How Do UX Practitioners Communicate AI as a Design Material? Artifacts, Conceptions, and Propositions

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

Maxwell James Coppock
University of Washington

David W. McDonald
University of Washington

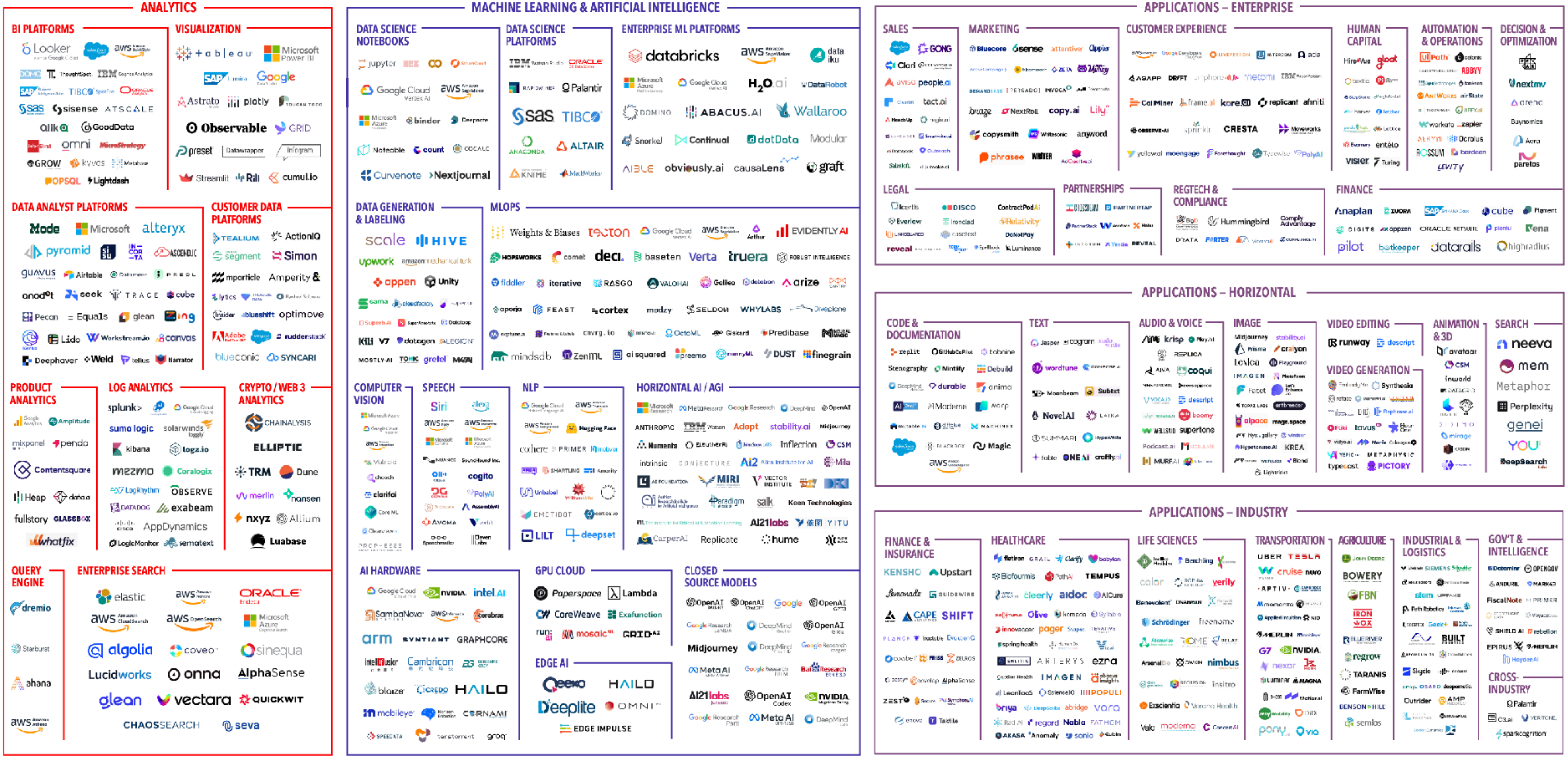
DIS 2023
July 10 – 14, 2023 | Pittsburgh, PA, USA

HUMAN CENTERED DESIGN & ENGINEERING

DESIGN USE BUILD

AI has taken the world by storm.

Illustration: Will Joel at The Verge



Credit: Matt Turck and Firstmark Capital

User experience practitioners (UXPs) should be well prepared to work with AI as a design material

AI as a design material

AI as a design material

Understanding how to creatively use AI, or decide when not appropriate to use it, to solve design problems.

AI as a design material

Understanding how to creatively use AI, or decide when not appropriate to use it, to solve design problems.

Technology & Adoption — CHI 2017, May 6–11, 2017, Denver, CO, USA

UX Design Innovation: Challenges for Working with Machine Learning as a Design Material

Graham Dove, Kim Halskov
CAVI, Aarhus University
Aarhus, Denmark
graham.dove@cc.au.dk, halskov@cavi.au.dk

Jodi Forlizzi, John Zimmerman
HCII, Carnegie Mellon University
Pittsburgh PA, USA
forlizzi@cs.cmu.edu, johnz@cs.cmu.edu

Machine Learning as a UX Design Material: How Can We Imagine Beyond Automation, Recommenders, and Reminders?

Qian Yang
yangqian@cmu.edu
Human-Computer Interaction Institute, Carnegie Mellon University
Pittsburgh, Pennsylvania

CHI 2020 Paper — CHI 2020, April 25–30, 2020, Honolulu, HI, USA

Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design

Qian Yang **Aaron Steinfeld** **Carolyn Rosé** **John Zimmerman**
Carnegie Mellon University, Pittsburgh PA, USA
{yangqian,steinfeld}@cmu.edu cp3a@andrew.cmu.edu johnz@cs.cmu.edu

AI as a design material

Understanding how to creatively use AI, or decide when not appropriate to use it, to solve design problems.

Difficult to gain concrete understanding of properties

Removed from material's fabrication and preparation

Technology & Adoption CHI 2017, May 6–11, 2017, Denver, CO, USA

UX Design Innovation: Challenges for Working with Machine Learning as a Design Material

Graham Dove, Kim Halskov
CAVI, Aarhus University
Aarhus, Denmark
graham.dove@cc.au.dk, halskov@cavi.au.dk

Jodi Forlizzi, John Zimmerman
HCII, Carnegie Mellon University
Pittsburgh PA, USA
forlizzi@cs.cmu.edu, johnz@cs.cmu.edu

Machine Learning as a UX Design Material: How Can We Imagine Beyond Automation, Recommenders, and Reminders?

Qian Yang
yangqian@cmu.edu
Human-Computer Interaction Institute, Carnegie Mellon University
Pittsburgh, Pennsylvania

CHI 2020 Paper CHI 2020, April 25–30, 2020, Honolulu, HI, USA

Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design

Qian Yang, Aaron Steinfeld, Carolyn Rosé, John Zimmerman
Carnegie Mellon University, Pittsburgh PA, USA
{yangqian,steinfeld}@cmu.edu cp3a@andrew.cmu.edu johnz@cs.cmu.edu

...but what if UXPs could fabricate or tinker with the material throughout the design process?

UXPs work in teams

UXPs work in teams

…so communicating design work is crucial.

UXPs work in teams

…so communicating design work is crucial.

How designs are communicated can tell us about how designers think and work.

RESEARCH QUESTION

How do UXPs communicate AI as a design material when they are given an opportunity to tinker with it during the design process?

Method

Method

1:1 task-based design sessions with 27 UXPs.
Each session was 2 hours long.

Most (20) had no prior ML design experience.

