Computational Notebooks as Co-Design Tools: Engaging Young Adults Living with Diabetes, Family Carers, and Clinicians with Machine Learning Models

Amid Ayobi
University College London, London, United Kingdom

Jacob Hughes
University of Bristol, Bristol, United Kingdom

Christopher J Duckworth
University of Southampton, Southampton, United Kingdom

Jakub J Dylag
University of Southampton, Southampton, United Kingdom

Sam James
University of Bristol, Bristol, United Kingdom

Paul Marshall
University of Bristol, Bristol, United Kingdom

Matthew Guy
University Hospital Southampton, Southampton, United Kingdom

Anitha Kumaran
University Hospital Southampton, Southampton, United Kingdom

Adriane Chapman
University of Southampton, Southampton, United Kingdom

Michael Boniface
University of Southampton, Southampton, United Kingdom

Aisling Ann O'Kane
University of Bristol, Bristol, United Kingdom

Engaging end user groups with machine learning (ML) models can help align the design of predictive systems with people's needs and expectations. We present a co-design study investigating the benefits and challenges of using computational notebooks to inform ML models with end user groups. We used a computational notebook to engage young adults, carers, and clinicians with an example ML model that predicted health risk in diabetes care. Through co-design workshops and retrospective interviews, we found that participants particularly valued using the interactive data visualisations of the computational notebook to scaffold multidisciplinary learning, anticipate benefits and harms of the example ML model, and create fictional feature importance plots to highlight care needs. Participants also reported challenges, from running code cells to managing information asymmetries and power imbalances. We discuss the potential of leveraging computational notebooks as interactive co-design tools to meet end user needs early in ML model lifecycles.

1 Introduction

Artificial intelligence (AI) is increasingly becoming ubiquitous in people's daily work and life. From healthcare to agriculture [32,45], AI has the potential to transform industry, but may also undermine human values and amplify structural inequalities [62,72,86]. As a response, academic institutions, corporations, and government bodies have focused attention on human-centred principles, such as fairness, accountability, trust, and ethics [1,43,58,63]. Research increasingly draws on human-centred and participatory approaches to empower different stakeholders in the design of AI-driven systems [60], including regulatory bodies and end consumers whose lives can be significantly affected by algorithmic outcomes [42]. However, responsible AI requires not only multidisciplinary collaborations but also novel interdisciplinary approaches that bridge applied disciplines, such as data science and user experience design [40,85].

Computational notebooks provide significant potential to facilitate multidisciplinary collaborations aimed at informing the design of human-centred AI algorithms. People from diverse professional backgrounds already use computational notebooks to combine code, text, and multimedia resources in a single document to explore and visualise data [75,80]. Platforms, such as Jupyter Notebook [47], have been shown to be particularly suitable to support data exploration, documentation, and collaboration within educational and professional settings [46,77,91,93]. Prior work has highlighted the need for user-centred and participatory design approaches to developing AI systems [60,90]. However, the potential of using computational notebooks, which are primarily designed as tools for data science, as part of human-centred design processes with anticipated end user groups is not well explored. We offer two methodological, research tool-focused contributions.

Firstly, we present the findings of a co-design study including a series of online workshops and post-workshop interviews conducted using a computational notebook that combines static illustrations, interactive ML model explanations, and participatory design fiction activities to foster mutual learning and collective creativity. Engaging with young adults with T1D, family carers, and clinicians as co-designers, we identify the perceived benefits and challenges of using computational notebooks as co-design tools during early ML model development stages. Participants particularly valued using the interactive data visualisation elements of the computational notebook to collaboratively learn about ML concepts, anticipate potential benefits and harms of the example ML model, and create fictional ML model feature importance plots to highlight misalignments between the example ML model and their individual care needs. Furthermore, participants highlighted challenges in working with the computational notebook, such as running code cells, dealing with information asymmetries, and managing power imbalances.

Secondly, we contribute implications for the design and use of future co-design tools to facilitate human-centred and participatory ML development with end users with varied ML literacy. Computational notebooks provide a powerful platform to explain ML models in illustrative ways, identify potential mismatches between people's lived experiences and ML model features, and collaboratively adapt ML model representations according to people's needs at early development stages. However, the design of interactive programming environments, such as computational notebooks, needs to be tailored to people's individual backgrounds, including their diverse information and learning needs, to avoid disempowering experiences throughout multidisciplinary co-design processes. While co-design

predominately draws on user experience (UX) design methods, we provide inspiration on how data science tools can be adopted and adapted as co-design tools to bridge gaps between ML and UX design processes. In doing so, we highlight the importance of developing more inclusive and participatory research environments that empower people from non-data science backgrounds in taking more seminal roles in aligning ML models with their individual needs and expectations in daily life.

2 Related Work

We first review related work on human-AI interaction and identify the importance of boundary objects, such as computational notebooks [85,98], in fostering multidisciplinary collaborations on human-centred AI. We then provide an overview of research on computational notebooks and situate our work within the context of type 1 diabetes care.

2.1 Human-AI Interaction

Scholars from different research fields have documented how individuals experience AI-driven technologies in daily life [62,97], and how people from different professional backgrounds use and develop AI systems at work [94,98]. For example, previous work has documented the situated and collaborative facets of ML practices [22,38]: data science work typically involves data processing, ML model building, and ML model deployment activities, from cleaning data and feature engineering to selecting, validating, monitoring, and optimising ML models [98]. While data scientists develop an intuitive sense of datasets and ML models [69], accurately describing output data visualisations and managing trust in ML model interpretability tools are potential challenges as part of data science workflows [51]. Kross et al. [56] describe data scientists' workflows in collaborating with clients in academia and industry, including developing trust, understanding constraints, collaboratively framing problem spaces, aligning data science with domain expertise, and supporting clients in coping with analytical outcomes. Importantly, Piorkowski et al. [76] identify "knowledge mismatches" between AI practitioners and stakeholders with different levels of data science literacy and educational efforts to overcome these gaps with the help of slide decks, data visualisations, and explanatory stories.

As consumer products and services increasingly provide AI-driven functionality, not only data scientists, but also stakeholders with diverse professional roles are expected to adopt ML concepts to productively inform ML model lifecycles [42,60]. In particular, user experience designers and researchers increasingly engage with AI as a design material [30,31]. For example, an academia- and industry-led study [96] identifies that user experience design professionals are particularly equipped to contribute to AI innovation at system and service levels (e.g., identifying human-AI goals and defining flows and processes) in addition to interaction design levels (e.g., designing user interfaces). However, designing human-AI interactions can bring about significant challenges [30,31,39,94,95], such as understanding data science jargon and appreciating ML capabilities [94], envisioning feasible AI experiences and rapidly prototyping realistic human-AI interactions [12,31,95], as well as developing a shared language and methodologies that can help align user-centred design and ML model development [39]. Moving on from how professionals in cross-functional roles work with ML, prior work has drawn on participatory design approaches to investigate human-centred principles and needs [36,68,70,89]. For example, Katan et al. [50] have demonstrated the utility of interactive ML to support people with disabilities in customising gesturally controlled musical interfaces and Loi et al. [59,60] have presented co-design approaches as promising directions to informing AI futures.

Overall, prior work has not only exemplified the benefits of drawing on ethnographic, user-centred, and participatory design methodologies to inform ethically aligned and responsible AI innovation, but it has also identified the importance of boundary objects in fostering fruitful multidisciplinary collaborations between data scientists, user experience designers, and anticipated end user groups [4,5,85,98].

2.2 Computational Notebooks

People from a wide range of professional backgrounds, including portfolio managers, astronomers, and journalists, use computational notebooks for data science [75,80]. One of the most widely used computational notebook platforms is

Jupyter Notebook [47], a tool that supports data exploration, analysis, and collaboration, with flexible building blocks, such as code, text, images, and interactive plots.

Initial lines of research illustrate the use and experience of computational notebooks. For example, Rule et al. [80]'s set of studies found (1) that one in four notebooks on GitHub had no explanatory text; (2) that the minority of academic notebooks documented reasoning and results; and (3) that the use of notebooks can be described as *"personal, exploratory, and messy."* Data scientists typically intertwine interactive code environments, such as Jupyter Notebook, with a wide range of different tools, including communication and presentation applications (e.g., Slack and Microsoft PowerPoint), to collaborate within cross-functional teams [98] and share information across expertise boundaries [85].

Computational notebooks can facilitate data exploration in one single document but can also cause significant challenges in daily work. Drawing on semi-structured interviews and a survey with data scientists, Chattopadhyay et al. [16] identify several pain points, from importing different data sources and handling large code corpuses to managing confidentiality and security challenges. Furthermore, collaborative work can bring about tensions between rapid data exploration activities, documentation requirements, and coordination needs [91]. These realities and intricacies of using computational notebooks [16,80] have informed design-led research on how the functionality of computational notebooks could be improved according to people's needs: the academic community has pushed technical boundaries to support end users in managing code [41], documenting code [92], creating data comics [49], and generating slide decks [99]. Furthermore, computational notebooks have been tailored to support interactive teaching and learning [23]. While computational notebooks can effectively support blended learning and, in particular, technical writing and communication skill exercises, there are also potential limitations in classroom settings, such as in managing hidden states, overcoming limited debugging capabilities, and facilitating software engineering best practices [46,93].

Computational notebooks provide significant potential to support research projects aimed at informing the design of human-centred ML models with affected stakeholder and anticipated end user groups. Although much attention has been paid to the use of computational notebooks within professional and educational settings [46,91], less is known about their use by people without significant data literacy. The benefits and limitations of using computational notebooks to support human-centred design processes with non-ML domain experts, such as people living with chronic health conditions, has received much less attention.

2.3 Towards Human-Centred and Participatory Machine Learning in Diabetes Care

In recent years there has been much interest in using ML for supporting diabetes management, including improved glucose prediction [28], classification of the impact of specific behaviours on health [29], and association of GPS location with blood glucose level variability [27]. Type 1 diabetes (T1D) is an auto-immune condition whereby the body does not produce sufficient insulin: people with T1D need to take daily insulin injections, and insulin doses rely on complex interdependent factors, including exercise, diet, stress, and others [2]. ML based approaches could help overcome some of the barriers to adoption of current self-care technology through increased use of sensors to reduce the burden of data collection [15,20,53] and the cognitive effort needed to gain actionable insights from data [54]. As T1D management depends primarily on self-care, it is crucial that systems are oriented towards people's needs [18,37,64,65,84]. Therefore, design approaches which integrate end users as co-designers of potential solutions provide a promising methodological framework for diabetes system development. Diverse user-centred design methodologies have often been applied to diabetes technologies, with researchers reporting that such methods helped answer research questions, gain knowledge on user mental models, clarify misconceptions, and inform the design of usable and useful self-care technology [57].

Although there is a significant body of literature on the use of human-centred design methods for diabetes care [57,84], there is little research specifically on how to involve people with T1D and their support networks into the design of ML models [4,5]. Here, we report the findings of a series of online workshops conducted with young adults with T1D, family carers, and clinicians to provide a qualitative and reflective account on the benefits and challenges of using computational notebooks as design probes to support affected stakeholders and anticipated end users in informing the design of human-centred ML models for desirable T1D care.

3 Method

Our work was informed by two core co-design principles: mutual learning and collective creatively [6,14,83]. During initial phases, we co-designed a computational notebook within a team of clinicians, data scientists, and HCI researchers. The iterative design of the computational notebook was a collective design effort of intertwining data science processes (e.g., implementing an example ML model) with clinical expertise (e.g., narrating ML explanations) and user experience design (e.g., designing data visualisations). The computational notebook was used as a co-designed artifact to facilitate a series of design workshops with anticipated end user groups. These workshops involved co-design teams with HCI researchers, data scientists, and an additional participant group (young adults with T1D, family carers, and clinicians). The computational notebook was intended to support mutual learning (e.g., learning about ML concepts and learning about people's lived experiences) and foster collective creativity (e.g., co-designing fictional ML feature importance plots). We would like to acknowledge up front that conducting the workshops with stakeholder groups separately could be considered a limitation. However, we decided to conduct the design workshops with HCI researchers, data scientists, and homogeneous groups of young adults, family carers, and clinicians to support each participant group in leveraging their shared lived experiences at initial project stages, given that the clinically informed example ML model could have led to tensions between participant groups. Our objective was to gain a detailed understanding of people's perceptions and the risk prediction needs of each participant group at early ML development phases.

