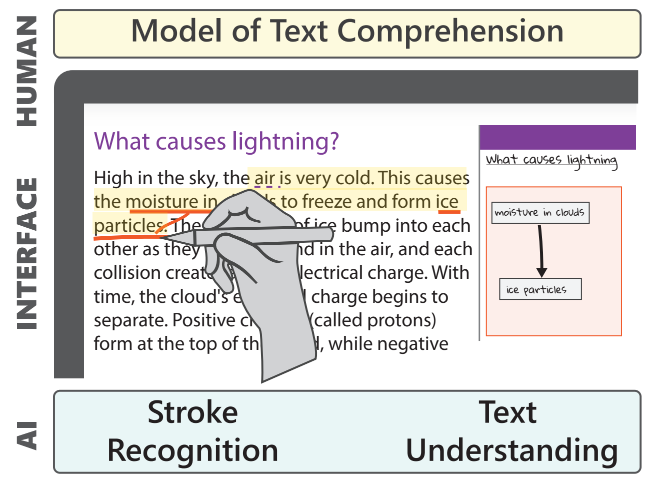

Figure 1. Florum: An active reading tool for causal diagramming and causal text comprehension.

The overarching goal across all these applications is to *make things easy* for end-users. However, if we consider the cognitive processes associated with these tasks such as human learning, creativity, or sensemaking, there is such a thing as too easy or too automated. For example, if AI replaces the creative agency of animators by directly animating from a script, the resulting experience would be less enjoyable for the animator. Or if learning to solve equations is the goal, automatically solving them for the student is less than ideal. In such cases, how do designers decide as to where along the continuum of **human-effort—augmentation—automation** lies the optimal user experience that is desirable to end-users but also meets their cognitive goals and expectations?

Current human-AI design principles, as well as general usability guidelines, prioritize automation and efficient task execution over human-effort (e.g., [7, 9]). For instance, mixed-initiative design takes a probabilistic approach to act on users' goals when the goals are clear and uses dialog to resolve any uncertainty. When applied to cognitive tasks, it is unclear how such a system should handle the trade-off between goal execution and disruption to cognition (e.g., the goal of the student to solve the equation, and the underlying learning objective). Similarly, Nielsen defines five usability goals for designing interactive systems: learnability, memorability, efficiency, low error rate, and satisfaction. By integrating AI features, designers can still meet all these goals. But obviously, the goals do not account for the negative impact of AI on cognition.

Therefore, developing guidelines for intelligent systems should require careful consideration of analogous cognitive models. In fact, many of the usability guidelines in use today are influenced by such models [3]. For example, Fitt's law is based on our understanding of human perception, cognition, and motor-movement. Design principles for grouping interface elements, determining the amount of information to display, and ways to direct user attention are derived from models of human memory, and other perceptual gestalts.

In my dissertation, I investigate how cognitive models can inform the design of human-AI applications. Specifically, in *Florum*, an active reading tool, I look at how machine understanding of causal texts can assist readers in constructing causal diagrams from the text. Through a strategy called "automate after human effort," Florum monitors learning outcomes and automatically generates diagrams for already learned concepts. Both *SmartCues* and *Affinity Lens* use machine perception to scaffold the top-down approach to sensemaking—a strategy I call "automation for last-mile optimization." In *TakeToons*, I look at coupling performance-based animation with automated animation output, i.e., "automation as complementary to human effort" to produce animated stories. Based on these systems, I am looking to develop a design framework to help designers identify potential threats to cognition and validate design features against those threats.

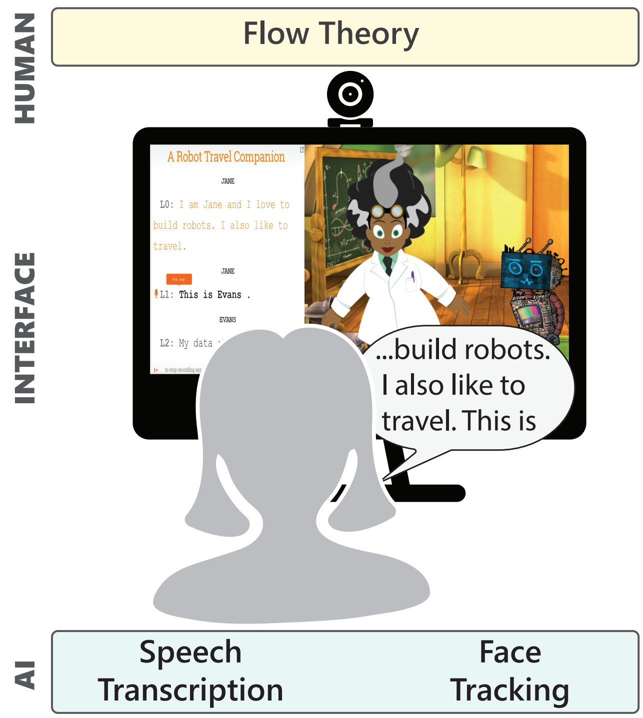

Figure 2. TakeToons: A tool for script-based performance animation.

