Human–Computer Interaction for AI Systems Design: Reflections on an Online Course on Human–AI Interaction for Professionals

Per Ola Kristensson[1,2], Emily Patterson[2]

[1]Department of Engineering, University of Cambridge, Cambridge, United Kingdom
[2]Cambridge Online Education, Cambridge University Press & Assessment, Cambridge, United Kingdom
pok21@cam.ac.uk, emily.patterson@cambridge.org

Abstract

Human–Computer Interaction for AI Systems Design is an eight-week short online course aimed at professional students. It is part of an online course platform called Cambridge Advance Online, which is a joint effort between Cambridge University Press & Assessment and the University of Cambridge. This course launched in July 2023 amidst a massive increase in interest in AI and its applications, and quickly became one of the platform's highest-enrolling courses, attracting about 50 students per quarterly course run. To date, more than 200 students have completed the course, and more than 90 percent have rated their experience 'good' or 'excellent'. This paper reports on our experiences in designing and teaching this course.

Introduction

This paper explains the course design philosophy, syllabus and design, outcomes, and principles for a course on AI that tackles human–AI system design. *Human–Computer Interaction for AI Systems Design* is an eight-week commercial short online course aimed at professional students. This course launched in July 2023 and has been a success—to date, more than 200 students have completed the course, and more than 90 percent have rated their experience 'good' or 'excellent'.

The central idea behind this course is to teach students to build human–AI *systems*. There is a rich body of teaching materials on machine learning and artificial intelligence (Russell and Norvig 2016; Bishop 2006; Flach 2012). Similarly, there is a plethora of information on human–computer interaction, including several textbooks (Dix 2004; Preece, Rogers, and Sharp 2015; Shneiderman and Plaisant 2010). This course addresses the gap in between—how do we build systems bestowed with AI that allow users to achieve their goals in ways that are effective, efficient, and safe?

We are surrounded by human–AI systems in our everyday life. For instance, many of us rely on spam filters, autocorrect, spelling and grammar checking, speech recognition, gesture recognition, chatbots, smart climate controls, and so on. Beyond immediate visibility, human–AI systems are gradually being introduced in workplaces, offering opportunities ranging from automating mundane and routine tasks to playing an active part in redefining work, for instance, assisting people in creative tasks.

Such human–AI systems are challenging to build for several reasons. First, they are *complex*—they involve not only many modules, subsystems, and people, but also demand consideration of data management, bias, professional ethics, and so on. Second, they are *tightly coupled*, that is, the components and people that comprise the system itself are brittle as they tend be heavily affected by changes in the system. Third, the systems have *emergent* properties. In particular, *interactivity* is a critical emergent quality of such a system that can only be indirectly created through user interface design. Since interactivity only manifests itself through user interaction, and it is impossible to fully anticipate every user's needs, wants, and goals, as well as every possible use context, it is challenging to design human–AI interaction systems that yield interactivity that allow users to achieve their goals in an effective, efficient, and safe manner.

The stance taken in this course is to view this design problem through the lens of *systems thinking*—our aim is to teach students to view human–AI system design holistically and accept the challenges and risks that arise when designing complex systems (Elliott and Deasley 2007). However, while many system design failures can be attributed to a lack of systems thinking (Monat and Gannon 2018), systems thinking itself does not present actionable principles or methods for systematic human–AI design.

The approach taken in this course to ensure we are teaching *actionable* human–AI system design through the lens of systems thinking is to use *design engineering* as a foundation on which we can infuse human-centered AI theories, frameworks, and ideas (Kristensson et al. 2020). Design engineering, sometimes referred to as engineering design, is a branch of engineering that provides systematic design methods for building products and some services that covers the entire design process, ranging from solution-neutral problem statements[1], concepts, embodiments, and detailed designs, to manufacturing, support, and disposal (Pahl and Beitz 2013). Design engineering allows us to teach human-AI systems thinking in an actionable manner. .

[1]A *solution-neutral problem statement* is a problem statement that is at such a high level of abstraction that it does not mention or indicate potential solutions to the problem.