3.1 Co-Design of a Computational Notebook

3.1.1 Design Objectives

While the optimisation of ML models is traditionally driven by data science expertise, human-centred ML is increasingly multidisciplinary, driven by domain experts from diverse backgrounds. Based on this background, we aimed to involve key end user groups in early ML-model design stages to avoid adverse downstream effects and to inform the design of clinically viable and desirable ML-models for T1D care. Our objective was to work with three anticipated end user groups: clinicians, young adults with T1D, and family carers. Our intention was to collaborate with clinicians to co-design a ML model that demonstrates the potential use case of risk stratification in T1D care. Young adults with T1D and family carers are domain experts with lived experience. Both stakeholder groups are anticipated data holders: they would need to provide informed consent for their data to be used and interact with a predictive risk stratification system in close partnership with clinicians. Whilst multi-stakeholder approaches can bring about multidisciplinary challenges, early engagement with anticipated end user groups is crucial to inform the design of human-centred AI systems that help bridge real world gaps between clinical and domestic settings. Our objective was to pursue the following design research objectives: (1) investigating the benefits and challenges of using computational notebooks as interactive co-design tools to engage anticipated end user groups in informing the design of ML models; (2) deriving implications for the design and use of future co-design tools to facilitate human-centred and participatory ML development with end users with varied ML literacy.

3.1.2 Design Process

The core co-design team included three academic data scientists, three clinicians, and three human-computer interaction researchers. We iteratively co-designed a computational notebook that illustrates an example ML model. The ML model predicts risk from hospitalisations due to hypoglycemia, hyperglycemia and diabetic ketoacidosis. We implemented this ML model with pre-selected features to demonstrate a plausible example and foster mutual learning and support participants in understanding ML models. We implemented the ML model based on a publicly available dataset provided by the T1D Exchange Registry. The dataset included routinely collected information for 22,697 people living with T1D in the United States [79]. A supervised ML approach drawing on the XGBoost framework was applied [17]. The outcome was a tree-based ML algorithm with eleven ML model features, including demographic related features (e.g., age) and clinically informed features (e.g., blood pressure). A feature importance plot shows the usefulness of each ML model feature in predicting the target. The most predictive ML model features were HbA1C, kidney function, and use of continuous blood monitors. The ML model feature importance plot was key in

understanding the example ML model within the multidisciplinary design team and in informing the design of the computational notebook.

We used the JupyterLab [47] to host the computational notebooks on a university server. JupyterLab provided a web-based environment to work with code consoles, terminals, and Jupyter Notebook [47]. The computational notebooks were made available to participants via an encrypted internet connection and password protected link. The design of the computational notebook draws on an incremental explanation approach to foster multidisciplinary discussion and learning, including a static illustration to elicit lived care experiences (Part I); an interactive table and graphs to discuss the structure and limitations of the dataset used to implement the ML model (Part II); an static illustration to mediate ML terminology (Part III); an interactive visualisation and static feature importance plot to exemplify the ML model (Part IV); and a participatory design activity to support participants in creating fictional ML model representations according to their individual needs (Part V). We structured the computational notebook along five open-ended questions:

3.1.3   Part I: How can data support our health and wellbeing?

The first part of the computational notebook (see Figure 1) showed an illustration intended to help understand a participant's current lived self-care experiences and roles of data-driven technologies. Based on the illustration, we asked open-ended questions, such as *"What types of technologies do you use to manage your health and wellbeing?"*, *"How would you describe your experience using digital technologies?"*, *"What are your thoughts on diabetes technologies that make predictions (e.g., current CGM devices predict blood glucose levels and provide alerts)?"*.
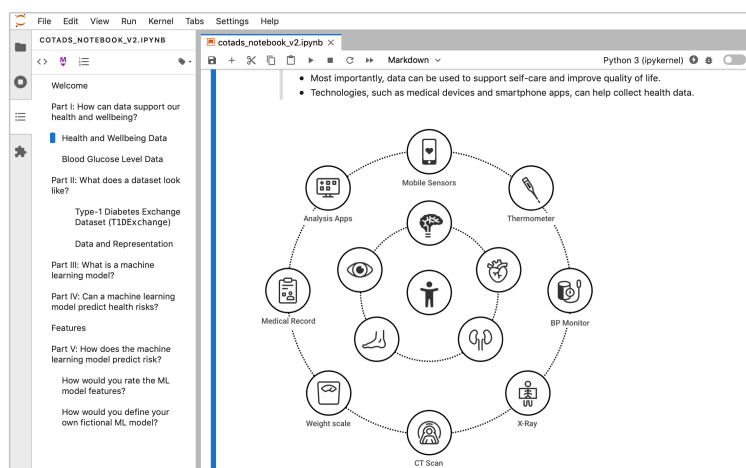

Figure 1: Part I showed an illustration intended to help understand people's data-driven care experiences.

3.1.4   Part II: What does a dataset look like?

The second part of the computational notebook (see Figure 2) aimed to exemplify a dataset and its potential limitations. It provided a table showing the data of four people according to the T1D Exchange dataset [79]. In addition, we provided an overview of how people are represented in the T1D Exchange dataset (i.e., race and gender). Probing questions included: *"How do you think this representation could affect any predictive technologies using this data?"*, *"How would you change the dataset?"*, *"What are your thoughts on contributing your personal health data to a public dataset?"*
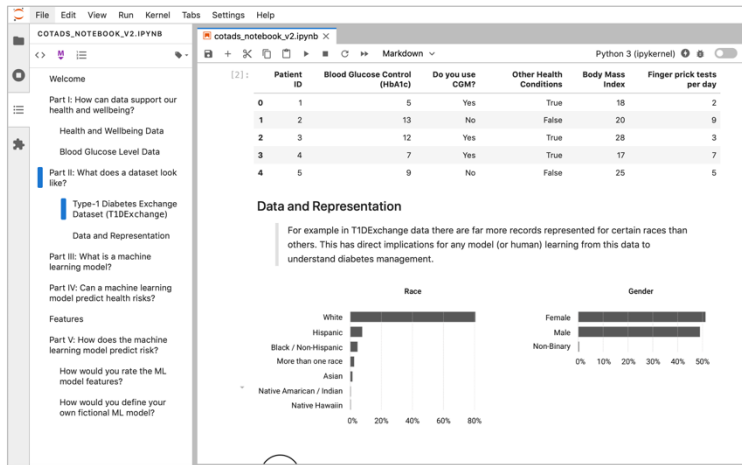
6

Figure 2: *Part II provided an overview of the T1D Exchange dataset with a breakdown regarding race and gender.*

3.1.5 Part III: What is a machine learning model?

As part of the next section (see Figure 3), we explained that ML models make use of data to make predictions. We introduced ML concepts, such training and testing phases, as well as basic ML terminology, including features and labels. We drew on the previously discussed T1D Exchange dataset and used a simplified example of predicting the number of clinical appointments based on HbA1c values. In the remainder of the session, we encouraged participants to ask questions on what a ML model is and how it makes predictions.
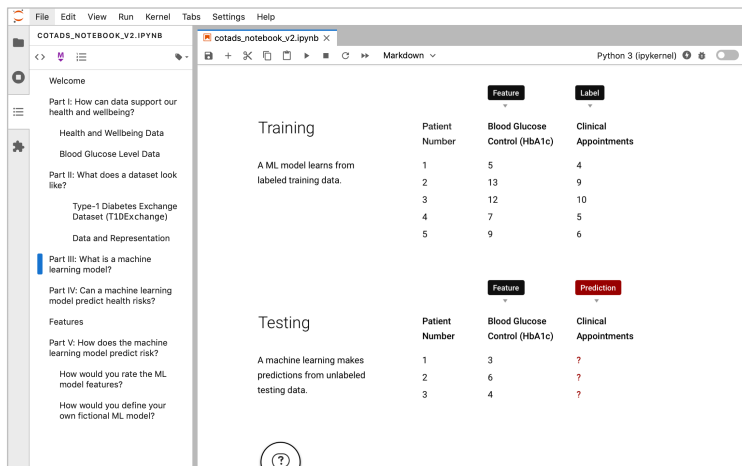

Figure 3: Part III introduced ML concepts, including training and testing phases and terms, such as features and labels.

3.1.6 Part IV: How does the machine learning model predict health risk?

The fourth section of the computational notebook (see Figure 4) presented an interactive risk score visualisation to exemplify how ML model features can influence ML model outcomes. Participants could change three example ML

7

model features (i.e., HbA1c, number of blood glucose tests per day, and CGM use) and observe how their changes led to different risk predictions (i.e., low risk, medium risk, high risk). We asked the following questions *"How would you describe your experience using the risk score visualisation?", "How would you describe how the risk score works?", "What are your thoughts on assigning a risk level to a person with T1D to make a prediction?"*
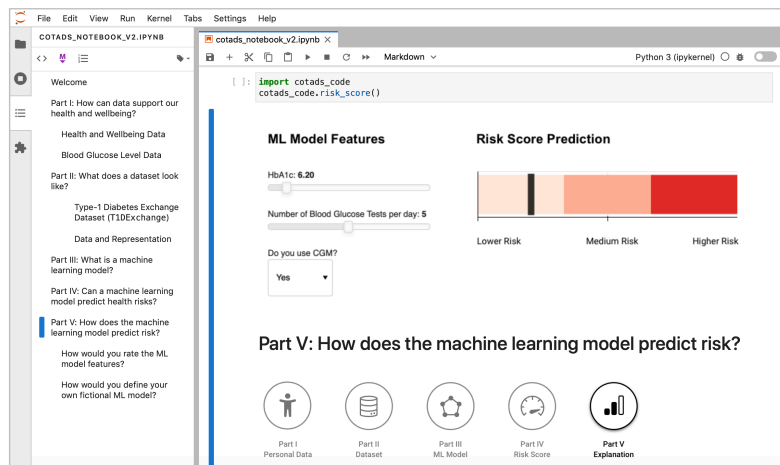

Figure 4: **Part IV included an interactive risk score visualisation with adaptable features, such as HbA1C.**

3.1.7 Part V: How would you define your own fictional machine learning model?

As part of the last activity (see Figure 5 and Figure 6), we revisited the example ML model feature importance plot and then encouraged participants to create their own fictional ML model feature importance plots according to their subjective needs. We scaffolded this activity by sharing our screen and demonstrating how a fictional feature importance plot could be created along three steps: (1) selecting the template code cell; (2) changing and/or adding new values to an array that defined fictional features (e.g., HbA1c); (3) changing and/or adding new values to a numeric array that defined the feature importance (e.g., 80); and (4) running the selected code cell.
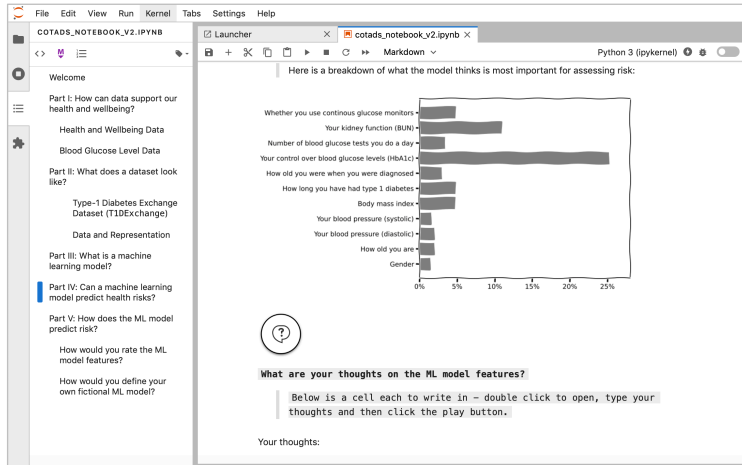
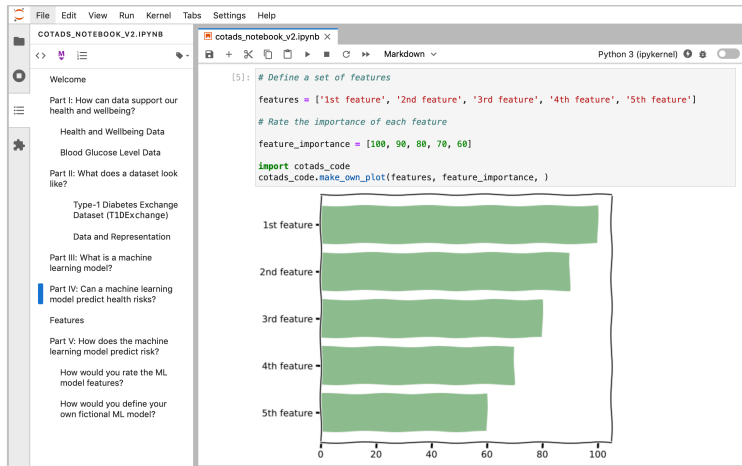**Figure 5:** Part V presents an example ML model and feature importance plot based on the T1D Exchange dataset.

Figure 6: Part VI of the computational notebook supported participants in creating fictional feature importance plots.

3.2 Evaluation of a Computation Notebook

In the following sections, we detail our data collection and analysis process.

