Creating Design Resources to Scaffold the Ideation of AI Concepts

Nur Yildirim
Carnegie Mellon University
Pittsburgh, PA, USA
yildirim@cmu.edu

Changhoon Oh
Yonsei University
Seoul, Korea
changhoonoh@yonsei.ac.kr

Deniz Sayar
Izmir Ekonomi Universitesi
Izmir, Turkey
deniz.sayar@ieu.edu.tr

Kayla Brand*
Wellesley College
Wellesley, MA, USA
kb102@wellesley.edu

Supritha Challa*
UW Madison
Madison, WI, USA
srchalla2@wisc.edu

Violet Turri*
Carnegie Mellon University
Pittsburgh, PA, USA
vmturri@sei.cmu.edu

Nina Crosby Walton*
Washington University
St. Louis, MO, USA
cnina@wustl.edu

Anna Elise Wong*
UC Santa Cruz
Santa Cruz, CA, USA
anewong@ucsc.edu

Jodi Forlizzi
Carnegie Mellon University
Pittsburgh, PA, USA
forlizzi@cs.cmu.edu

James McCann
Carnegie Mellon University
Pittsburgh, PA, USA
jmccann@cs.cmu.edu

John Zimmerman
Carnegie Mellon University
Pittsburgh, PA, USA
johnz@cs.cmu.edu

ABSTRACT

Advances in artificial intelligence have enabled unprecedented technical capabilities, yet making these advances useful in the real world remains challenging. We engaged in a Research through Design process to improve the ideation of AI products and services. We developed a design resource capturing AI capabilities based on 40 AI features commonly used across various domains. To probe its usefulness, we created a set of slides illustrating AI capabilities and asked designers to ideate AI-enabled user experiences. We also incorporated capabilities into our own design process to brainstorm concepts with domain experts and data scientists. Our research revealed that designers should focus on innovations where moderate AI performance creates value. We reflect on our process and discuss research implications for creating and assessing resources to systematically explore AI's problem-solution space.

CCS CONCEPTS

• **Human-centered computing → Interaction design process and methods**.

KEYWORDS

User experience, artificial intelligence, human-centered AI, ideation

ACM Reference Format:
Nur Yildirim, Changhoon Oh, Deniz Sayar, Kayla Brand, Supritha Challa, Violet Turri, Nina Crosby Walton, Anna Elise Wong, Jodi Forlizzi, James McCann, and John Zimmerman. 2023. Creating Design Resources to Scaffold the Ideation of AI Concepts. In *Designing Interactive Systems Conference (DIS '23), July 10–14, 2023, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3563657.3596058

*Work done as a research intern at Carnegie Mellon University, equal contribution.

DIS '23, July 10–14, 2023, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-9893-0/23/07.
https://doi.org/10.1145/3563657.3596058

1 INTRODUCTION

Advances in artificial intelligence (AI) have enabled many unprecedented capabilities: AI systems drive cars, translate between languages, and discover new drugs. The prevalence of AI in everyday products and services suggests that our community has a robust AI innovation process. Interestingly, research indicates the opposite. Today, more than 85% of AI innovation projects fail; they fail to co-create value for users and services for a variety of reasons [25, 43, 84]. Many breakdowns stem from a lack of human-centered design; HCI is often not involved until the choice of what innovation to make has already happened [50, 62, 70]. Practitioners report repeatedly experiencing AI project failures due to working on the wrong problem – solutions that do not address real needs [94].

Researchers point out that many AI failures can be traced back to problem selection and formulation [68, 94]. Data science teams often do not systematically elicit needs from domain experts (users) and product managers. Without this input, they envision AI systems users do not want [50, 52, 89, 94]. Practitioners in product roles (e.g. designers, product managers) lack an understanding of what AI can reasonably do. They envision AI concepts that cannot be built [19, 89, 91]. Teams tend to envision complex solutions and seem to miss low hanging fruit – situations where simple AI would improve user experience (UX) [91]. In addition, engaging domain stakeholders in early phase AI development remains a great challenge [50, 76, 97].

In recent years, resources in the form of human-AI guidelines and design patterns have become available [1, 2, 67]. However, practitioners report that guidelines mostly help with prototyping and refining – *making the thing right* [11]. What designers and product managers most strongly lack are resources to help with ideation and problem framing: *"What are the problems that we can solve for these users by employing AI?"* [94]. Designers also reported tensions when following a user-centered approach to create AI innovations [86, 94, 96]. To overcome these challenges, internal resources capturing AI capabilities and examples that demonstrate how AI gets utilized in existing products and services have been created [93, 94]. While the resources seem to be useful for brainstorming and envisioning AI innovations, they are only internally available for a few, select teams, and they are limited to capabilities and examples relevant to an individual team's product domain.

As a team of HCI researchers and designers, we set out to improve the process of envisioning AI products and services. We thought that more effective ideation would reduce the risk of producing AI innovations that cannot be built or that do not address real needs. We took a Research through Design approach, engaging in an expansive design process that spanned four years. We conducted three design experiments where the outcome of each experiment reframed our research goals and questions:

Design Experiment 1 focused on creating a resource that captures AI capabilities and examples. We focused on capabilities repeatedly found in commercial products and services to keep design ideation within a space of what is possible. We curated a collection of 40 AI features used in many products and services across a wide range of domains *(e.g., spam filter, language translation, product review analytics)*. Using a bottom-up process, we iteratively analyzed this collection of examples. This resulted in a resource of 8 high-level capabilities, 40 AI examples with many detailed capabilities, and a grammar for describing and extending the resource with new capabilities and examples. This experiment led us to speculate on how the resource might impact ideation.

Design Experiment 2 focused on understanding the usefulness of this resource. How and when should it be considered in the design process? What should more successful ideation and related outcomes look like? To explore these questions, we created a set of slides documenting high-level AI capabilities and examples. We asked designers to ideate new AI features before and after reviewing the slides. Our resource helped them to consider more AI capabilities, but designers still generated ideas that would be difficult to build. They mostly focused on situations that require near-perfect model performance. This experiment revealed *AI model performance* as a key consideration, implying that innovators should search for places where moderate model performance is useful. We also observed that a user-centered approach (identifying pain points prior to considering what AI can do) unintentionally limited the ideation of buildable concepts and the exploration of the problem-opportunity space.