AUTOMATION AFTER HUMAN-EFFORT

In the learning domain, broadly, there are two types of intelligent tools: productivity tools, and intelligent tutoring systems. Productivity tools aim to improve learner efficiency by reducing the time and cognitive cost of learning (e.g., dictation tool). On the other hand, intelligent tutoring systems focus on successful learning through student modeling, personalized content and learning interventions. The reality is, there is a need for hybrid learning tools that adhere to learning goals but also makes the process efficient for learners.

My ongoing work explores this idea in the context of science text comprehension. Scientific text is often causal in nature (e.g., *what causes coral bleaching?*). But many readers find it challenging to understand causal relationships in scientific text. Prior work in cognitive psychology has shown that knowledge externalization strategies such as student-generated causal diagrams could facilitate comprehension [4]. But studies have also shown that students perform poorly due to the cognitive load involved in making diagrams while reading [15]. In *Florum*, which is a pen and ink tool for reading and diagramming, I look at ways to make diagramming easy, but also facilitate causal understanding (i.e., maintain desirable difficulty). Florum combines NLP models of causal text understanding with active reading techniques for diagram construction.

As shown in Figure 1, as opposed to a dedicated diagramming interface which is inefficient, readers construct diagrams by using pen and ink annotations over the text. They start by highlighting important sentences while reading. Under the hood, Florum uses a text understanding model to extract key causal phrases from the text and displays them as hints to the reader. Then through rereading, they underline individual causal phrases which get added to the diagram view. Florum uses a combination of stroke recognition and the context of the text to determine how to represent the text as a diagram object (e.g., nodes, icons, etc.). Further readers can build edges between causal phrases by drawing a connecting line between them. Florum also allows readers to create multiple sub-diagrams, and then combine them using summarization strategies such as deletion and generalization. These features are directly modeled after the constructive approach to text comprehension in which readers start by identifying individual phrases, and then build sentence-level propositions, and finally combine them to form a coherent mental model [8].

In addition to direct assistance through NLP hints, Florum automatically generates diagrams from text using a strategy I call automation after human-effort. To determine when diagrams can be automatically generated, I consider evidence from past diagramming activities. If there is evidence that the reader has successfully represented certain causal relationships (e.g., cold air causes water to freeze), when reading subsequent explanations, if exists, those relationships are automatically rendered in the diagram view. This facilitates building new causal relationships from prior causal models (i.e., prior knowledge). My work extends the mixed-initiative model [7] by factoring diagram construction and learning causal relationships as separate goals with weighted utility functions. When the utility of learning is higher (i.e., no prior evidence that the student understands a causal relationship), the system favors assistive hints. But when the utility from diagramming is higher,

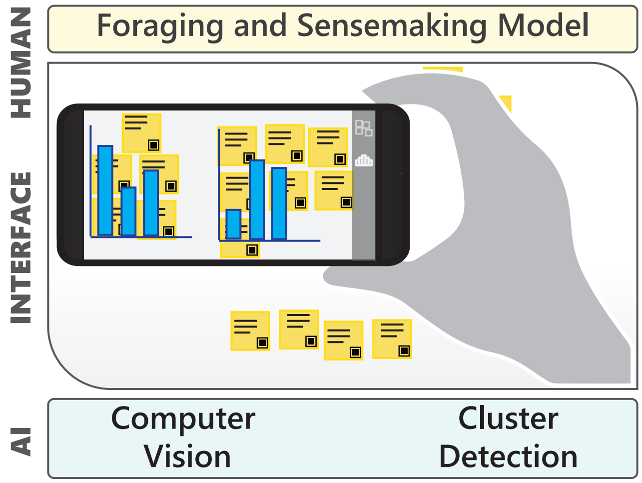

Figure 3. Affinity Lens: A tool for data assisted affinity diagramming using augmented reality.

Florum takes action by automatically rendering those causal relationships.