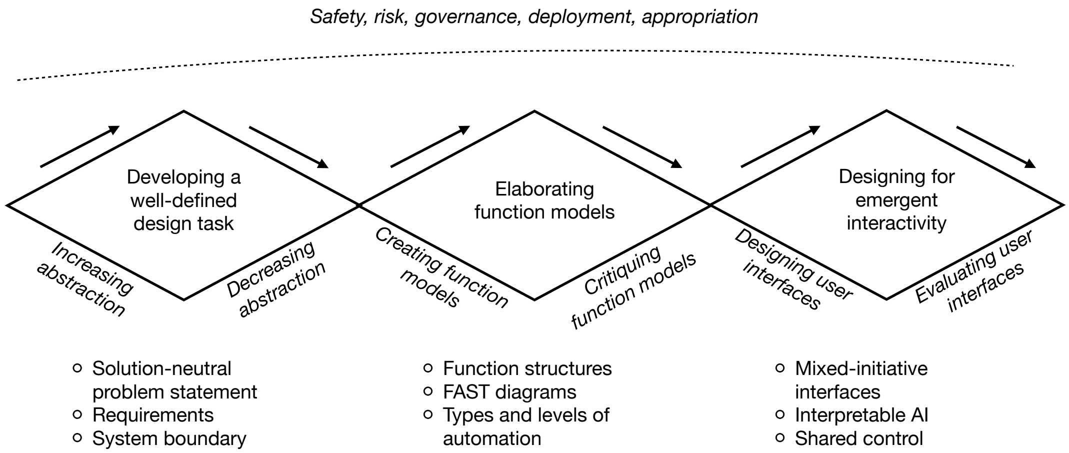

Figure 1: An overview of the online course Human–computer Interaction for AI Systems Design. The course teaches students human–AI system design using a triple diamond model. The first diamond focuses on arriving at a well-defined design task that can be meaningfully tackled as a design problem. The second diamond explores the creation of meaningful function models that can describe the system design problem at the functional level. The third diamond concentrates on translating functions into solutions by focusing on designs that yield meaningful interactivity between users and AI systems. Safety, risk, governance, deployment, and appropriation cut across all three diamonds in this model.

We use methods from design engineering, such as solution-neutral problem statements, system mapping, and function modeling, to teach students how to systematically elaborate on a system design that is sufficiently abstract to avoid design fixation, but also sufficiently detailed to enable infusion of human–AI frameworks and approaches, such as reasoning about the types and levels of automation in the system, interaction with automated services, system interpretability, shared control, and safety and risk.

Course Design Philosophy

The course is specifically designed for working professionals who may need to design human–AI systems but who may not be machine learning specialists. As such, it covers the entirety of the human–AI systems design cycle, from writing solution-neutral problem statements, carrying out user research and mapping systems to considering adoption, appropriation, and risk.

The course material itself is presented using a rich variety of text, custom imagery, video, audio, and interactive exercises. The exercises serve both to give students the opportunity to apply what they have learned and to allow them to self-assess how well they master the content.

Course Model

Figure 1 illustrates the high-level design of the course. We view the human–AI system design process as a sequential triple diamond process. Each diamond represents a phase of divergence and convergence. In a divergence phase the designer expands the space of solutions through design activities. In a convergence phase, the designer decreases the space of solutions through evaluation and critique.

Developing a Well-Defined Design Task

The first diamond is focused around arriving at a *well-defined design task*, that is, a design task that has (1) design decisions that can be made; (2) a design space that can be explored; (3) objectives describing what the design should achieve; and (4) constraints, such as requirements or limits. This is a view of design as a problem solving activity (Newell and Simon 1972).

To teach students how to do this, we teach them the technique of iteratively reformulating problem statements into increasingly solution-neutral terms by arriving at a *solution-neutral problem statement*, we teach them how to elicit *requirements*, and we teach them how to define a *system boundary* that encompasses all relevant aspects of the system, including components, subsystems, networks, people, regulation, and so on.

This diamond illustrates a process of initially increasing the level of abstraction by arriving at a problem statement in solution-neutral terms, followed by a subsequent decrease in abstraction by elaborating on requirements and setting a system boundary.

Elaborating Function Models

One of the outputs of the first diamond is an *overall function* of what the human–AI is supposed to carry out. The second diamond decomposes this overall function into one or several *function models*. Function models, such as function structures (Pahl and Beitz 2013) or Function Analysis Systems Technique (FAST) diagrams, allow a designer to gradually decompose an overall function into the key subfunctions. For example, the overall function `Auto-Correct Word` may be initially decomposed into the subfunctions `Type Key`, `Infer Word` and `Replace Word`.