3.2.1 Research Question

We aimed to qualitatively evaluate the benefits and limitations of using computational notebooks as co-design tools with anticipated end user groups (i.e., young adults with T1D, family carers, and clinicians). We investigated the following research question: what are the perceived benefits and challenges of using computational notebooks as co-design tools to inform the design of ML models with anticipated end user groups in T1D care?

3.2.2 Data Collection

This research study received institutional ethical approval. We conducted a series of five online workshops with young adults living with type 1 diabetes (n=6), family carers (n=4), and clinicians (n=3) (see Table 1). Family carers were not carers of the young adult participants, but carers of children with T1D.

All participants were proficient in using digital communication technology, including video conferencing systems and slide deck-based presentation applications. However, participants had different levels of prior experience with using computational notebooks. While none of the three clinicians had previously interacted with computational notebooks, they used different clinical systems for reviewing patient data in table and graph-based formats. Two of the six young adults had used computational notebooks while studying computer science, but all young adults relied on different T1D technologies and data visualisations to self-manage their health and wellbeing on daily basis. One of the four family carers worked in the data science and management domain and was familiar with computational notebooks. Similarly to the clinicians and young adults with T1D participant groups, all family carers had lived experience of making sense of data visualisations for managing T1D, in particular blood glucose level measures.

To support participants in building on their shared lived experiences within a homogeneous group of peers, workshops were conducted consecutively in three separate groups; one with young adults living with T1D; one with family members; and one with clinicians. In the young adult workshops, when there were six participants present, the workshops were split into breakout rooms with three participants in each room. If there were fewer than six participants, then the workshop remained as one group. Each of the five online workshops lasted one hour and covered one of the five sections of the computational notebook. In the first workshop for each group, the notebook was introduced, and the leading facilitator used screen sharing to demonstrate the usage of the Jupyter notebook. At the start of subsequent workshops, participants were reminded how to interact with the notebook, particularly the notion of executing a code cell. HCI researchers and data scientists took the roles of workshop facilitators and design coaches.

In addition, we conducted 13 post-workshop interviews, including the young adults (n=6), family carers (n=4), and clinicians (n=3). The interviews focused on people's subjective experiences of taking part in the online workshops and using the computational notebook. We asked open-ended questions, such as: (1) how would you describe your experience of taking part in the workshops?; (2) what was your favourite part?; (2) how would you describe your experience using the computational notebook?; (3) how do you feel about the workshop outcomes considering your expectations? Young adults with T1D and family carers received up to £90 value of shopping vouchers for taking part in the five workshops and follow-up interview.

Table 1: Overview of Research Participants

#	ID	Role	Gender	Age Range	#	ID	Role	Gender	Age Range
1	C1	Clinician	Female	35-44	8	Y1	Young adult	Female	18-24
2	C2	Clinician	Female	45-54	9	Y2	Young adult	Female	18-24
3	C3	Clinician	Male	45-54	10	Y3	Young adult	Male	18-24
4	F1	Family Carer	Female	45-54	11	Y4	Young adult	Male	18-24
5	F2	Family Carer	Female	45-54	12	Y5	Young adult	Female	18-24
6	F3	Family Carer	Female	45-54	13	Y6	Young adult	Female	18-24
7	F4	Family Carer	Male	45-54					

3.2.3 Data Analysis

Data collection and analysis was conducted in a staggered way according to the three participant groups (i.e., young adults with T1D, family members, and clinicians.). Qualitative data analysis software was used to thematically code workshop and interview data [10]. At initial data analysis stages, we focused on the thematic analysis of the interview data, as our research question centred on participants' methodological reflections on using the computational notebooks as part of the workshops. At later stages of the data analysis, interview codes helped not only to develop overarching high-level themes but also select situated workshop data to exemplify and strengthen participants'

reflective accounts. With a focused thematic analysis, we intertwined interview and workshop data and iteratively shared identified themes within the multidisciplinary project team. In the following findings sections, we identify young adults with *"Y"*, family carers with *"F"*, clinicians with *"C"*, followed by the participant identification number, and the suffix *"W"* (i.e., workshop quote) or *"I"* (i.e., post-workshop interview quote).

4   Findings

Based on the analysis of the online workshops and retrospective interviews, we describe how the computational notebooks supported participants in applying a sensitive approach to sharing lived experiences as part of group discussions, iteratively learning about ML concepts, anticipating potential benefits and harms of an example ML model, and collaboratively informing the design of alternative ML models according to personal health and wellbeing needs. We then unpack how the user experience of the computational notebook also caused critical incidents and led to information asymmetries and power imbalances.

4.1 Perceived Benefits of Using Computational Notebooks as a Co-Design Tool

4.1.1   Supporting Individual Ideas and Collective Conversations

The computational notebook enabled participants to formulate individual ideas by working through each of the five sections by themselves, before group discussion occurred and in parallel to group discussion. Although workshop participants were referencing the same content, everyone having their own notebook allowed space for original and uninfluenced thought. Y6 reported enjoying the fact that they had a "*private*" environment "*to test those ideas before sharing*", without a fear of "*judgment*" of the way they interacted with workshop activities (Y6I). When faced with interactive activities, participants had the freedom to input their own ideas, such as personal parameters in the risk score slider, which enriched their personal understanding of an activity. Private use of the notebooks was complemented by the fact that participants had different lived experiences of T1D. Thus, they drew conclusions from an activity as an individual, prior to discussion (*"I thought it was interesting, I put in my parameters on it, and saw what I would have."* (Y1I)). By affording participants more agency in their comprehension, they had a deeper understanding of the concept at hand before discussing with the group. This was a powerful affordance because it approach meant that participants were not bound to the pacing of the group and could take their time with an activity if desired, which would not be the case if, for example, the group was following a slideshow presented by the host (*"there's just something nice about giving each individual the capability of, in their own time, flicking through each day and having to think about it."* (C1I)). A young adult with T1D also identified that there would have been less potential for them to bring their own ideas into an activity had they been sharing one notebook as a group, which would have limited their creativity and ability to explore the activity:

> *"I think there's a potential that people could be overly dominant, if it's just one notebook, [...], I think that there is a possibility that people wouldn't be as creative in trying out ideas." (Y6I)*

Features of the computational notebooks kept conversation on track and provided stimulus to allow fruitful discussion. The visual elements of the notebook provided prompts for the participants to talk about, which was much more effective than discussing a topic without something tangible to compare it to (*"I think having that information there in front of us focused us and focused the discussion."* (C3I)). The static and interactive visualisations were used by participants as "*jumping off points*" (Y2I) to ponder deeper questions and discuss what was not shown in the computational notebook. Y1I pointed out that group discussion gave them time to think deeper on a subject and further develop their own ideas as a result.

> *"I think it was nice to be in a group rather than individual because sometimes you could sit back and properly think about it. [...] the discussion probably brought more out than if you had just asked me on my own, because sometimes what they were saying helped to push me to say more." (Y1I)*

4.1.2 Facilitating a Calm Approach to Sensitive Topics

The workshops approached sensitive and controversial topics surrounding T1D, including: health risk perceptions; roles of metrics, such as BMI and HbA1c; individuals' choice of technology; and the individual nature of diabetes. While many of these topics have stigma attached to them, the computational notebook was appropriated as a ticket-to-talk about a given topic, with control over concealing and revealing personal experiences:

> *"I think [the dataset activity] was a really good way to sort of hammer down the point that everybody's diabetes and their diabetes control is so vastly different, with people using different tech as well. I think it was a really good conversation starter for that."* (Y6I)

The notebook helped to facilitate the discussion of these sensitive topics by clearly displaying information, and then inviting open discussion, rather than railroading the conversation down a specific narrative path. For example, a young adult with T1D challenged the acceptability and significance of the ML feature *"body mass index"* (BMI) and a family carer referred to conversations on how inappropriate representation of population groups can lead to health inequalities as part of the dataset session:

> *"It was really clear that making false assumptions and making decisions, particularly when it comes to medical issues and clinical need, based on a dataset that doesn't represent a particular subset, and applying that decision to that subset is not just unhelpful, but potentially harmful."* (F4I)

Additionally, participants would find familiarity not only in the way that data was presented as charts and tables, but in the content of the data itself, explaining that "*there isn't any embarrassment"* when discussing sensitive topics, but in fact *"there is commonality"* between participants (Y6I). It is a positive that participants felt empowered by the content contained in the notebook and felt comfortable addressing sensitive topics:

> *"I think the way that people opened up individually meant that everybody felt pretty comfortable sharing their different perspectives. I think these three charts definitely facilitated and opened the conversation because of the way it represented things that people recognise from their lives."* (F2I)

4.1.3 Anticipating Potential Benefits and Harms of an Example ML Model

A further benefit of deploying the computational notebook as a design probe was that it supported participants in anticipating potential benefits and harms of the example ML model and the example risk score visualisation. Participants valued the ML model feature importance plot as a visual way to learn how the ML model predicted risk and appreciated the simplicity of a risk score as a boundary object to foster discussion.

Considering that hospital admissions often highlight a need for clinical intervention and that resources are typically limited, clinicians gauged whether the ML model could not only help identify most vulnerable patient groups and conduct targeted interventions but also help allocate resources more efficiently and equitably. If aligned with clinical workflows, the ML model could inform the quality assessment of clinical practice and effectiveness of *"education"* (C1W referring to session that covered self-management). Short-term and longer-term changes in risk score measures could support clinical decision-making between appointments and inform self-care in daily life:

> *"It should be something that can be looked at, uhm, can be looked at by the patient whenever they want, and the clinical teams when they want, but also proactively, you know, flagging changes in the risk or to just highlight people who maybe on the move to somewhere that's not as safe or as pleasant as where they are at the moment."* (C3W)

Participants highlighted the importance of aligning the ML model with life transitions and evolving priorities in daily life. For example, a young adult highlighted their priorities in daily life, such taking part in social life, and explained that a ML model that predicts health risk could help coping with the emotional challenges of managing long-term effects of T1D, if the AI system was personalised and provided customisable health risk alerts:

> *"Most people who are young diabetics they're trying to juggle their social life, their work, life, education. They do think well, one hypo, they might have one three times a week and think: Yeah, it's not ideal, but it's what it is. But it's when you look at it from the long run, that's doing irreversible damage. So, it's kind of preventing these risks in the first place is important, you know. And even the idea of ketoacidosis that is scary. So, to have the peace of mind that you're not going to get to that stage with the machine doing it [predicting health risk], it just takes pressure away off."* (Y4W)

In contrast, participants also anticipated potential harms of the example ML model. Taking a constrictive attitude, participants addressed the limitations of the dataset used to train the example ML model, such as a lack of transferability due to structural and cultural differences between the geographical origin of the dataset and participants' local context. Clinicians highlighted that derived ML model features from the dataset, such as HbA1c, were already integral part of existing clinical practices and that ML model features, such as BUN (kidney function indicator), were not tailored to certain patient groups, such as teenagers and young adults.

All participants felt that the emotive nature of the applied terminology 'risk' could potentially cause emotional distress when presented to young adults and family carers outside clinical settings. People with T1D and their families may feel uncomfortable in sharing their personal health records to inform the ML model, in particular, when personal health predictions and clinical decisions are being made without their ongoing consent and input. The risk score could be *"another number to be judged by"* (C1W) and *"another metric to live up to"* (C3W). People's perception of the risk score might change over time and depend on the accuracy of predictions according to changes in daily life.

Additionally, participants focused attention to mismatches between binary ML model features and their lived experiences. Both, clinicians and young adults made clear that ML model features, such as whether a person has a chronic health condition or whether they use a CGM device, cannot (and should not) define people's health risk and quality of care. Participants highlighted the idiosyncratic nature of T1D to express scepticism of implementing ML models according to the realities of living with T1D (e.g., *"Diabetes is so unpredictable - a ML model most likely won't be able to account for everything!"* (Y5W)). They anticipated that an aggregated health risk score could lead to unintended health behaviour, including adverse data dependencies, over-obsession, and feelings of guilt, shame, and blame. A young adult exemplified this theme by explaining that the *"management of diabetes varies between every person, a model that classifies people as high risk may stress newly diagnosed people into overcompensating with insulin doses"* (Y3W).

