Approachable modeling and smart methods: a new methods field of study

Corey Schimpf & Brian Castellani

Published online: 17 Aug 2022.

Submit your article to this journal ↗

Article views: 1511

View related articles ↗

View Crossmark data ↗

Citing articles: 3 View citing articles ↗

Routledge
Taylor & Francis Group

Approachable modeling and smart methods: a new methods field of study

Corey Schimpf [ID][a] and Brian Castellani [ID][b]

[a]Department of Engineering Education, University at Buffalo, SUNY, Buffalo, NY, USA; [b]Department of Sociology, Durham University, Durham, UK

ABSTRACT

Advances in the integration of smart technology with interdisciplinary methods has created a new genre, approachable modeling and smart methods – AM-Smart for short. AM-Smart platforms address a major challenge for applied and public sector analysts, educators and those trained in traditional methods: accessing the latest advances in interdisciplinary (particularly computational) methods. AM-Smart platforms do so through nine design features. They are (1) bespoke tools that (2) involve a single or small network of interrelated (mostly computational) methods. They also (3) embed distributed expertise, (4) scaffold methods use, (5) provide rapid and formative feedback, (6) leverage visual reasoning, (7) enable productive failure, and (8) promote user-driven inquiry; all while (9) counting as rigorous and reliable tools. Examples include R-shiny programmes, computational modeling and statistical apps, public-sector data management platforms, data visualisation tools, and smart phone apps. Critical reflection on AM-Smart platforms, however, reveals considerable unevenness in these design features, which hamper their effectiveness. A rigorous research agenda is vital. After situating the AM-Smart genre in its historical context and introducing a short list of platforms, we review the above nine features, including a use-case on how AM-Smart platforms ideally work. We end with a research agenda for advancing the AM-Smart genre.

Introduction

Advances in the integration of smart technology with interdisciplinary methods (particularly computational) has created a new methods genre, *approachable modeling and smart methods*. Or *AM-Smart* for short. A critical review of existing AM-Smart platforms, however, reveals unevenness in design, making them less accessible or useful for social inquiry or for using the latest advances in interdisciplinary methods. A rigorous research agenda is vital. In addition to naming this new field, our goal is to establish AM-Smart platforms as an avenue of methodological innovation and research. We begin with a basic summary of what makes AM-Smart platforms unique, including several 'best practice' platforms. We provide a bit of historical context, grounding this genre in the smart technology movement of which it is a part, as well as the methodological challenges it was created to address. Next, we outline the nine design features crucial for AM-Smart platforms to be more accessible and useful. We then present a use-case of how AM-Smart platforms ideally work. We end with a critical reflection on the potential impact of AM-Smart platforms for social inquiry

and the new research directions the field raises – all with the hope that others will likewise critically engage the field, including the development of novel AM-Smart platforms.

The new world of approachable smart methods

In the vocabulary of actor network theory and the new materialism (e.g. Hayles 1999, Hayles, 2017), what makes *AM-Smart* platforms unique is that they employ the latest advances in nonconscious machine cognition to create a methods environment in which the platform acts as an expert guide for social inquiry. AM-Smart platforms do this by design: by allowing users to *cognitively offload* the challenges of running otherwise complex methods (as in the case of computational modeling) they increase non-expert access to highly novel forms of methods-driven inquiry. More specifically, these platforms draw on the concept of distributed cognition or the idea that cognition does not simply happen within the mind but is shared across a system of entities including other individuals, artifacts, and digital tools (Hutchins, 2006; Liu et al., 2008).

Given this approach, expertise is built directly into the platform. Traditional statistical software, such as SPSS (Statistical Packages for Social Science), Statistica, SAS (Statistical Analysis System) and MATLAB get close to doing something similar, but for most forms of methodological inquiry, particularly computational modeling techniques, still require a high level of user expertise. Most statistical software are also general packages: their strength is in the access they provide to a wide variety of methods.

In contrast, from our survey of the field (which we explain next) AM-Smart platforms demonstrate some combination of nine core (italicised) features. AM-Smart platforms are *bespoke tools* designed to focus on *a single technique or small network of closely interrelated methods* (mostly computational in focus), which help users to simultaneously use and learn new methods. They facilitate methods learning by *scaffolding* both routine and difficult tasks, providing *rapid and formative feedback*, enabling *productive failure*, and leveraging *visual reasoning skills*, while still being *rigorous, authentic, and reliable*. The AM-Smart emphasis on methodological approachability by *embedding distributed expertise* is key because, like most smart technology, it allows for a quicker, more seamless, and responsive approach to information gathering, analytical execution and, in this case, social inquiry. AM-Smart platforms specifically target learning through doing, as they facilitate *user-driven learning* of a topic, primarily through intuitive, tailored supports within a no-fault learning environment designed to solve specific user-identified tasks. AM-Smart platforms are mainly designed for applied and public sector analysis, those trained in traditional methods, and educators to access wider interdisciplinary methods.

AM-Smart methods can be stand-alone platforms, downloadable computer programmes or apps on the web or mobile. They can also be built from existing programming environments (e.g. R or Python) while maintaining a standalone interface. Moreover, AM-Smart platforms tend to embrace the open-source, open-access movement which seeks to create transparent and accessible software, where the codebases behind the platforms are shared so that others may build upon it and software is distributed without costs to users (Carillo & Okoli, 2008). This ensures maximal distribution for new methodological innovations.