Function models allow designers to elaborate on the key functions in a design and describe how they interrelate. Function structures are function models that connect functions by input and output signals while FAST diagrams are function models that describe relationships between functions in terms of abstraction. Regardless of which function model is used, once it has been created it is possible to apply frameworks to reason about AI systems design. The course introduces the types and levels of automation framework (Parasuraman, Sheridan, and Wickens 2000), which considers the type of automation, such as action or decision automation, and level of automation that is appropriate for a particular function.

The second diamond represents the iterative process of divergence that arise due to further elaborations of function models and automation considerations, and convergence emerging through the elimination of less useful function models and automation solutions that are unlikely to succeed.

Designing for Emergent Interactivity

The output of the second diamond is a *functional description* of a human–AI system that is, to some extent, *solution neutral*. That is, the function model describes *what* should be done but does not prescribe precisely *how* to do it.

The third diamond captures the central idea of design engineering: that once we know which functions we desire and understand how they interrelate, we can proceed to translate functions into *function carriers*, sometimes called solutions.

The third diamond illustrates this idea by considering designing for emergent interactivity as the central activity. When we select solutions to functions, such as a specific machine learning algorithm to carry out inference or a particular visualization technique to illustrate system status, we either implicitly or explicitly give rise to interactivity, which is an emergent property of a human–AI system.

We teach students to consider such interactivity from several perspectives. First, by teaching students the principle of direct manipulation (Shneiderman 1982) and the model of interaction as goal-directed dialogue (Norman 1988), it is possible to introduce principles of mixed-initiative interfaces (Horvitz 1999), which enable the design of dialogues between users and automated services by the application of the principle of direct manipulation. This gives rise to a range of human–AI design considerations, such as identifying opportune times to interrupt users, balancing tradeoffs between automation benefits and costs, and designing solutions that enable users to initiate, pause, terminate, and configure automation.

Further, building on the idea of interaction as dialogue, the course teaches students how to design for interpretability, including visualizing and otherwise making sense of data underpinning AI services, and explainable AI techniques for translating design objectives, such as mitigating human cognitive biases, to effective explainable AI techniques (e.g. (Wang et al. 2019)). Other issues that are discussed here include the nature of uncertainty and why it is not frequently not possible to rely on confidence scores outputted by AI systems (e.g. Vertanen and Kristensson (2008)), and why large language models (LLMs) tend to generate false information (so-called “hallucination”).

Finally, starting with the H-metaphor[2] (Flemisch et al. 2003), we introduce the notion of *shared control*, to allow students to experience and reflect on different models of sharing control between users and AI services. The third diamond captures this process by introducing a wide array of approaches and principles for designing human–AI user interfaces while simultaneously discussing how to evaluate such user interfaces, either by simulation, using heuristics, or by engaging in user studies.

Safety, Risk, Governance, Deployment, and Appropriation

Overarching the triple diamond are concerns that span the entire design process—safety, risk, governance, deployment, and appropriation. AI safety and risk are approached through three lenses: (1) the human factors view of human error, including the idea of normal accidents (Perrow 1999); (2) risk management, as applied in engineering when building products, systems, and services; and (3) governance of AI systems, including principles and guidelines for professional ethics, human control of technology, fairness, accountability, and so on (Fjeld et al. 2020).

Another overarching aspect in the course is deployment, as ultimately human–AI systems are only successful if they are adopted by users. The research literature indicates that the study of deployed systems is a rich source of design know-how that can assist in identifying barriers and new design opportunities (e.g. (Kristensson, Mjelde, and Vertanen 2023)). The course teaches a range of methods of carrying out deployment studies, ranging from unobtrusive methods, such as log analysis, to *in situ* methods, such as contextual inquiry.

Finally, for human–AI systems to be adopted users have to appropriate them, that is, users have to learn interaction strategies that allow them to carry out their goals when using the systems. In practice, such appropriation is frequently social, that is, influenced by colleagues and friends (Orlikowski 1995). The course teaches that appropriation is inevitable because designers cannot fully predict all future interaction contexts and, even if they could, users' needs, wants, and values may change over time. We use an established set of appropriation guidelines (Dix 2007) to create awareness among students about this issue while simultaneously enabling them to apply guidelines to reason about how users may appropriate human–AI systems.

Course Syllabus and Design

The course consists of eight modules, each of which are designed to take the average student around six hours to work through. Students are expected to complete one module per week.