METHOD

Session Structure

2 hrs

METHOD

Session Structure

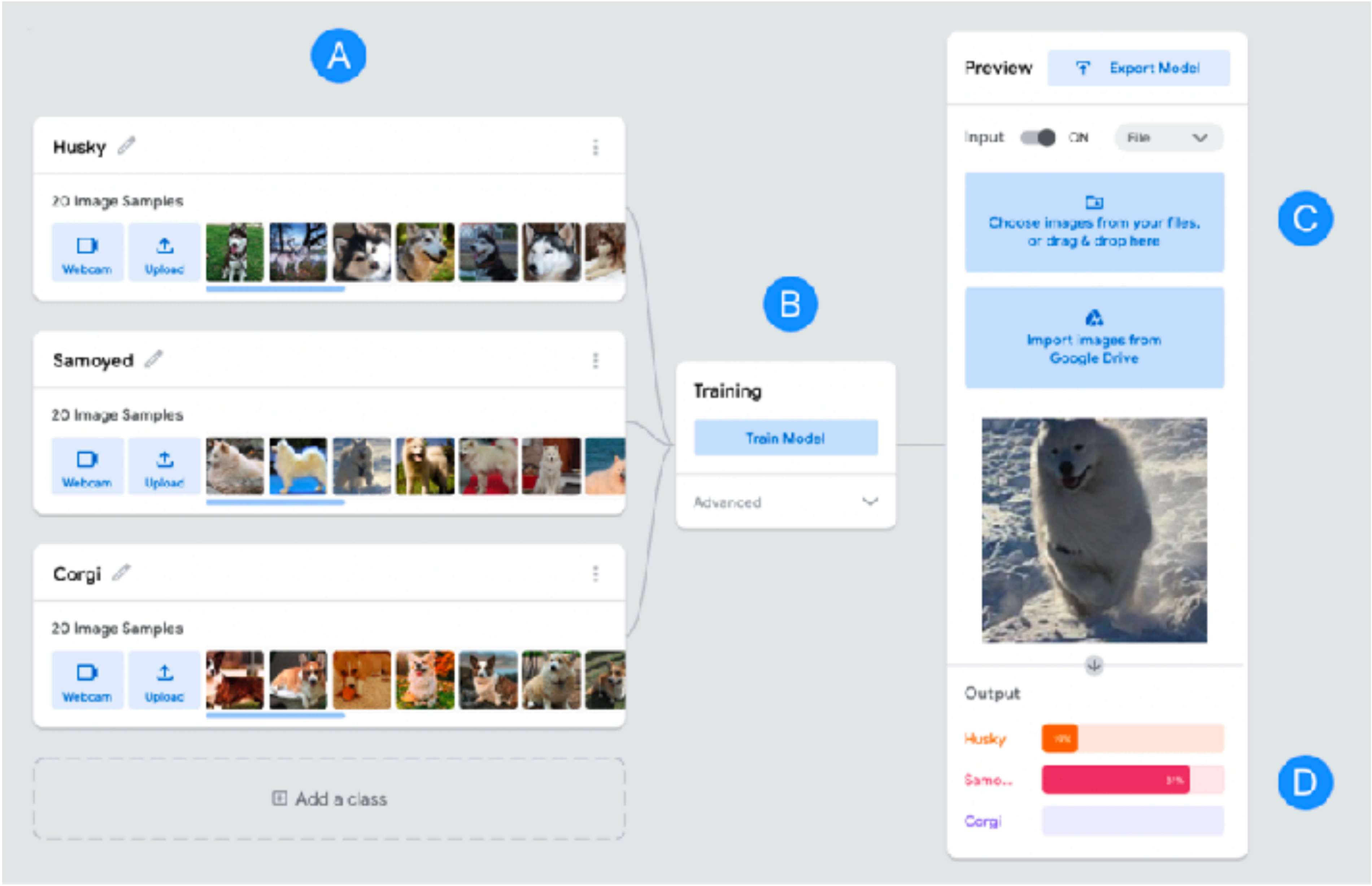
METHOD

Session Structure

15 mins | 1 hr 15 mins | 30 mins

Tutorial

- Gave tutorial of simple interactive AI tool
- Chose Google's Teachable Machine for simplicity and low learning curve



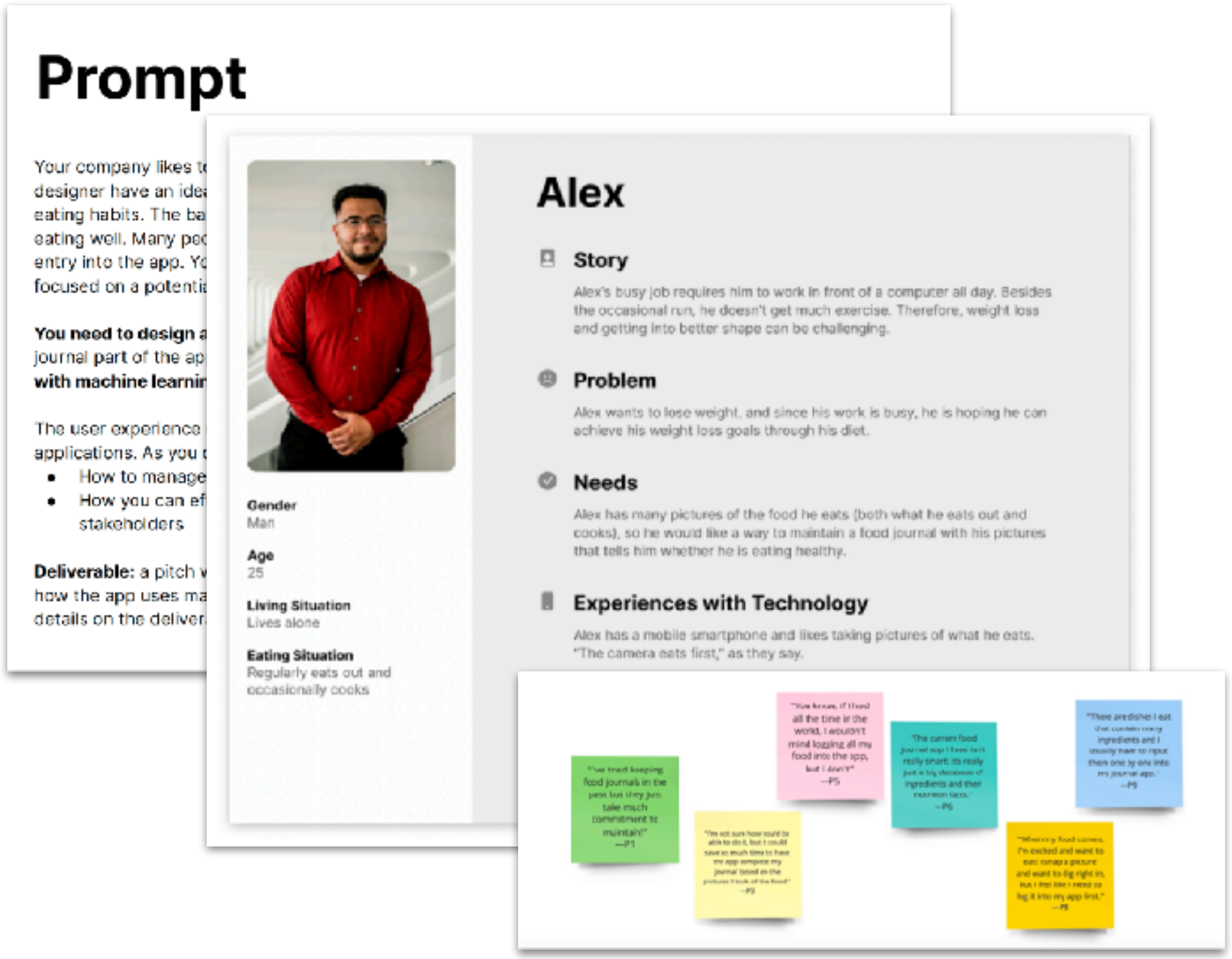
METHOD

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 some materials (personas, research insights) to get started
- Deliverable: short slide deck/ document with their concept