4.1.4 Scaffolding Step-by-Step Learning and Multiple Learning Preferences

Due to the structure of the notebook, an environment was created that encouraged a step-by-step approach to learning ML concepts, and how to use the notebook itself. While they were separate learning outcomes, they do appear to be intertwined, as a positive user experience with the notebook seemed to have resulted in a positive understanding of ML concepts being taught. In terms of developing understanding of ML concepts, multiple participants identified that the notebook "*built up into thinking what would your ideal model be*" (F2I), and that this learning process meant that "*people really were engaged in thinking of everything they could that would possibly have an influence*" on their design for the final activity: creating a fictional feature importance plot. Since the final task produced important empirical data, it was vital that participants had a thorough understanding of ML before attempting the activity. Y4 described how the explanations provided by the notebook elements throughout the sessions had supported their understanding of ML applications to diabetes care. They reported that completing the previous activities was important, in order to consolidate their understanding, before moving on to the final activity:

> *"I think it built up quite nicely. [The explanations] helped me to think about what sort of inputs would go into [the model], and how it was important to ensure that that was right before starting out with a big model that could be clinically used. I would have got used to being thrown straight into machine learning, but I guess it was useful to start more basic."* (Y4I)

The other aspect of step-by-step learning was building confidence in the use of Jupyter Notebook itself. For those who had no prior knowledge of using a computational notebook, it was important to teach the participants how to interact with the notebook and internalise the process of running code cells first, to not overwhelm them when introducing the more confounding coding task. F4 explained that their first impression of the notebook was total unfamiliarity, and that the notebook appeared "*alien*" to them. They reinforced the finding that introducing the coding task too early would have overwhelmed participants, as it would have been too many new things to do in one session, and they appreciated the "*gradual*" approach to using the notebook.

Participants reported that they enjoyed the fact that the notebook catered to multiple 'learning styles'. Despite this phrase not being part of the interview guide, multiple participants used the term "*visual learner*" (Y3, F3) to describe themselves or others, while other participants suggested ideas of "*implicit*" (Y5, Y6) and "*explicit*" (Y5, Y6) learning. Y5 identified that the feature importance plot "*wasn't interactive and it was a bit more explicit rather than implied*", unlike

the risk score slider which facilitated implicit learning due to its interactive nature. Y5, who had prior knowledge of ML as a computer science student, expressed that they preferred the explicit feature importance plot because "*it was a bit more in depth at showing that these models can be made of plenty of different features with various weightings*" (Y5I).

It became clear that it was important to participants to have a blend of both implicit and explicit coaching, because the learning styles complement each other and allow for a broad, but also deep understanding. It was important for them to first establish a baseline understanding through implicit learning, before explicitly "*digging into it*" (Y6I) by looking at more detailed examples that pushed the boundary of the participant's understanding. As someone without prior knowledge of ML, Y6 explained how the interactive, implicit activities provided necessary foundational knowledge, before that knowledge could be expanded on by studying more explicit information: "*Having the implicit first to understand how it works and then digging into it was the right approach for me.*" (Y6I).

C3 praised the way that the notebook incorporated and catered towards these different ways of digesting information, "*I think different ways of putting information together suits different people differently*" (C3I). Y3 identified themselves as a visual learner and described how the visual information combined with the aural aspect of discussion, supported them in learning about how ML models learn how to make predictions:

> "*[The notebook elements] obviously gave me visuals, which I said I learned by [...]. I found the combination of using the notebook, and then speaking with the team and obviously the [researchers] who were leading the sessions, the combination of the two helped me fully understand it and then I was able to give the best answers I could.*" (Y3I)

Even though F3 identified themselves as a "*reading person*", they still stressed the importance of including visual representations to support understanding:

> "*Without anything visual - and I'm not a visual learner, in all honesty, I'm a reading person - if you'd [explained] to me a machine learning model without those things, I would have been [sic] "What on earth are you talking about?" It would have made literally no sense.*" (F3I)

Interactive visualisations seemed to fulfil the criteria for hands-on, experimental learning, and participants were vocal about how the interactivity of the notebook was the standout feature; the aspect that elevated it when comparing computational notebooks to other presentation tools, such as a PowerPoint slideshow ("*The interactive models where we could enter parameters and see the changes ourselves – I thought that was great.*" (F2I)). When describing the interactive risk score slider, F2 enthused, "*I'm a sucker for things you can slide around and see the impact of, it's almost like gaming, isn't it, which I think is one of the engaging elements of using the notebook approach*" (F2I).

4.1.5 Harnessing of Coding by Non-Data Science Experts

Presenting the computational notebook, a technical tool, to non-data science experts was understandably difficult and met with apprehension. Participants, such as F4, who had no prior experience with coding, described their first impressions of the notebook as "*daunting*" (F4I), and F1 expressed, "*when I opened it for the first time, I was maybe worried a little bit and not sure how I was going to cope with it*" (F1I.) However, with guidance from the workshop facilitators, and by approaching the notebook slowly and with step-by-step principles, participants who were initially put off were still confident in giving the notebook their best attempt. A young adult described the importance of having a coach to help when coding in the following way: "*it was really good that we had a supervisor in to nudge us in the right direction if we needed help with it*" (Y6I).

F3 described how they struggled, at first, to assign numerical weightings to features because of their fictional nature. Upon initially coding values and generating a graph, F3's understanding was rapidly developed by the visual representation of their figures; they could now visualise how the model's features were weighted, relative to each other, and iteratively adjust values until it looked right to them: "*When I initially did the features of importance, I remember thinking all these are just pie in the sky figures, not really sure, and I went back and found I could change them quite easily and quite quickly as well*" (F3I). While they noted that the coding was not necessarily the most intuitive way of creating a graph, F4 also described the benefit of the instant feedback from changing the code:

> "*It's the visualization of it that when you type it in it, it shows the impact of changing the value immediately. Whereas I think other ways are more intuitive in terms of, yeah, we all know how to fill in a form or a Google doc or whatever, but then you wouldn't necessarily see what the graph then looks like until you have done it all, and then you'd have to go back and start all over again.*" (F4I)

F4 reiterated that by the time the coding task came around, they found the task "*relatively straightforward*" and although they did make mistakes while creating code, they persevered and managed to succeed due to the gradual approach to the task:

"*I think if this had come further up in the workshops and it was in the first or second [session], I think I'd have really struggled. It was better being in the last one and it was a little daunting, but relatively straightforward to do. I think I got tripped up and I don't know if I was the only one tripped up by adding in features and getting caught out by not putting commas or quotation marks in the right places.*" (F4I)

A positive finding was the reactions of participants who had never coded before, after succeeding in the task. Upon their first impression of the notebook, F3 noted frustration and despair at the perceived technicality of the environment, "*when I first saw it, I thought 'Oh hell, I'm never going to get to grips with this,' because it looks very scientific to me and a bit geeky, computer science-y, and nerdy*" (F3I). However, when eventually tasked with writing code, F3 discovered that the task was not as difficult as they first thought, "*I was just following the steps and when I tried it the first time I thought, 'oh I can do it, that's good!'*" F3 went on to report that the coding task ended up being their favourite activity of all the workshops, and that they gained satisfaction by being able to code successfully.

"*I actually quite liked doing that coding thing when it came down to it. I thought I was going to hate it, but I quite liked the interactivity of it [...] I had quite a little bit of a sense of satisfaction about being able to do it. I know it's ridiculous, but I was like, 'Oh yeah, this is easy actually.'*" (F3I)

Y6, who had no prior coding experience, also reported similar feelings of empowerment and jubilation when successfully coding. Their success with their coding enabled them to produce the most comprehensive feature importance graph out of all participants, with fourteen features in total. Because the example code only had placeholders for five features, this demonstrates that they had an adept understanding of altering code and adding elements to an array:

"*For me I had a couple of issues at the start, because I wanted to do more factors compared to what was initially put into the code, but I felt like I did alright with it, and as I got into it, I picked it up pretty quickly with how to stick the right things in. [...] [I felt] really, really proud of myself because I'm not techy. I can work a computer - IT is absolutely fine - but coding is something that I never thought I'd be able to do. Because I don't understand it, [...] it is quite intimidating. Especially with a big lack of women in STEM, sometimes it's hard to pique that interest of, 'actually, it's not as intimidating as you think it is, you can do it.'*" (Y6I)

4.1.6   Collaboratively Informing the Design of Alternative ML Models

Creating a fictional feature importance plot was a participatory and collaborative learning experience (see Figure 7). Participants were positive in describing how the group discussion about the notebook content furthered their own understanding of the task, and how sharing ideas helped to inform their own activities. For example, Y3 described how they were inspired by other participants' feature importance plots, and how they used other people's ideas to improve their own fictional ML model features:

"*I thought it was really insightful to see other people's priorities, because sometimes people have very different priorities in terms of their condition [...]. It helped me fill out my graph a bit more - there was things I didn't even think I wanted, and I was like, 'oh, that's interesting, I'll put that in mine.'*" (Y3I)

Participants were reassured that they were on the right track with their activity when sharing their ideas and outcomes within their group. Since their child had only been diagnosed with diabetes for a year at the time of the workshops, F3 reported concerns that they initially felt that they didn't have the knowledge to complete a feature importance plot. They felt that they were missing something that other family carers, "*whose kids had been diagnosed like five or seven years,*" would have, and that the other participants were "*in a different place and they have a better understanding of [diabetes]*". Thus, when the participants were sharing their feature importance plot, and F3 found that "*despite so many different levels of experience, people were pulling out the same things,*" they were reassured that their opinion, experience, and position as a participant in the workshops, was justified.

Receiving inspiration from workshop peers, participants used their fictional ML model plots to (1) articulate and share their personal health and wellbeing needs; (2) highlight the limitations of current medical devices and consumer health technologies; and (3) describe desirable functionalities of future AI-driven systems for T1D management. Most participants did not take over the features of the example ML model: one young adult, one family carer, and two clinicians reused two of the seven features of the example ML model, namely HbA1c and gender.
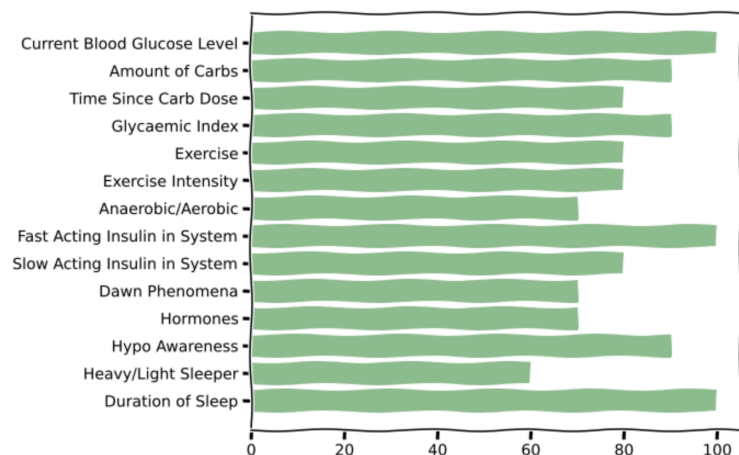
Figure 7: Fictional feature importance plots created by a young adult (Y1).

Rather than adopting the example ML model, participants defined a wide range of holistic ML model features (see Table 2), from blood glucose levels, insulin measures, and adverse health events (i.e., hypos) to diet, physical activity, sleep, stress, and hormonal changes. Participants' fictional ML model features showcase different writing styles, including conversational (e.g., Y6W: *"any hypos during day?"*), technology savvy (e.g., F4W: *"External temp: auto-calc"*), and clinical (e.g., C1W: *"presence of complications"*). Importantly, these ML model features address potential tensions between automation and people's agency, as shown by C1 who highlighted clinician's roles in making decisions (e.g., C1W: *"clinical concerns"*).
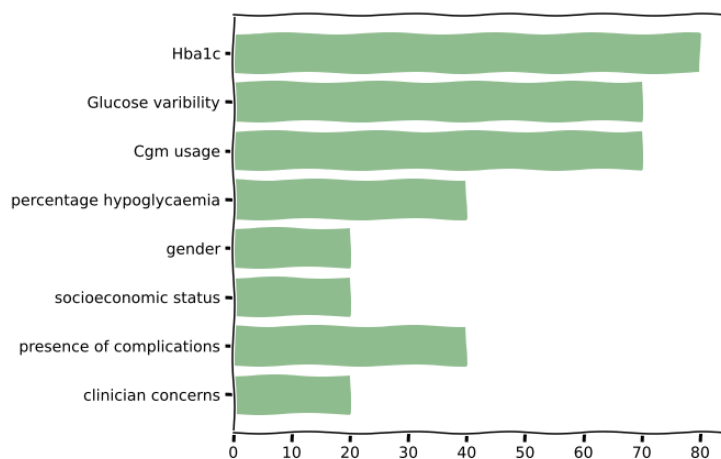

Figure 8: Fictional feature importance plots created by a clinician (C1).

The fictional ML model feature importance plots display participant's creativity in appropriating the activity in personally meaningful ways. They defined not only feasible ML model features but also articulated the need for specific functionalities of T1D technology, such as a reminder to inject (Y2), highlighted important considerations for conducting digital health inventions in real world setting, such as addressing stigma in school setting (F1), and envisioned futuristic data collection technology, such as under-skin monitors to track blood glucose direction and speed (F4).