Design Experiment 3 explored a different ideation process. Designers facilitated ideation with domain experts and data scientists instead of ideating on their own. The process blended user-centered and technology-centered approaches to simultaneously consider both AI capabilities (what AI does well) and user needs. We conducted ideation sessions exploring how AI might improve critical care medicine practiced in the intensive care unit (ICU). We drew from our resource a subset of AI examples where moderate model performance produced value. We probed domain experts (i.e., physicians, nurses, fellows) to identify needs that matched AI capabilities, and we probed data scientists to understand if the concepts could be built. This approach yielded many ideas that were both low-risk in terms of technical feasibility and medium to high value for clinicians. The process seemed to provide many better ideas that could function as the starting place of an AI innovation development effort.

Our paper makes four contributions:

(1) an extensible resource of AI examples, capabilities, and a grammar for describing the capabilities and extending the resource;
(2) discovery of model performance as a critical yet overlooked consideration when ideating AI concepts and mapping AI's problem-opportunity space;
(3) an improved ideation approach for overcoming the tension between user-centric and technology-centric innovation approaches;
(4) and a first-person case demonstrating the potential impact of this improved brainstorming approach to generate low-risk, high-value AI concepts.

The resources we produced – examples, capabilities, and grammar – are released as open-source[1] for the research community to create new design tools, methods, and exercises that support designers in envisioning AI concepts. We discuss the implications for future research aiming to support the ideation and problem selection within the context of AI innovation.

2 RELATED WORK

HCI distinguishes sketching, generating many ideas for *making the right thing*, and prototyping, iterative refinement for *making the thing right* [11]. In the early days of technology development, many software products failed as they did not address a real human need [11, 14]. Human-centered design became widely adopted as it effectively reduces the risk of developing technology people do not want through sketching many different solutions. Recent work revealed that similar to early software development, AI projects increasingly fail due to a lack of human-centered design [84, 94]. Against this backdrop, our work draws from research that explores sketching methods and approaches for new technologies, and research that investigates designing with data and AI as design materials.

2.1 Ideating with Technology

HCI employs many methods and activities for *sketching*, such as brainstorming, wireframing interaction flows, and writing scenarios [11]. HCI and design practitioners gain an understanding of the capabilities and limitations of a technology as they explore novel design spaces using these methods [71, 85]. Sketching and ideating with new or partially understood technologies remains challenging [11]. To ease these challenges, UX designers and researchers have generally followed one of two approaches. First, they engage emerging technologies (e.g., internet of things [80], haptics [60], software [66]) through design-led inquiry, including research through design and speculative design [30]. These first-person accounts of envisionment typically result in design exemplars, case studies, and concepts that illustrate the technology's capabilities and experiential possibilities to practitioners. For example, Moussette investigated the design space of haptics through haptic sketches, a set of physical prototypes that embodied a wide range of haptic sensations [60].

A second, complementary approach is when design researchers develop a conceptual understanding of technology as a design material, often through a meta-analysis of multiple design-led inquiries.

[1] aidesignkit.github.io.

Here, researchers analyze the experiential attributes of technology as it relates to interaction design, and provide abstractions of the technology's capabilities and experiential qualities. The result is intermediate-level design knowledge in the form of conceptual frameworks and abstractions that support designers in effectively envisioning with emergent technologies [17, 38, 55, 63, 65].

2.2 Ideating with Data and AI

Data and AI are challenging materials for UX design. HCI literature has explored AI's design challenges, including explainability [53], trust [46, 73], algorithmic bias that creates harm [6, 87], privacy [18], and inference errors [36, 59]. In response, practitioner-facing resources in the form of design patterns and guidelines have became available to help address many of these challenges (e.g., [1, 2, 67]). A recent study investigating how product teams use these resources shows that guidelines mostly help with prototyping and refining – *making the thing right* [11]. What designers and product managers need now are new resources that help with ideation and problem framing, asking *"how can we discover problems where AI might offer an effective solution?"* [94].

Researchers have investigated AI's design challenges around envisioning and ideation, mainly by conducting design-led inquiry [4, 8, 9, 20, 54, 56, 64, 69, 78]. For example, Yang et al. [89] ideated with NLP capabilities to generate many novel concepts for an intelligent writing assistant. These first-person accounts of sketching with data and AI provide case studies and design concepts that offer generative lenses. Other research investigated how design practitioners engage data and AI as design materials [13, 19, 90, 93, 99]. These studies show that designers find it difficult to grasp what AI can and cannot do, and they frequently envision ideas that exceed AI's capabilities and cannot be built. Instead of leveraging AI capabilities that are immediately available, designers seem to often focus on emergent AI capabilities where there might not be any existing AI libraries, pre-built models, or labeled datasets [91].

A growing number of studies investigated the emergent industry best practices for designing with data and AI [16, 31, 44, 79, 86, 90, 93, 94, 96]. This line of work revealed that designers who effectively envision AI products and services work with *designerly abstractions* of AI capabilities and product examples that embody a capability (e.g., *predicting user intent*, as in chatbots). Some design teams created internally available resources detailing AI capabilities and examples to scaffold brainstorming [93, 94]. These resources captured AI capabilities as action verbs (e.g., *discover, identify, create, recommend*) instead of technical AI terms (e.g., *supervised learning, neural networks*). For example, they described an AI capability where the system could "see" text on packaging, and "read" text to find ingredients that the user might be allergic to. These abstractions and exemplars supported designers in gaining an understanding of *what* AI can do. AI capability resources also made brainstorming sessions more accessible, allowing designers to collectively brainstorm with data scientists, AI engineers, and domain stakeholders. Studies also noted tensions with user-centered design and technology-centered development (e.g., matchmaking [7]), observing emergent design practices that blend these two approaches [91, 94, 96].

Current designer-facing AI resources developed often detail how AI functions [24, 35, 45], or operationalize AI capabilities to enable designers to "play with" AI, for example, allowing users to build their own classifiers to recognize gestures [12, 23, 27, 57, 81]. Other work focused on creating tools, processes, or datasets to facilitate design-oriented data exploration [26, 32, 51, 64, 74]. Building on evidence from studies of practitioners, researchers investigated extracting AI capabilities from HCI literature [88] and from patents [41]. A recent study proposed a tangible AI capability toolkit to support design students in learning and ideation [39]. However, it is unclear whether these capability abstractions can support practitioners or how, when, and in what form they might be integrated into AI product development. Our work makes an advance by capturing AI capabilities and examples that are commonly used in real world products and services, and detailing how these might be useful for designers.