Cataloguing AM-Smart methods

Given the fast-changing, endemic nature of smart app life today, it is presently difficult to bracket, count, or create a definitive catalogue of the AM-Smart methods currently in play. Examples range from computational modeling suites and statistical apps to digital research environments and smart phone apps to public-sector data management platforms and visualisation tools, such as those that flourished during the COVID pandemic.[1]

To gain a basic impression of the field, we did the following. First, we reviewed the gallery of apps on *R Shiny*.[2] 'Shiny is an R package that makes it easy to build interactive web apps straight from

R. You can host standalone apps on a webpage or embed them in R Markdown documents or build dashboards. You can also extend your Shiny apps with CSS themes, htmlwidgets, and JavaScript actions.'[3] Given its open-source flexibility, a significant number of AM-Smart apps are made using R. Second, we did a *Google* search, using such terms as 'computational modeling and app' and 'shiny and machine learning,' which yielded most platforms we found. Third, we searched for AM-Smart platforms on the *Apple App Store*, which were primarily statistical or data management in nature. Finally, we put out a call on *Twitter* asking colleagues for examples, to which we received a handful of replies.

Two caveats are important to note from our basic review. First, the majority of AM-Smart platforms are in the natural, engineering and computational sciences and applied mathematics. Second, we could not find a rigorous AM-Smart platform for qualitative inquiry. The closest we found were some of the R *COMPASSS*[4] packages for running qualitative comparative analyses. But these were rather conventional.

Based on our initial survey of the field, we identified a handful of 'best example' platforms[5] for social inquiry and, along with them, the nine key design features we listed earlier. Platforms include *COMPLEX-IT*[6] for computational modeling and data visualization; *Radiant*[7] for statistics and machine learning; *JASP*[8] for Bayesian statistical modeling; *PRSM*[9] for participatory systems mapping; *SAGEMODELER*[10] for learning systems dynamics through designing models; *MAIA*[11] and *NetLog*o[12] for designing and exploring agent-based models; *Cytoscape*[13] for modeling complex networks; and *ExPanD*[14] for visually exploring your data. All these platforms are online and include tutorials, datasets, and published examples to explore (See footnotes for links).

Setting the historical background

AM-Smart methods are part of the wider shift in the *knowledge economy* toward smart technology. The disciplines and fields involved in the formation of smart technology are significant, including computer science, software engineering, information science, the learning sciences, artificial intelligence in education, technology start-ups and global technology companies. More recent fields include computational science, data science, and machine learning. Smart technology builds on and adds to advances in smart environments, ubiquitous computing, smart devices, and the internet of things.

While situated at the nexus of fields involved in smart technology, AM-Smart platforms draw, in varying degree, on two interdisciplinary areas: the learning sciences and human-computer interaction. *The learning sciences* examine learning across all ages. A key focus is supporting the development of the complex and adaptive skills needed for the smart globalised worlds in which we now live (Sawyer, 2014). Another focus is how computational technologies may be leveraged to support learning (e.g. see, Bransford et al., 2000; Hardy et al., 2020; Reiser, 2004). *Human-Computer Interaction* (HCI) focuses on understanding, designing, and evaluating the interface between people and computational technologies (Carroll, 2003). While HCI is used extensively in the development of many types of software, including those dedicated to research methods, its integration with the learning sciences to support the development of methods software is a new trend. Concepts, principles, and theories from both fields undergird the AM-Smart genre. Integrating these two fields, AM-Smart methods espouse the philosophy of learning through engaging with methodological innovation.

Why AM-Smart methods are important

AM-Smart platforms respond to three key methodological challenges. The first is the massive growth in computational methods, from agent-based modeling and artificial intelligence to network analysis and computational mathematics (Castellani & Rajaram 2022). Burrows and Savage notably covered this challenge in a series of articles (see 2014), as have other scholars (Kovalchuk et al.,

2021; Marres, 2017). This challenge is fuelled by advances in computer science and applied mathematics and a deep research interest for new approaches, fundamental and applied, for examining our complex, multifaceted, and dynamic world (Barbrook-Johnson & Carrick, 2021).

Considering the rate of computational innovations, this voluminous material on methods cannot be sufficiently covered in most social science training. Even for the most impactful or widely adopted methodological innovations, it takes time to see substantial uptake into the classroom, public and private organizations or applied settings. As a result, most social scientists and those in the public and applied research are not well trained in these methods, leaving them outside the methodological mainstream of leading-edge methods and limiting their capacity to influence these techniques in application to social science and associated topics (Castellani and Rajaram 2022).

This methodological lag is particularly concerning given the 'big data' world in which we now live and the datafication of everything – the second major challenge. Big data, digital data, it is all the same issue: social science today is confronted with a massive level of data that challenge conventional methods, be they statistical, qualitative, or historical. Social science is also confronted with the conflation of our material and digital worlds, via social media and online life, creating a posthuman (computationally enhanced) existence for most people (Hayles, 1999), which, to make sense of, requires new forms of digital methods and social science (Marres, 2017).

The third challenge is complexity. The world has simply become more complex (both in terms of data and topics) such that studying any issue often requires the usage of interdisciplinary or transdisciplinary teams, as well as the involvement of not only academics and methodologists but also public and private sector stakeholders (Bernstein, 2015). Such interdisciplinary work means that even the best of methods experts will be, at certain points, outside their areas of expertise, requiring them to work with new methods. AM-Smart platforms respond to this challenge, and the other two, through their design approach, which aims at making advanced interdisciplinary methods more accessible, primarily by drawing upon the learning sciences, human-computer interaction advances, and the general philosophy of learning through methodological innovation.

The nine design features of AM-Smart platforms

Despite the early nature of the field and unevenness in design, our initial survey of AM-Smart platforms identified nine design features central to the AM-Smart genre. Going forward, any future research agenda needs to address these features, as well as others we may have not identified.

Features 1 and 2: Bespoke methods

AM-Smart platforms are not like statistical packages such as SPSS or wide-breadth platforms such as MATLAB. AM-Smart platforms are bespoke tools that increase access and approachability by focusing on a single method or small network of closely interrelated methods. The 'bespoke methods' approach has value in the world of social science inquiry as it reduces the cognitive load required for running highly complex methods environments, which can negatively impact non-expert ability to perform or learn from a task (Hollender et al., 2010; Paas & Sweller, 2014). As bespoke methods, AM-Smart platforms are not a package or library in a general programme. They have a distinct interface and can be accessed as unique applications. All of the platforms we reviewed meet these two criteria.