Moreover, participants suggested alternative directions for implementing agency and wellbeing supportive AI systems to *"get on with life"* (F3W), including predicting time in range (C3), recognising patterns in long-term CGM data (C1), predicting hypoglycaemia and hyperglycaemia (Y4), recommending suitable injection sites (Y1), recommending intensity of physical exercises (Y4), assessing foot health (Y1), and predicting adverse mental health states in family carers (F2).

Table 2: **Fictional ML model features created by young adults, family carers, and clinicians**

1. Blood Glucose Levels	2. Blood Glucose Measures	3. Blood Glucose Monitoring
Blood glucose level (Y1, Y2, Y3, Y4)	HbA1c (Y3, F2, C1)	CGM (F2, F3, C1, C2)
Time of blood sugar levels (Y4)	Effect on HbA1c (A5)	CGM est HbA1c for last 24 hrs (C2)
Levels over night (Y5)	Time in range (F1, C2)	Blood glucose direction and speed: under
BG fluctuations (F2)	Glucose variability (C1)	skin monitor (F4)
	Std deviation (F2)	External temp: auto-calc (F4)
	Percentage of low BG (F2)	Mechanic failure alert and auto-
	Percentage of high BG (F2)	replacement (F4)
	Percentage of very high BG (F2)	Pump site (N/A because internal) (F4)
4. Insulin	**5. Hypo/Hyper**	**6. Diet**
Insulin taken (Y4, Y6, F1, F3)	Any hypos during day? (Y6)	Amount of carbs (Y1, Y4. F3, P4)
Time of dosages (Y4)	Percentage hypoglycaemia (C1)	Carbs consumed (and type) (Y6)
Short-term insulin (Y1, Y2, Y3)	Predicted hypo times (Y2)	Time since carb dose (A1)
Long-term insulin (Y1, Y2, Y3)	Chance of hypo based on daily	GI of carbs: auto-calc (F4)
Type of short-term insulin (Y3)	activities/events (Y5)	Carb=insulin calculations (Y2)
Type of long-term insulin (Y3)	Change of hypers based on daily	Accuracy in carb counting (F1)
Reminder to inject (Y2)	activities/events (Y5)	Glycaemic index (Y1, F3)
Age of infusion set (C2)	Hypo awareness in sleep (F1)	Carb counting capacity (known carbs +
Awareness if cannula falling off (F1)	Dawn phenomena (Y1)	education) (C2)
All factors that the pancreas uses to	Hypo awareness (Y1)	Ability to adjust (Y2) Gluten checker (F1)
manage insulin and BG control (F4)	Hypo/hyper alert (Y2)	
7. Physical Activity	**8. Sleep**	**10. Medical History**
Exercise (Y1, Y2, Y3, P1)	Heavy/light sleeper (Y1)	Presence of complications (C1)
Expected exercise level (C2)	Duration of sleep (Y1)	Clinician concerns (C1)
Exercise type (Y6)		Effect of illness (Y5)
Physical activity duration (Y4, Y6)	**9. Stress**	
Physical activity intensities (Y1, Y4)	Stress levels (adrenaline) (F3)	**11. Demographics**
Anaerobic/aerobic (Y1)	Adrenaline levels under skin monitor	Gender (C1)
Blood glucose at start (Y6)	(F4)	Socioeconomic status (C1)
Blood glucose at end (Y6)	Previous day emotional rating (C2)	
Activity level from Fitbit (F3)	Expected stress levels (C2)	**12. Other**
Physical activity: under skin device (F4)		Hormones (Y1, Y2, F4)
		Explaining the differences between type 1
		and 2! (F1)

4.2 Perceived Challenges of Using a Computational Notebooks as Co-Design Tool

4.2.1   Technical User Experience Can Contribute to Power Imbalances

The user interface of the computational notebook was unfamiliar for many participants and it did not meet their expectations for a user-friendly, aesthetically pleasing tool. F3 criticised the visual design of the notebook, explaining that the layout generated by the markdown file was not "*very visually appealing*" and that "*it kind of looked like a textbook*" (F3I). They explained how this 'textbook' was off putting to non-technical users: "*Textbooks scare me a bit because I'm not a maths or science person!*" (F3I) Although all participants proactively shared their experiences and created their own fictional feature importance plot, it was difficult for, particularly, two family carers, to interact with the computational notebook because of its technical user experience. For example, F2 explained that the notebook could be perceived as exclusionary:

> "*If you want to be inclusive, make it easy to consume. The best phone apps that we have are the simple ones, all the hard work is done behind the scenes, and the participants just have to focus on the question that's been asked, rather than how they interact with the notebook itself.*" (F2I)

Difficulty with aspects of the software, such as navigating the document and executing code cells, made participants initially feel like they had not the required expertise and agency to engage with the computational notebook. F3 explained how the presentation of the environment, such as the code in cells, made them "*feel like a complete fish out of water with this notebook because of this coding that I can see that I don't understand*"(F3I). The code in the cells had been abstracted away as much as possible, it was still unfamiliar to the participant to the extent that it had a negative effect on their experience, and consequently, their understanding. F3 continued, "*If [the code] was taken out I'd have probably quite a different feeling about it,*" and went on to discuss how the design of the notebook didn't seem to have much consideration for non-data science experts:

> "*It looked very scientific and that did put me off, if I'm honest. I just thought some of the things that were on there didn't really make sense to me, and it felt like there hadn't been any consideration for the way that some of the terminology might make the user feel.*" (F3I)

Furthermore, participants struggled with running code cells in the notebook. The notebook had an issue where code cells would not execute on first load, and the notebook kernel required restarting. Young adults who had no prior experience with computational notebooks demonstrated fluency with the notebook by restarting the bugged kernel without problems. However, in the post-workshop interviews, family carers (F1, F3, F4) lacked the confidence to fix the issue on their own and required guidance to restart the kernel.

By not offering a similar user experience as typical digital tools, such as a shared document or whiteboard, the notebook environment strengthened data explanation, exploration, and visualisation affordances. However, this came at the cost of the participants' abilities to digest and understand the information. F4 discussed how the computational notebook was harder to use than other collaborative tools that they had used before, because of the technical nature:

> "*It was very different to anything I'd used previously. For someone who doesn't have any kind of coding or particular IT expertise, it did feel more technical than a lot of the other ways in which you can get people to share comments. I've been to other online workshops that are more sort of focus group-like where there are virtual post it boards and that kind of thing, which is easier to get to grips with - the computational workbook felt far more technical.*" (F4I)

4.2.2   Ambiguity Can Lead to Information Asymmetries

The notebook sometimes failed to communicate the intentions of activities. This miscommunication resulted in breakdowns in different ways, such as participants not taking away potential learning outcomes from activities, not understanding the information being presented in an activity, and not performing an activity as intended. Whilst most of the notebook was successful at communicating its intentions, these occasional breakdowns must be reported on.

The notebook failed to communicate the context of the risk score slider activity and explain that it was an example. During a conversation with Y6 about the risk score slider, they reported:

"*I was so frustrated with [the risk score slider] because it's just not representative of diabetes management. You could be testing six times a day, but if you don't act upon it and don't inject, you're still at a high risk of going into hospital. There needed to be three different outcomes: hospitalisation for hypoglycaemia, hyperglycaemia, and DKA*" (Y6I).

Y6 had missed, understandably so, a bullet point in a list of text above the risk score slider that said, '*The model shown is an example of what can be done and does not represent diabetes education or advice.*' By missing this one line in the notebook, Y6 became disillusioned with the workshop activity. For Y6, the context of this activity completely changed from foundational learning about the concept of ML, to frustration at the perceived simplicity of the model and lack of awareness from the designers (*"I might have not picked up that this wasn't a fully-fledged thing at first, so I think that's on me."* (Y6I)). Y5 also expressed how they were unsure if the risk score slider was an example: *"I wasn't really sure if this was a genuine model [...], so I was slightly concerned because I'm not sure how good this model really is. I feel like just three features and a value between 0 and 1 wouldn't actually produce any meaningful results."* (Y5I).

The explorative and open-ended nature of design probes presented challenges for participants in sharing their experiences and taking part in design activities. For example, Y1 needed the notebook to go further when explaining what contribution it expected from them, to feel encouraged to participate fully:

"*I would have been able to contribute more with a greater understanding of the project and where it was trying to go. I wouldn't say that I held loads back, but sometimes I felt like what we were discussing was a bit vague. I didn't really know what to say and what not to say.*" (Y1I)

4.2.3 'One Size Fits All' Approach to Explaining ML Does Not Meet Individual Information Needs

The notebook provided different forms of ML explanations, and while some participants were satisfied with the amount of detail that the notebook went into, a handful of participants expressed interest in more complex explanations. C3, an advocate for the given level of detail, reported: "*I think anything more perhaps would have been a little bit too much, I think it might have lost my focus from the information that I was there to provide*" (C3I). Y1 recognised that there was more to be said on the topic of ML, but deemed that the notebook had struck a healthy balance of detail and simplification:

"*There's obviously a lot more that goes on, but I don't know what that is and how you'd visualise that, so for me personally, that's a decent level of detail and explanation, but also a good level of simplification because I feel like if you put too much maths in it, I wouldn't understand.*" (Y1I)

While our intention was to keep the notebook elements simple to facilitate understanding in all users, F2 thought that making the content of the risk score slider more complex, by adding more variables, would be insightful: "*With just the three dials, it is pretty easy to get to the bottom of what you're trying to say. Would I have got more or less insight out of having more or less of these [variables] to choose from? I'd have probably gone for more just to see.*" (F2I)

Y5 observed that the simplicity of the examples could be counter-intuitive and create a narrative that ML was not capable of handling more than a few parameters, because none of the early examples incorporated more than three variable inputs. This would cause participants without prior understanding of ML to underestimate its power. Y5 suggested that some examples could afford to be more complex:

"*In the explanation of a model, this training and testing model is very simple as well. Because this isn't very interactive, [...] it can be a bit more of a complicated image of a model to show the extent of them; rather than 'OK, these models are always just taking in one, two or three features, and any more than that, it might not be able to handle.' I feel like people who wouldn't be comfortable with machine learning wouldn't know [otherwise] if they haven't touched it at all before. I would assume it's quite obvious, but it was only because all the examples in the notebook are very simple, that I wasn't really sure how simple the actual models potentially were.*" (Y5I)

While most of the participants who were eager for more detailed content had prior experience with ML, some participants without prior experience were still open to deeper ML explanations. When asked if they thought the explanations given in the workshop were sufficient, Y4 recognised the benefit of more information:

"*I think in the workshop, perhaps because it's not like a fully-fledged finalized model that's going to be harder to do and I understand that, but I think maybe a little bit more information is a starting*

point. Just the basics of how these things are integrated would be useful, but I guess without a complete understanding of how a model is going to work, you can't really give a lot of information." (Y4I)

5 Discussion

Recent human-centred calls have highlighted the importance of finding new ways to support participation in AI design with stakeholders from diverse backgrounds [25,26,59]. Empowering different stakeholder groups to inform the design of AI/ML systems has significant potential to avoid adverse and unintended consequences, from inaccessible user interface design to racial discrimination and health and safety hazards [ibid]. Prior work has demonstrated the benefits of involving stakeholder groups in the design of predictive systems in different contexts, including child welfare [19,82], mental health [87,88], and clinical decision making [13,74]. However, it remains unclear what and how research tools can be effectively used to engage people from non-data science backgrounds in the design of ML models. We presented a qualitative account of the perceived benefits and challenges of using computational notebooks as co-design tools to inform the design of ML-based systems for T1D management. In the following sections, we discuss implications for facilitating participation in AI/ML design and developing appropriate co-design tools within this context.

5.1 Exploring the Feasibility of Using Computational Notebooks as Co-Design Tools

We explored the feasibility of deploying computational notebooks as co-design tools. Computational notebooks are increasingly being adopted not only to perform data science tasks at work [80,98] but also to conduct data science courses within educational setting [46,93]. However, it is unclear how non-data science domain experts, such as clinicians, carers and people living with chronic health conditions, experience computational notebooks as part of participatory research aimed at informing the design of human-centred AI algorithms. As part of this work, we conducted a series of online workshops through JupyterLab hosted computational notebooks with young adults with T1D, family carers, and clinicians.

Prior work documented the challenges of designing AI systems, including methodological discrepancies between data science and user experience design workflows, such as time-intensive ML model development processes and ambiguity on user interface levels at early project phases [31,95]. We found that drawing on an existing dataset to implement an example ML model and sharing ML model outputs with computational notebooks can provide a time sensitive approach to intertwining ML model development and user experience design workstreams. In particular, when project priorities shift from optimising the accuracy of ML models to iteratively sharing and discussing ML model explanations (such as feature importance plots) at early ML model lifecycles within multidisciplinary teams. Reusing an existing dataset to implement a ML model (1) supported a relatively rapid approach to human-centred and iterative design processes, (2) provided realistic design material for developing a computational notebook with static and interactive ML model explanations, and (3) fostered discussion of feasible, plausible, and futuristic directions for ML-based systems (see Table 2).