[2] The H-metaphor is a model of shared control between a human and and an automation system modeled on horse riding. It introduces notions such as *tight rein control* when more control is taken by the human and *loose rein control* when more control is delegated to the automation system.

Module 1: Human–AI Interaction

This first module explains what is meant by *human–AI interaction* and why it is a distinct problem in comparison to human–computer interaction, and AI and machine learning. It first observes that human-AI systems are actually all around us: auto-correct on a capacitive touch mobile phone keyboard, email spam filters, streaming service recommendations, and search engines and chatbots are all examples of deployed human-AI system which we increasingly rely on in our daily lives.

The module explains why we need to design, what is meant by design, and how people carry out design. It then explains why human–AI system design is uniquely difficult. It introduces a technique for deriving a solution-neutral problem statement from a problem context and subsequently describes how to evolve a requirements specification for a desired system by considering a variety of sources, including user-elicited needs and wants.

Module 2: Function Modeling

This module explains how to use an overall function and decompose it into subfunctions using two easy-to-understand, yet versatile, approaches to functional modeling: function structures (Pahl and Beitz 2013) and FAST diagrams. The former allows modeling a human–AI system as a series of functions connected via flows of signals while the latter allows discovering all critical functions by the successive elaboration of abstract functions into more concrete functions.

Both models can be parameterized and these parameters can be classified into controllable and uncontrollable parameters. Controllable parameters can then be tuned or optimized while the effects of uncontrollable parameters can be studied through sensitivity analysis.

This module builds heavily on recent research in human–computer interaction on how to use design engineering methods to systematically design and analyze interactive AI-infused systems at an early stage in the design (Kristensson et al. 2020; Kristensson and Müllners 2021; Yang and Kristensson 2023; Kristensson 2024).

Module 3: Automation

This module introduces the automation problem and explains what is meant by automation and why automation is not a silver bullet. It explains what is known as the *ironies of automation* (Bainbridge 1983), where increasing automation levels can in fact amplify rather than reduce demands on human attention and skill.

It teaches the types and levels of automation framework (Parasuraman, Sheridan, and Wickens 2000), a well-established framework for automation analysis, and demonstrates how to apply it to functional descriptions of human–AI systems. It teaches how each function can be assigned a type (such as action automation) and level (such as full automation) of automation and how these initial types and levels of automation can then be evaluated using user-centric criteria, such as the risk of complacency, and system-centric criteria, such as the reliability of automation.

Module 4: Mixed Initiative Systems, Teaming and Partnerships

This module teaches approaches for allowing users to interact with AI. It begins by teaching the principle of direct manipulation (Shneiderman 1982) and ideas around viewing interaction as dialogue. It then introduces principles for mixed initiative interaction (Horvitz 1999), which couples direct manipulation with automated services. Systems supporting mixed initiative interaction allow both users and the automation service to take initiative for automation but also bestows the user with substantial control to pause, resume, terminate, and configure aspects of automation. In addition, the principles of mixed-initiative interaction raise issues about understanding the value of automation for users, the challenge of identifying opportune moments to interrupt users with automation, and the importance of mixed initiative systems to learn from user interaction.

It then discusses how human–AI interaction can be seen as interaction between users and AI working in teams and partnerships. This view gives rise to different concerns, such as ensuring users and AI have a shared understanding of the objectives. This leads to the introduction of the notion of alignment and the high-level computational ideas that can enable it.

Module 5: Understanding and Interpreting AI

A fundamental problem in human–AI systems is the difficulty for users to interpret the actions or results of AI. This module teaches the fundamentals of explaining AI to users.

It first introduces information visualization and explains how visualization techniques translates abstract information into visual variables. It provides an overview of visualization methods for AI, in particular uncertainty visualization. The central idea here is that data-driven AI is only as useful as the data underpinning it, hence it is frequently important to understand the nature of the underlying data.

In then covers a range of approaches to explainable AI, including non-visual approaches such as dialogues ("Explain why..."). To guide design, it introduces the theory-driven user-centric explainable AI framework (Wang et al. 2019), which links explainable AI techniques to objectives for guiding human decision-making and avoiding cognitive biases.

An important aspect of this module is to get students to realize the importance of data and uncertainty. First, data-driven AI is reliant on high-quality data and thus poor quality data introduces risk. Second, there is inherent uncertainty in both the underlying data and in any predictions or recommendations generated by an AI system. It is also frequently not possible to trust the confidence of an AI system due to the way such systems tend to calculate the probabilities for their predictions or recommendations.