Feature 3: Building distributed expertise systems

Most computational methods require a high level of user expertise. AM-Smart platforms address this issue by building expertise into the software, allowing the platform to become part of the user's distributed cognition system, primarily by acting as a skilled guide for social inquiry (Hollan et al., 2000; Hutchins, 2006). Two examples are *COMPLEX-IT*[15] and *JASP*,[16] which provide non-expert

access to such interdisciplinary techniques as machine learning and Bayesian statistics. Both embed expertise by pre-setting method parameters, decomplicating tasks, transporting data and output from one technique to the next, and providing tools to compare results from different methods.

Feature 4: Scaffolding practice

AM-Smart platforms are designed to increase effective usage of new methods. To do so, AM-Smart methods employ guides or supports, referred to as *scaffolding* in the learning sciences (Quintana et al., 2004; Reiser, 2004). Scaffolding involves a more knowledgeable entity (e.g. teacher, peers, or a tool) supporting a novice or new user to engage in practices or processes they may not otherwise be able to perform (Kim & Hannafin, 2011). Scaffolding can be deployed in a variety of ways in AM-Smart platforms. Below we discuss two prominent approaches.

The first type of scaffolding – which overlaps with Feature 3, Embedding Expertise – supports methods usage and exploration by minimizing or removing low-level, tedious, routine, or overly complicated tasks from the methodological process inside the platform (Kim & Hannafin, 2011; Quintana et al., 2004). This could involve merging or reducing the number of methods (as in the case of creating bespoke methods); reducing the amount of input or controls a user needs to manipulate; providing alternative ways to setup a method; or reducing data cleaning or transformation steps before or during the method. Through scaffolding, users are compelled to focus on key steps without being bogged down by routine practices that are necessary but do not assist in developing a conceptual understanding of a method. Two examples are *Radiant*[17] and *PRSM*.[18]

The second type is *procedural scaffolding*, which guides users through the operational aspects of a platform (Kim & Hannafin, 2011). This could involve prompts after executing a method step; embedded explanation of a method input control or other considerations; or step-by-step 'worked' examples of using the methods. Procedural scaffolds may be static or dynamic; embedded or nonembedded (Devolder et al., 2012). Static procedural scaffolds are fixed guidelines, whereas dynamic scaffolds change depending on the user or software state. Embedded scaffolds are within the platform. Non embedded, such as software documentation or a supporting website, are outside the platform. Non embedded and static procedural scaffolds require the user to seek them out, whereas embedded and dynamic are more apparent and directive in the system. Often, combinations of both are useful. Whatever the combination, the key here is that users are given just-in-time support when they become unsure about the execution of a method. Examples include *JASP*,[19] *MAIA*[20] and *Cytoscape*.[21]

Feature 5: Rapid and formative feedback

AM-Smart platforms employ learning science strategies to provide rapid feedback that facilitates user understanding – what scholars call formative, as opposed to summative, feedback (Dixson & Worrell, 2016). Summative feedback relays finalized results or outcomes for some completed process, for example, grades on an exam. Formative feedback relays more intermediate information to a user about their current state or performance on a task, which can be delivered through multiple modes or timing (Shute, 2008). Formative feedback is particularly potent for supporting methods learning, as it indicates what is happening at a specific point for a task, which a user can use to adjust or rethink their actions (Bente & Breuer, 2009; Shute, 2008).

AM-Smart platforms typically use formative feedback. This is accomplished, first, by distilling the results of a method or method step into the most central or key details that a user needs to understand the outcome, assuming minimal prior experience. Using scaffolding, for example, full technical details of a method or method step will be condensed into a succinct subset. Deciding on this subset requires expert judgment for that method and what constitutes the minimum set of key details. Second, feedback in AM-Smart is delivered quickly to users once they have completed

a task, so they can swiftly ascertain the results of their actions and analysis – which makes AM-Smart rapid feedback similar to game-based environments (Bente & Breuer, 2009). Examples from our list include *SAGEMODELER*,[22] *MAIA*,[23] and *NetLogo*.[24]

Feature 6: Leveraging visual reasoning

People often excel at processing and analysing visual information over other information formats (Arnheim, 1997; Damasio, 2012; Kirk, 2016; Ware, 2003). In our present data saturated world, visualisation has become a core area of methods study, contributing to several fields including data visualization, software design, visual complexity, and data science. Computational methods are intentionally visual in output – from fractals and complex network diagrams to systems maps and agent-based model simulations. Visualizations tap strongly into distributed cognition. Users can reason more powerfully with external visual representations (Liu et al., 2008; Ware, 2003), be it through pictures or videos or diagrams, often allowing otherwise complex information to be processed quickly and more thoroughly – from showing the critical aspects of results to displaying comparisons across multiple outputs to highlighting outliers. Almost all the AM-Smart methods we have mentioned use some form of visualization. The best examples are *ExPanD*[25] and *PRSM*.[26]

Feature 7: Enabling productive failure

Failure is often viewed as a negative, unwanted outcome in research practice or learning. Failure can also be productive (Kapur & Bielaczyc, 2012). Within a research methods platform, 'failure' could entail incorrectly specifying method parameters, selecting inappropriate factor types (e.g. categorical versus numerical), executing method-steps out of order, or misinterpreting results. For those unfamiliar with a given method innovation, these possible missteps are understandable. New users will not have a clear mental model or internal working representation of how the method works, what factors or data it requires, or how to interpret the results (Payne, 2003). While it may be tempting to scaffold these possible missteps, over-scaffolding can create an inauthentic and unrepresentative interaction with a method, where users are able to execute a method but not really learn how to use it correctly. AM-Smart platforms balance scaffolding with productive failure. Users can run a method-step with limited guidance, for example, and receive formative feedback if the results are outside typical ranges or expectations. By striking a balance, users recognize gaps in their knowledge of a method (Loibl & Rummel, 2014) and begin to develop their mental model of it (Kapur & Bielaczyc, 2012). Examples include *SAGEMODELER*,[27] *MAIA*[28] and *ExPanD*.[29]

Feature 8: Supports user-driven learning and inquiry

While multiple models for guiding learning through scientific inquiry exist, Pedaste et al. (2015) synthesized them into a meta-model involving five stages. They are orientation, conceptualization, investigation, conclusions, and discussion.