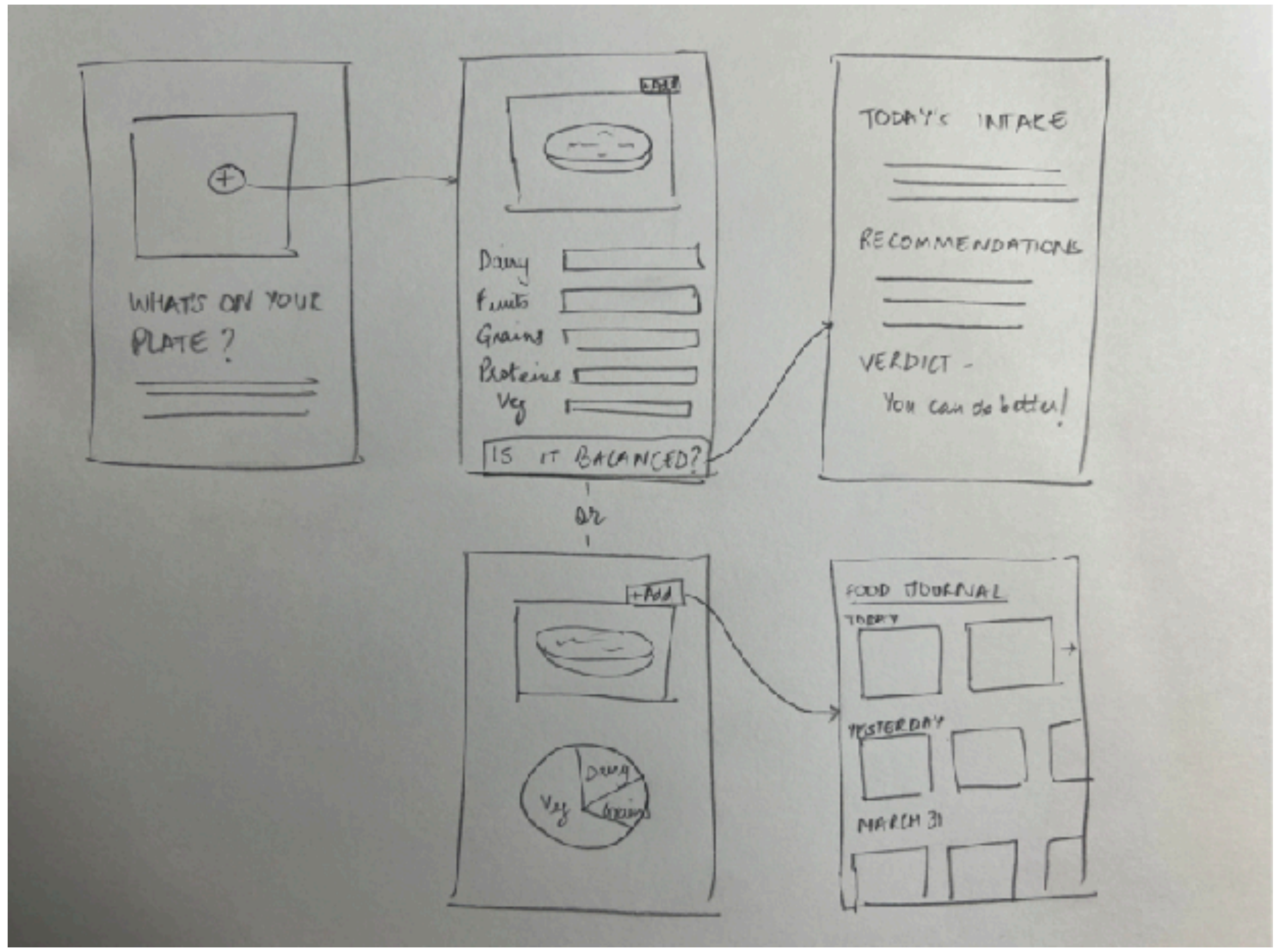
METHOD

Session Structure

15 mins | 1 hr 15 mins | 30 mins

Interview

- Participants reflected on experiences tinkering with AI during the design process
- Walked through certain design decisions they made in their deliverable



Generic
Not Actionable
Informative
01
Time Efficient
Help people quickly upload photos before eating
02
Smart
Provide intelligent features to detect ingredients
03
Food Classification
Classification of breakfast, Lunch, Dinner and snack
Inaccurate analysis
Hard to educate users how to take better pictures
2- class blindly categorizes food into healthy/ unhealth can cause the user to end up having an all or nothing approach relationship with food where they categorize into good and bad and can demonize certain types of groups without fully understanding the role it plays).
3- class is just a categorization of food- it may not alig the goals of the app.
It is important to not have a judgemental approach to diet- explain opportunities to improve- track daily intak carbs/ protein, etc and meet the goals for the same ins using negative phrasing like inadequate, unhealthy, et
ML MODEL - 3 CLASSES
Only categorizing meals, does not contribute to knowing / documenting the type of food people eat
Risk of bias - due to various eating habit and culture difference, some food might be put into a wrong bucket.
Garlic Bread Detected
750 Calories
23g Carbs
13g Fats
1g Protein
3 Jul (detected 3 meals)
Fruits
Dairy
Grains
Meat
Vegetables
Your Goal :
3 cups of Milk
2 cups of fruits
2½ cups of veggies
5½ oz of Meat
6 oz of grains
Today's diet
Yesterday's balance
Add food
apple
Banana
Cake
Smart Food
CONCEPT
Based on our dataset, we estimate your meal to be:
Carbohydrates
Protein
Fat
Estimated calories:
Proof of concept design
Eatr
A smart food journalling app for all your eats
MEAT
VEGETABLES
FRUITS
MILK
GRAINS
How it Works
Dairy
Fruits
Grains
martDiet
ck what you eat,
omatically
Classification models explored
Two-class model: "healthy" vs "unhealthy"
The design team recommends against this approach in general; our app's goal is not to make broad value judgments towards the user's specific dietary needs and requirements.
Three-class model: appetizers; desserts; entrees
Pros:
In preliminary testing, this model was the most accurate according to our UXD's personal classification of the test images
Cons:
Does not allow for cultural variation of different cuisine types
Does not allow for other meals such as breakfast, snacks, etc.
Five-class model: dairy, fruits, grains, proteins, vegetables
Pros:
These categories will be fairly appropriate across different cultural contexts
Allows users to decide which types of foods they want to consume more or less of
Cons:
In preliminary testing, the classification seems to rely on image color as a primary identifier, which may not be effective in different cuisines or even beyond certain types of meals in normative "American" cuisine
MI Food
\ mī fūd \
Re-imagine food journaling

METHOD

Data Analysis

Qualitative coded design presentations as a team.

Supplemented data with interview transcripts.

Findings

FINDING

UXPs considered accuracy to be important, but struggled to communicate accuracy in their designs.

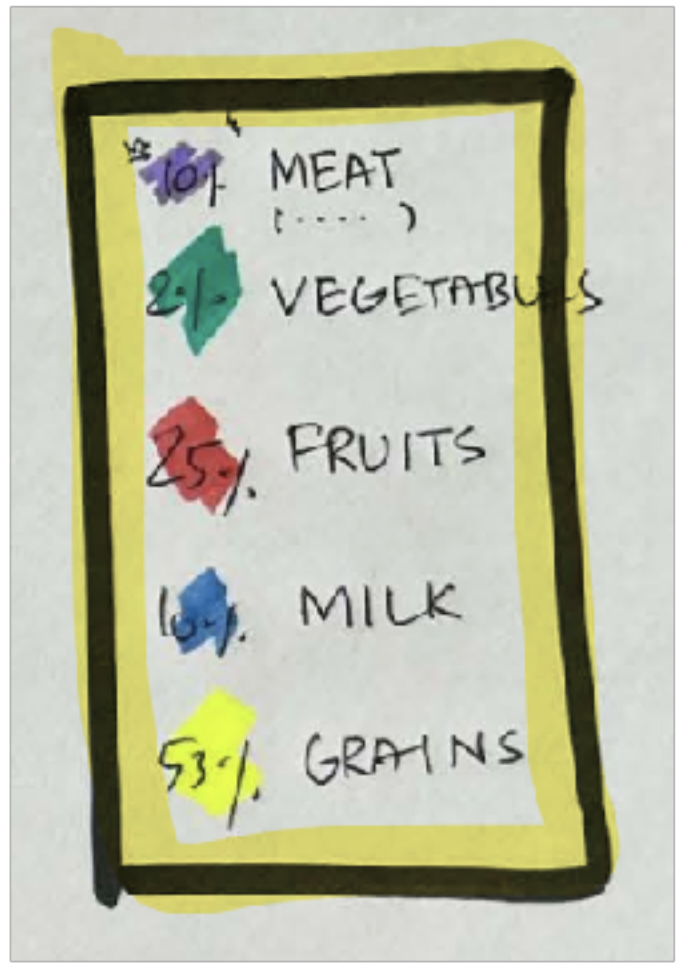
“One of the design goals that I had for the app was that it gives accurate and reliable information.” (P19).

“If [the AI]’s not accurate, then the rest [of the design] is meaningless.” (P17).