Module 6: Managing Control and Agency

This module addresses three central interaction challenges induced by human–AI systems. First, to ensure users have sufficient control over a system. Second, to ensure users are able to understand and interpret system output and behavior. Third, the need to ensure users maintain their *sense of* *agency*: their sense of ownership of their actions.

It builds on research around metaphors, e.g. the H-metaphor (Flemisch et al. 2003), and recent research on how to objectively measure agency (e.g. Coyle et al. (2012)) to discuss issues of shared control and agency in human–AI system design.

An important aspect that is emphasized in this module is just how challenging it can be to control a dynamic system. To make this point, in the very beginning of this module students are introduced to the challenge of balancing a simple isolated ecosystem consisting solely of foxes and rabbits. Even though this dynamic system can be controlled to reach equilibrium by solving the underlying differential equations (Lotka 1910; Volterra 1926), research has shown that users find this system very challenging to control (Jensen and Brehmer 2003).

Module 7: Governance of Human–AI Systems

The highly successful data-driven machine learning techniques that underpin most of the human–AI systems we see in our daily lives, crucially, rely on data, and such data has to be collected and processed. The treatment of such data, including collecting, storing and processing it, has ethical implications. In addition, data is used to train machine learning algorithms; if the data is biased, it can lead the algorithms to make biased decisions. Finally, AI has to be safe. This module introduces these aspects of governance and teaches the main challenges of principled management of human–AI systems using an established set of principles (Fjeld et al. 2020).

It then introduces systematic risk management from an engineering perspective. It begins by teaching students to define a system boundary to ensure any risk assessment is bounded and thereby tractable. It then teaches a range of system mapping techniques to allow students to describe the human-AI system in terms of its processes, people, data flows, and so on. It then teaches risk assessment and risk visualization by introducing the structured what-if technique (SWIFT), failure mode and effects analysis (FMEA), fault trees, and risk matrices.

This module emphasizes that risk is pervasive and thus a level of risk has to be accepted and managed. First, a risk-free system cannot do any work. Second, complex tightly-coupled systems with emerging properties induce normal accidents (Perrow 1999), that is, in such systems it is inevitable that there are eventually undesirable outcomes. Thus, what is important is to arrive at an acceptable level of risk and consequently design processes to monitor and manage the acceptable level of risk throughout the lifespan of the human-AI system.

Module 8: Evaluating Human–AI Systems

It is vital to evaluate any system to ensure we have built the thing right—*verification*—and built the right thing—*validation*. This module introduces the verification cross-reference matrix (VCRM) as a tool to ensure requirements have been met, with a particular emphasis on the unique challenges in verifying human–AI system requirements. Such requirements frequently need to be situated within carefully considered use contexts and user populations for verification to be meaningful.

This module also discusses validation strategies that can be used to ensure a system is fit for purpose either immediately before deployment or at the deployment stage. To make the most out of deployment it also teaches how to carry out deployment studies, such as log studies or surveys.

Finally, for human-AI systems to be successful they have to be adopted by users. However, users only adopt systems if such systems enable them to achieve their goals. When faced with new systems, users appropriate such systems, that is, they learn how to use these systems to achieve their goals. It is therefore important to design for successful appropriation. This module explains why appropriation is inevitable—designs cannot anticipate all needs, wants, and use contexts—and introduces an established set of appropriation design guidelines (Dix 2007) to assist students in thinking about appropriation for their human-AI system design.

Assessment

Students must achieve a mark of 70 percent to pass the course. This is obtained through a mixture of participating in various interactive activities and assignments throughout the course, submitting six end-of-module project assignments, and submitting a final project report.

Each module project assignment asks students to implement learning from that module in order to complete a section of a design of a human–AI system. Students share their work in each module with one another and have the opportunity to offer and receive feedback, and learn from one others' work.

For the final project assignment, students assemble their module tasks into a full design report of a human–AI system, using the feedback and learning they have gained throughout the course to update their previous efforts. Required elements include:

- solution-neutral problem statement and requirements
- function model and morphological chart
- automation strategy
- interaction strategy
- brief discussion on any interpretability issues arising in the system
- issues around sharing of control and user agency
- system boundary and an analysis of the key risks in the system
- verification and validation

These submissions are marked by the course tutors using a rubric and all students receive detailed feedback. Students receive a digital badge and certificate upon successful completion of the course.

