Addressing UX Practitioners' Challenges in Designing ML Applications: an Interactive Machine Learning Approach

K. J. Kevin Feng
University of Washington
kjfeng@uw.edu
@kjfeng_

David W. McDonald
University of Washington

IUI 2023
March 27 – 31, 2023 | Sydney, Australia

HUMAN CENTERED
DESIGN & ENGINEERING

DESIGN
USE
BUILD

Addressing UX Practitioners' Challenges in Designing ML Applications: an Interactive Machine Learning Approach

K. J. Kevin Feng
University of Washington
kjfeng@uw.edu
@kjfeng_

David W. McDonald
University of Washington

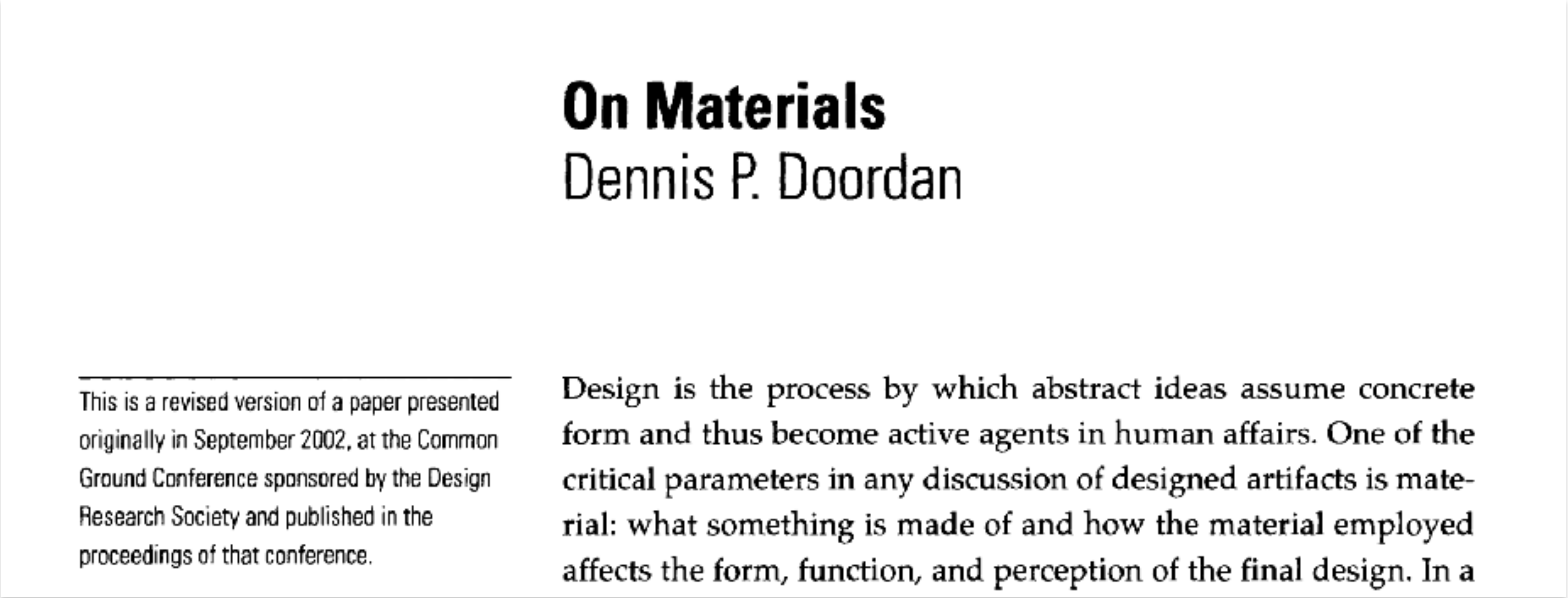
IUI 2023
March 27 – 31, 2023 | Sydney, Australia

HUMAN CENTERED DESIGN & ENGINEERING

DESIGN USE BUILD

ML as a design material

ML as a design material


On Materials
Dennis P. Doordan

This is a revised version of a paper presented originally in September 2002, at the Common Ground Conference sponsored by the Design Research Society and published in the proceedings of that conference.

Design is the process by which abstract ideas assume concrete form and thus become active agents in human affairs. One of the critical parameters in any discussion of designed artifacts is material: what something is made of and how the material employed affects the form, function, and perception of the final design. In a

Doordan. *On Materials*. Design Issues, Autumn 2003.

Fabrication

Application

Appreciation

Fabrication

Preparing the raw material for use

- Data + models typically prepared by ML practitioners [1]
- ML abstractions can be divorced from UX concepts [2]
- UXPs treat ML as a blackbox [3]

[1] Subramonyam et al. *Towards A Process Model for Co-Creating AI Experiences*. DIS 2021.
[2] Yang et al. Investigating How Experienced UX Designers Effectively Work with Machine Learning.. DIS 2018.
[3] Dove et al. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. CHI 2017.

Application

Transforming materials into usable products

- Blackboxed nature makes it difficult to calibrate UX [4]
- Unable to anticipate ethical and fairness issues [5]

[4] Benjamin et al. *Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry*. CHI 2021.
[5] Holmquist. *Intelligence on tap: artificial intelligence as a new design material*. Interactions 2017.

Appreciation

Gathering feedback on material from users

- ML can constantly evolve and shift UX [6]
- Challenging to grasp evolutions with blackboxed understanding [7]

[6] Subramonyam et al. *ProtoAI: Model- Informed Prototyping for AI-Powered Interfaces*. IUI 2021.
[7] Yang et al. Re- Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. CHI 2020.

Fabrication

Preparing the raw material for use

- Data + models typically prepared by ML practitioners [1]
- ML abstractions can be divorced from UX concepts [2]
- UXPs treat ML as a blackbox [3]

[1] Subramonyam et al. *Towards A Process Model for Co-Creating AI Experiences*. DIS 2021.
[2] Yang et al. Investigating How Experienced UX Designers Effectively Work with Machine Learning.. DIS 2018.
[3] Dove et al. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. CHI 2017.

Application

Transforming materials into usable products

- Blackboxed nature makes it difficult to calibrate UX [4]
- Unable to anticipate ethical and fairness issues [5]

[4] Benjamin et al. *Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry*. CHI 2021.
[5] Holmquist. *Intelligence on tap: artificial intelligence as a new design material*. Interactions 2017.