3 OVERVIEW OF THE DESIGN PROCESS

We wanted to improve the ideation of AI products and services. We wanted to overcome problems of envisioning things users do not want and envisioning things that cannot be built. We chose to use Research through Design (RtD), a reflective approach to research that focuses on reframing a problematic situation through making and critiquing [29, 72, 98]. RtD generates knowledge as a proposal. HCI literature has a rich history of design research proposing new methods and approaches that improve the practice of design (e.g., [15, 28]). Similarly, we set out to advance design practice by improving the ability of designers to engage AI – a new, challenging design material.

Two important concepts when capturing and articulating knowledge while using RtD are *design experiments* and *drift* [5, 48, 49, 100]. A design experiment includes any design move researchers make to explore, investigate, and gain insight into their research questions. RtD programs often involve several design experiments that repeatedly probe the same problem space. Design experiments often create friction with the research question. They cause RtD researchers to reframe, to change their perspective and ask new questions. In this way, RtD programs *drift with intention* [48, 100].

Our team involved HCI researchers and designers with backgrounds in interaction design, service design, and computer science who had many years of experience designing human-AI interactions. We set out to investigate the preferred future for envisioning and ideating AI products and services. To explore this problem space, we conducted three design experiments over the course of four years. Each experiment caused drift, leading us to reframe our research goals. Design Experiments 1 & 2 draw from prior research that explored the experiential qualities of new technology materials to produce abstractions and conceptual frameworks that support ideation [63, 65]. Design Experiment 3 builds on design-led inquiry that investigated AI as a design material to provide first-person accounts of ideation [18, 89]. Below, we provide a brief overview of each design experiment and a summary of findings. We unpack each experiment in subsequent sections by detailing the research goals, our design process, and our reflections on the insights gained.

Design Experiment 1: *Can we identify AI capabilities in ways that are useful for designers?* In this section, we detail how we created a resource capturing AI capabilities and examples. This experiment resulted in a resource of 8 high-level capabilities, collection of 40 AI examples with granular capabilities, and a grammar for capturing and extending this resource with new capabilities and examples. The experiment led us to further probe the usefulness of the resource we collected and curated.

Design Experiment 2: *Can designers use AI capability abstractions and examples to improve their ideation process? How can we assess whether ideation is better?* We detail a failed pilot study involving ideation sessions with HCI students. This experiment revealed the importance of *AI model performance*, and resulted in a Task Expertise-AI Performance matrix. The analysis of AI examples on the matrix suggested the need to search for situations where moderate performance creates value. The experiment also surfaced tensions around the user-centered design process when designers have predetermined that AI is the solution.

Design Experiment 3: *Can designers sensitize innovation teams to look for opportunities where moderate model performance might be valuable?* We assembled an interdisciplinary team of data scientists and critical care clinicians, and we brainstormed AI concepts for the intensive care unit (ICU). We found that starting with examples of AI systems that create value with moderate model performance helped the team generate concepts that were valuable and low-risk. The experiment revealed that an innovation process blending user-centered and technology-centered approaches leads to better ideation.

4 DESIGN EXPERIMENT 1: COLLECTING AI EXAMPLES AND CAPABILITIES

We wanted to create a collection of AI examples and capabilities to help designers more effectively ideate AI concepts. We wanted to create a public resource, building on successful resources used in industry practice [93, 94]. We had three requirements:

(1) **Capabilities over mechanisms.** We wanted to capture capabilities; *what* AI can do. This is in contrast to most AI literature that focuses on describing mechanisms; *how* AI makes an inference (e.g., deep neural networks, etc).
(2) **Useful.** We wanted a *useful collection* that could guide designers and non-data scientists away from envisioning things that cannot be built. We cared less about capturing *everything* AI might possibly do. We wanted this resource to capture what designers could reasonably ask AI to do.
(3) **Extensible.** We wanted the resource to be extensible. AI keeps growing and changing, and we wanted to make it easy to add new examples and capabilities to keep up with the advances in technology.

4.1 Design Process

Our iterative process involved three main activities: collecting examples, drawing out and abstracting capabilities, and critiquing our emerging resource. This process took four years and involved several complete restarts. We kept working on this until we achieved what felt like a stable collection of AI examples, detailed low-level capabilities, and links showing how these lead to eight high-level capability abstractions. As part of the process, we met with AI experts to discuss and critique our collection in order to discover gaps and missing capabilities.

One continuous challenge was defining what counts as AI, a point the experts repeatedly raised. Prior research noted an absence of discussion on *"what AI means as it relates to HCI or UX design"* [91]. Unfortunately, there is no agreed upon definition of AI, even within the AI research community. While the term is broadly used, it is also disputed and its meaning remains in flux [77]. In our search for AI examples and capabilities, we chose not to employ any specific definition of AI. Instead, we used "artificial intelligence" as a search term and accepted the search results we got back. We view this as an *operational definition of AI* [91] that collectively comes from people writing about and discussing AI.

Table 1: Our collection of 40 AI Examples across 14 domains.

Domain	AI Example
Education	Automated Essay Scoring
	Personalized Lesson Plans
Energy & Infrastructure	Home Energy Optimization
	Predictive Maintenance
Finance	Robotic Invoice Processing
	Stock Trading Recommendations
Governance & Policy	Child Welfare Risk Assessment
	Infectious Disease Forecasting
Healthcare	Drug Discovery
	Medical Imaging Analysis
	Synthetic Health Data Generation
	Smartwatch Workout Detection
Hospitality	Review Analytics
	Smart Pricing
Human Resources	Resume Screening
	HR Chatbot
	Workforce Scheduling
Leisure, Content & Media	Smart Speaker Question Answering
	Media Feed
	Game Player
	Image Style Transfer
	Mobile App Face Filter
	Deepfakes
Manufacturing & Agriculture	Crop Monitoring
	Defect Detection
	Robotic Pick and Place
Marketing & Sales	AR Item Viewer
	Personalized Advertisements
	Web Usage Analytics
Office Productivity & Business Workflow	Text Generation
	Spam Filter
	Language Translation
	Meeting Summarization
Risk Mitigation & Security	Biometric Security
	Fraudulent Transaction Detection
Science	Aerial Wildlife Monitoring
	Weather Prediction
Transportation	Lane Departure Prediction
	Navigation Route Planner
	Autonomous Parking

4.1.1 *Collecting AI Examples.*

We first generated a small set of AI examples by drawing on our personal experience designing human-AI interaction. The initial set included 15 products and product