Course Design Principles and Practice

The course was developed in tandem with a university academic course lead and a learning designer. The course is primarily asynchronous and online, allowing students to work through the material at their own pace. However, the course is also cohort-based, so all students in a cohort begin the course simultaneously and each module unlocks for all students at the same time. This allows students to make use of interactive features, such as discussion boards and online whiteboards, while ensuring that students will be working through the material in roughly the same time frame, forming a community of practice (Wenger 1999).

The learning design framework applied to scaffold the course material uses a research-backed approach to structuring interactive online learning, ensuring that constructive alignment is maintained between learning material, learning outcomes, and assessment. The framework also facilitates a balance between different types (acquisition, practice, inquiry, production, etc.) of online learning activities.

The course learning design emphasizes an active learning methodology to ensure that students leave the course with actionable skills that they can take forward into their workplace and beyond. In practice, this is done through smaller interactive and reflective exercises embedded in the course material which allow students to check for understanding, engage in independent research, and connect new learning to prior experience. The course also assigns larger, module-level project assignments to allow students to integrate and apply new knowledge to a topic or to relate learning to professional practice. These module-level project assignments ultimately lead to a course-level project which brings everything together. Tutors mark all submissions and moderate interactive exercises, such as discussion forums, and are available to students to help clarify material or assist with queries.

In addition to the asynchronous material, each week the academic course lead and tutor hold a one-hour live session to explore that module's topic in greater depth and to answer questions. Although attendance at this live session is not mandatory, the session is recorded and the recording is made available for all students to view. These live sessions last an hour each and typically 50–60% of the cohort attends each live session.

Student Outcomes and Feedback

By the end of the course students are able to:

- derive a solution-neutral problem statement that motivates a human–AI system and arrive at a requirements specification that can be used to test the system
- design a function model of a human–AI system and analyze the types and levels of automation that can be used to address the solution-neutral problem statement
- perform a risk analysis and determine the types of risks that are inherent in the human–AI system and propose mitigation activities
- create a verification cross-reference matrix that can be used to verify that system requirements have been met for all deployment contexts relevant to the human–AI system
- develop a strategy for managing the risks and governance issues of a human–AI system
- create a validation strategy to ensure the human–AI system is fit for purpose and addresses the overall function it is intended to perform.

Selected Student Feedback

To date more than 200 students have completed the course (about 80 percent of enrollees), and more than 90 percent have rated their experience 'good' or 'excellent' on a five-point scale ('terrible', 'poor', 'average', 'good', and 'excellent'). Here we share indicative anonymous qualitative student feedback.

In response to the question, "What new knowledge or skills have you acquired by taking this course?"':

- "The base theory of human–AI that as a designer I should know. It made me stronger and now I am aware how to balance human [and] AI and what process or tool to use to evaluate the ideas."—September 2023
- "Before beginning the course I had very little to no knowledge on what is AI is and the difference between automation, AI and ML. This course helped me to comprehend each term in its own and I could also learn the design process to design an AI system."—January 2024
- "Methodical approach to designing a human AI System—including how to describe the problem, define requirements, model the functions, primary and secondary evaluation criteria, automation and interpretability strategies, address control and agency concerns; verification and validation methods, and post deployment study strategies."—January 2024
- "I learnt how to take a conceptual AI system apart and build it in a way that augments human lives. This course provided me several insights interesting insights into different factors that affect the development and implementation of AI. For e.g. I did not know how to build function structures. It has been such a fascinating process learning about these new tools. I had no knowledge of automation strategies either. I think the first 4 modules introduced me to several new concepts!"—January 2024
- "Most of the course [materials] were entirely new, but thanks to the practical exercises, I learned to do them independently. I also learned to look at the system design process holistically and see the importance, interdependencies, and consequences of the different steps."—January 2024

In response to the question, "What did you find most enjoyable about this course?":

- "I found the interaction sections and quizzes particularly enjoyable. They provided hands-on experience and reinforced the course material in an engaging and practical way."—April 2024
- "I thoroughly enjoyed the interactive live sessions, and the quizzes and assignments helped me stay focused."—April 2024
- "Engaging with our tutor and other participants in the weekly live lectures was a motivating way to learn. I also

particularly enjoyed modules relating to interaction strategy for the Human–AI team, as well as considering Control & Agency."—Jan 2024