Participant Presentations

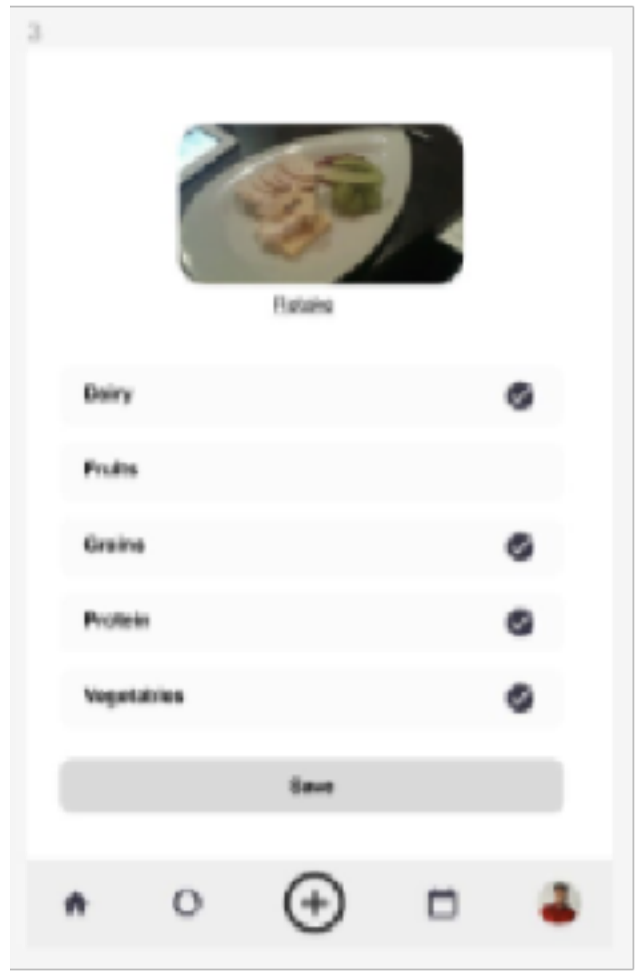
P4

P6

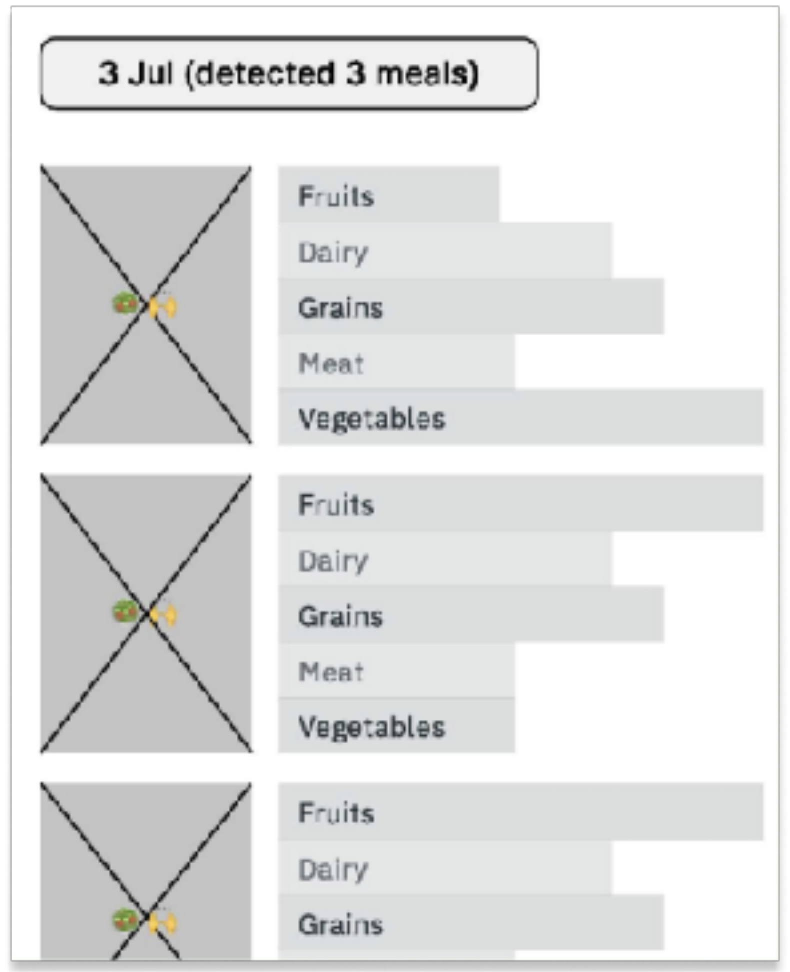


P14

P15



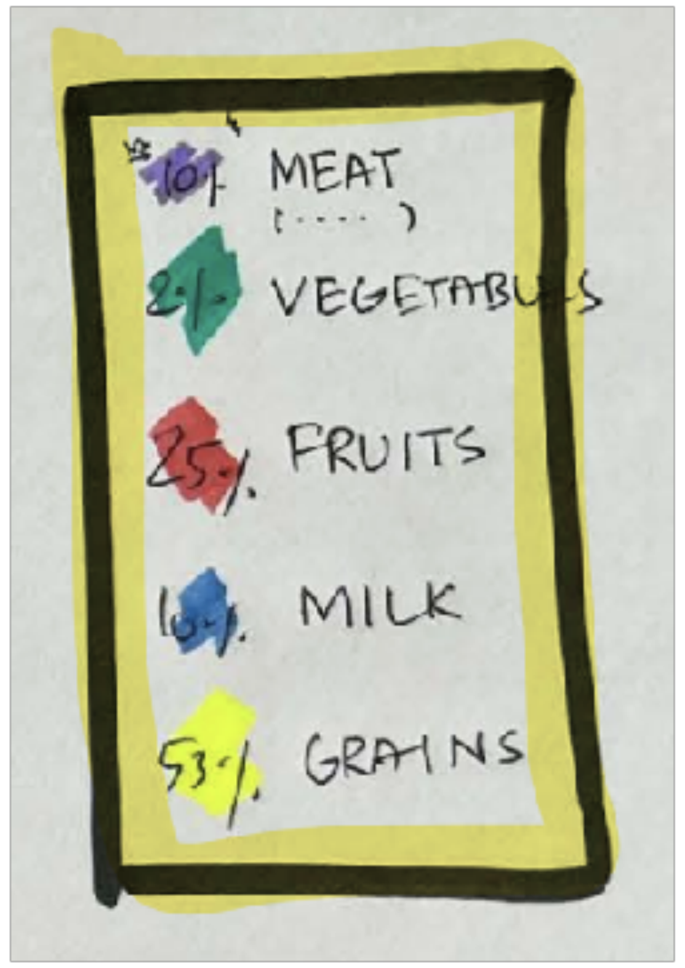
P26



Participant Presentations

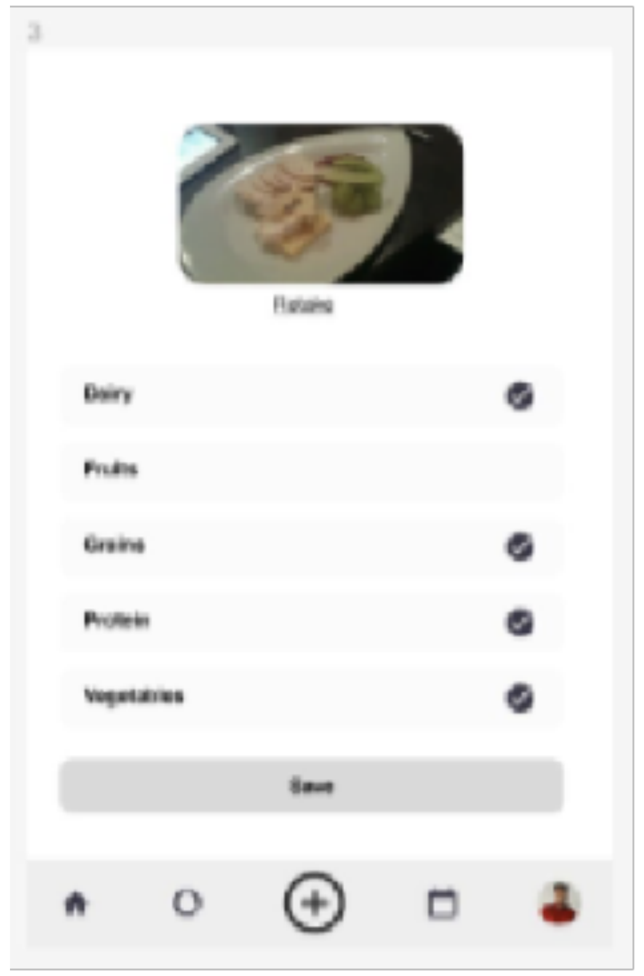
P4

P6

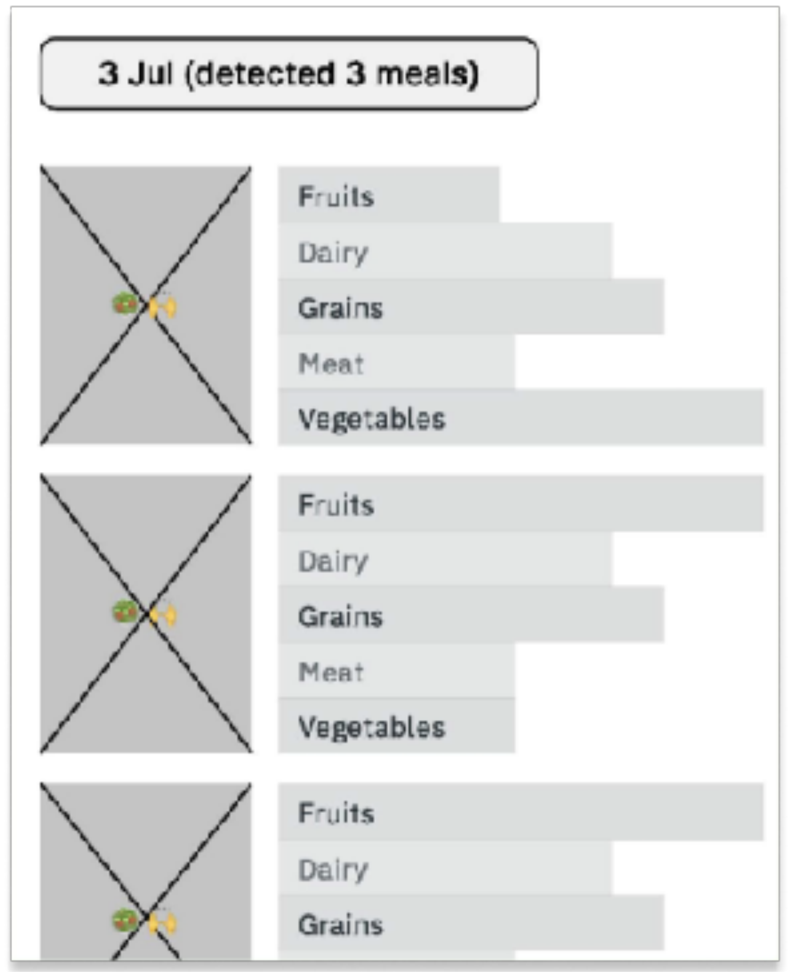


P14

P15

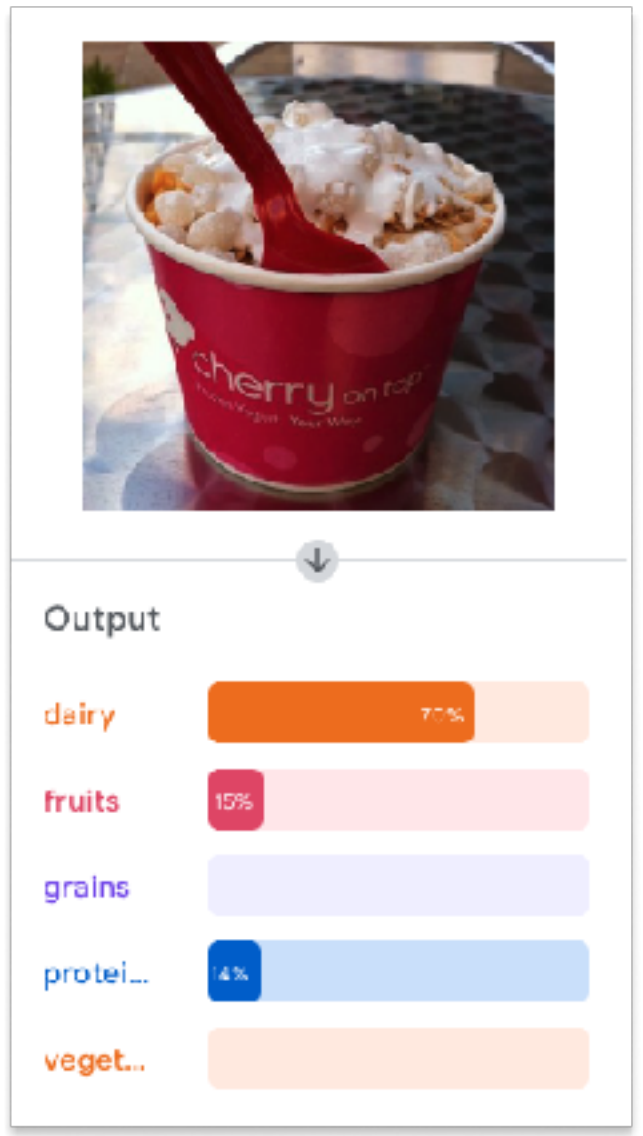


P26

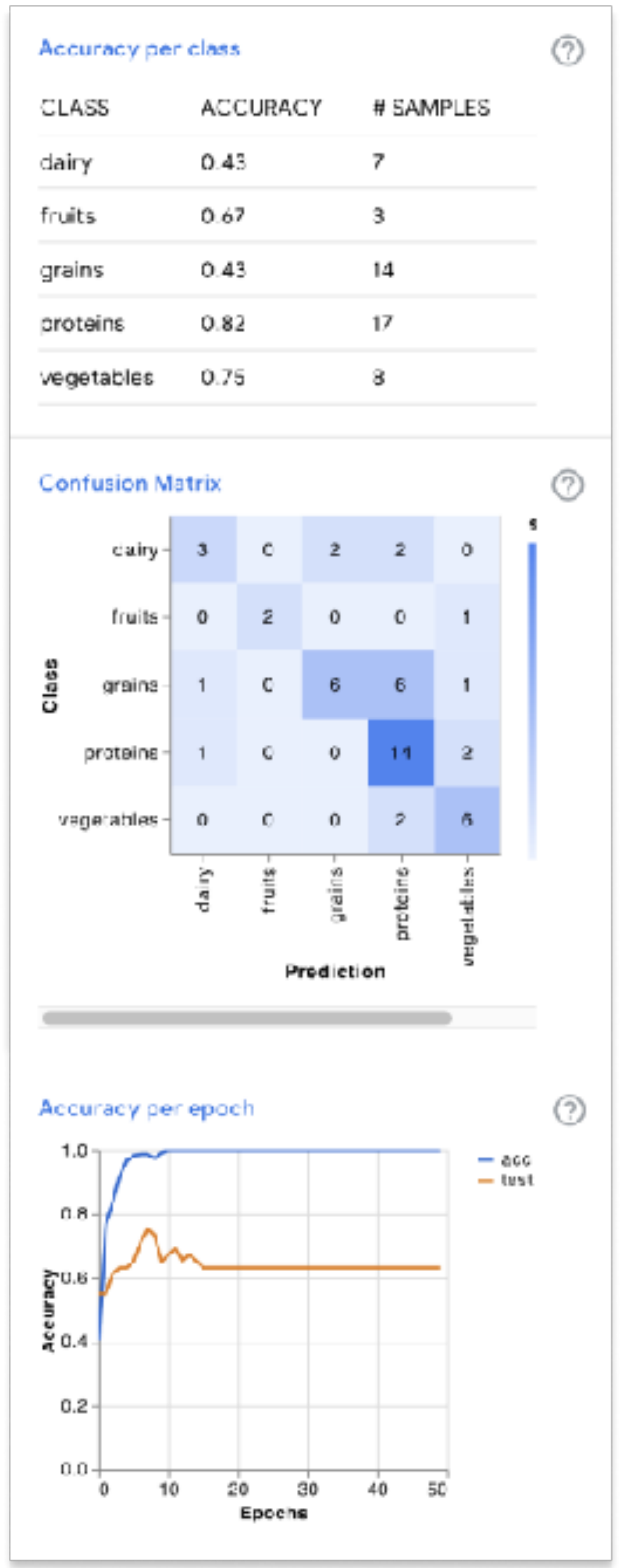


Teachable Machine

Output module



Accuracy in Advanced menu



CLASS	ACCURACY	# SAMPLES
dairy	0.43	7
fruits	0.67	3
grains	0.43	14
proteins	0.82	17
vegetables	0.75	8

Designerly evaluation

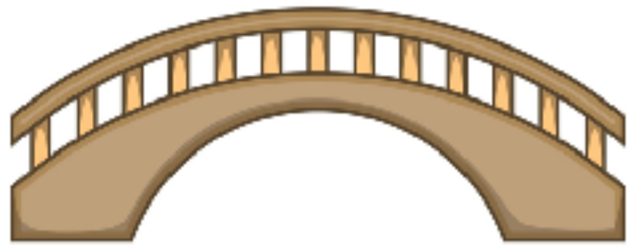
- Compatibility of accuracy with user needs.
- Alignment with design goals.
- Calibration of user trust.

Technical evaluation

- Accuracy = (TP + TN) / (TP + TN + FP + FN).
- Benchmark comparisons.
- Precision, recall, F1 score, ROC AUC, etc.

Designerly evaluation

- Compatibility of accuracy with user needs.
- Alignment with design goals.
- Calibration of user trust.



Technical evaluation

- Accuracy = (TP + TN) / (TP + TN + FP + FN).
- Benchmark comparisons.
- Precision, recall, F1 score, ROC AUC, etc.

FINDING

Tinkering with AI models shows promise in improving communication with other stakeholders.

FINDING

Tinkering with AI models shows promise in improving communication with other stakeholders.

“The teams I work with are so called ‘machine learning’ teams, but to be honest, looking at our current site experience, I have ***zero faith in what this team does****, or their capability of using machine learning to deliver something relevant. [...] so I always question* ***what’s the purpose of the machine learning team?****”* (P5).

FINDING

Tinkering with AI models shows promise in improving communication with other stakeholders.

“I don’t want to be a data scientist, like I don't want to have to train [the production-level model] or edit it.” (P23).

FINDING

Tinkering with AI models shows promise in improving communication with other stakeholders.