AUTOMATION COMPLEMENTARY TO HUMAN-EFFORT

Creativity is a key trait of human intelligence. According to Boden, it is our ability to produce novel and valuable ideas by applying a set of generative rules to any given conceptual space (e.g., painting, music, poetry, etc.) [2]. But due to the exponential number of variables that can be manipulated within these creative spaces, creativity tools are highly complex. AI can reduce some of this complexity by abstracting low-level tasks, but the trade-off is reduced control for the human. Therefore a key challenge is to determine ways to reduce interface complexity while still maintaining human creative agency.

In *TakeToons* [14], I explore this design problem in the context of performance animation. Performance animation tools such as Adobe's Character Animator [1] allow actors to directly map their performances on to puppets by using speech and head-pose tracking. But in addition to performance mapping, the interface also supports other types of 'trigger' based animations such as the use of props, special effects, scene change, camera angle change, etc. Collectively these operations enhance the 'value' of the final animation output, but they have to be manually triggered by the actor during the performance. This disrupts the 'flow-state' (intense concentration and enjoyment) of the actor in performance [6].

My approach in TakeToons is to lower the interaction cost by using automation as complementary to human performance. Specifically, TakeToons maps spoken dialog to a pre-defined script that is annotated with animation triggers (by the actor). As shown in Figure 2, as the actor performs the dialog, TakeToons automatically triggers non-performance animation in real-time. This is accomplished by transcribing the spoken dialog and aligning it to the script (and metadata). This approach reduces the interaction cost while allowing actors to maintain agency and perform at their own pace and style. In addition, TakeToons provides speech-based commands to support just-in-time edits and retakes, and automatically compiles the final animation for the actor. The design decisions implemented in TakeToons are based on flow theory and the need to minimize disruption to the creative process.

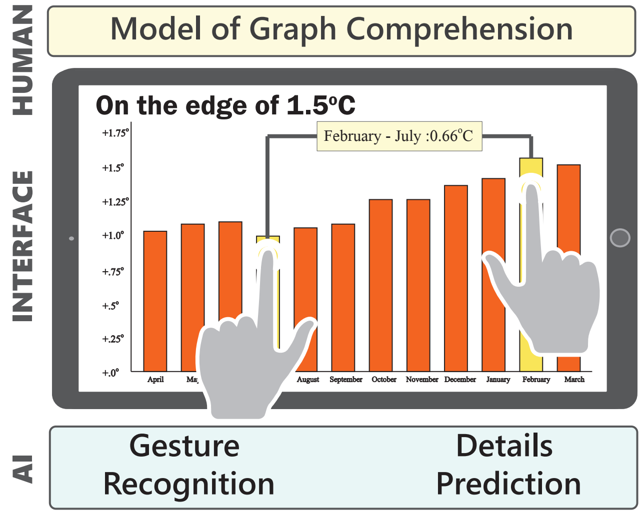

Figure 4. SmartCues: A gestural interfaces for details-on-demand through dynamically computed queries

AUTOMATION AS LAST-MILE OPTIMIZATION

In the case of sensemaking, humans engage both top-down and bottom-up processes to search raw data, generate hypotheses, gather insights and eventually build up to a mental model about the data [11]. But to deal with increasingly large and complex data sources, intelligent tools have been developed that offer assistance by automating the foraging and insight generation process. The problem is that analysts often find it difficult to understand the context or relevance of such insights [5]. In my work, I look at ways to optimize the top-down search process by combining human inputs with automation, as an alternative to replacing it. I have developed two systems: *Affinity Lens* [13] and *SmartCues* [12] in which AI facilitates querying over data and visualizations.

Affinity Lens is a tool that allows designers to incorporate mixed data sources in the process of affinity diagramming. In traditional affinity diagramming, designers use physical sticky notes that represent qualitative data points and engage in sensemaking by forming affinity clusters of those notes. When incorporating quantitative data into this process, current approaches favor automated clustering using quantitative data. This is problematic because designers may find it hard to interpret the resulting clusters. As shown in Figure 3, in Affinity Lens, I allow designers to use notes and clusters as query inputs to fetch quantitative insights. By using computer vision to identify individual notes, Affinity Lens generates and overlays data insights on top of physical sticky notes. In addition, based on current (human-generated) clusters, my system makes recommendations for clustering other notes. In Affinity Lens, designers have control over the top-down search process and queries are generated from affinity clusters (i.e., last mile optimization).