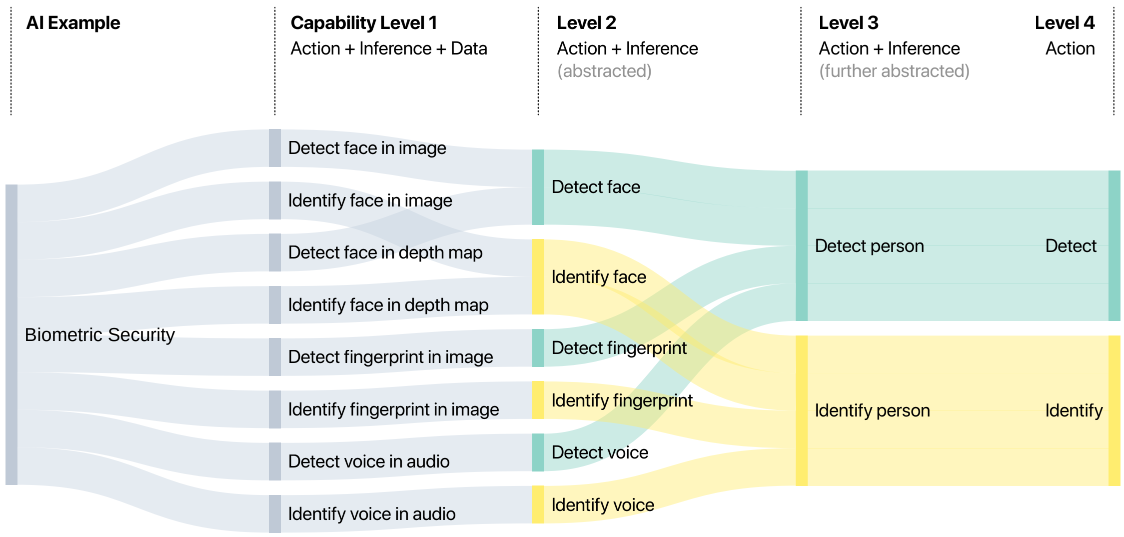


Figure 1: AI example *Biometric Security* has eight unique capabilities converging into high-level capabilities *Detect* and *Identify*.

features across various tech companies *(e.g. Amazon Alexa, Google Translate, etc)*. Our critique of this initial set raised several criteria we used for the remainder of the project to improve our selection of examples: granularity, generality, and breadth.

Granularity. Our initial set of examples mixed whole products, like Amazon's Alexa, and product features, like a spam filter found in most email clients. Products proved to be way too complicated. They often involved many unrelated AI capabilities as well as lots and lots of non-AI technology. We refocused on AI-enabled features within products and services *(e.g. email spam filter, smart speaker question answering, fraudulent transaction detection)*. For the remainder of the project, when we critiqued the examples, we focused on if the feature felt self-contained and if it matched the level of granularity of the other examples.

Generality. Our initial set of examples included AI specific to a single company and AI features found across companies. For example, Alexa was specific to Amazon while spam filters could be found in many email applications. Given our focus on supporting ideation, we decided to focus on AI features that were not bound to any specific company. The fact that a feature repeatedly showed up across companies and products offered a soft guarantee that it could be financially viable and technically achievable. We felt this quality could increase the ideation of buildable AI concepts.

Breadth. We noticed our initial set of examples almost exclusively contained consumer-facing products and services. We had examples from mostly mobile apps and online services. We did not have things like *Defect Detection* used in manufacturing nor any business-to-business services. We realized we needed to broaden our search beyond our personal AI experience to better capture more of the ways AI could co-create value for different stakeholders.

We shifted our search strategy, first focusing on identifying a set of industrial domains. We conducted online searches for industries most impacted by AI. We synthesized the various lists we found. Our lists came from industry-focused news and media, research articles, and white papers. Our synthesis resulted in a list of 14 domains (Table 1). Next, we searched for the most common AI applications and features for each domain. From these lists, we selected two to four examples for each domain. We then searched across all of the examples and eliminated ones that had a large overlap. Our process was impacted by critiques to examine granularity and generality, and by our meetings with AI experts to find gaps. Our final list included 40 examples related to the 14 domains. For each example, we created a short definition, described how value was co-created between the service and the customer, and we classified the example as being either business-to-business or business-to-consumer [see supplementary materials].

4.1.2 Extracting Capabilities from AI examples. We conducted a bottom up analysis of the examples, identifying the specific capabilities each required. We searched for explanations, triangulating across various sources including research papers, business and news articles, marketing product descriptions, and API documentation. In deciding what should count as a capability, we made distinctions between the *inference*, and the *reaction* an application has following the inference. For example, email applications classify emails as spam or not spam and then sort them into the inbox and spam folder. We considered the classification of the email as an AI capability. We did not include automatically sorting classified documents, viewing this as disconnected from the AI capability. Similarly, we worked to separate the user interface presentation of AI output (its form) from the capability. For example, we captured that a retail service's recommender compares and ranks all items for sale as an AI capability. However, we viewed the choice to present these as product recommendations as a design choice and not as an AI capability.

We searched for an appropriate form to capture capabilities by writing terse descriptions. As we worked across examples and critiqued our efforts, a simple grammar emerged: *Action + Inference + Data/Entity/Metric*. Each example had several capabilities captured