Relative to these stages, AM-Smart platforms support a self-directed learning approach (Brookfield, 2009; Song & Hill, 2007). *Orientation* involves becoming acquainted with the topic being studied. In a self-directed AM-Smart learning context, this is accomplished by the user identifying what interest or part of their topic or method they want to examine. *Conceptualization* involves generating specific questions or hypotheses. In an AM-Smart platform, conceptualization can be supported using embedded or non embedded scaffolds that prompt users to think about questions and hypotheses for the data they have, in context of the methods the AM-Smart supports. *Investigation* involves research design as well as collecting and analysing data. In the context of AM-Smart, investigation is primarily supported through the methods embedded in the platform with scaffolds as mentioned before, under the assumption some prior research design and data collection has happened. *Conclusions* involves drawing inferences or implications from

investigation results. In AM-Smart, this is facilitated by formative feedback, leveraging users visual reasoning when possible. *Discussion* involves sharing and communicating results of the study. In AM-Smart this is facilitated by these platform's emphasis on interoperability with other tools and result exporting functionality. Two examples are *COMPLEX-IT*[30] and *PRSM*.[31]

Feature 9: Rigorous, authentic, and reliable method platforms

Whenever a method is simplified for non-expert usage, there is the immediate tension around issues of rigor and reliability. While useful, is the platform dependable? While informative, are its algorithms accurate? While its results lead to new insights, can they be published or shared with others? And, while it facilitates learning, can the platform actually be used to guide decision making? AM-Smart platforms actively embrace this tension, seeking to support accessibility with highly rigorous programming. Case in point are the R packages and programming out of which many AM-Smart methods are built. Still, given the field is just emerging, unevenness does exist, making it critical that any AM-Smart app be vetted and field tested by experts in those methods.

The other issue is *task authenticity*, which is particularly important to applied researchers and public sector analysts (Strobel et al., 2013). Task authenticity refers to the degree to which a learning environment is sufficient complex to effectively model the real-world problem being studied (Schimpf et al., in press). Task authenticity exists on a spectrum. In the low extreme, a platform could support a user to perform a complicated procedure by simply uploading data and 'pressing a button', with no options for adjusting the task or its parameters. At this end of the continuum, models can lack real-world value. At the high extreme users could implement their own custom functions in some programming language to execute the method, which results in platforms that are rigorous but less accessible. AM-Smart platforms try to solve this tension by keeping the methods rigorous but their complexity of functions in the background through scaffolding (Quintana et al., 2004; Reiser, 2004). Examples include *Radiant*,[32] *JASP*,[33] and *ExPanD*.[34]

AM-Smart target audiences

AM-Smart methods are largely designed for applied researchers, public sector analysts, those trained in traditional methods, and those seeking to use interdisciplinary methods outside their expertise. There are two other audiences, however, to consider: students and methods innovators.

AM-Smart platforms in the classroom

Many methods classes aim to give students a broad overview of relevant methods. These classes generally employ existing statistical packages or wide-breadth methods programming platforms. Relying on these platforms is effective for teaching about well-established methods but not newer and emerging methods. The nine design features of AM-Smart platforms make it easier for an instructor to bring these methods into their class or create meaningful projects or assignments around it, without requiring extensive instructional pre-planning or supplemental education before their students can use it.

We do recognise, nonetheless, that it is highly useful if students learn some programming, particularly in an age of big data, machine learning, AI, and datafication. Still, scripting in R or MATLAB is not the same as developing a methods package. The expertise required for the latter remains insurmountable for many researchers, which makes clear room for the need for AM-Smart platforms.

Method innovators

The creators of methodological innovations also can benefit from AM-Smart platforms.

While methods innovators can publish their new ideas, if their method requires a high level of expertise, which they usually do, a wide segment of users are immediately cut off from it. The only hope is for the method to be picked up by one of the major statistical or programming platforms, or for the method to be taught in the classroom – all of which often takes years or even decades. AM-smart platforms give method innovators an immediate means to quickly minimize or mitigate such risks. By creating an AM-platform, innovators can make their method innovations more readily available and avoid the pitfalls of obscurity or being largely inaccessible.

Use-Case: Policy evaluation using case-based computational modeling

All of the platforms we have reviewed are online, with tutorials, datasets, examples and even, in some instances, publications for readers to explore. To give a sense of how an AM-Smart method ideally works, we will describe a use-case with COMPLEX-IT.[35]

You are policy analyst working for an international non-profit energy monitor. You have struggled creating a public report about energy trends in the United States due to wide variation in the different configurations of state-level policies and resources. In your role, while you do not have much time for learning new methods, you are painfully aware that you need access to some combination of configurational techniques such as classification, clustering, and machine intelligence.

Recently you learned about the configurational approach, case-based modeling (CBM). You also heard from a colleague about the AM-Smart platform COMPLEX-IT, which aims to classify any dataset of cases into a set of clusters or cluster-based trajectories, given key differences in their respective factor-based configurations. This is accomplished through a *bespoke method suite* that integrates k-means clustering, the neural network self-organizing map (SOM) and data visualisation (Castellani & Rajaram, 2022; Schimpf & Castellani, 2020). You therefore decide to use COMPLEX-IT with a dataset from the Energy Information Agency (EIA), which contains state-level energy use by all common non-renewable (e.g. coal and natural gas) and renewable sources (e.g. wind and solar), see, (Energy Information Agency, 2021).

To get started you go to the online tutorial webpage that supports COMPLEX-IT. While watching the video tutorials, you notice several guidelines that act as *static scaffolds*, prompting you to think about the number of clusters you might find in your data and also how you most likely will need to compare multiple solutions to find an adequate model. You decide, as a first go, you will look for seven clusters.