"A good start to getting our knowledge on the same grounds." (P13).

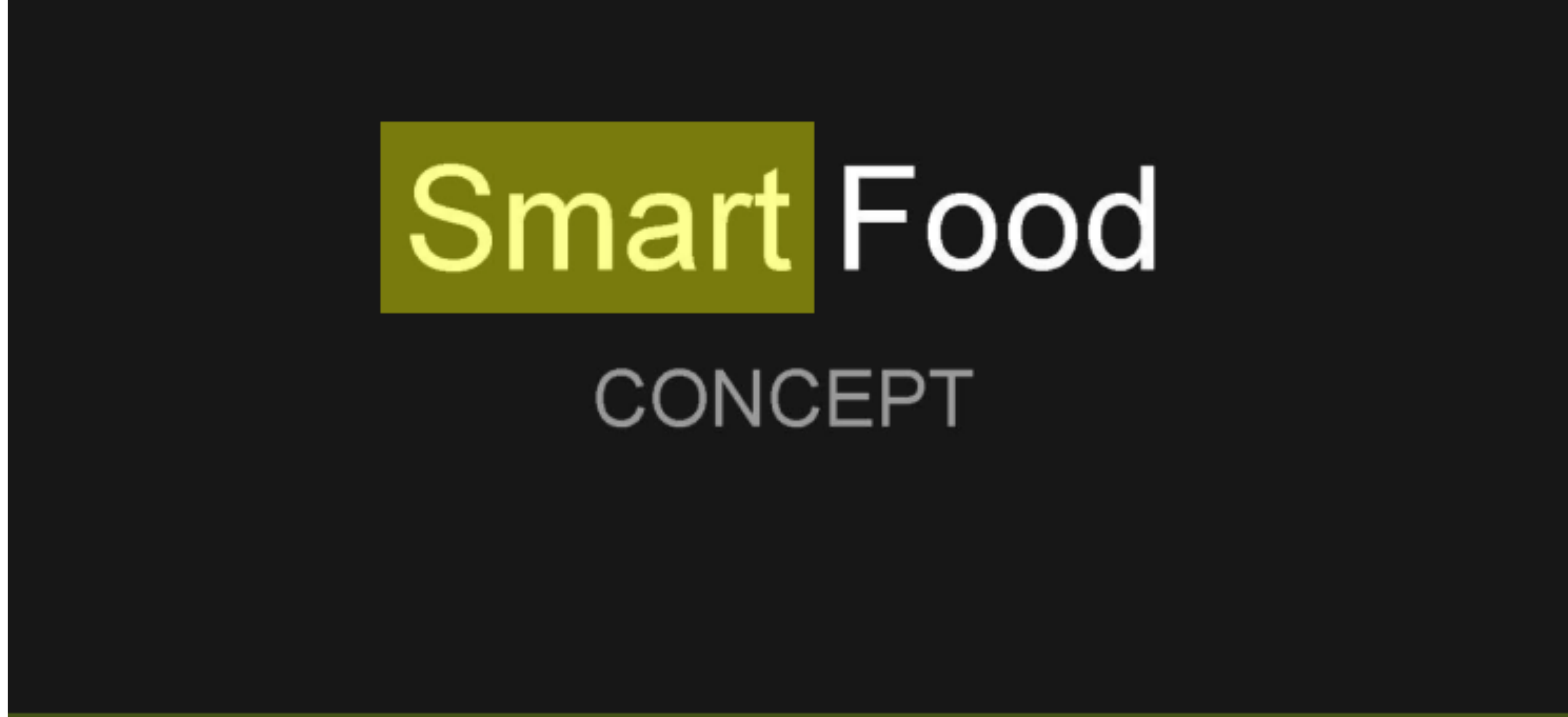
"Maybe I can [draft up a model] myself, test it, and give some comments to my engineers and we can discuss together." (P12).

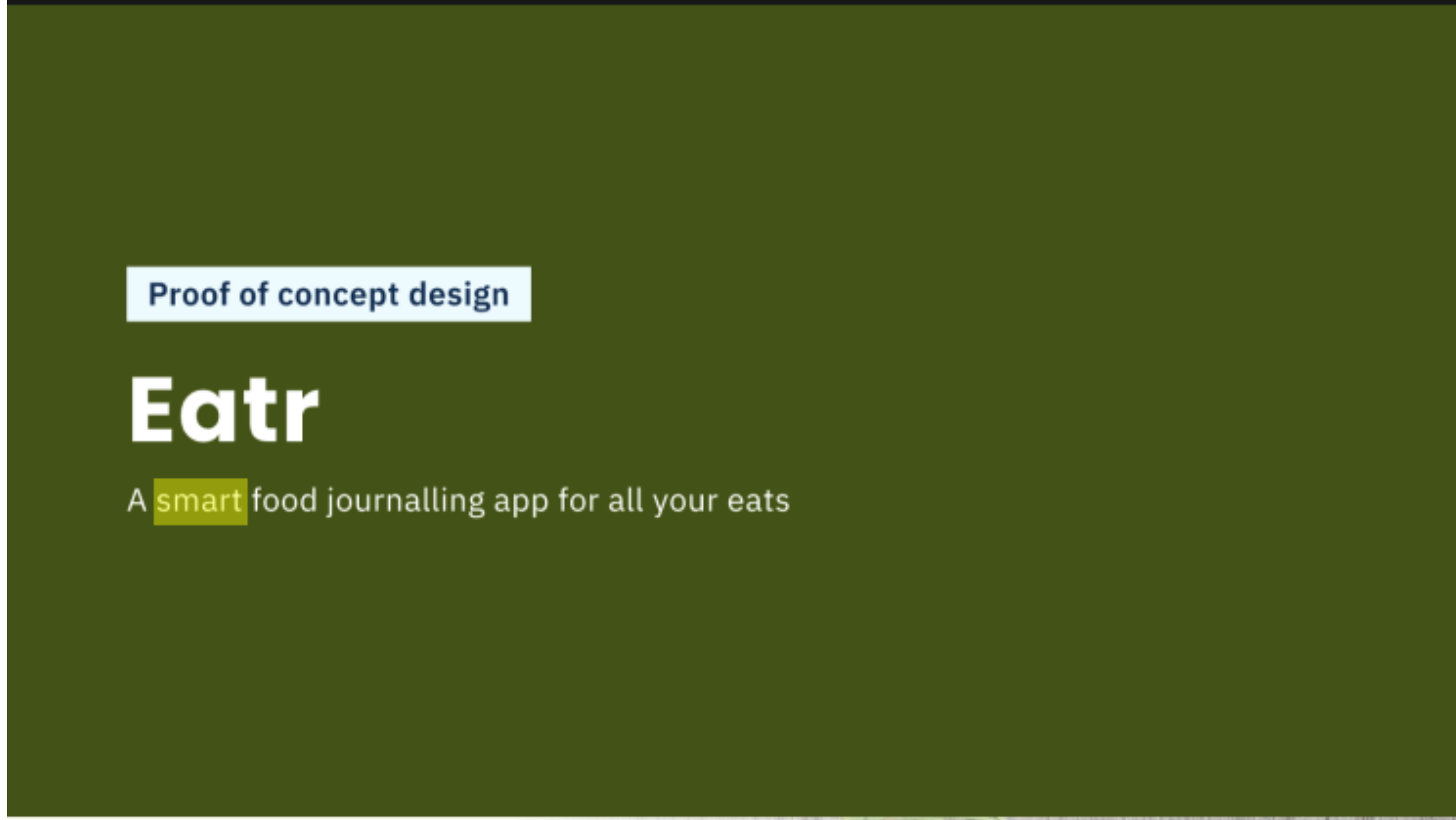
"Usually executives have a hard time imagining what [the AI-enabled design] could look like if you don't show a visual." (P5).

FINDING

UXPs identified promises and perils of AI in the context of their designs.

AI can make the app “smart.”





“Smart”

Make personalized recommendations based on user data.

Automating repetitive tasks (e.g., manually inputting ingredients).

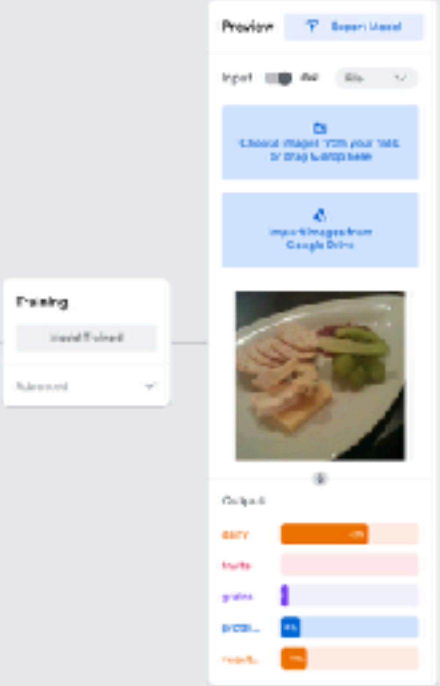
“Instead of describing the general ‘rules’ to eating healthy, [AI can] try to throw in a [personalized] health tip like ‘it is recommended that __ g of protein must be consumed every day.” (P16).