Appreciation

Gathering feedback on material from users

- ML can constantly evolve and shift UX [6]
- Challenging to grasp evolutions with blackboxed understanding [7]

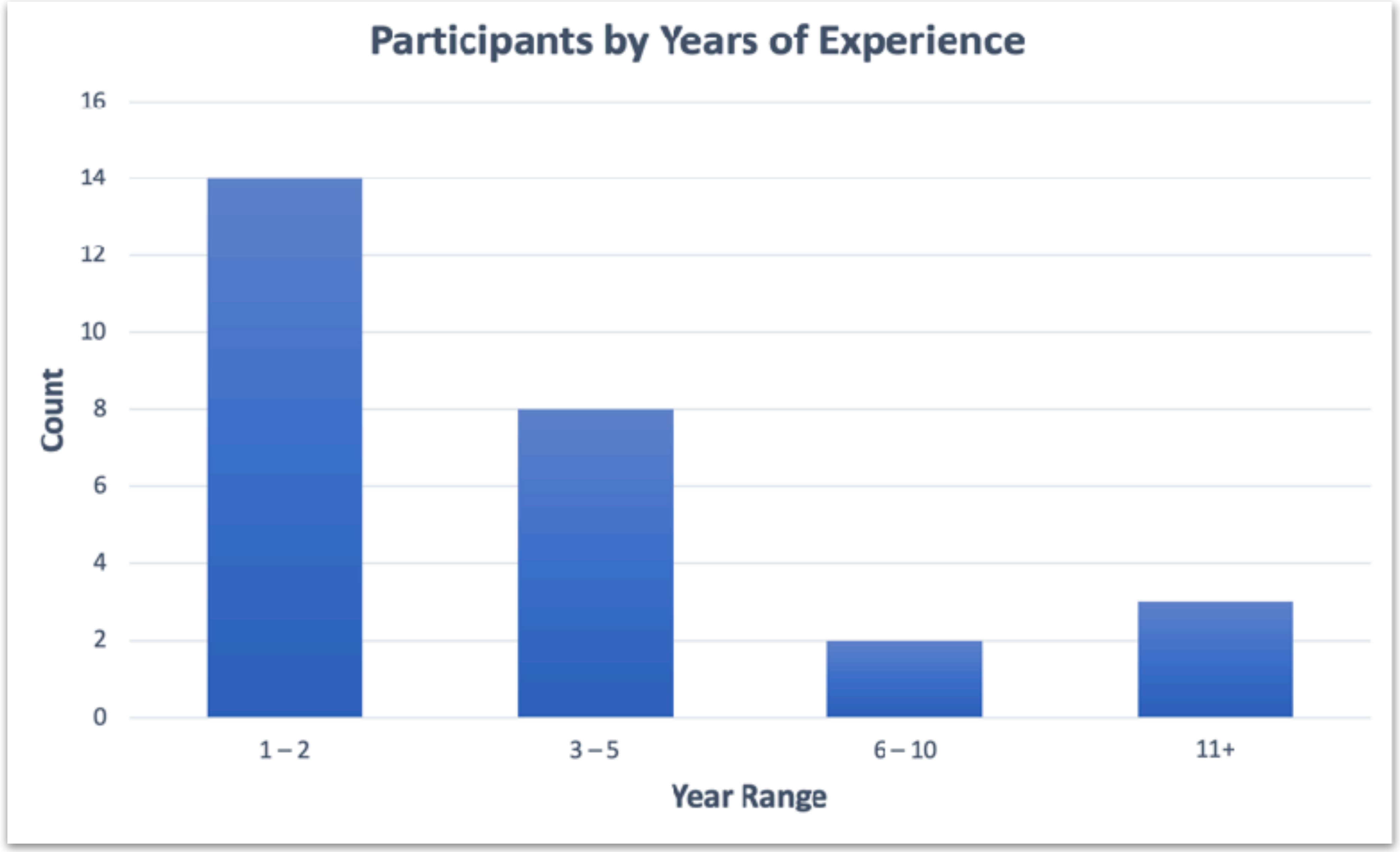
[6] Subramonyam et al. *ProtoAI: Model- Informed Prototyping for AI-Powered Interfaces*. IUI 2021.
[7] Yang et al. Re- Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. CHI 2020.

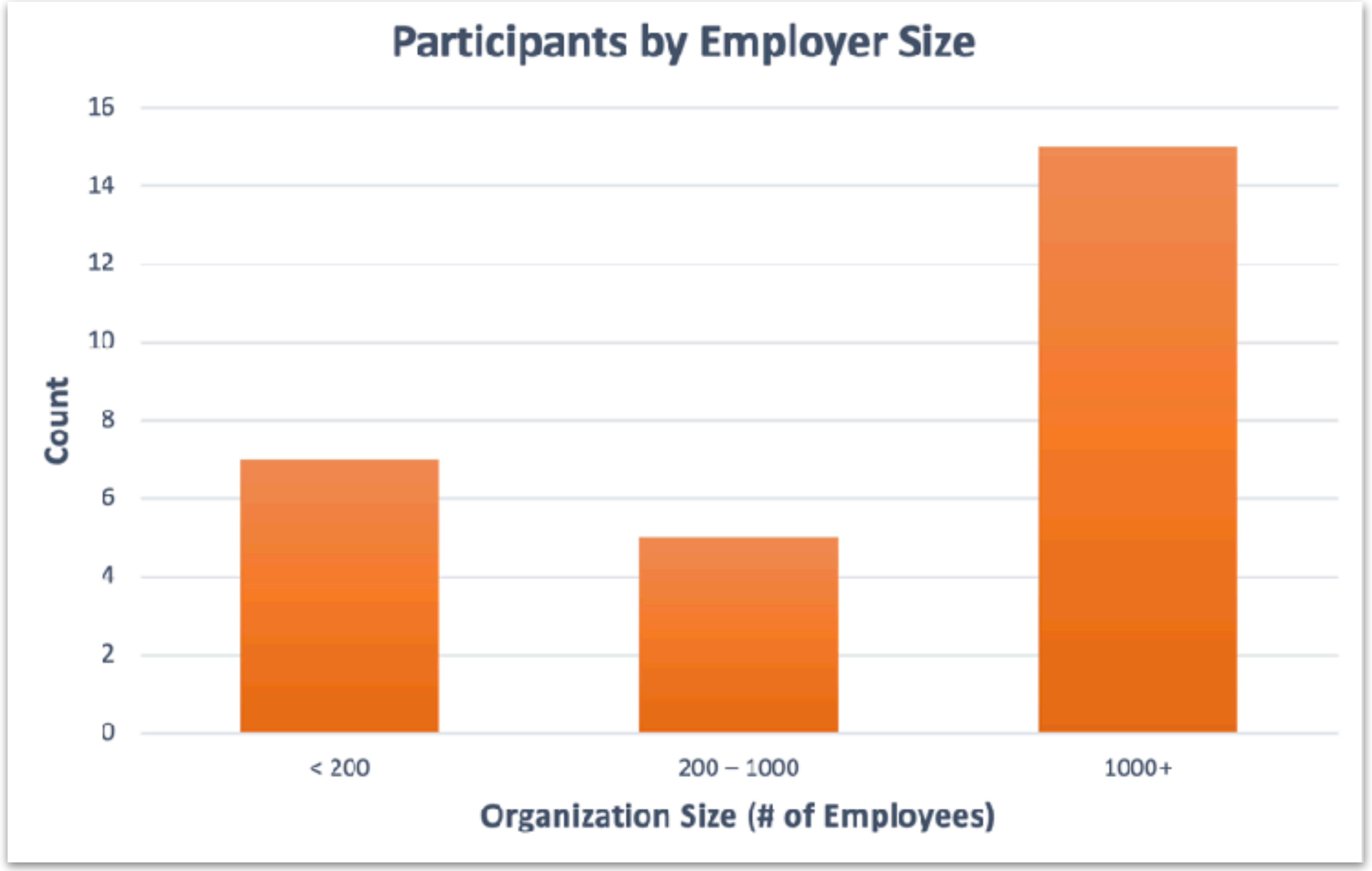
What if UXPs could fabricate ML as a design material *as part of the design process?*