After uploading your data using the first tab – see, Figure 1 – you turn to the clustering tab. Upon running your seven-cluster k-means, you receive *immediate feedback* on the case configurations and their statistical performance. The solution does not seem that great, with some clusters containing only a few cases. So, you run a few more cluster models, including a six-cluster, then five-cluster solution. During these runs, you also notice that some factors like geothermal and biomass are consistently low and exhibit similar values across clusters, making you decide to remove these factors from your model. These are examples of *productive failure*, the AM-Smart platform is helping you to learn about your data and the model you are building.

Settling on a five-cluster solution, you move to the third tab to corroborate your results. The third tab is an example *of rigor and reliability* and *embedding expertise*, as it allows you to run a form of semi-supervised machine learning, called the self-organizing map algorithm (SOM) to corroborate your k-means solution. In short, Tab 2 lets you supervise the design your model and Tab 3 compares your design to a semi-supervised machine solution. The SOM is pre-set with a number of defaults for immediate use, but as you learn more about the method, you can take over more control, including downloading the SOM solution into a programming package such as MATLAB or R.

With your SOM solution complete, it is time to explore and compare your k-means and SOM solutions via *visual reasoning*. Tab 4 projects your results onto a two-dimensional grid,
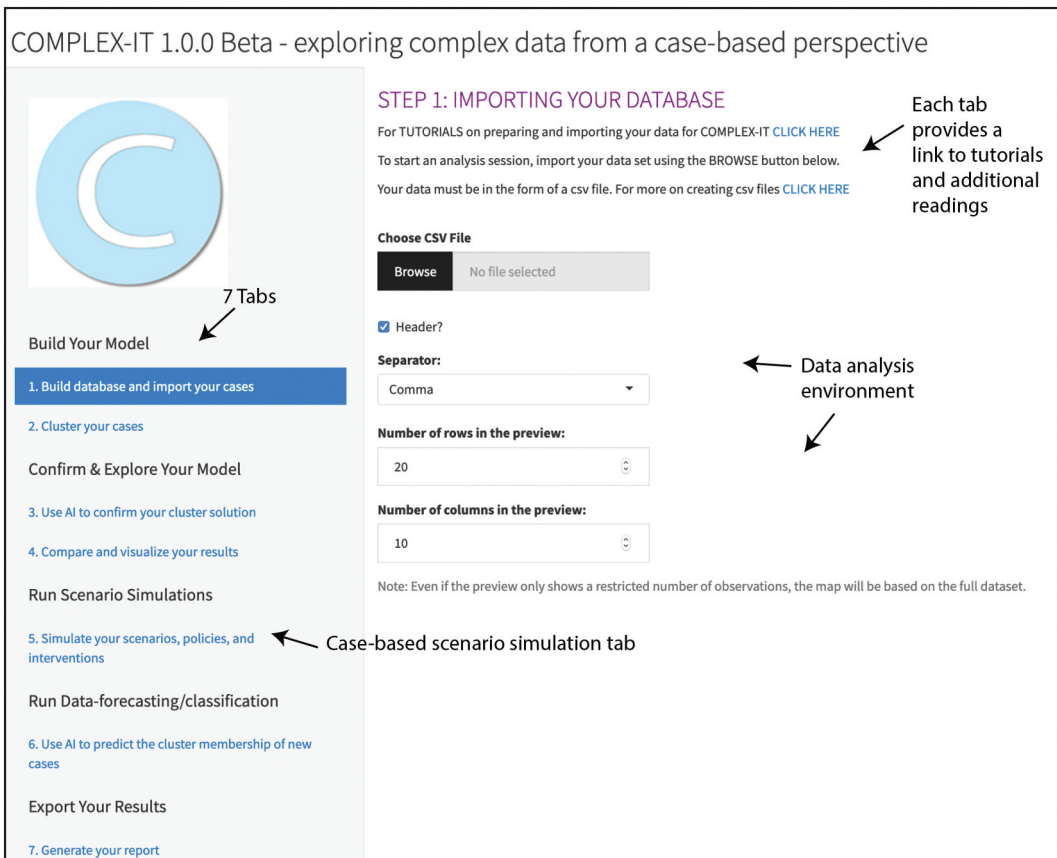
Figure 1. COMPLEX-IT Interface and Layout

allowing you to explore a list of different outputs, from boxplots and barplots of all of your cases, based on cluster membership, to plotting the identification of each case based on its k-means and SOM cluster solution. The 'names' plot is particularly important. It leverages your *visual reasoning* to inspect the agreement between the k-means and SOM solutions. The 'names' plot is also a form of *expertise built into the system*, as you do not have to manually write a function that combines the output of a k-means solution in SPSS, for example, with a SOM solution in MATLAB. Engaged in *user-driven inquiry*, you also learned a lot about k-means and the SOM through an *exploratory and relatively authentic analysis* of a relevant analytical challenge in your daily work.

Having completed your analyses, you are ready to write your report. Here the *rigor and reliability* of COMPLEX-IT is extended, as the last tab provides a downloadable file that not only contains many of the visual outputs from your analyses, but also EXCEL files of your results, which you can upload and run in a statistical package or programming environment or use to confer with colleagues. If needed, one can start all over and run COMPLEX-IT again with your colleagues, using your results and output as the new dataset.