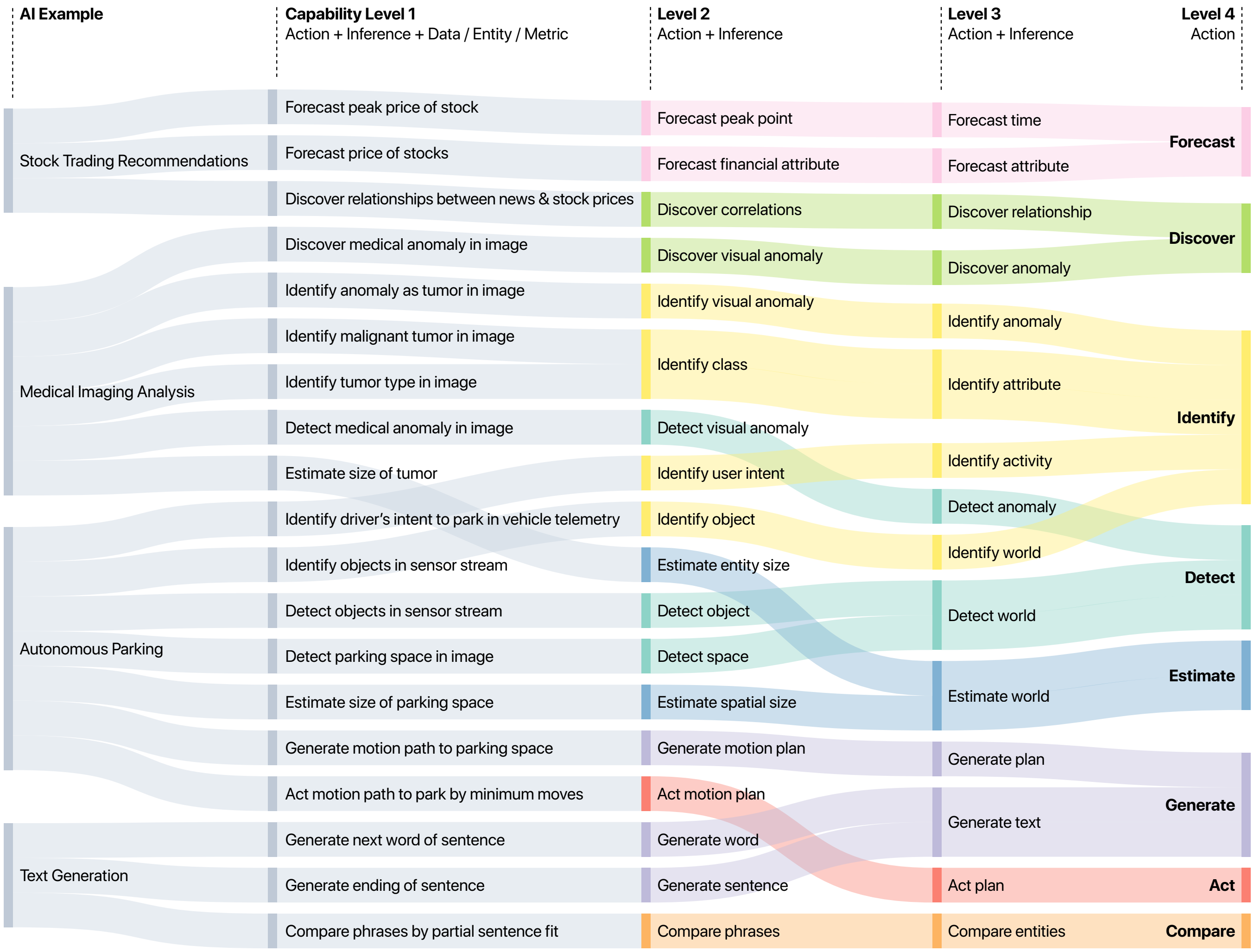


Figure 2: An excerpt of the examples and four capability levels rendered as a Sankey diagram.

Level	Count	Grammar	Description
1	209	Action + Inference + Data/Entity/Metric	Captures all of the distinct capabilities for each example.
2	120	Action + (Abstracted) Inference	Captures a more abstracted inference. Clusters at this level reveal AI capabilities disconnected from specific data. Following the links back to Level 1 reveals different kinds of data that might provide this capability. For example, following ***identify face*** from Level 2 to Level 1 shows that either an image or a depth map can be used to identify a face.
3	44	Action + (Further Abstracted) Inference	Captures a further abstracted inference. Clusters at this level reveal higher level capabilities. Following links back to Level 2 shows different ways to achieve the higher level inference. For example, following ***identify person*** from Level 3 to Level 2 reveals that people can be identified by their face, voice, name, or finger print.
4	8	Action	Captures the eight distinct, high-level capabilities. The size of the cluster at this level offers an indication of how frequently this capability is used across all of the features.

Table 2: Details the four levels of the AI capabilities. See appendix for all AI examples and capabilities.

Capability and Synonyms	Definition	Examples
Estimate Rate, Grade, Measure, Assess	Infer a value (e.g., position, size, duration, cost, impact) related to the current situation. This is about making an inference about now.	Estimate driving time (navigation planner) Estimate chances this is spam (email) Estimate direction sound came from (smart speaker)
Forecast Predict, Guess, Speculate	Infer a value that will be true or some attribute or impact of a future situation that may or may not happen (e.g., stock price, sales, weather, chance of something being true).	Forecast best time to buy stock (financial planner) Forecast tomorrow's weather (weather app) Forecast max price for my house (real estate app)
Compare Rank, Order, Find Best, Find Cheapest, Recommend	Compare a collection of like items based on a metric (e.g., a set of social media ads based on the likelihood a user might click). Allows services to select, rank, or curate a collection of things.	Compare items by likelihood of purchase (online store) Compare posts by likely engagement (social media) Compare movies by likelihood of watching (media)
Detect Monitor, Sense, Notice, Classify, Discriminate	Notice if a specific kind of a thing is in a data set or if it shows up in a sensor stream.	Detect human voice in audio (smart speaker) Detect face in image (camera) Detect step in motion sensor stream (smartwatch)
Identify Recognize, Discern, Find, Classify, Perceive	Notice if a specific item or class of items shows up in a set of like items.	Identify if message is spam (email) Identify if Steve's face (security) Identify the type of cancer (medical imaging)
Discover Extract, Notice, Organize, Cluster, Group, Connect, Reveal	Analyze a dataset and notice a pattern that allows clustering of similar things or identification of outlying entitites.	Discover how people use this site (usage mining) Discover unusual bank transactions (fraud detection) Discover person's routine (energy optimization)
Generate Make, Compose, Construct, Create, Author	Generate something new (message, image, sound) based on knowledge of similar things.	Generate chat response (chat agent) Generate detail in image (photo retouching) Generate synthetic medical records (medical data)
Act Do, Execute, Play, Go, Learn, Operate	Execute a strategy to achieve a specific goal and continue to update the strategy based on advance towards the goal.	Act: Park the car (autonomous parking) Act: Play poker (gambling agent) Act: Fly drone to location (drone pilot)

Table 3: Eight high-level AI capabilities with synonyms, definitions, and examples.

in this terse structure. For example, *Biometric Security* lets users unlock things with their face. The example has the capability to Detect *(action)* + a face *(inference)* + in an image *(data)*. *Detecting* things (e.g., is there a person or an object in this image?) is different from *Identifying* things (e.g., is this Jane's face?). Each individual capability captured a distinct inference or data type (e.g. *face in image, face in depth map, voice in audio*) (Figure 1).