Method

Study Overview

- 1:1 task-based design sessions with 27 UXPs
- Most (20) had no prior ML design experience





Session Structure

2 hrs

Method

Session Structure

15 mins | 1 hr 15 mins | 30 mins

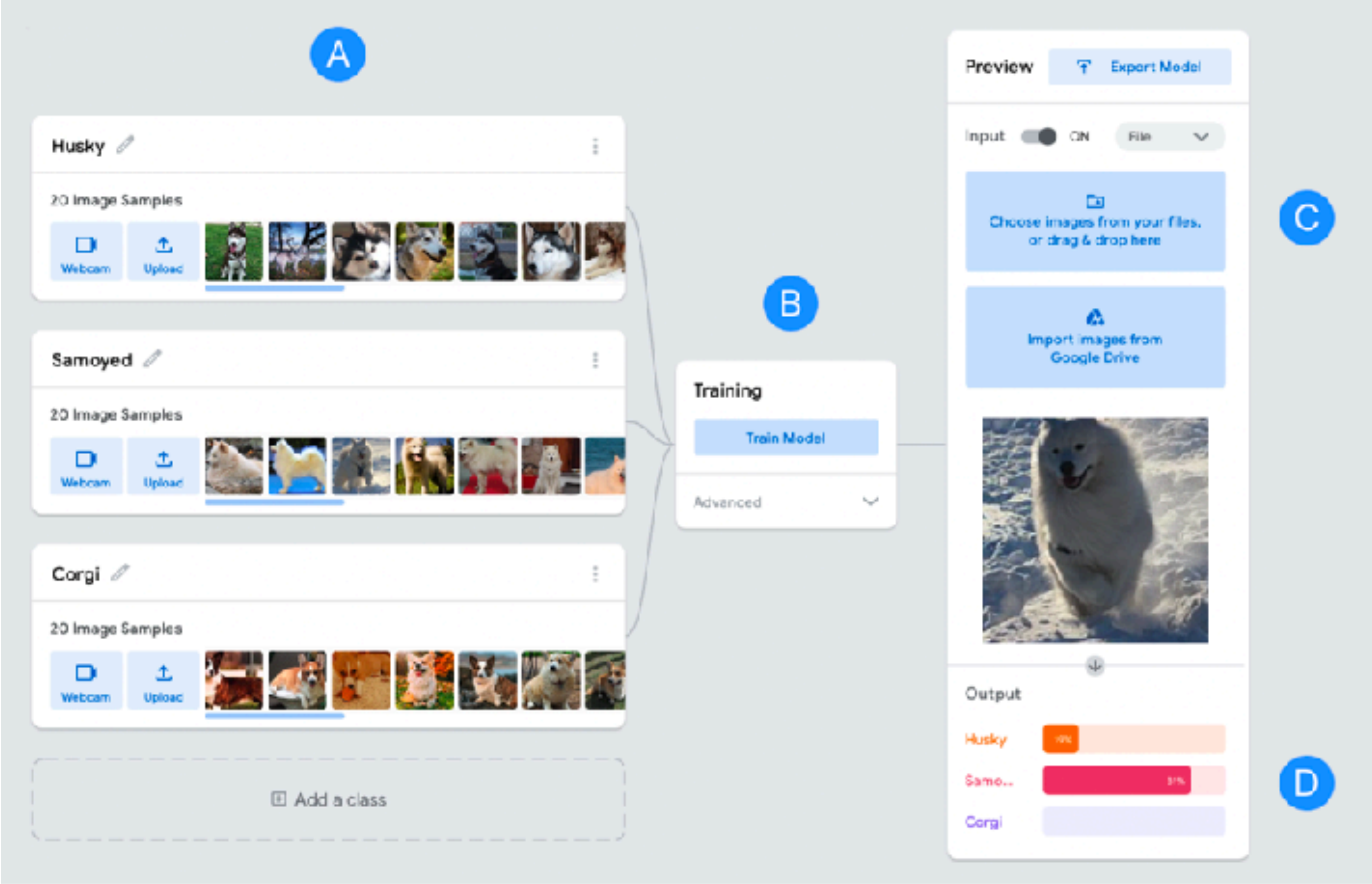
Method

Session Structure

15 mins | 1 hr 15 mins | 30 mins

Tutorial

- Gave tutorial of simple interactive ML tool by walking through 2 examples
- Chose Google's Teachable Machine for simplicity and low learning curve