Setting the research agenda

While the present work sets out the vision, core features, and supported audiences for the AM-Smart genre, considerable work remains. Several fields associated with smart technologies may have

interest in broadening the AM-Smart research agenda – from the learning sciences and computational modeling to social science education to software engineering and HCI to methods training in the public, private and third sectors. Given the learning sciences focus on understanding and improving learning environments, the AM-Smart genre may be of heightened interest as a hybrid technology that supports learning and practice simultaneously. Possible research directions include:

(1) Engaging with stakeholders to co-design the types of methods they need.
(2) Cataloguing scaffolding strategies for learning different research methods (Kim & Hannafin, 2011; Reiser, 2004). While we explored two approaches, others may prove more useful, which goes to the next point, critique.
(3) Exploring the ways in which AM-Smart methods may not be useful or hinder methods training or learning. For example, do AM-Smart platforms foster a button-pushing culture? Does running an AM-Smart platforms with limited grounding in the relevant epistemological or methodological framing lead to inappropriate data selection or usage or an ill-informed interpretation of results and their potential implications? AM-Smart risks will be an important area of future research to better understand where, when, and how these risks emerge as well as whether the risks can be lessened through better design of AM-Smart platforms.
(4) Studying if AM-Smart platforms lead to increased methods uptake and greater insights into complex social inquiry.
(5) Investigating what combinations of scaffolding and productive failure are effective at supporting learning-by-doing through AM-Smart platforms.
(6) Exploring what types and formats of feedback best support different learners (Shute, 2008).

Research within these areas may also draw on a variety of theories, frameworks, or methodologies commonly used in the learning sciences. There is constructivism, for example, which focuses on helping learners build on their past knowledge to incorporate new ideas or skills (Nathan & Sawyer, 2014). There is also design-based research, which seeks to iteratively develop interventions based on specific learning contexts and generates theories from the local context outward (Barab & Squire, 2004). And there is situated cognition, just to name a few, which focuses on how people learn the practices and norms of a given area through social interaction (Lave & Wenger, 1991).

Given HCI's focus on designing interfaces between people and computational technology (Carroll, 2003), AM-Smart may also be of heightened interest as a new means to facilitate interactions between analysts and supporting technology. Possible research directions of interest may include:

(1) Comparing the effectiveness of standard method software and AM-Smart platforms for learning complex skills.
(2) Exploring ways to support collaborative research through AM-Smart platforms.
(3) Tracking user behaviour throughout these platforms and identifying common user tasks (Crystal & Ellington, 2004) or interaction scenarios (Carroll, 2000).
(4) Evaluating the effectiveness of static or interactive visualizations to support user reasoning (ElTayeby & Dou, 2016).
(5) Recently within the HCI literature there has also been a growing interest in human and AI interaction (Amershi et al., 2019; Yang et al., 2020). AI applications could be promising for supporting users to learn method innovations through AM-Smart, and further solidify the platforms built-in expertise. An AI system could act as a collaborator with a human user (Wang et al., 2020), for instance, using approaches like the genetic algorithm or simulated annealing to search for more optimal method parameters when the parameter space is expansive.

(6) Developing AM-Smart methods for qualitative inquiry and for mixed-methods research.

Research within HCI may also draw on a variety of frameworks or methodologies, such as information-processing that views human cognition as an information processing system (John, 2003); eye-tracking approaches that use eye-tracking technology to gather data on how and what users pay attention to while completing tasks (Olsen et al., 2011) or think-aloud protocols where users verbally express their thoughts while using some technology to create a detailed record of their cognition and decisions (Olmsted-Hawala et al., 2010).

Conclusion

To advance a useful and critical engagement with the wider social science methods community, we end our article challenging others to use, edit, expand, repurpose, or develop their own AM-Smart platforms, following the nine software features outlined above. This includes AM-Smart platforms for qualitative and mixed-methods inquiry. We also challenge smart technology researchers broadly and those in the learning sciences and HCI to begin to research existing and newly created AM-Smart platforms. Special sessions, workshops or panel discussions on this field and its research directions could be explored at well-established conferences such as HCI's Human Factors in Computer Systems hosted by the Association for Computing Machinery (ACM) or the learning sciences International Society of the Learning Sciences (ISLS) Conference, hosted by ISLS.

AM-Smart platforms hold tremendous potential for enhancing broader access, development, and interdisciplinary exchange around methods innovations. They also have the potential to open a new field of inquiry and development within the space of smart technologies and contributing fields. Expansion of AM-Smart platforms and broader and deeper research from the learning sciences, HCI, and other fields could revolutionize the way method innovations are created, distributed, and deployed and how researchers, practitioners, and students tackle the complexities and digitization of our times.

Notes

1. The COVID pandemic tipped daily global life into the AM-Smart world as researchers, software designers, public health officials, newspapers, social media and governments created a myriad of online platforms for engaging in data-driven analysis and mapping of the pandemic as it spread locally, nationally and globally. One example is the World Health Organization's COVID dashboard, https://covid19.who.int/
2. To view a gallery of R Shiny examples, see https://shiny.rstudio.com/gallery/#user-showcase
3. See https://shiny.rstudio.com/
4. https://compasss.org/
5. There are AM-Smart methods for other sciences, including engineering, biology and chemistry, but we avoided those, focusing only on social inquiry.
6. To run COMPLEX-IT and read through tutorials and publications, etc, visit www.art-sciencefactory.com/complexit.html
7. To run Radiant, visit https://shiny.rstudio.com/gallery/radiant.html For more on Radiant, visit https://radiant-rstats.github.io/docs/
8. To run JASP and read tutorials, visit https://jasp-stats.org/
9. To run PRSM, visit https://www.prsm.uk/
10. To run SAGEMODELER and learn more about it, visit https://sagemodeler.concord.org/
11. To run MAIA and read tutorials and published examples, visit http://maia.tudelft.nl/index.html
12. To run NetLogo online or download App and explore model gallery, visit www.netlogoweb.org/
13. To run Cytoscape, read through tutorials and publications, and explore suite of apps, visit https://cytoscape.org/index.html
14. To run ExPanD, visit https://jgassen.shinyapps.io/expand/ To read about the app and explore tutorials and examples, visit https://shiny.rstudio.com/gallery/explore-panel-data.html