“[Automating manual input] reduces the amount of energy and increases their motivation to stay on track, so I think that's really powerful.” (P26).

Risks of AI

Risks

- Classifying photos may not be accurate enough, and it may take too much time to correct
    - Should get better over time, as more people make corrections?
    - Assumption: Training data from one user can be shared to improve the overall classification
- Not sure what the laws/regulations are with respect to sharing data



Potential risks in UX

Inaccurate analysis

Hard to educate users how to take better pictures

Considerations

- We need to train our model with photos from many environments to account for variability in angle, lighting, etc.
- We can recommend the percentages of the food pyramid, but does that add enough value?
- We can add more value in the journaling aspect, adding tags at upload to track emotions
- Approve language with Legal – this is not medical advice, consult a medical professional for any health issues

Risks

- Bias - Care should be taken to ensure that the model is trained with foods from as many cultures as possible (cultural bias)
- Bias - Definitions of "healthy" vary from person-to-person and from culture-to-culture. The model will favor the USDA definition of healthy.
- Mis-categorization - Food is a sensitive topic for some and there could be shame/fear in being told you are eating poorly
- Mis-categorization - Users over/under eat certain foods because the app gave them misleading information

Data Inaccuracies

For the most part, the AI was able to determine what is healthy and what is unhealthy. Some results, however, are inaccurate and require further training.

SWOT

S	W
Benefits/ Strengths Uses photos instead of writings- visuals are more engaging No need to write	**Weaknesses** may not have much other use other than food journalling

O	T
Opportunities community engagement motivational articles	**Risks/ Threats** Gives a narrow way of viewing nutrition feels more like a food diary Relies on Habit-forming behaviors

Risks of AI

Risks associated with properties of AI:

Poor accuracy
Lack of explainability
Not trustworthy and reliable

Risks of applying AI to food & nutrition:

Misleading health advice
No input from medical experts
Inability to account for cultural context

Implications

Low fidelity

- Does not accurately represent final product
- Easy to create
- Cheap to iterate

High fidelity

- Accurate representation of final product
- Takes effort to create
- Expensive to iterate

Low fidelity

- Does not accurately represent final product
- Easy to create
- Cheap to iterate

Test assumptions

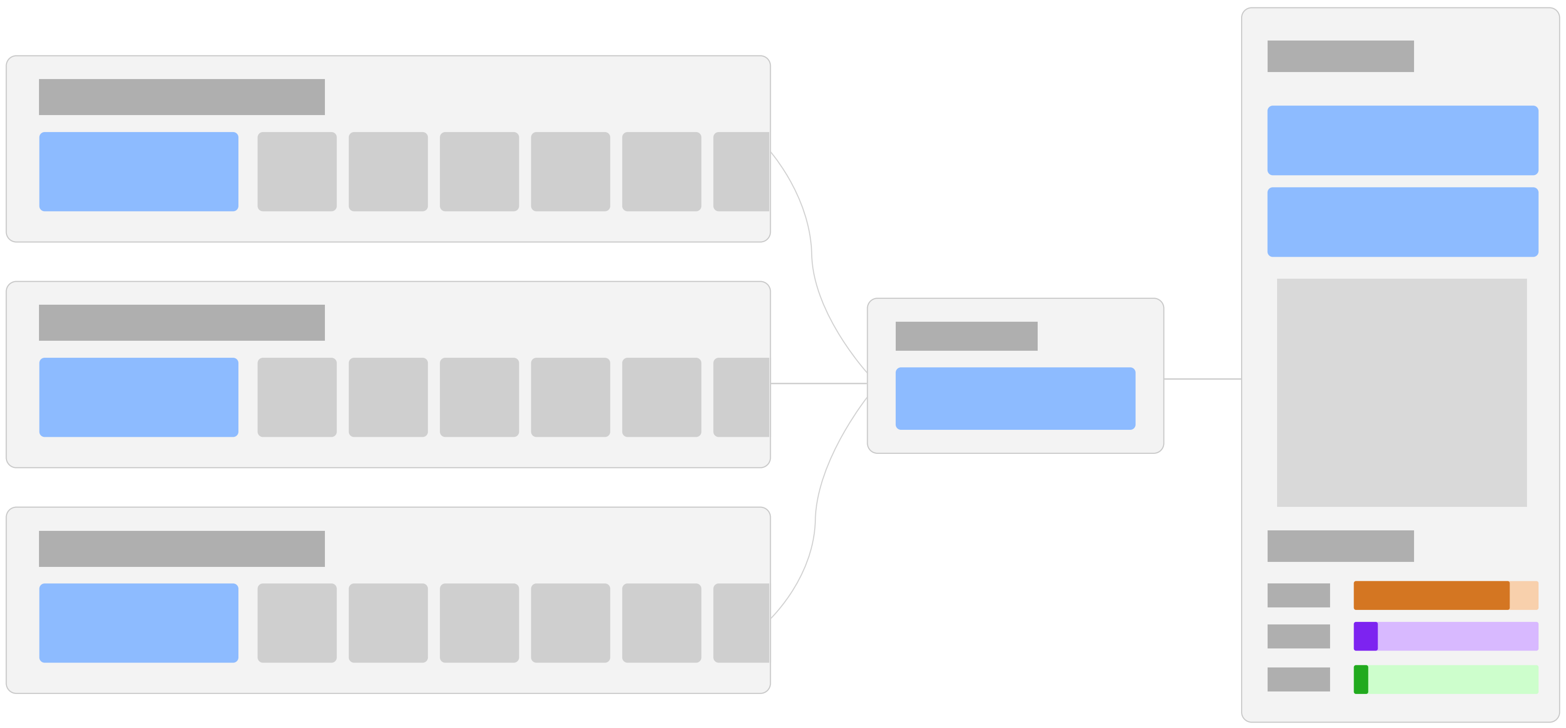
Include user feedback

Identify usability flaws

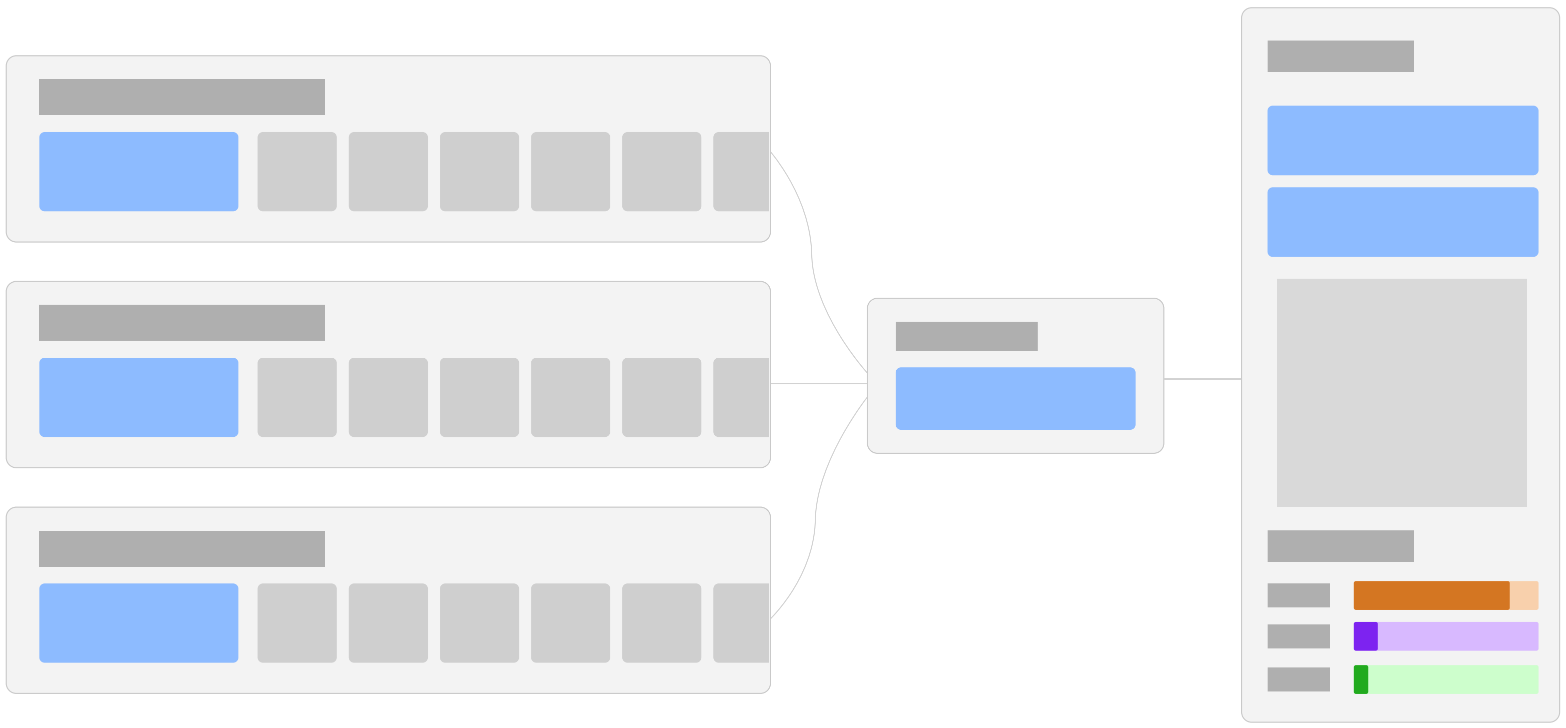
High fidelity

- Accurate representation of final product
- Takes effort to create
- Expensive to iterate