- "Probably function modelling and solution neutral problem statements. I enjoyed the quizzes as they really helped make the concepts solidify in my mind. I felt like I was learning a lot."—September 2023
- "I enjoyed working on the tasks and the final project, which were well-designed and helped reinforce my course knowledge. The discussion assignments and breakout sessions during live sessions challenged my thinking on the given topics, and I appreciated hearing different perspectives from peers who come from diverse industries."—July 2023

In response to the question, "What did you find most challenging about this course?":

- "It was difficult to grasp some themes while working at pace, but I have several areas that I will revisit to become more comfortable."—September 2023
- "The topic itself is a challenging one. Now I have a better knowledge and looking forward to implement."—July 2023
- "The concept of systems and the technicalities associated with it was difficult for me to grasp as a designer coming from UX /visual design background."—January 2024
- "The most challenging aspect was balancing automation with human control to ensure usability and user satisfaction. Developing comprehensive risk mitigation strategies and ensuring all relevant factors were considered in the system boundary required careful thought and analysis."—April 2024
- "The most challenging aspect of this course was deriving flow charts and diagrams. Integrating these visual elements required a deep understanding of complex concepts and meticulous attention to detail."—April 2024
- "Addressing ethical considerations and ensuring privacy and security in AI-driven systems added layers of complexity."—April 2024

Negative feedback generally centers on the theme of time management, which is often a concern for professional learners trying to balance continuing education against the demands of work and non-work life, or quirks of the online learning management system. In a widely available professional development course, it is difficult to anticipate the prior knowledge and application needs of every potential learner, and it is therefore to be expected that some learners will need more time than advertised to work through and apply the material to their particular industry or organization. It would also be difficult to significantly reduce the learning material or assignments and still achieve the same level of learning outcomes. However, for future iterations of the course we could consider content management options such as learning pathways or optional activities which allow learners more control over their individual learning journeys.

Conclusion and Outlook

This paper has introduced the course Human–Computer Interaction for AI Systems Design, an eight-week short online course aimed at professional students. The course launched in July 2023 and quickly became one of the platform's highest-enrolling courses, attracting about 50 students per quarterly course run. To date more than 200 students have completed the course, and more than 90 percent have rated their experience 'good' or 'excellent' on a five-point scale ranging from 'terrible' to 'excellent'.

This paper has reported on our experiences in designing and teaching this course. From student feedback, we conjecture that the most critical aspect in explaining the success of the course is the focus on *actionable* system design techniques, which allow us to inject human–AI design insights into the course in a way that readily allows the students to apply their knowledge to their own human–AI designs.

This course demonstrates how recent research in human–computer interaction on early stage functional modeling adapted from design engineering (Kristensson et al. 2020; Kristensson and Müllners 2021; Yang and Kristensson 2023; Kristensson 2024) can be used to provide students with systematic design methods for human–AI system design. The outcomes of these methods, such as solution-neutral problem statements, system boundaries, function models, and system maps, can subsequently be used as a canvas for injecting AI-specific interactivity, such as analyzing opportunities for automation, reasoning around shared control, and identifying opportunities for visualizing uncertainty or using explainable AI techniques.

In terms of outlook, we see two developments on the horizon in the landscape of human–AI system design courses. First, human–computer interaction is becoming increasingly sophisticated in modeling interactivity using various computational models, such as closed-form mathematical models and data-driven machine learning (Oulasvirta et al. 2018; Williamson et al. 2022). As user interfaces become increasingly sophisticated and reliant on inferring user intent from noisy observations, there will be increasing demand in upskilling in this area.

Second, there is a lot of excitement about AI but a growing realization that AI needs to be infused into systems in ways that add value to users. This necessitates design approaches that take more holistic system perspectives on design to ensure such systems are adopted by users and provide effective, efficient, and safe interaction. There are a plethora of design methods that tackle some of these aspects in the human–computer interaction, design, and design engineering literature, but these developments need to be brought together into course offerings that equip students with versatile toolboxes, principles, and techniques for human–AI system design. Only then is it possible to realize the full potential of state-of-the-art developments in the AI field to improve the human condition.

Acknowledgements

The authors would like to thank Ariana Marnel Laureta and Ellen Collier for their assistance.