We worked on two additional tasks in parallel to our efforts to capture a precise set of capabilities: 1) We developed consistent terms for everything labeled as an *Action*, *Inference*, or *Data /Entity/Metric*; 2) We worked to move up to higher levels of abstraction from the terse, detailed description of the capabilities. We tried many different verbs to describe the actions, many different terms to describe the inference, and many terms to describe the data, entity (the subject of an inference), and the metric. For example, the capability *Estimate size of tumor* has an entity (tumor) that is the subject of the inference (size). Through an interactive process, we consolidated these into a non-overlapping set. This resulted in 8 high-level Actions (Level 4) and 17 inference clusters (Level 3). See appendix for a table of AI examples and capabilities.

We were inspired by scientific work on taxonomies. We felt having a similar hierarchy for the AI capabilities would make them more understandable and useful. We tried various ways of visualizing the connections between the AI examples, the first level of the AI capabilities, and the higher level capabilities, eventually settling on a Sankey diagram. This made it easy to see clusters forming at different levels. For instance, *Identifying a face* (Level 2) is ultimately about *identifying a person* (Level 3). A person can be identified by their face in an image, or by their name in text, or by their voice in audio. All these low-level inferences would abstract to "person" (Figure 1).

Table 2 provides a description of the different levels. Figure 2 provides an excerpt of the AI example-capability relationships rendered as a Sankey diagram. A complete Sankey diagram can be found in supplementary materials along with definitions for all of the Actions, Inferences, and Data/Entity/Metrics. A description of the eight high-level capabilities can be found in Table 3.

4.2 Reflection

Our design experiment produced several artifacts that collectively provide a resource of AI examples and capabilities. These are captured in the following documents (see supplementary materials):

(1) A detailed list of all AI examples documenting the example description, service type (B2B or B2C), how they co-create value for customers and services,
(2) Detailed definitions of all the *Actions*, *Inferences*, *Data types*, *Entities*, and *Metrics*. These definitions make it easier to add new AI examples and to recognize when new examples will require the creation of new AI capabilities,
(3) A Sankey diagram that visualizes how the features connect to specific AI capabilities and how the capabilities abstract across four levels.
(4) A Github repository hosting the resource files along with a dedicated project website[2].

[2]aidesignkit.github.io.

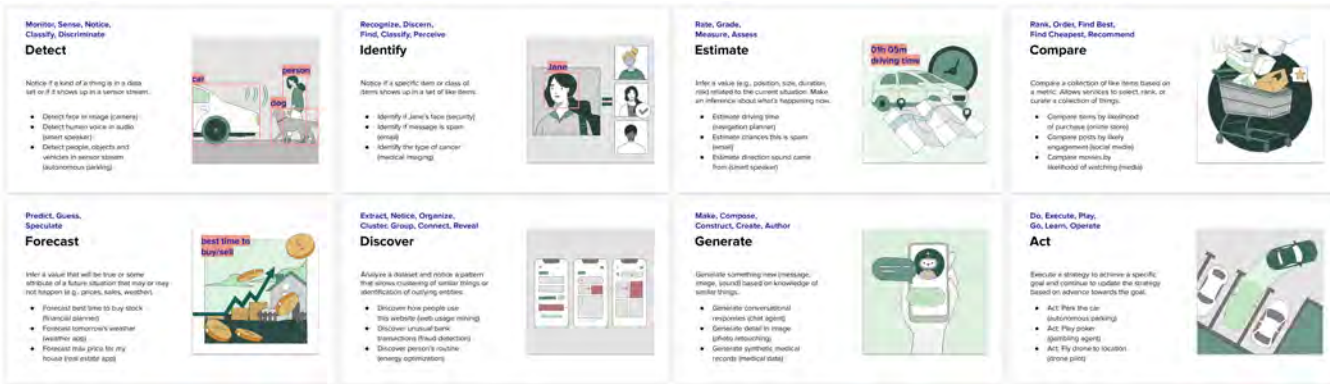


Figure 3: Slides meant to communicate AI capabilities and examples to help designers ideate.

We viewed our resource of examples and capabilities as a stable initial version that offered *good enough* coverage of what AI can reasonably be expected to do. We felt the collective content we produced could aid non-data scientists in understanding AI capabilities, in ideating new product/service concepts that leverage AI capabilities, and in collaborating with data scientists. Developing this resource surfaced many new research questions. *What communicative forms might make these examples and capabilities useful and actionable in support of ideation? How could we integrate this resource in a design process? How does access to this resource impact ideation? How can we assess the impact on ideation?* Our reflection set us up for a new design experiment to explore these challenges.

5 DESIGN EXPERIMENT 2: MAKING THE CAPABILITIES AND EXAMPLES USEFUL

We wanted to explore if our collection of AI examples and capabilities might help designers when they ideate. Would it help them envision concepts that were buildable and valuable? Our design process included three main activities:

(1) **New Forms.** We developed new forms to make the resource useful for ideation.
(2) **Assessing Impact.** We conducted an informal assessment to gain insight on how access to the resource transformed ideation. This forced us to consider how to assess the quality of the concepts created during ideation.
(3) **Reflection.** We reflected on why ideation with the support of our resource did not change ideation in the way we expected. It did not produce more buildable ideas. This reframed the problem, and it offered insights on what makes some concepts easier or more difficult to build.

5.1 Exploring communicative forms

Our resource of examples and capabilities provided a hierarchical, extensible structure. However, the collection of artifacts making up this resource seemed too abstract and overwhelming to effectively sensitize designers to AI capabilities. To jump start the process of exploring different forms, we first created a one-page table (Table 3). This functioned as a sort of cheat sheet for thinking about new forms. The table holds a listing of the eight high-level capabilities along with synonyms commonly used. They are organized in subgroups, using color to visually group similar capabilities. For example, *Detect* and *Identify* both address how AI can classify things. Next, the table holds a brief, high-level definition that describes the types of inferences this capability might make. Finally, it holds a small set of examples, illustrating common forms this capability takes in current products and services.

Next, we sketched various communicative forms. We explored making a deck of capability cards, an interactive website where visitors could explore the connections between capabilities and examples, mood boards, high-level capability posters, and slides. Based on recent research that documented practitioner-created AI design resources in the form of playbooks and slide decks [94], we decided to focus on slides. Our set of slides included capability definitions (Table 3) and each high-level capability as a slide within a 10-page slide deck (see Figure 3 and supplementary materials).