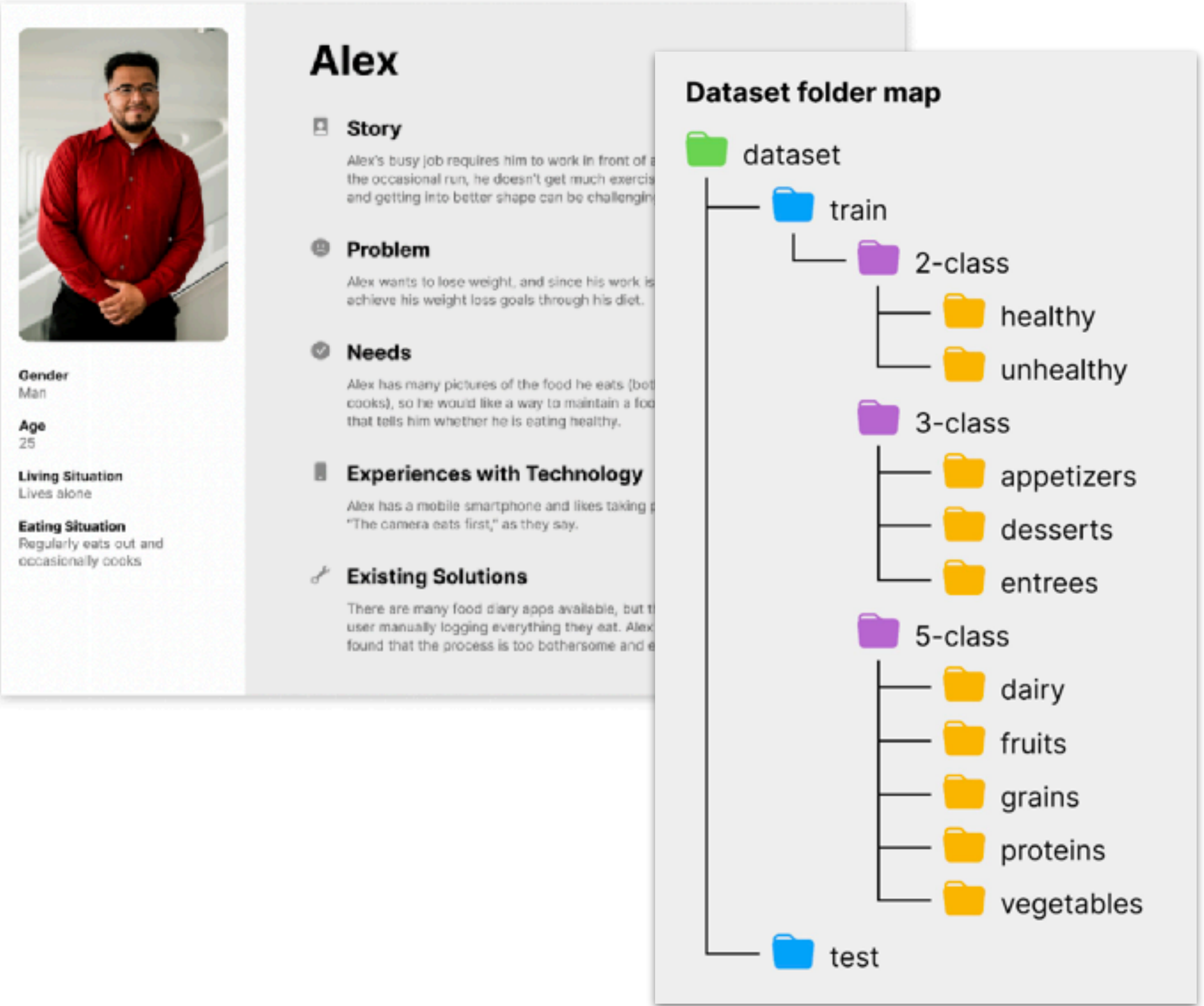


Session Structure

15 mins | 1 hr 15 mins | 30 mins

Activity

- Main design task: create a proof-of-concept for a food classification app
- Provided necessary materials, including 3 datasets for training models
- Deliverable: short slide deck/ document with their concept



Session Structure

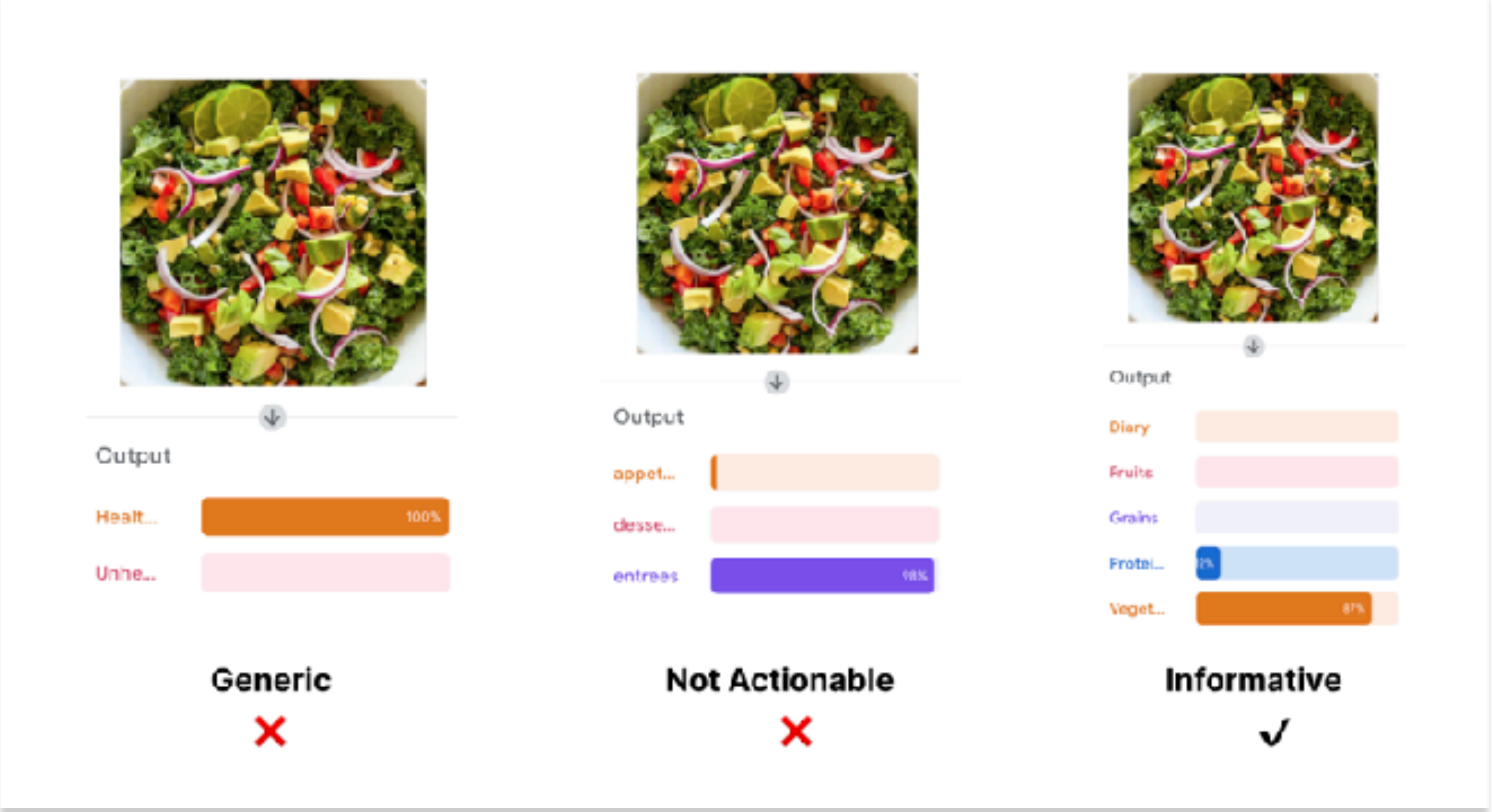
15 mins

1 hr 15 mins

30 mins

Interview

- Asked about participants' experiences using Teachable Machine as a design aid
- Asked about certain design decisions they made in their deliverable