References

Bainbridge, L. 1983. Ironies of automation. *Automatica*, 19(6): 775–779.

Bishop, C. 2006. *Pattern Recognition and Machine Learning*. Springer.

Coyle, D.; Moore, J.; Kristensson, P. O.; Fletcher, P.; and Blackwell, A. 2012. I did that! Measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2025–2034.

Dix, A. 2004. *Human-Computer Interaction*. Pearson.

Dix, A. 2007. Designing for appropriation. In *Proceedings of the 21st British HCI Group Annual Conference*, 27–30.

Elliott, C.; and Deasley, P. 2007. *Creating Systems that Work: Principles of Engineering Systems for the 21st Century*. Royal Academy of Engineering.

Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; and Srikumar, M. 2020. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*. Berkman Klein Center.

Flach, P. 2012. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge University Press.

Flemisch, F. O.; Adams, C. A.; Conway, S. R.; Goodrich, K. H.; Palmer, M. T.; and Schutte, P. C. 2003. The H-Metaphor as a Guideline for Vehicle Automation and Interaction. Technical report, Langley Research Center.

Horvitz, E. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 159–166.

Jensen, E.; and Brehmer, B. 2003. Understanding and control of a simple dynamic system. *System Dynamics Review*, 19(2): 119–137.

Kristensson, P. O. 2024. Designing virtual and augmented reality user interfaces using parameterized function structure models. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, 419–423. IEEE.

Kristensson, P. O.; Lilley, J.; Black, R.; and Waller, A. 2020. A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paper 398.

Kristensson, P. O.; Mjelde, M.; and Vertanen, K. 2023. Understanding adoption barriers to dwell-free eye-typing: Design implications from a qualitative deployment study and computational simulations. In *Proceedings of the International Conference on Intelligent User Interfaces*, 607–620.

Kristensson, P. O.; and Müllners, T. 2021. Design and analysis of intelligent text entry systems with function structure models and envelope analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Article No. 584.

Lotka, A. J. 1910. Contribution to the theory of periodic reactions. *Journal of Physical Chemistry*, 14(3): 271–274.

Monat, J. P.; and Gannon, T. F. 2018. Applying systems thinking to engineering and design. *Systems*, 6(3): 34.

Newell, A.; and Simon, H. A. 1972. *Human Problem Solving*. Prentice-hall.

Norman, D. A. 1988. *The Psychology of Everyday Things*. Basic books.

Orlikowski, W. J. 1995. Learning from notes: Organizational issues in groupware implementation. In *Readings in Human–Computer Interaction*, 197–204. Elsevier.

Oulasvirta, A.; Kristensson, P. O.; Bi, X.; and Howes, A. 2018. *Computational Interaction*. Oxford University Press.

Pahl, G.; and Beitz, W. 2013. *Engineering Design: A Systematic Approach*. Springer.

Parasuraman, R.; Sheridan, T. B.; and Wickens, C. D. 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 30(3): 286–297.

Perrow, C. 1999. *Normal Accidents: Living with High Risk Technologies*. Princeton University Press.

Preece, J.; Rogers, Y.; and Sharp, H. 2015. *Interaction Design: Beyond Human Computer Interaction*. Wiley.

Russell, S. J.; and Norvig, P. 2016. *Artificial Intelligence: A Modern Approach*. Pearson.

Shneiderman, B. 1982. The future of interactive systems and the emergence of direct manipulation. *Behaviour & Information Technology*, 1(3): 237–256.

Shneiderman, B.; and Plaisant, C. 2010. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson.

Vertanen, K.; and Kristensson, P. O. 2008. On the benefits of confidence visualization in speech recognition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1497–1500.

Volterra, V. 1926. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Memoria della Reale Accademia Nazionale dei Lincei*, 2: 31–113.

Wang, D.; Yang, Q.; Abdul, A.; and Lim, B. Y. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paper 601.

Wenger, E. 1999. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.

Williamson, J. H.; Oulasvirta, A.; Kristensson, P. O.; and Banovic, N. 2022. *Bayesian Methods for Interaction and Design*. Cambridge University Press.

Yang, B.; and Kristensson, P. O. 2023. Imperfect surrogate users: Understanding performance implications of augmentative and alternative communication systems through bounded rationality, human error, and interruption modeling. *Proceedings of the ACM on Human-Computer Interaction*, 7(MHCI): Article No. 213.