5.2 Assessing the Impact on Ideation

We discussed what it meant to improve ideation, and different ways of measuring the impact of the AI capability slides. We focused on the general idea of envisioning "better" AI concepts. As the discussions progressed, three specific criteria emerged:

- **Breadth.** Researchers noted that designers learning to work with AI often had a very limited range of ideas. Many seemed to consider only familiar applications, such as chatbots or recommenders [90]. More effective ideation produces a diverse range of alternatives and solutions [11, 22]. We wanted to assess if access to the slides helped designers envision concepts that drew upon more of the capabilities.
- **Effort.** Designers tend to envision AI concepts that cannot be built, and they fail to notice situations where simple, low-risk inferences co-create value for customers and service providers [19, 91]. We wanted to assess how much effort would be needed to create the envisioned AI concepts. While our example and capability resource did not capture any information about development effort, we felt our choice to limit examples to things that had been commercially viable could guide designers towards more buildable ideas.

- **Impact.** One of the main reasons AI initiatives fail is that they do not generate enough value for the service provider; they do not generate more revenue than it costs to develop and deploy [25, 43, 84]. Similarly, AI initiatives also fail when they do not generate enough value for users, and users do not accept and use the technology as intended [84, 94]. We wanted to assess the impact of an envisioned concept. How much value might it co-create?

We designed a within-subjects study to assess the impact of the slides on ideation. We asked designers to first ideate solutions to a design challenge without the slides. Next, they ideated solutions to a different design challenge with the slides. We chose within-subjects over between-subjects for two reasons. First, prior literature evaluating design resources with between-subjects studies noted that it is challenging to control the variance between experiment and control groups [21]. Second, we wanted the designers to compare and reflect on their experiences after brainstorming. One limitation of the within-subjects approach is the session order: we could not switch the order of the conditions. Once designers had seen the slides with the capabilities, they would not be able to forget this when ideating without the slides.

We created two similar design briefs: designing AI-enabled interactions for a ride hailing service and for a vacation rental service. We chose to focus on designing for predefined services – as opposed to imagining new service concepts from scratch – as it more closely resembled the majority of day-to-day design practice. We selected the services based on people's familiarity with them as users. We did not want to select a service that would require additional domain expertise, such as healthcare. We created a single slide for each brief that detailed the available data that could drive the potential AI-enabled features. It also listed a set of pain points, something that typically drives a human-centered design process.

We conducted a literature review to gain insights into the needs and pain points. We looked at the needs of drivers and riders (ride sharing) as well as the needs of hosts and travelers (vacation rental). We prepared personas and user journey maps for each design brief, detailing current experience (e.g., *before, during, and after a ride*) [see supplementary materials]. We created a Figma workspace, displaying the design brief, persona, and the user journey as well as sticky notes for ideation.

Before running a full study with professional designers, we first conducted a pilot with 10 HCI students. We conducted 2-hour ideation sessions consisting of a brief study introduction, two consecutive ideation sessions, and a post study interview. Participants were asked to generate as many ideas as they could for each phase of the user journey (20 min), and select and refine five concepts (10 min). Next, a member of the research team introduced the slides and the next brief for participants to ideate using slides. After this session, we interviewed participants about their experience, probing on whether they felt the slides impacted their ideation.

We analyzed the interviews, our observation notes, and the AI concepts pilot participants generated using affinity diagramming. We specifically looked at breadth and quality (impact and effort). To assess breadth, we compared the capabilities in the concepts during the first and second sessions. To assess quality, we created impact-effort matrices [33], a standard prioritization tool commonly used in innovation [93]. We looked at the five concepts delivered at the end of each session, and rated how difficult they would be to make (effort) and how much value they might co-create (impact). We paid attention to the availability and reliability of data, and how easy it would be to produce good enough inferences. We considered the relevance and usefulness of the AI for the user. We then worked to agree on where a concept should go on the matrix.

5.3 Pilot Findings

Access to the slides seemed to increase the breadth of AI capabilities incorporated into the concepts participants generated. Their interviews echoed this finding. Almost all shared that the slides helped them come up with a larger variety of ideas. Several participants shared that the structure (i.e. *action + inference*) helped them to both generate and communicate concepts. Most found the detailed capabilities (Level 1) the most useful.

Surprisingly, we saw no real difference in the quality of concepts between the two sessions. Almost none of the concepts were easily buildable. The impact-effort matrices showed mostly high effort-low impact ideas: things that are difficult or impossible to build with unclear value co-creation. Interestingly, participants who had the most experience with AI were more able to ideate low effort-medium impact concepts for both sessions.

We noticed that most concepts were created without an awareness of whether the data needed was available. Concepts also generally focused on difficult problems, situations where AI would not likely perform well, and near-perfect performance was needed for an AI system to be valuable. Interestingly, two participants shared that the examples in the slides sensitized them to consider situations where AI would still be useful with moderate model performance. They noted that AI could make things faster with moderate performance and still create value. We found this observation interesting.

We observed that a human-centered design approach — the inclusion of the design brief, persona, and journey map — seemed to conflict with effective AI ideation. Most participants gave their greatest attention to user needs and spent less time considering what AI can do and do well. For example, several participants came up with the idea of predicting rider or traveler reliability (i.e., whether they will cancel) based on historical data. This pain point captured in the user journey would not be easily addressed with an AI prediction. It has too much uncertainty. The human-centered materials seemed to push participants to think of AI as magic and to ignore the value it might generate for users that was not specifically documented in the materials. Similar to recent literature that reported tensions between user-centered design (UCD) and AI development process [94, 96], some participants reflected that the ideation process felt different compared to UCD as they had to consider both AI capabilities and human needs.

5.4 Reflection

This design experiment exemplifies Krogh et al.'s claim that RtD is often about *drifting with intention*, the idea that experiments often challenge and change research questions more than they answer them [48]. On the surface, our pilot study failed. We did not get designers to generate more buildable AI concepts. However, it revealed the importance of *AI model performance* and the tensions

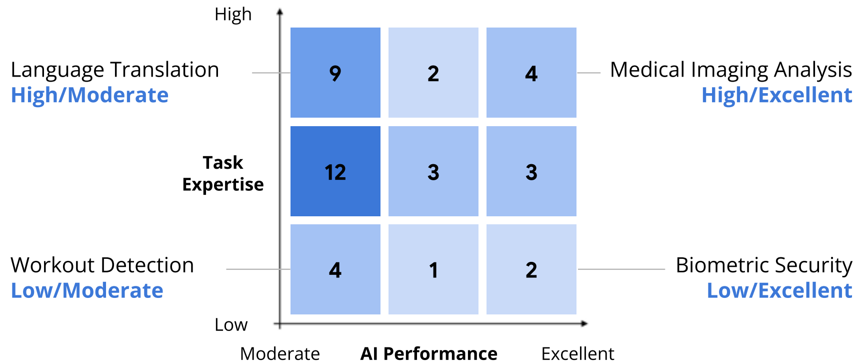


Figure 4: The Task Expertise-AI Performance matrix analysis of 40 AI examples.

between UCD and AI ideation as two unarticulated challenges that we need to overcome. Prior work investigating the challenge of engaging with AI as a design material noted that designers struggle to understand AI capabilities [19], and that they seem to focus on situations where there is both great uncertainty around a capability and great complexity in the output of an AI system [91]. Our design experiment managed to get a bit more below the surface of this problematic situation.