Participants with prior experience of using computational notebooks were fluent in completing workshop activities and proactive in sharing their knowledge and skills with their workshop peers who had never used computational notebooks before. Participants with no prior experience with computational notebooks described initial barriers to learning how to use the Jupyter environment, in particular family carers from non-data science backgrounds. However, all participants managed to create a fictional feature importance plot. A key finding was that participants who learned how to use computational notebooks as part of the workshop series reported a sense of empowerment, being pleased and proud to have created a fictional feature importance plot according to their personal health and wellbeing values.

However, reusing an existing dataset to implement an example ML model to inform the development of a co-design tools presented challenges as well. For example, it became clear that an existing dataset can significantly predefine the scope of ML model predictions, such as in our case risk from hospitalisations due to adverse health events. This

technology centric dimension of dataset-based co-design tools can be at odds with bottom-up approaches, such as ethnographic design work [9,24] and creative approaches, such as exploring ML through *"monster"* metaphors [30].

Although we attempted to use lay terminology, reduce code visibility, and scaffold learning through making time for questions during online workshops, some participants described the user interface of JupyterLab as overwhelming and experienced critical incidents when navigating their notebooks and running code cells. These experiences highlight not only the importance of providing personalised explanations and coaching but also present promising directions for developing novel extensions, such as customisable user interfaces to support administrators in reducing the perceived complexity of available functionalities. In addition, our findings suggest supporting seamless interactions with digital design applications for sketching and presenting artefacts to support more accessible and collaborative experiences.

5.2 Collaboratively Aligning ML Models with Lived Experiences

Using computational notebooks as co-design tools helped to understand the lived care experiences and health risk prediction needs of young adults with T1D, family carers, and clinicians. We drew in an example ML model that predicted risk from hospitalisations and an example risk score to elicit people's personal and shared views. Prior work has documented the viability of risk stratification within clinical settings [21,67]. In our case, all participants groups identified, both, potential benefits and harms of the example ML model and the associated risk score. On the one hand, participants anticipated benefits, such as managing clinical resources more effectively and equitably. On the other hand, participants highlighted potential harmful consequences of applying the ML model, from fostering emotional distress and obsessive self-care, to amplifying human bias and health inequalities. Importantly, participants' accounts revealed tensions between their lived experiences, biomedical views, and ML model features. Personal data are commonly collected to predict the diagnosis and prognosis of health and wellbeing conditions. However, our findings highlight that using certain scientifically constructed and normative measures, such as BMI, to predict health outcomes might be perceived as unacceptable from the points of view of those who live with chronic health conditions. We argue that anticipating benefits and harms of ML models with relevant stakeholders is a key strength of utilising computational notebooks as co-design tools. In this way, multidisciplinary teams can critically reflect on the purpose and features of ML models to iteratively avoid cascading issues [81] and pursue human-centred AI innovation [58].

While prior work on developing ML models for diabetes care paid much attention to the feasibility and optimisation of ML models [66], our work exemplified a human-centred and participatory approach to eliciting the lived experiences and health risk prediction needs of anticipated end user groups. For example, the first part of the computational notebook helped gaining empathy for diabetes care [73], life transitions [78], and roles of social circles [11], including family carers, teachers, and sport coaches.

Similar to storywriters' needs to remain in control over AI writing companions [8], clinicians highlighted the importance of being able to enact their agency in making diagnostic decisions with predictive risk stratification systems. Participants' fictional feature importance plots provided inspiration for alternative directions by exhibiting a holistic array of feasible and futuristic ML model features, from biomedical phenomena (e.g., hypos), to daily experiences (e.g., stress) and bodily transitions (e.g., hormonal changes). These accounts highlight the importance of co-designing datasets that are not only clinically viable but also personally meaningful to people with T1D and their social circles. A promising direction is to explore personalised approaches to implementing ML models [38] by combining data from manual self-tracking tools [3], CGM devices [48], and consumer health technology [33].

5.3 Informing the Design of Accessible Co-Design Tools for AI/ML Development

Our second research objective was to derive methodological implications for using interactive user interfaces, such as computational notebooks, as co-design tools to facilitate human-centred and participatory AI/ML development with anticipated end user groups from non-data science backgrounds. There are different use cases for adapting and adopting computational notebooks to inform human-centred ML, from sharing data science outcomes ad-hoc within

multi-functional teams in industry settings, to using computational notebooks to foster shared decision making, mutual learning, and collective creativity as part of longer-term academic projects. While a ML model can be implemented relatively quickly based on an existing dataset, the development and deployment of human-centred ML-based systems in real-world settings is a collective and time-extensive endeavour. Our findings exemplify one approach to using computational notebooks as co-design tools to explain ML concepts and support participants in expressing their needs in interactive ways.

We drew on an incremental and holistic explanation approach in designing a computational notebook with static illustrations and interactive activities to help researchers in explaining an example ML model and participants in (1) understanding the example ML model; (2) sharing their views and experiences; and (3) creating fictional ML model feature importance plot to express their needs. Our work exemplifies how computational notebooks can be adopted in illustrative ways to support multidisciplinary learning and foster discussion at early ML model lifecycles. While we did not systematically evaluate learning, participants alluded to learning preferences according to Neil Fleming's VARK model [34,35]. VARK stands for *'Visual, Aural, Reading/Writing, Kinaesthetic'*. Computational notebooks are particularly suitable to blend these modalities: (1) visual; through illustrative explanations; (2) aural; from the workshop coaches presenting the notebook, and from group discussion of the content; (3) reading; through the text contained within the notebook and writing through editable cells; and (4) kinaesthetic; from interactive *'learn by doing'* data visualisations. By combining different modalities, computational notebooks provide significant potential to support non-data science experts in learning about ML concepts and engaging in data science work. A promising research direction is to investigate in which ways different approaches to explaining ML concepts (e.g., example-based or no ML concept explanations at all) can influence co-design outcomes (e.g., usefulness of co-design artifacts in meeting the needs of different stakeholders, such as supporting mutual learning, fostering trust, and informing ML model development).

In particular, our work exemplifies how computational notebooks can be used to support participatory engagement: all participants particularly valued the interactive elements of the notebook and participatory coding activity that empowered them to create their own fictional ML feature importance plot to share their personal needs and care priorities. These positive experiences could be reinforced, for example, by participatory pair-programming activities based on participants' personal health and wellbeing data.

Furthermore, it is useful to look at prior work on the use of computational notebooks in classroom settings. Scholars derived best practice guidance on using computational notebooks to deliver data science education specifically for students from scientific, technological, engineering, and mathematical (STEM) backgrounds [46]. Our work involved primarily participants from non-STEM domains across different age groups and professional backgrounds. An implication from our work for using computational notebooks in higher education is to leverage co-design as a human-centred and participatory approach to fostering mutual learning and collective creatively between student groups, rather than delivering content in traditional modes, including unidirectional lectures and demonstrations. Furthermore, it would be plausible to offer accessible and participatory professional development courses based on computational notebooks for knowledge workers from non-data science backgrounds who work on AI/ML in multidisciplinary teams, such as UX designers and researchers. An important implication from higher education practice for using computational notebooks as co-design tools as part of research projects is to leverage formative and summative assessment approaches to evaluate learning outcomes in systematic ways. While we applied an iterative approach to explaining ML concepts and viewed the final activity of creating fictional feature importance plot as the main learning outcome indicator, future work could draw on summative assessment approaches, such as quizzes, surveys, and reflective writing assignments, to evaluate learning outcomes of different explanation approaches across diverse participant groups.

However, we need to highlight that choosing a data science tool, such as JupyterLab, in research and educational settings can also lead to disempowering experiences, in particular, for those using a computational notebook as a non-data scientist for the first time. A promising research direction is to intertwine participatory design [44,52,70] with human-centred ML approaches [7,38,50] to support inclusive experiences where participants from non-data science backgrounds can feel and act as equal co-designers throughout the implementation of personally meaningful ML models. Future design projects could draw on data-enabled design [55,61,71] as a guiding framework to leverage the

design of interactive co-design tools based on existing datasets. Alternatively, future projects could take bottom-up approaches to engage with end users and stakeholders to co-design novel datasets, potentially helping to inform the design of ethically-aligned AI algorithms [58,60,70]. In this ways, computational notebooks could be adopted and adapted as new interfaces for participation in research environments with participatory design and co-design being routes to supporting human-centred principles, from mutual learning and trust to shared decision-making and fairness.

6   Limitations

We have explored the feasibility of using computational notebooks as co-design tools with young adults (n=6), family carers (n=4), and clinicians (n=3). The participant numbers might be considered a limitation. However, we have conducted 15 workshops and 13 post-workshop interviews. Through a series of five workshops with three participants groups we focused on gaining an in-depth understanding of the experience and needs of each individual at the early stages of this research project rather than systematically analysing outcome measures to establish the generalisability of findings. While our thematic analysis draws on retrospective interview and situated workshop data, findings on how the computational notebook supported multidisciplinary learning would be stronger with quantitative measures. Furthermore, it needs to be acknowledged that this work was potentially leading as we presented an example ML model and risk score to elicit people's health risk prediction needs and inform the design of predictive T1D technology. Mixing participants from each participant group and envisioning potential user interfaces could have helped to explore alternative directions for the design of T1D technology and services. However, the strength of our qualitative account is that it identified the perceived benefits and challenges of using computational notebooks as co-design tools with end user groups to inform design of human-centred ML models for T1D care.

7   Conclusion

A lack of appropriate domain expertise and end user feedback can adversely affect ML model lifecycles right from the onset [81]. We presented a qualitative account based on a series of online workshops and post-workshop interviews with young adults with T1D, family carers, and clinicians to exemplify how computational notebooks can be used as co-design tools at early ML development stages to (1) explain ML concepts with static and interactive data visualisations; (2) foster discussion on potential benefits and harms of an ML models with anticipated end user groups and support multidisciplinary learning; (3) and collaboratively inform the design of human-centred ML models according to people's health and care needs. A promising research direction is to design novel multidisciplinary research support tools to bridge human-centred ML and co-design processes to inform ethically aligned and socially meaningful AI innovation with anticipated end user groups.

References

<bib id="bib1"><number>[1]</number>Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, ACM Press, Montreal QC, Canada, 1–18. DOI:https://doi.org/10.1145/3173574.3174156</bib>

<bib id="bib2"><number>[2]</number>American Association of Diabetes Education. 2009. AADE Guidelines for the Practice of Diabetes Self-Management Education and Training (DSME/T). *Diabetes Educ* 35, 3_suppl (November 2009), 85S-107S. DOI:https://doi.org/10.1177/0145721709352436</bib>

<bib id="bib3"><number>[3]</number>Amid Ayobi, Tobias Sonne, Paul Marshall, and Anna L. Cox. 2018. Flexible and Mindful Self-Tracking: Design Implications from Paper Bullet Journals. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), Association for Computing Machinery, Montreal QC, Canada, 1–14. DOI:https://doi.org/10.1145/3173574.3173602</bib>

<bib id="bib4"><number>[4]</number>Amid Ayobi, Katarzyna Stawarz, Dmitri Katz, Paul Marshall, Taku Yamagata, Raúl Santos-Rodríguez, Peter Flach, and Aisling Ann O'Kane. 2021. Machine Learning Explanations as Boundary Objects: How AI Researchers Explain and Non-Experts Perceive Machine Learning. Joint Proceedings of the ACM IUI 2021 Workshops, April 13-17, 2021, College Station, USA. DOI:http://ceur-ws.org/Vol-2903/IUI21WS-TExSS-3.pdf</bib>

[5] Amid Ayobi, Katarzyna Stawarz, Dmitri Katz, Paul Marshall, Taku Yamagata, Raul Santos-Rodriguez, Peter Flach, and Aisling Ann O'Kane. 2021. Co-Designing Personal Health? Multidisciplinary Benefits and Challenges in Informing Diabetes Self-Care Technologies. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (October 2021), 1–26. DOI:https://doi.org/10.1145/3479601

[6] Pascal Béguin. 2003. Design as a mutual learning process between users and designers. *Interacting with Computers* 15, 5 (October 2003), 709–730. DOI:https://doi.org/10.1016/S0953-5438(03)00060-2

[7] Francisco Bernardo, Michael Zbyszynski, Rebecca Fiebrink, and Mick Grierson. 2016. Interactive machine learning for end-user innovation. American Association for Artificial Intelligence (AAAI).

[8] Oloff C. Biermann, Ning F. Ma, and Dongwook Yoon. 2022. From Tool to Companion: Storywriters Want AI Writers to Respect Their Personal Values and Writing Strategies. In *Designing Interactive Systems Conference*, ACM, Virtual Event Australia, 1209–1227. DOI:https://doi.org/10.1145/3532106.3533506

[9] Jeanette Blomberg and Mark Burrell. 2009. An ethnographic approach to design. *Human-computer interaction: Development process* (2009), 71–94.