Findings

Findings

- Interactive fabrication improved participants' ability to reason about model and training data

Improved Ability to Reason About ML

Considering data labels/model classes alongside UX goals

"[The athletic persona] may really want to see how much protein she got" — P12

"We want to start with something minimal that we can actually test with users, and so I think it's easier to do that with the two category model." — P23

Improved Ability to Reason About ML

Considering data labels/model classes alongside UX goals

"[The athletic persona] may really want to see how much protein she got" — P12

"We want to start with something minimal that we can actually test with users, and so I think it's easier to do that with the two category model." — P23

Identifying new potential for ML by extending current labels/classes

Nutrition labels, combining existing labels, detecting specific nutrients

"I think it would be helpful for me to discuss with [the data scientist on my team] on what's the right categorization, and what are we trying to achieve." — P14

Findings

- Interactive fabrication improved participants' ability to reason about model and training data
- Participants saw opportunities to design UI affordances to help users interact with ML

UI Affordances for Interaction with ML

Guiding users to take higher quality pictures

“[referencing phone camera’s night mode] ‘you need a little bit more light’ it’s guiding you.” — P8

Findings

UI Affordances for Interaction with ML

Guiding users to take higher quality pictures

“[referencing phone camera’s night mode] ‘you need a little bit more light’ it’s guiding you.” — P8

Ingredient-based photo-taking for higher accuracy

P25’s approach: guide users to take photos of individual ingredients to make predictions from a combination of photos.

UI Affordances for Interaction with ML

Guiding users to take higher quality pictures

“[referencing phone camera’s night mode] ‘you need a little bit more light’ it’s guiding you.” — P8

Ingredient-based photo-taking for higher accuracy

P25’s approach: guide users to take photos of individual ingredients to make predictions from a combination of photos.

“High touch” opportunities for user customization & correction

High touch: human service in care-based jobs

“Regardless of info presented, users are able to modify [outputs] and have high touch opportunities” — P19

On the Importance of Customization

“It would be interesting for users to be able to choose what categories they wanted to use. I know some people are really into micro and macro foods or whatever, and so, if you could choose that as the user can take control of how they want to track their food, that could be really cool” — P1.

Findings

- Interactive fabrication improved participants' ability to reason about model and training data
- Participants saw opportunities to design UI affordances to help users interact with ML
- Participants identified key ethical considerations of their designs

Ethical Considerations (Societal/Topical), cont.

Human biases may seep in

"Maybe there's someone who's vegan and they think eating meat is not particularly healthy, and that is completely up to their own beliefs. They may classify the model to be unhealthy, and that could further inform the consumers of the app that if they're eating meat, then it's unhealthy or dangerous." — P26.

Concern of lack of flexibility

"What about cultures that don't eat a lot of dairy but eat a lot of rice? Like their stuff's going to be wrong if we go from just this dataset" — P21.

Ethical Considerations (Technical)

ML may not deliver on promises

“Ethically I think it’s important to tell [users] that this is not the gospel truth.” — P16.

ML can come with privacy risks

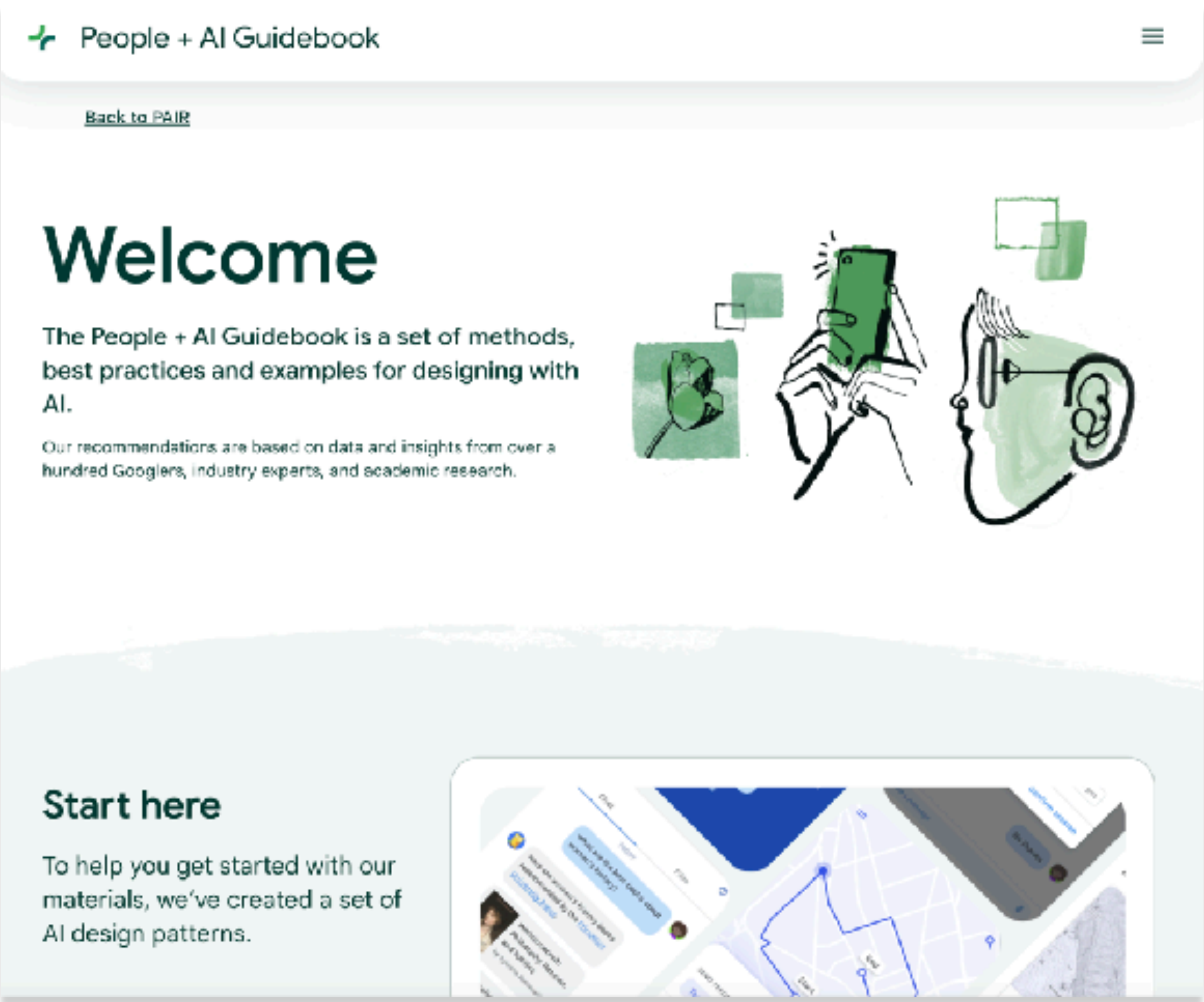
Let users know how their photos are processed.