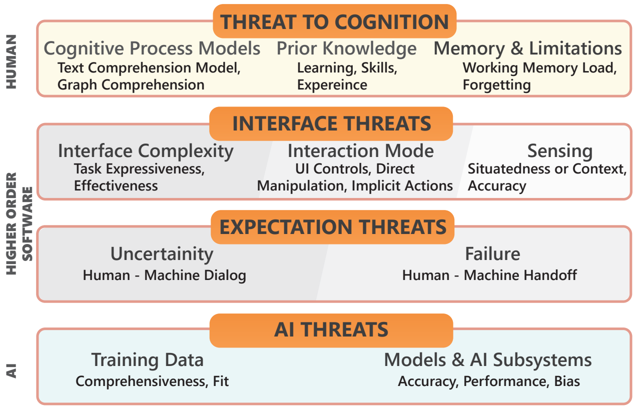

Figure 5. A preliminary framework for evaluating Human-AI system design.

In SmartCues, I look at facilitating the process of chart comprehension. As an alternative to automatically annotating insights over charts, SmartCues allows chart readers to request specific details through direct manipulation gestures. The design is motivated by the theory of graph comprehension [10] in which readers acquire insights through a series of extraction that involve one or more data points. What is difficult for readers is extracting values of encoded visual marks (bar, point, line, etc.) by reading corresponding axes values—last-mile problem. As shown in Figure 4, the reader expresses the query to retrieve the difference between two bars by performing simultaneous tap gesture over those bars. SmartCues converts this gesture into a search query and looks for all details that correspond to those bars. Here SmartCues uses a combination of prediction, and mixed-initiative dialog to render the final query result as an annotation. In this case, the result is the difference between the two bars.

CONCLUSION

In summary, my dissertation work looks at using cognitive models to design interactive intelligent systems for human critical tasks such as learning, creativity, and sensemaking. I have developed systems in all three domains that implement novel strategies for combining AI with human effort: (1) Automation after human-effort, (2) Automation complementary to human effort, and (3) Automation as last-mile optimization. Based on the design of these systems, I have a preliminary idea for a design framework that models the potential 'threats' from each of the components in the Human-AI pipeline (Figure 5). I would love to brainstorm with the UIST DC committee about further developing this framework, and get advice on ways to evaluate the framework. I also welcome feedback on the themes I have generated across these different systems to strengthen the contributions of my dissertation.

ACKNOWLEDGEMENTS

I am grateful to my thesis advisor Eytan Adar for his guidance and feedback. I would also like to thank my collaborators Colleen Seifert, Priti Shah, Steven Drucker, Mira Dontcheva, and Wilmot Li for their support and inputs.

REFERENCES

[1] Adobe. 2018. Character Animator CC. (2018).

[2] Margaret A Boden. 2009. Computer models of creativity. *AI Magazine* 30, 3 (2009), 23–23.

[3] Stuart K Card. 2018. *The psychology of human-computer interaction*. Crc Press.

[4] Michelene TH Chi, Nicholas Leeuw, Mei-Hung Chiu, and Christian LaVancher. 1994. Eliciting self-explanations improves understanding. *Cognitive science* 18, 3 (1994), 439–477.

[5] Alex Endert, M Shahriar Hossain, Naren Ramakrishnan, Chris North, Patrick Fiaux, and Christopher Andrews. 2014. The human is the loop: new directions for visual analytics. *Journal of intelligent information systems* 43, 3 (2014), 411–435.

[6] Jawaid A Ghani and others. 1995. Flow in human computer interactions: Test of a model. *Human factors in information systems: Emerging theoretical bases* (1995), 291–311.

[7] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 159–166.

[8] Walter Kintsch and CBEMAFRS Walter Kintsch. 1998. *Comprehension: A paradigm for cognition*. Cambridge university press.

[9] Jakob Nielsen. 1994. *Usability engineering*. Elsevier.

[10] Steven Pinker. 1990. A theory of graph comprehension. *Artificial intelligence and the future of testing* (1990), 73–126.

[11] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.

[12] Hariharan Subramonyam and Eytan Adar. 2018. SmartCues: A Multitouch Query Approach for Details-on-Demand through Dynamically Computed Overlays. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 597–607.

[13] Hariharan Subramonyam, Steven M Drucker, and Eytan Adar. 2019. Affinity Lens: Data-Assisted Affinity Diagramming with Augmented Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 398.

[14] Hariharan Subramonyam, Wilmot Li, Eytan Adar, and Mira Dontcheva. 2018. TakeToons: Script-driven Performance Animation. In *The 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, 663–674.

[15] Peggy Van Meter and Joanna Garner. 2005. The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review* 17, 4 (2005), 285–325.