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (January 2006), 77–101. DOI:https://doi.org/10.1191/1478088706qp063oa

[11] Nicola Brew-Sam, Madhur Chhabra, Anne Parkinson, Kristal Hannan, Ellen Brown, Lachlan Pedley, Karen Brown, Kristine Wright, Elizabeth Pedley, Christopher J Nolan, Christine Phillips, Hanna Suominen, Antonio Tricoli, and Jane Desborough. 2021. Experiences of Young People and Their Caregivers of Using Technology to Manage Type 1 Diabetes Mellitus: Systematic Literature Review and Narrative Synthesis. *JMIR Diabetes* 6, 1 (February 2021), e20973. DOI:https://doi.org/10.2196/20973

[12] B. Buxton. 2010. *Sketching User Experiences: Getting the Design Right and the Right Design.* Elsevier Science. Retrieved from https://books.google.co.uk/books?id=2vfPxocmLh0C

[13] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (November 2019), 1–24. DOI:https://doi.org/10.1145/3359206

[14] Mirian Calvo, Leon Cruickshank, and Madeleine Sclater. 2022. Exploring mutual learning in co-design. *DISCERN: International Journal of Design for Social Change, Sustainable Innovation and Entrepreneurship* 3, 2 (November 2022), 79–96.

[15] Betty P. I. Chang, Thomas L. Webb, and Yael Benn. 2017. Why Do People Act Like the Proverbial Ostrich? Investigating the Reasons That People Provide for Not Monitoring Their Goal Progress. *Front. Psychol.* 8, (February 2017). DOI:https://doi.org/10.3389/fpsyg.2017.00152

[16] Souti Chattopadhyay, Ishita Prasad, Austin Z. Henley, Anita Sarma, and Titus Barik. 2020. What's Wrong with Computational Notebooks? Pain Points, Needs, and Design Opportunities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–12. DOI:https://doi.org/10.1145/3313831.3376729

[17] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, San Francisco California USA, 785–794. DOI:https://doi.org/10.1145/2939672.2939785

[18] Yunan Chen. 2010. Take It Personally: Accounting for Individual Difference in Designing Diabetes Management Systems. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems* (DIS '10), ACM, New York, NY, USA, 252–261. DOI:https://doi.org/10.1145/1858171.1858218

[19] Hao-Fei Cheng, Logan Stapleton, Ruiqi Wang, Paige Bullock, Alexandra Chouldechova, Zhiwei Steven Steven Wu, and Haiyi Zhu. 2021. Soliciting Stakeholders' Fairness Notions in Child Maltreatment Predictive Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–17. DOI:https://doi.org/10.1145/3411764.3445308

[20] Eun Kyoung Choe, Saeed Abdullah, Mashfiqui Rabbi, Edison Thomaz, Daniel A. Epstein, Felicia Cordeiro, Matthew Kay, Gregory D. Abowd, Tanzeem Choudhury, James Fogarty, Bongshin Lee, Mark Matthews, and Julie A. Kientz. 2017. Semi-Automated Tracking: A Balanced Approach for Self-Monitoring Applications. *IEEE Pervasive Computing* 16, 1 (January 2017), 74–84. DOI:https://doi.org/10.1109/MPRV.2017.18

[21] Charles M. Clark, James W. Snyder, Robert L. Meek, Linda M. Stutz, and Christopher G. Parkin. 2001. A Systematic Approach to Risk Stratification and Intervention Within a Managed Care Environment Improves Diabetes Outcomes and Patient Satisfaction. *Diabetes Care* 24, 6 (June 2001), 1079–1086. DOI:https://doi.org/10.2337/diacare.24.6.1079

[22] Michael F. Clarke, Joseph Gonzales, Richard Harper, David Randall, Thomas Ludwig, and Nozomi Ikeya. 2019. Better Supporting Workers in ML Workplaces. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, ACM, Austin TX USA, 443–448. DOI:https://doi.org/10.1145/3311957.3359429

[23] CoCalc. CoCalc – Collaborative Calculation. Retrieved August 3, 2022 from https://cocalc.com/

[24] A. Crabtree, M. Rouncefield, and P. Tolmie. 2012. *Doing Design Ethnography.* Springer London. Retrieved from https://books.google.co.uk/books?id=Irm2KKegDjQC

[25] Fernando Delgado, Solon Barocas, and Karen Levy. 2022. An Uncommon Task: Participatory Design in Legal AI. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1 (March 2022), 1–23. DOI:https://doi.org/10.1145/3512898

[26] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2021. Stakeholder Participation in AI: Beyond" Add Diverse Stakeholders and Stir". *arXiv preprint arXiv:2111.01122* (2021).

[27] Sean T. Doherty and Stephen P. Greaves. 2015. Time-Series Analysis of Continuously Monitored Blood Glucose: The Impacts of Geographic and Daily Lifestyle Factors. *Journal of Diabetes Research* 2015, (2015), 1–6. DOI:https://doi.org/10.1155/2015/804341

[28] Klaus Donsa, Stephan Spat, Peter Beck, Thomas R. Pieber, and Andreas Holzinger. 2015. Towards Personalization of Diabetes Therapy Using Computerized Decision Support and Machine Learning: Some Open Problems and Challenges. In *Smart Health: Open Problems and Future Challenges*, Andreas Holzinger, Carsten Röcker and Martina Ziefle (eds.). Springer International Publishing, Cham, 237–260. DOI:https://doi.org/10.1007/978-3-319-16226-3_10

[29] Afsaneh Doryab, Mads Frost, Maria Faurholt-Jepsen, Lars V. Kessing, and Jakob E. Bardram. 2015. Impact factor analysis: combining prediction with parameter ranking to reveal the impact of behavior on health outcome. *Pers Ubiquit Comput* 19, 2 (February 2015), 355–365. DOI:https://doi.org/10.1007/s00779-014-0826-8

[30] Graham Dove and Anne-Laure Fayard. 2020. Monsters, Metaphors, and Machine Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–17. DOI:https://doi.org/10.1145/3313831.3376275

[31] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, Denver Colorado USA, 278–288. DOI:https://doi.org/10.1145/3025453.3025739

[32] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 7639 (February 2017), 115–118. DOI:https://doi.org/10.1038/nature21056

[33] G. Eysenbach. 2000. Recent advances: Consumer health informatics. *BMJ* 320, 7251 (June 2000), 1713–1716. DOI:https://doi.org/10.1136/bmj.320.7251.1713

[34] Neil Fleming and David Baume. 2006. Learning Styles Again: VARKing up the right tree! *Educational developments* 7, 4 (2006), 4.

[35] Neil D. Fleming. 1995. I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom. In *Research and development in higher education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA), HERDSA*, 308–313.

[36] Zhiyong Fu and Yuyao Zhou. 2020. Research on human–AI co-creation based on reflective design practice. *CCF Trans. Pervasive Comp. Interact.* 2, 1 (March 2020), 33–41. DOI:https://doi.org/10.1007/s42486-020-00028-0

[37] Martha M. Funnell and Robert M. Anderson. 2004. Empowerment and Self-Management of Diabetes. *Clinical Diabetes* 22, 3 (July 2004), 123–127. DOI:https://doi.org/10.2337/diaclin.22.3.123

[38] Marco Gillies, Bongshin Lee, Nicolas d'Alessandro, Joëlle Tilmanne, Todd Kulesza, Baptiste Caramiaux, Rebecca Fiebrink, Atau Tanaka, Jérémie Garcia, Frédéric Bevilacqua, Alexis Heloir, Fabrizio Nunnari, Wendy Mackay, and Saleema Amershi. 2016. Human-Centred Machine Learning. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, ACM Press, San Jose, California, USA, 3558–3565. DOI:https://doi.org/10.1145/2851581.2856492

[39] Fabien Girardin and Neal Lathia. 2017. When User Experience Designers Partner with Data Scientists. In *AAAI Spring Symposia*.

[40] Richard H. R. Harper. 2019. The Role of HCI in the Age of AI. *International Journal of Human–Computer Interaction* 35, 15 (September 2019), 1331–1344. DOI:https://doi.org/10.1080/10447318.2019.1631527

[41] Andrew Head, Fred Hohman, Titus Barik, Steven M. Drucker, and Robert DeLine. 2019. Managing Messes in Computational Notebooks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, Glasgow Scotland Uk, 1–12. DOI:https://doi.org/10.1145/3290605.3300500

[42] Michael Hind. 2019. Explaining explainable AI. *XRDS* 25, 3 (April 2019), 16–19. DOI:https://doi.org/10.1145/3313096

[43] Ben Hutchinson and Margaret Mitchell. 2019. 50 Years of Test (Un)fairness: Lessons for Machine Learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, ACM Press, Atlanta, GA, USA, 49–58. DOI:https://doi.org/10.1145/3287560.3287600

[44] Ole Sejer Iversen, Kim Halskov, and Tuck W. Leong. 2012. Values-led participatory design. *CoDesign* 8, 2–3 (June 2012), 87–103. DOI:https://doi.org/10.1080/15710882.2012.672575

[45] Kirtan Jha, Aalap Doshi, Poojan Patel, and Manan Shah. 2019. A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture* 2, (June 2019), 1–12. DOI:https://doi.org/10.1016/j.aiia.2019.05.004

[46] Jeremiah W. Johnson. 2020. Benefits and Pitfalls of Jupyter Notebooks in the Classroom. In *Proceedings of the 21st Annual Conference on Information Technology Education*, ACM, Virtual Event USA, 32–37. DOI:https://doi.org/10.1145/3368308.3415397

[47] Jupyter. Project Jupyter. Retrieved August 2, 2022 from https://jupyter.org

[48] Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. 2009. Factors Predictive of Use and of Benefit From Continuous Glucose Monitoring in Type 1 Diabetes. *Diabetes Care* 32, 11 (November 2009), 1947–1953. DOI:https://doi.org/10.2337/dc09-0889

[49] DaYe Kang, Tony Ho, Nicolai Marquardt, Bilge Mutlu, and Andrea Bianchi. 2021. ToonNote: Improving Communication in Computational Notebooks Using Interactive Data Comics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–14. DOI:https://doi.org/10.1145/3411764.3445434

[50] Simon Katan, Mick Grierson, and Rebecca Fiebrink. 2015. Using Interactive Machine Learning to Support Interface Development Through Workshops with Disabled People. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, ACM Press, Seoul, Republic of Korea, 251–254. DOI:https://doi.org/10.1145/2702123.2702474

[51] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–14. DOI:https://doi.org/10.1145/3313831.3376219

[52] Finn Kensing and Jeanette Blomberg. 1998. Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW)* 7, 3–4 (1998), 167–185.

[53] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. 2017. OmniTrack: A Flexible Self-Tracking Approach Leveraging Semi-Automated Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (September 2017), 28. DOI:https://doi.org/10.1145/3130930

[54] David C. Klonoff and David Kerr. 2018. Overcoming Barriers to Adoption of Digital Health Tools for Diabetes. *J Diabetes Sci Technol* 12, 1 (January 2018), 3–6. DOI:https://doi.org/10.1177/1932296817732459

[55] Janne van Kollenburg, Sander Bogers, Heleen Rutjes, Eva Deckers, Joep Frens, and Caroline Hummels. 2018. Exploring the Value of Parent Tracked Baby Data in Interactions with Healthcare Professionals: A Data-Enabled Design Exploration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), ACM, New York, NY, USA, 297:1-297:12. DOI:https://doi.org/10.1145/3173574.3173871

[56] Sean Kross and Philip Guo. 2021. Orienting, Framing, Bridging, Magic, and Counseling: How Data Scientists Navigate the Outer Loop of Client Collaborations in Industry and Academia. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (October 2021), 1–28. DOI:https://doi.org/10.1145/3476052

[57] Cynthia LeRouge and Nilmini Wickramasinghe. 2013. A Review of User-Centered Design for Diabetes-Related Consumer Health Informatics Technologies. *J Diabetes Sci Technol* 7, 4 (July 2013), 1039–1056. DOI:https://doi.org/10.1177/193229681300700429

[58] David Leslie. 2019. Understanding artificial intelligence ethics and safety. *arXiv preprint arXiv:1906.05684* (2019).