“[We should know] both if it's legal and also what people are comfortable doing.” — P2.

Discussion

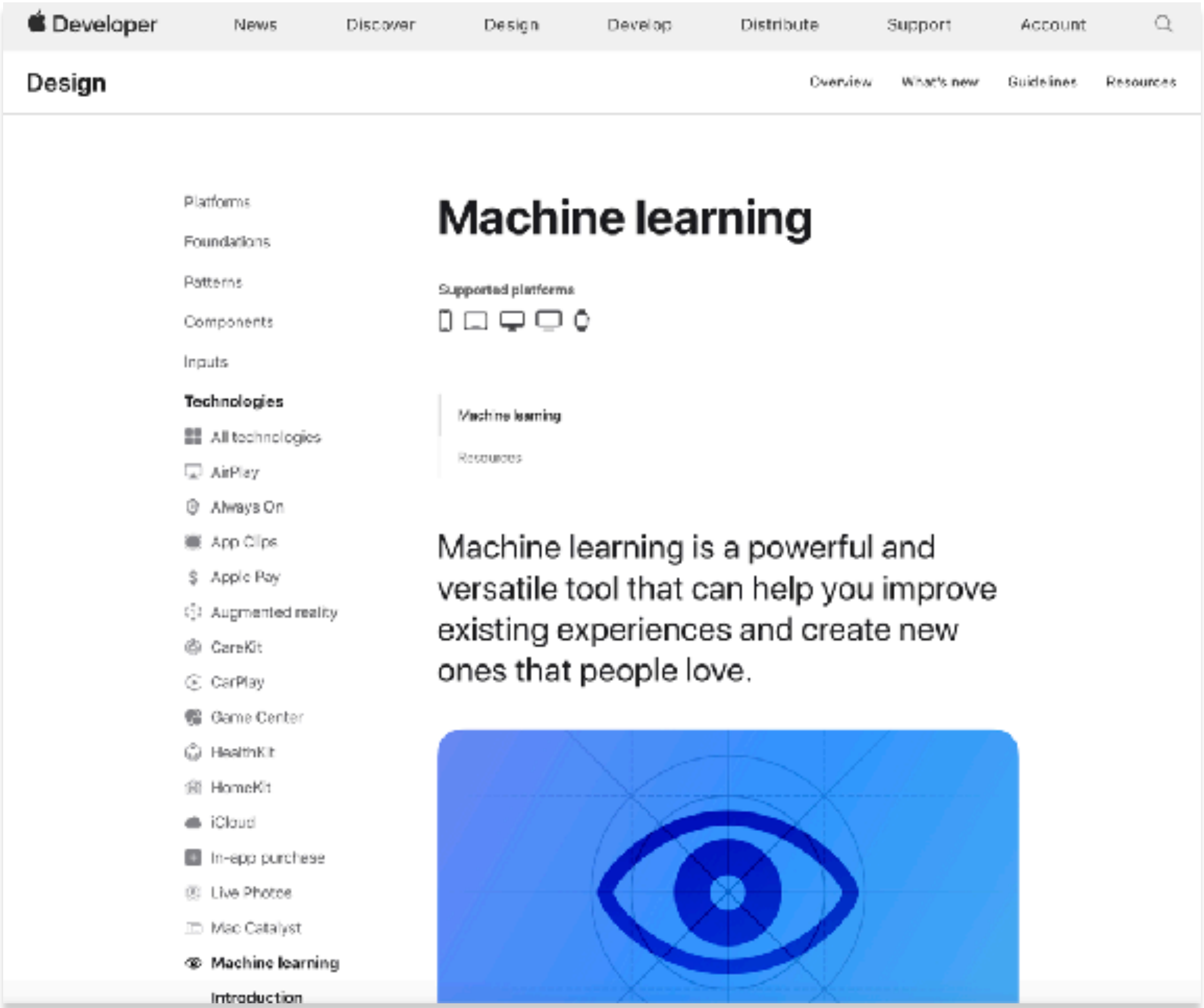
Discussion

Human-AI Guidelines



Google PAIR guidebook

Microsoft HAX toolkit



Apple Human Interface ML guidelines

Discussion

ID (from [5])	Guideline
G1	Make clear what the system can do.
G2	Make clear how well the system can do what it can do.
G4	Show contextually relevant information.
G5	Match relevant social norms.
G6	Mitigate social biases.
G8	Support efficient dismissal.
G9	Support efficient correction.
G10	Scope services when in doubt.
G11	Make clear why the system did what it did.
G13	Learn from user behaviour.
G15	Encourage granular feedback.
G17	Provide global controls.

From Amershi et al.

Discussion

ID (from [5])	Guideline	Example UXP Insight From Design Session
G1	Make clear what the system can do.	Informing users that AI may make errors, particularly in earlier periods of usage (P18).
G2	Make clear how well the system can do what it can do.	Use hedging language and tell users to consult health experts for conclusive advice (P3).
G4	Show contextually relevant information.	Guide users to take better-lit photos in low-light environments (P8).
G5	Match relevant social norms.	Avoid subjective labelling of meal courses as they may vary across cultures (P19).
G6	Mitigate social biases.	Avoid irrelevant model classes for users with dietary restrictions—e.g. having a dairy class for vegan users (P16).
G8	Support efficient dismissal.	Allow users to manually label images that the AI inaccurately classified (P4).
G9	Support efficient correction.	Provide sliders in the model output UI so users can adjust as needed (P15).
G10	Scope services when in doubt.	Reduce or remove recommendations on how users should eat and live (P13).
G11	Make clear why the system did what it did.	Incorporate short strings that briefly describe features the model is using to make a decision (P21).
G13	Learn from user behaviour.	Observe signs of dietary restrictions early on and ask users if they would like to eliminate unobserved classes (P12).
G15	Encourage granular feedback.	Provide an interface for users to adjust numerical model outputs (P2).
G17	Provide global controls.	Allow users to define, label, and train on their own classes (P23).