15. To run COMPLEX-IT, visit www.art-sciencefactory.com/complexit.html. See, also: Schimpf and Castellani (2020). COMPLEX-IT: A Case-Based Modeling and Scenario Simulation Platform for Social Inquiry. *Journal of Open Research Software*, 8: 25. DOI: https://doi.org/10.5334/jors.298
16. To run JASP and read tutorials, visit https://jasp-stats.org/
17. To run App, visit https://shiny.rstudio.com/gallery/radiant.html For more on platform, visit https://radiant-rstats.github.io/docs/
18. To run PRSM, visit https://www.prsm.uk/
19. To run JASP and read tutorials, visit https://jasp-stats.org/
20. To run MAIA and read tutorials and published examples, visit http://maia.tudelft.nl/index.html
21. To run Cytoscape, read through tutorials and publications, and explore suite of apps, visit https://cytoscape.org/index.html
22. To run SAGEMODELER and learn more about it, visit https://sagemodeler.concord.org/
23. To run MAIA and read tutorials and published examples, visit http://maia.tudelft.nl/index.html
24. To run NetLogo online or download App and explore model gallery, visit www.netlogoweb.org/
25. To run ExPanD, visit https://jgassen.shinyapps.io/expand/
26. To run PRSM, visit https://www.prsm.uk/
27. To run SAGEMODELER and learn more about it, visit https://sagemodeler.concord.org/
28. To run MAIA and read tutorials and published examples, visit http://maia.tudelft.nl/index.html
29. To run ExPanD, visit https://jgassen.shinyapps.io/expand/
30. To run COMPLEX-IT, visit www.art-sciencefactory.com/complexit.html
31. To run PRSM, visit https://www.prsm.uk/
32. To run Radiant, visit https://shiny.rstudio.com/gallery/radiant.html For more on Radiant, visit https://radiant-rstats.github.io/docs/
33. To run JASP and read tutorials, visit https://jasp-stats.org/
34. To run ExPanD, visit https://jgassen.shinyapps.io/expand/
35. To explore the tutorials, videos and datasets for COMPLEX-IT, visit www.art-sciencefactory.com/complexit.html

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Corey Schimpf is an Assistant Professor in the Department of Engineering Education at the University at Buffalo, SUNY, USA. He is the Chair-Elect of the Design in Engineering Education Division of the American Society of Engineering Education. His lab focuses on engineering design, advancing research methods, and technology innovations to support learning in complex domains. Major research strands include: (1) analyzing how expertise develops in engineering design across the continuum from novice pre-college students to practicing engineers, (2) advancing engineering design and applied education or social science research by integrating new theoretical or analytical frameworks (e.g., from data science or complexity science) and (3) conducting design-based research to develop scaffolding tools for supporting the learning of complex skills like design and social inquiry.

Brian Castellani is the Director of the Research Methods Centre and Co-Director of the Wolfson Research Institute for Health and Wellbeing at Durham University, UK. He is also Adjunct Professor of Psychiatry (Northeastern Ohio Medical University, USA), Editor of the Routledge Complexity in Social Science series, CO-I for the Centre for the Evaluation of Complexity Across the Nexus, and a Fellow of the UK National Academy of Social Sciences. Brian also runs InSPIRE, a UK policy and research consortium for mitigating the impact places have on air quality, dementia and brain health across the life course. Brian is trained as a public health sociologist, clinical psychologist, and methodologist and take a transdisciplinary approach to his work. His methodological focus is primarily on computational modelling and mixed-methods. He and his colleagues have spent the past ten years developing a new case-based, data mining approach to modelling complex social systems and social complexity, called COMPLEX-IT, which they have used to help researchers, policy evaluators, and public sector organisations address a variety of complex public health issues. Staff Profile

ORCID

Corey Schimpf http://orcid.org/0000-0003-2706-3282
Brian Castellani http://orcid.org/0000-0002-8763-8052

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

Arnheim, R. (1997). *Visual thinking*. Univ of California Press.

Barab, S., & Squire, K. (2004). Design-Based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, *13*(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1

Barbrook-Johnson, P., & Carrick, J. (2021). Combining complexity-framed research methods for social research. *International Journal of Social Research Methodology*, 1–14. https://doi.org/10.1080/13645579.2021.1946118

Bente, G., & Breuer, J. (2009). *Making the implicit explicit. Embedded measurement in serious games. Serious games: Mechanisms and effects*. Routledge.

Bernstein, J. H. (2015). Transdisciplinarity: A review of its origins, development, and current issues. *Journal of Research Practice*, *11*(1), 1–20.

Bransford, J., Brophy, S., & Williams, S. (2000). When computer technologies meet the learning sciences: Issues and opportunities. *Journal of Applied Development Psychology*, *21*(1), 59–84. https://doi.org/10.1016/S0193-3973(99)00051-9

Brookfield, S. D. (2009). Self-directed learning. In Maclean, Rupert, Wilson, David, Chinien, Chris (Eds) *International handbook of education for the changing world of work* (pp. 2615–2627). Springer.

Burrows, R., & Savage, M. (2014). After the crisis? Big data and the methodological challenges of empirical sociology. *Big Data & Society*, *1*(1), 2053951714540280. https://doi.org/10.1177/2053951714540280

Carillo, K., & Okoli, C. (2008). The open source movement: A revolution in software development. *Journal of Computer Information Systems*, *49*(2), 1–9. https://doi.org/10.1080/08874417.2009.11646043

Carroll, J. M. (2000). Making Use: Scenarios and scenario-based design. *Proceedings of the 3rd conference on designing interactive systems: processes, practices, methods, and techniques*, 4. https://doi.org/10.1145/347642.347652

Carroll, J. M. (2003). Introduction: Toward a multidisciplinary science of human-computer interaction. In Carroll, John M. (Ed.) *HCI models, theories, and frameworks* (pp. 1–9). Elsevier. https://doi.org/10.1016/B978-155860808-5/50001-0

Castellani, B., & Rajaram, R. (2022). Big Data mining and complexity. *SAGE – Volume 11 of the SAGE Quantitative Research Kit*. London: SAGE.