<bib id="bib59"><number>[59]</number>Daria Loi, Thomas Lodato, Christine T. Wolf, Raphael Arar, and Jeanette Blomberg. 2018. PD manifesto for AI futures. In *Proceedings of the 15th Participatory Design Conference on Short Papers, Situated Actions, Workshops and Tutorial - PDC '18*, ACM Press, Hasselt and Genk, Belgium, 1–4. DOI:https://doi.org/10.1145/3210604.3210614</bib>

<bib id="bib60"><number>[60]</number>Daria Loi, Christine T. Wolf, Jeanette L. Blomberg, Raphael Arar, and Margot Brereton. 2019. Co-designing AI Futures: Integrating AI Ethics, Social Computing, and Design. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion - DIS '19 Companion*, ACM Press, San Diego, CA, USA, 381–384. DOI:https://doi.org/10.1145/3301019.3320000</bib>

<bib id="bib61"><number>[61]</number>Peter Lovei, Renee Noortman, and Mathias Funk. 2022. Introduction to Data-Enabled Design. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, ACM, New Orleans LA USA, 1–3. DOI:https://doi.org/10.1145/3491101.3503749</bib>

<bib id="bib62"><number>[62]</number>Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, San Jose California USA, 5286–5297. DOI:https://doi.org/10.1145/2858036.2858288</bib>

<bib id="bib63"><number>[63]</number>Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–14. DOI:https://doi.org/10.1145/3313831.3376445</bib>

<bib id="bib64"><number>[64]</number>Lena Mamykina, Andrew D. Miller, Elizabeth D. Mynatt, and Daniel Greenblatt. 2010. Constructing identities through storytelling in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1203–1212. Retrieved January 31, 2016 from http://dl.acm.org/citation.cfm?id=1753507</bib>

<bib id="bib65"><number>[65]</number>Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. 2008. MAHI: Investigation of Social Scaffolding for Reflective Thinking in Diabetes Management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), ACM, New York, NY, USA, 477–486. DOI:https://doi.org/10.1145/1357054.1357131</bib>

<bib id="bib66"><number>[66]</number>Md. Maniruzzaman, Md. Jahanur Rahman, Md. Al-MehediHasan, Harman S. Suri, Md. Menhazul Abedin, Ayman El-Baz, and Jasjit S. Suri. 2018. Accurate Diabetes Risk Stratification Using Machine Learning: Role of Missing Value and Outliers. *J Med Syst* 42, 5 (May 2018), 92. DOI:https://doi.org/10.1007/s10916-018-0940-7</bib>

<bib id="bib67"><number>[67]</number>M. Monteiro-Soares, E. J. Boyko, J. Ribeiro, I. Ribeiro, and M. Dinis-Ribeiro. 2011. Risk stratification systems for diabetic foot ulcers: a systematic review. *Diabetologia* 54, 5 (May 2011), 1190–1199. DOI:https://doi.org/10.1007/s00125-010-2030-3</bib>

<bib id="bib68"><number>[68]</number>Cecily Morrison, Edward Cutrell, Anupama Dhareshwar, Kevin Doherty, Anja Thieme, and Alex Taylor. 2017. Imagining Artificial Intelligence Applications with People with Visual Disabilities using Tactile Ideation. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Baltimore Maryland USA, 81–90. DOI:https://doi.org/10.1145/3132525.3132530</bib>

<bib id="bib69"><number>[69]</number>Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, ACM Press, Glasgow, Scotland Uk, 1–15. DOI:https://doi.org/10.1145/3290605.3300356</bib>

<bib id="bib70"><number>[70]</number>Michael Muller and Q. Vera Liao. 2017. Exploring AI Ethics and Values through Participatory Design Fictions. *Human Computer Interaction Consortium* (2017).</bib>

<bib id="bib71"><number>[71]</number>Renee Noortman, Peter Lovei, Mathias Funk, Eva Deckers, Stephan Wensveen, and Berry Eggen. 2022. Breaking up data-enabled design: expanding and scaling up for the clinical context. *AIEDAM* 36, (2022), e19. DOI:https://doi.org/10.1017/S0890060421000433</bib>

<bib id="bib72"><number>[72]</number>Ziad Obermeyer and Sendhil Mullainathan. 2019. Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70 Million People. In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, ACM Press, Atlanta, GA, USA, 89–89. DOI:https://doi.org/10.1145/3287560.3287593</bib>

<bib id="bib73"><number>[73]</number>Aisling Ann O'Kane, Yvonne Rogers, and Ann E. Blandford. 2015. Concealing or Revealing Mobile Medical Devices?: Designing for Onstage and Offstage Presentation. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, ACM, 1689–1698.</bib>

<bib id="bib74"><number>[74]</number>Cecilia Panigutti, Andrea Beretta, Fosca Giannotti, and Dino Pedreschi. 2022. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. In *CHI Conference on Human Factors in Computing Systems*, ACM, New Orleans LA USA, 1–9. DOI:https://doi.org/10.1145/3491102.3502104</bib>

<bib id="bib75"><number>[75]</number>Jeffrey M. Perkel. 2018. Why Jupyter is data scientists' computational notebook of choice. *Nature* 563, 7729 (October 2018), 145–146. DOI:https://doi.org/10.1038/d41586-018-07196-1</bib>

<bib id="bib76"><number>[76]</number>David Piorkowski, Soya Park, April Yi Wang, Dakuo Wang, Michael Muller, and Felix Portnoy. 2021. How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 1–25. DOI:https://doi.org/10.1145/3449205</bib>

<bib id="bib77"><number>[77]</number>Luigi Quaranta, Fabio Calefato, and Filippo Lanubile. 2022. Eliciting Best Practices for Collaboration with Computational Notebooks. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1 (March 2022), 1–41. DOI:https://doi.org/10.1145/3512934</bib>

<bib id="bib78"><number>[78]</number>Bodil Rasmussen, Glenn Ward, Alicia Jenkins, Susan J King, and Trisha Dunning. 2011. Young adults' management of Type 1 diabetes during life transitions: Young adults' management of Type 1 diabetes. *Journal of Clinical Nursing* 20, 13–14 (July 2011), 1981–1992. DOI:https://doi.org/10.1111/j.1365-2702.2010.03657.x</bib>

<bib id="bib79"><number>[79]</number>David Rodbard. 2019. State of Type 1 Diabetes Care in the United States in 2016–2018 from T1D Exchange Registry Data. *Diabetes Technology & Therapeutics* 21, 2 (February 2019), 62–65. DOI:https://doi.org/10.1089/dia.2019.0008</bib>

<bib id="bib80"><number>[80]</number>Adam Rule, Aurélien Tabard, and James D. Hollan. 2018. Exploration and Explanation in Computational Notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, Montreal QC Canada, 1–12. DOI:https://doi.org/10.1145/3173574.3173606</bib>

<bib id="bib81"><number>[81]</number>Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–15. DOI:https://doi.org/10.1145/3411764.3445518</bib>

<bib id="bib82"><number>[82]</number>Logan Stapleton, Min Hun Lee, Diana Qing, Marya Wright, Alexandra Chouldechova, Ken Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022. Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, ACM, Seoul Republic of Korea, 1162–1177. DOI:https://doi.org/10.1145/3531146.3533177</bib>

<bib id="bib83"><number>[83]</number>Marc Steen. 2013. Co-design as a process of joint inquiry and imagination. *Design Issues* 29, 2 (2013), 16–28.</bib>

<bib id="bib84"><number>[84]</number>Cristiano Storni. 2013. Design challenges for ubiquitous and personal computing in chronic disease care and patient empowerment: a case study rethinking diabetes self-monitoring. *Pers Ubiquit Comput* 18, 5 (August 2013), 1277–1290. DOI:https://doi.org/10.1007/s00779-013-0707-6</bib>

<bib id="bib85"><number>[85]</number>Hariharan Subramonyam, Jane Im, Colleen Seifert, and Eytan Adar. 2022. Solving Separation-of-Concerns Problems in Collaborative Design of Human-AI Systems through Leaky Abstractions. In *CHI Conference on Human Factors in Computing Systems*, ACM, New Orleans LA USA, 1–21. DOI:https://doi.org/10.1145/3491102.3517537</bib>

<bib id="bib86"><number>[86]</number>Franziska Tachtler, Konstantin Aal, Tanja Ertl, Daniel Diethei, Jasmin Niess, Mohammed Khwaja, Reem Talhouk, Giovanna Nunes Vilaza, Shaimaa Lazem, Aneesha Singh, Marguerite Barry, Volker Wulf, and Geraldine Fitzpatrick. 2021. Artificially Intelligent Technology for the Margins: A Multidisciplinary Design Agenda. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–7. DOI:https://doi.org/10.1145/3411763.3441333</bib>

<bib id="bib87"><number>[87]</number>Anja Thieme, Danielle Belgrave, and Gavin Doherty. 2020. Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Trans. Comput.-Hum. Interact.* 27, 5 (August 2020), 34:1-34:53. DOI:https://doi.org/10.1145/3398069</bib>

<bib id="bib88"><number>[88]</number>Anja Thieme, Maryann Hanratty, Maria Lyons, Jorge E. Palacios, Rita Marques, Cecily Morrison, and Gavin Doherty. 2022. Designing Human-Centered AI for Mental Health: Developing Clinically Relevant Applications for Online CBT Treatment. *ACM Trans. Comput.-Hum. Interact.* (October 2022), 3564752. DOI:https://doi.org/10.1145/3564752</bib>

<bib id="bib89"><number>[89]</number>Shari Trewin, Sara Basson, Michael Muller, Stacy Branham, Jutta Treviranus, Daniel Gruen, Daniel Hebert, Natalia Lyckowski, and Erich Manser. 2019. Considerations for AI fairness for people with disabilities. *AI Matters* 5, 3 (December 2019), 40–63. DOI:https://doi.org/10.1145/3362077.3362086</bib>

<bib id="bib90"><number>[90]</number>Greg Walsh and Eric Wronsky. 2019. AI + Co-Design: Developing a Novel Computer-supported Approach to Inclusive Design. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, ACM, Austin TX USA, 408–412. DOI:https://doi.org/10.1145/3311957.3359456</bib>

<bib id="bib91"><number>[91]</number>April Yi Wang, Anant Mittal, Christopher Brooks, and Steve Oney. 2019. How Data Scientists Use Computational Notebooks for Real-Time Collaboration. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (November 2019), 1–30. DOI:https://doi.org/10.1145/3359141</bib>

<bib id="bib92"><number>[92]</number>April Yi Wang, Dakuo Wang, Jaimie Drozdal, Michael Muller, Soya Park, Justin D. Weisz, Xuye Liu, Lingfei Wu, and Casey Dugan. 2022. Documentation Matters: Human-Centered AI System to Assist Data Science Code Documentation in Computational Notebooks. *ACM Trans. Comput.-Hum. Interact.* 29, 2 (April 2022), 1–33. DOI:https://doi.org/10.1145/3489465</bib>

<bib id="bib93"><number>[93]</number>Alistair Willis, Patricia Charlton, and Tony Hirst. 2020. Developing Students' Written Communication Skills with Jupyter Notebooks. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, ACM, Portland OR USA, 1089–1095. DOI:https://doi.org/10.1145/3328778.3366927</bib>

<bib id="bib94"><number>[94]</number>Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. 2018. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18*, ACM Press, Hong Kong, China, 585–596. DOI:https://doi.org/10.1145/3196709.3196730</bib>

<bib id="bib95"><number>[95]</number>Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376301</bib>

<bib id="bib96"><number>[96]</number>Nur Yildirim, Alex Kass, Teresa Tung, Connor Upton, Donnacha Costello, Robert Giusti, Sinem Lacin, Sara Lovic, James M O'Neill, Rudi O'Reilly Meehan, Eoin Ó Loideáin, Azzurra Pini, Medb Corcoran, Jeremiah Hayes, Diarmuid J Cahalane, Gaurav Shivhare, Luigi Castoro, Giovanni Caruso, Changhoon Oh, James McCann, Jodi Forlizzi, and John Zimmerman. 2022. How Experienced Designers of Enterprise Applications Engage AI as a Design Material. In *CHI Conference on Human Factors in Computing Systems*, ACM, New Orleans LA USA, 1–13. DOI:https://doi.org/10.1145/3491102.3517491</bib>

<bib id="bib97"><number>[97]</number>Yue You and Xinning Gui. 2020. Self-Diagnosis through AI-enabled Chatbot-based Symptom Checkers: User Experiences and Design Considerations. *AMIA Annu Symp Proc* 2020, (2020), 1354–1363.</bib>

<bib id="bib98"><number>[98]</number>Amy X. Zhang, Michael Muller, and Dakuo Wang. 2020. How do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (May 2020), 1–23. DOI:https://doi.org/10.1145/3392826</bib>

<bib id="bib99"><number>[99]</number>Chengbo Zheng, Dakuo Wang, April Yi Wang, and Xiaojuan Ma. 2022. Telling Stories from Computational Notebooks: AI-Assisted Presentation Slides Creation for Presenting Data Science Work. In *CHI Conference on Human Factors in Computing Systems*, ACM, New Orleans LA USA, 1–20. DOI:https://doi.org/10.1145/3491102.3517615</bib>