From Amershi et al.

Discussion

ID (from [5])	Guideline	Example UXP Insight From Design Session
G1	Make clear what the system can do.	Informing users that AI may make errors, particularly in earlier periods of usage (P18).
G2	Make clear how well the system can do what it can do.	Use hedging language and tell users to consult health experts for conclusive advice (P3).
G4	Show contextually relevant information.	Guide users to take better-lit photos in low-light environments (P8).
G5	Match relevant social norms.	Avoid subjective labelling of meal courses as they may vary across cultures (P19).
[illegible]	[illegible] social biases.	[illegible] restrictions—e.g. having a dairy class for vegan users (P16).
G8	[illegible]	[illegible]
G9	Support efficient correction.	Provide sliders in the model output UI so users can adjust as needed (P15).
[illegible]	[illegible]	[illegible] should eat and live [illegible].
G11	Make clear why the system did what it did.	Incorporate short strings that briefly describe features the model is using to make a decision (P21).
G13	Learn from user behaviour.	Observe signs of dietary restrictions early on and ask users if they would like to eliminate unobserved classes (P12).
G15	Encourage granular feedback.	Provide an interface for users to adjust numerical model outputs (P2).
G17	Provide global controls.	Allow users to define, label, and train on their own classes (P23).

Experienced-based enrichment of human-AI guidelines

Guidelines are naturally realized and operationalized within context-specific design problems and user needs.

From Amershi et al.

Discussion

ID (from [5])	Guideline	Example UXP Insight From Design Session
G1	Make clear what the system can do.	Informing users that AI may make errors, particularly in earlier periods of usage (P18).
G2	Make clear how well the system can do what it can do.	Use hedging language and tell users to consult health experts for conclusive advice (P3).
G4	Show contextually relevant information.	Guide users to take better-lit photos in low-light environments (P8).
G5	Match relevant social norms.	Avoid subjective labelling of meal courses as they may vary across cultures (P19).
G6	Mitigate social biases.	Avoid irrelevant model classes for users with dietary restrictions—e.g. having a dairy class for vegan users (P16).
G8	Support efficient dismissal.	Allow users to manually label images that the AI inaccurately classified (P4).
G9	Support efficient correction.	Provide sliders in the model output UI so users can adjust as needed (P15).
G10	Scope services when in doubt.	Reduce or remove recommendations on how users should eat and live (P13).
G11	Make clear why the system did what it did.	Incorporate short strings that briefly describe features the model is using to make a decision (P21).
G13	Learn from user behaviour.	Observe signs of dietary restrictions early on and ask users if they would like to eliminate unobserved classes (P12).
G15	Encourage granular feedback.	Provide an interface for users to adjust numerical model outputs (P2).
G17	Provide global controls.	Allow users to define, label, and train on their own classes (P23).

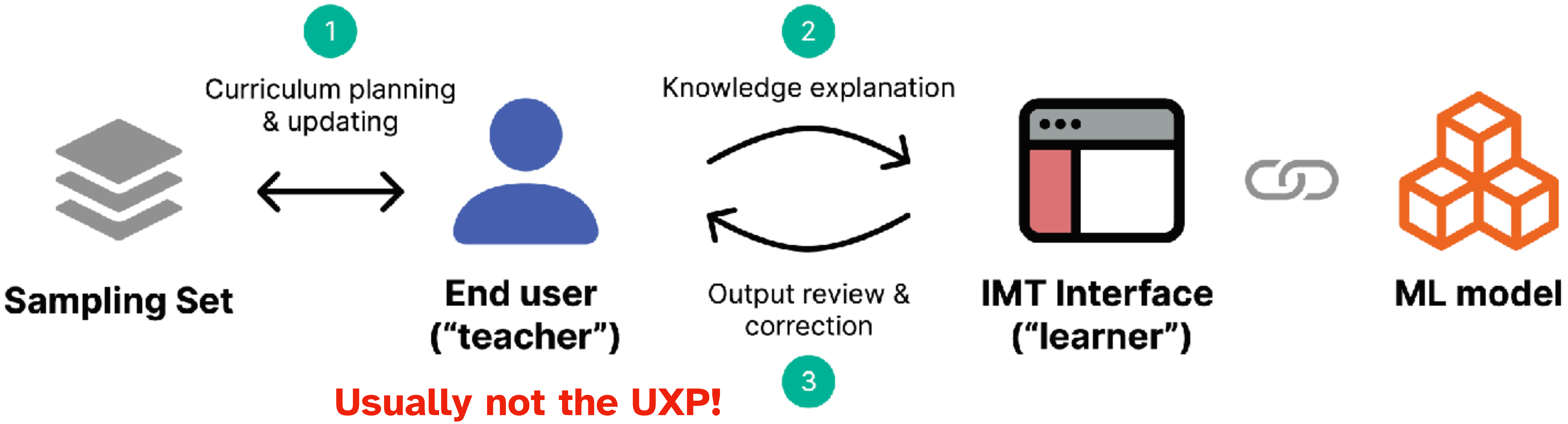
From Amershi et al.

Interactive Machine Teaching

Interactive machine teaching: a human-centered approach to building machine-learned models

Gonzalo Ramos[a], Christopher Meek[a], Patrice Simard[b], Jina Suh[a], and Soroush Ghorashi[b]