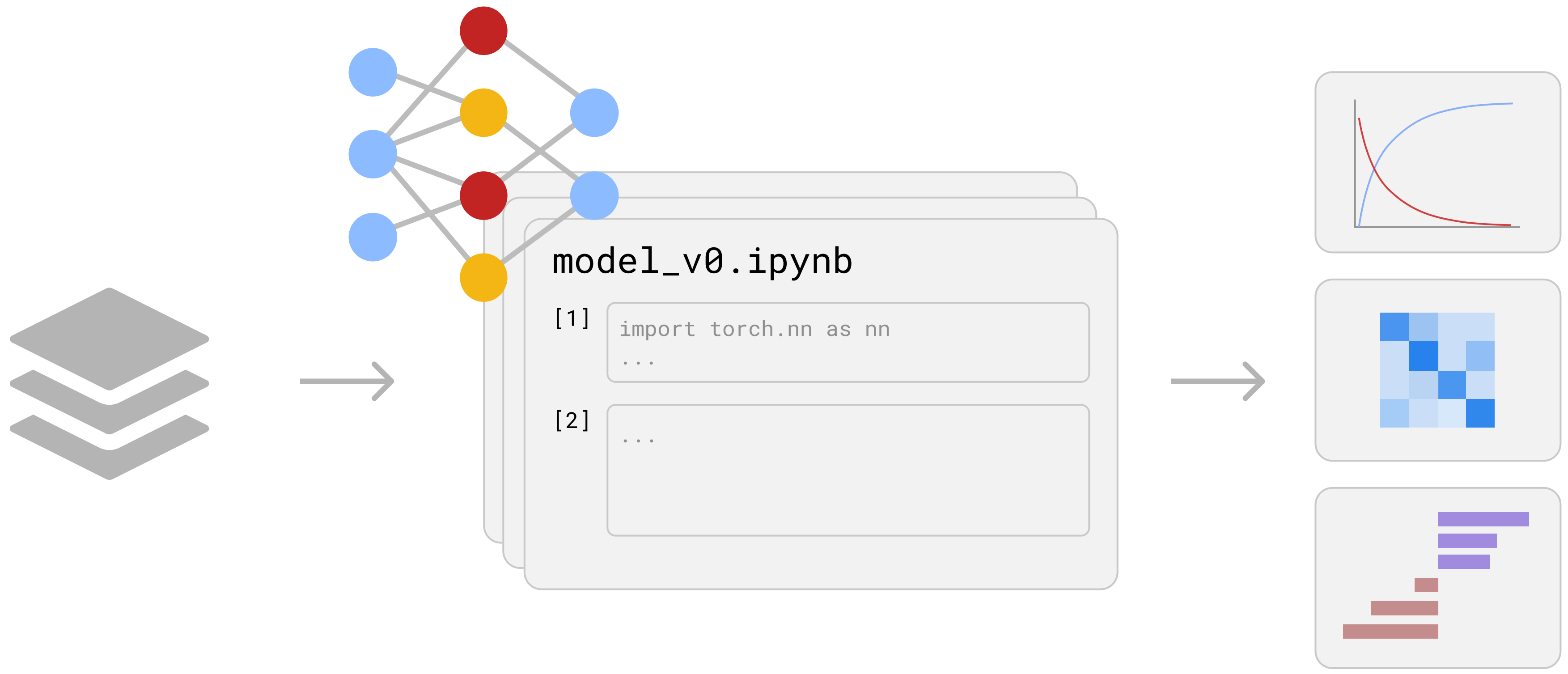
Low fidelity

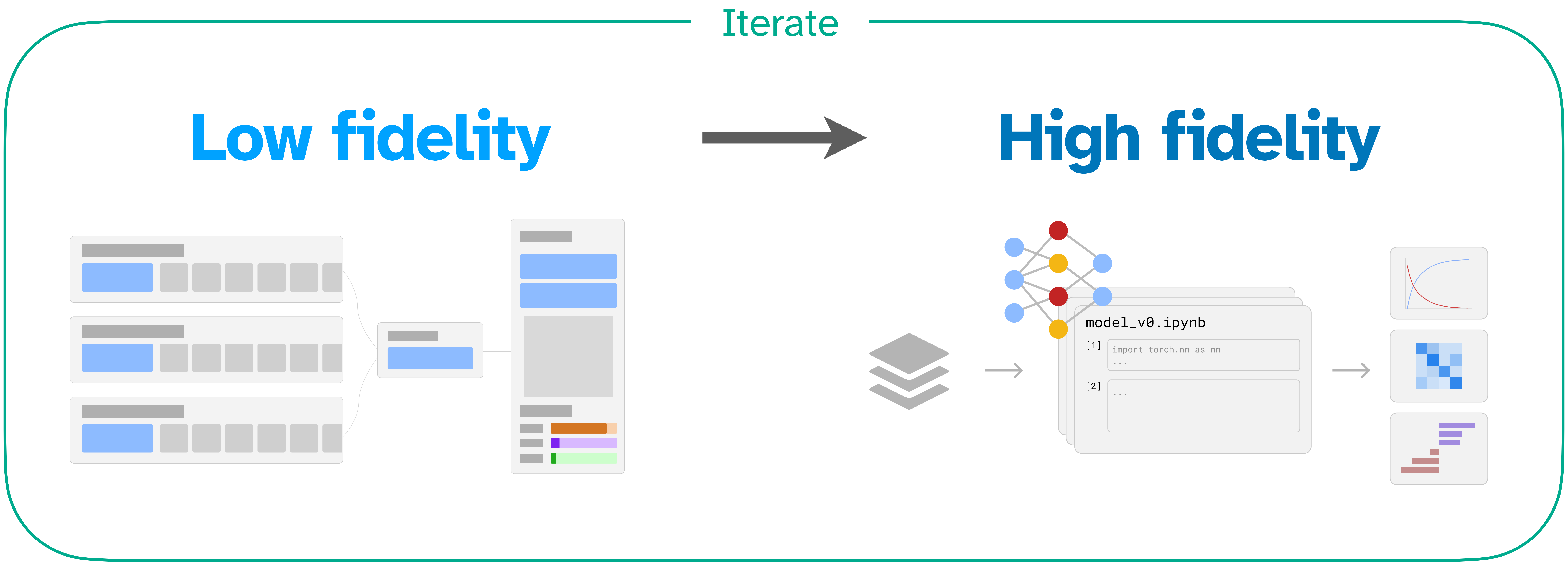


Low fidelity



High fidelity




Iterate
Low fidelity
High fidelity
model_v0.ipynb
[1]
import torch.nn as nn
...
[2]
...


Iterate
Low fidelity
High fidelity
model_v0.ipynb
[1]
import torch.nn as nn
...
[2]
...

LOOKING AHEAD

Directions for Future Work

Will communication strategies vary depending on who the stakeholder is?

Engineer vs domain expert vs business executive.

LOOKING AHEAD

Directions for Future Work

Will communication strategies vary depending on who the stakeholder is?

Engineer vs domain expert vs business executive.

How might AI tinkering change with different AI paradigms?

We only used supervised image classification models in our study.

Thanks!

How Do UX Practitioners Communicate AI as a Design Material? Artifacts, Conceptions, and Propositions



K. J. Kevin Feng
Maxwell James Coppock
David W. McDonald
University of Washington

Scan for full paper

We'd like to extend a special thanks to Amodini Khade and Katrina Vergara, along with all our participants and reviewers.

DIS 2023
July 10 – 14, 2023 | Pittsburgh, PA, USA

HUMAN CENTERED DESIGN & ENGINEERING

DESIGN USE BUILD