Crystal, A., & Ellington, B. (2004). Task analysis and human-computer interaction: Approaches, techniques, and levels of analysis. *AMCIS 2004 Proceedings*, 391.

Damasio, A. R. (2012). *Self comes to mind: Constructing the conscious brain*. Vintage.

Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: Systematic review of effects of scaffolding in the domain of science education: Scaffolding self-regulated learning with CBLES. *Journal of Computer Assisted Learning*, *28*(6), 557–573. https://doi.org/10.1111/j.1365-2729.2011.00476.x

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the Classroom. *Theory Into Practice*, *55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

ElTayeby, O., & Dou, W. (2016). A survey on interaction log analysis for evaluating exploratory visualizations. *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization*, 62–69. https://doi.org/10.1145/2993901.2993912

Energy Information Agency. (2021). *Net generation by state by type of producer by energy source (EIA-906, EIA-920, and EIA-923)*. (Data Set). EIA. https://www.eia.gov/electricity/data/state/

Hardy, L., Dixon, C., & Hsi, S. (2020). From data collectors to data producers: Shifting students' relationship to data. *Journal of the Learning Sciences*, *29*(1), 104–126. https://doi.org/10.1080/10508406.2019.1678164

Hayles, N. K. (1999). *How we became posthuman: Virtual bodies in cybernetics, literature, and informatics*. University of Chicago Press.

Hayles, N. K. (2017). *Unthought: The power of the cognitive nonconscious*. University of Chicago Press.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*(2), 174–196. https://doi.org/10.1145/353485.353487

Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior*, *26*(6), 1278–1288. https://doi.org/10.1016/j.chb.2010.05.031

Hutchins, E. (2006). The distributed cognition perspective on human interaction. In N. J.Enfield & S. C.Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 375–398). Berg.

John, B. E. (2003). Information processing and skilled behavior. In Carroll, John M. (Ed.)*HCI models, theories, and frameworks* (pp. 55–101). Elsevier. https://doi.org/10.1016/B978-155860808-5/50004-6

Kapur, M., & Bielaczyc, K. (2012). Designing for Productive Failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Kim, M. C., & Hannafin, M. J. (2011). Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice. *Computers & Education*, *56*(2), 403–417. https://doi.org/10.1016/j.compedu.2010.08.024

Kirk, A. (2016). *Data visualisation: A handbook for data driven design*. Sage.

Kovalchuk, S. V., Krzhizhanovskaya, V. V., Paszyński, M., Závodszky, G., Lees, M. H., Dongarra, J., & Sloot, P. M. (2021). *20 years of computational science: Selected papers from 2020 international conference on computational science*.

Lave, J., & Wenger, E. (1991). *Situated learning: legitimate peripheral participation (learning in doing: social, cognitive and computational perspectives*. Cambridge University Press.

Liu, Z., Nersessian, N., & Stasko, J. (2008). Distributed cognition as a theoretical framework for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *14*(6), 1173–1180. https://doi.org/10.1109/TVCG.2008.121

Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction*, *34*, 74–85. https://doi.org/10.1016/j.learninstruc.2014.08.004

Marres, N. (2017). *Digital sociology: The reinvention of social research*. John Wiley & Sons.

Nathan, M. J., & Sawyer, R. K. (2014). Foundations of the learning sciences. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 21–43). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.004

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010). Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. *Proceedings of the 28th international conference on human factors in computing systems - CHI '10*, 2381. https://doi.org/10.1145/1753326.1753685

Olsen, A., Schmidt, A., Marshall, P., & Sundstedt, V. (2011). Using eye tracking for interaction. *CHI '11 Extended abstracts on Human Factors in Computing Systems*, 741–744. https://doi.org/10.1145/1979742.1979541

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 27–42). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.004

Payne, S. J. (2003). Users' mental models: The very ideas. In Carroll, John M. (Ed,) *HCI models, theories, and frameworks* (pp. 135–156). Elsevier. https://doi.org/10.1016/B978-155860808-5/50006-X

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, *13*(3), 337–386. https://doi.org/10.1207/s15327809jls1303_4

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*(3), 273–304. https://doi.org/10.1207/s15327809jls1303_2

SageModeler [Computer software]. (2020). The Concord Consortium and the CREATE for STEM Institute at Michigan State University.

Sawyer, R. K. (2014). Introduction. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 1–18). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.002

Schimpf, C., & Castellani, B. (2020). COMPLEX-IT: A case-based modeling and scenario simulation platform for social inquiry. *Journal of Open Research Software*, *8*(1), 25. http://doi.org/10.5334/jors.298

Schimpf, C., Purzer, S., Quintana, J., Sereiviene, E., & Xie, C. (2022). What does it mean to be authentic? Challenges and opportunities faced in creating K-12 engineering design projects with multiple dimensions of authenticity. In K. Sanzo, J. P. Scribner, J. A. Wheeler, & K. W. Maxow (Eds.), *Design thinking: Research, innovation, and implementation*. Information Age Publishing, 27.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Song, L., & Hill, J. R. (2007). A conceptual model for understanding self-directed learning in online environments. *Journal of Interactive Online Learning*, *6*(1), 27–42. http://www.ncolr.org/issues/jiol/v6/n1/a-conceptual-model-for-understanding-self-directed-learning-in-online-environments.html

Strobel, J., Wang, J., Weber, N. R., & Dyehouse, M. (2013). The role of authenticity in design-based learning environments: The case of engineering education. *Computers & Education*, *64*, 143–152. https://doi.org/10.1016/j.compedu.2012.11.026

Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., & Wang, Q. (2020). From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, 1–6. https://doi.org/10.1145/3334480.3381069

Ware, C. (2003). Design as applied perception. In Carroll, John M. (Ed.) *HCI models, theories, and frameworks* (pp. 11–26). Elsevier. https://doi.org/10.1016/B978-155860808-5/50002-2

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining whether, why, and how Human-AI interaction is uniquely difficult to design. *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–13. https://doi.org/10.1145/3313831.3376301