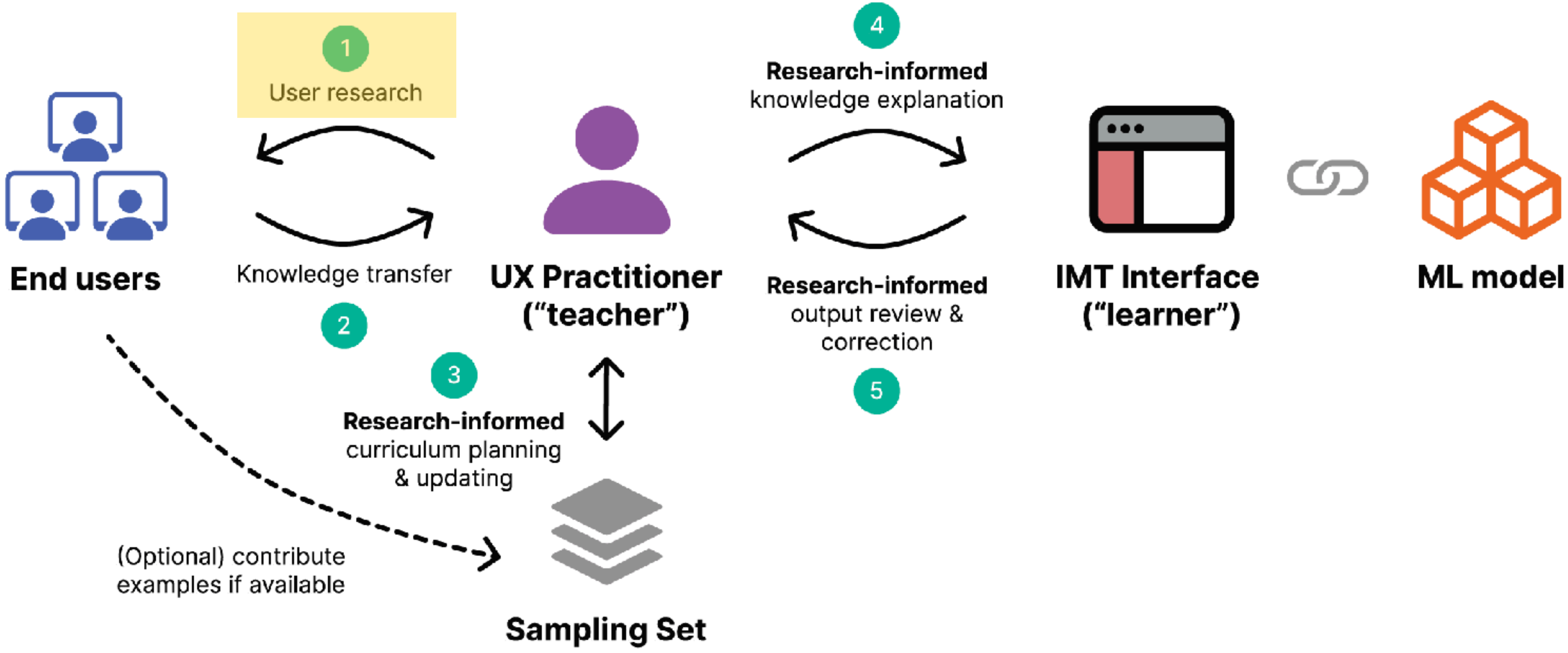
[a]Microsoft Research, Redmond, WA, USA; [b]Microsoft



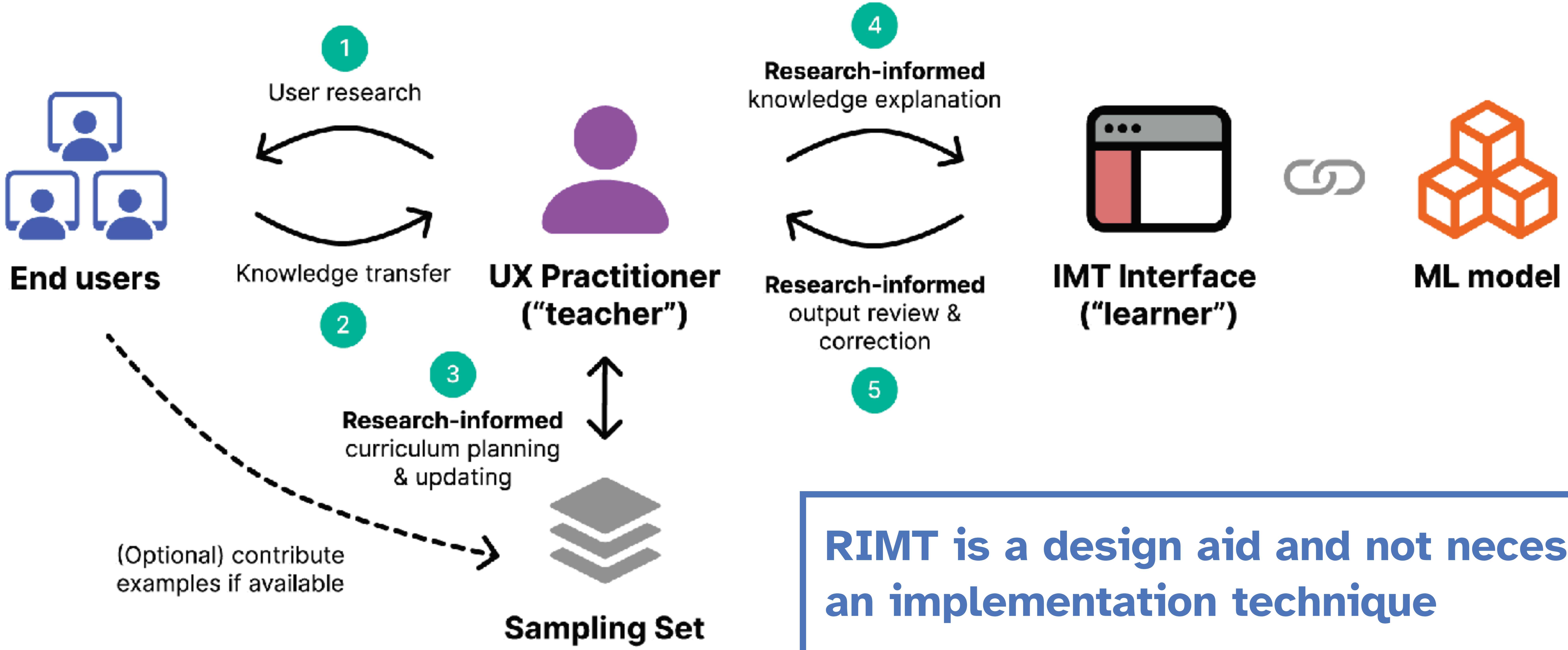
Usually not the UXP!

Discussion

Research Informed Machine Teaching (RIMT)



Research Informed Machine Teaching (RIMT)



RIMT is a design aid and not necessarily an implementation technique

Exciting Directions for Future Work

What does fabrication look like with other ML techniques?

Our study only considered supervised image classification.

How can UXPs work with domain experts in more specialized applications of ML?

Food and diet is a universally relatable domain.

Thanks!

Addressing UX Practitioners' Challenges in Designing ML Applications: an Interactive Machine Learning Approach

K. J. Kevin Feng
University of Washington
kjfeng@uw.edu
@kjfeng_

David W. McDonald
University of Washington

IUI 2023
March 27 – 31, 2023 | Sydney, Australia

HUMAN CENTERED DESIGN & ENGINEERING

DESIGN USE BUILD