5.4.1 AI Model Performance. Our discussions about the pilot led us to an interesting metaphor used by Google for training product teams on how to search for AI use cases [47]. Their internal course asks teams to think of "AI is an island of drunk people". AI can do things quickly and handle an inhuman quantity of information, because there are a lot of people. But drunk people can make mistakes, so teams should not expect a lot of intelligence. This motivated us to go back and re-examine the examples in our resource.

We noticed that many AI examples did not have excellent model performance, but they were still valuable to users and service providers. For instance, *Smart Speaker Question Answering* captured that AI can detect human speech and convert the speech into words. But it did not capture that the generated transcript has errors. Automatic speech recognition has typically about 90-95% accuracy [3], so around one word per sentence will be incorrect. However, this was good enough to find an answer the user wanted from a corpus of pre-written answers [61]. Applications such as voicemail transcripts or video captions provided other examples where moderate model performance is good enough. These are situations where there is currently no person performing the task, so a moderate quality transcript is better than no transcript. We realized that our resource did not capture how well the AI system needs to perform for co-creation of value.

We revisited each AI example. We decided that in addition to capturing the model performance, we also wanted to capture the human expertise required for the task. Through discussion, we broke each of these dimensions into three bins. For model performance, we chose to categorize examples as excellent (e.g., above 99% accuracy), good (e.g., 90-99% accuracy), or moderate (e.g., below 90% accuracy). In creating these bins our focus was not on capturing the maximum quality an AI system might produce, nor on the technical assessment of performance using certain metrics (e.g., precision, recall, F1 scores, etc). Instead, we captured performance from a UX perspective to understand "the minimum quality needed for users to experience AI as useful" [61]. *What is the minimum amount of accuracy or performance needed for this to be acceptable?* Similarly, we captured how much expertise each task would require for people to perform. Based on *the drunk island metaphor*, we ignored issues of speed and scale. We discussed if the task required more expertise than a typical adult (e.g., *diagnosing cancer*); expertise of a typical adult (e.g., *parking a car*); or less expertise than a typical adult, meaning a child could complete the task (e.g., *recognizing the exercise someone was doing*). We added task expertise and model performance to our description of each example.

To gain new insight on our resource, we developed the Task Expertise-AI Performance matrix (Figure 4). We thought of this as AI's opportunity space. *When ideating, do designers come up with ideas that cover the entire space, or do they largely focus on envisioning things that are difficult tasks and need near-perfect model performance?* The vertical axis represents the level of expertise, not counting issues of speed and scale. The horizontal axis represents how well the AI system must perform in order to co-create value. The upper left region holds AI applications such as *Language Translation*. These are tasks that require people to have high expertise, and moderate quality output has proven useful (while highly context dependent, often better than nothing). The upper right holds examples such as detecting cancer in a medical image. This requires a highly trained professional, and the performance must be excellent for AI systems to be useful. The lower right holds examples such as Biometric Security. This is fairly easy for people (match a person's face to their driver's license photo), and the model performance must be high for things like unlocking someone's phone. The lower left holds examples like smartwatch step counters. A child could count someone's steps (if they could maintain their attention). The quality only needs to be good enough to compare days. It does not need to be accurate to the individual step.

When we viewed all forty of our AI examples as a heat map, it revealed that only a few examples were in the upper right corner (Expertise-High/Performance-Excellent). Most examples (25 out of 40) appeared on the left side (Performance-Moderate). This suggested we needed a new approach to brainstorming, one that encourages people to envision situations where moderate model performance creates value.

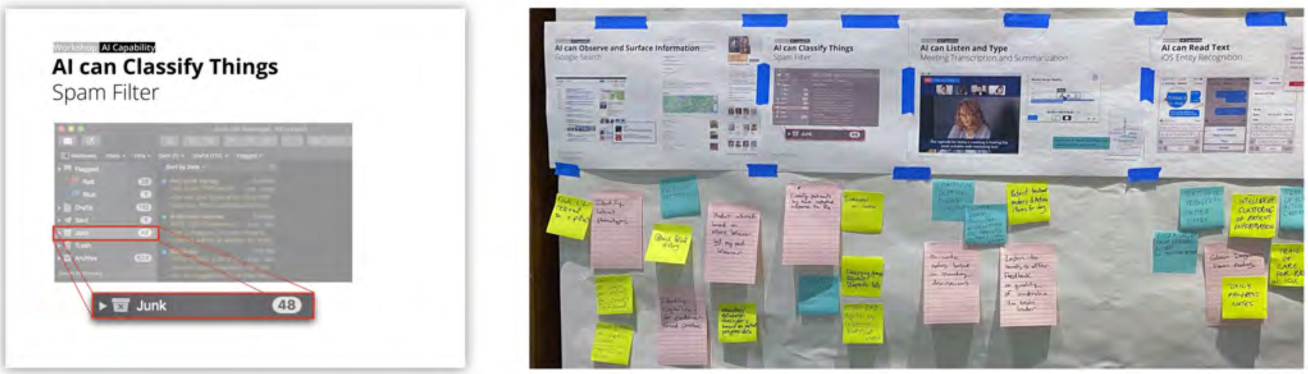


Figure 5: A slide communicating an AI capability and example (left), slide printouts used for ideation (right).

5.4.2 *Tension between AI Ideation and Human-Centered Design.*

Reflecting on the struggles people had in ideating AI concepts, we realized that user-centered design brainstorming does not work well when the solution must utilize AI. Needs uncovered in user research most often point to issues where AI will not help. In addition, this approach does not privilege what AI can do or what it does well. We considered matchmaking [7], a technology-centered innovation approach that starts with a technical capability and systematically searches for the best customer across many domains. However, work on AI innovation almost always focuses on a single domain, as the dataset that is available points to a specific set of users and contextual issues. This pre-selection of users and contexts seemed to conflict with matchmaking.

What we needed was a new innovation approach, one that blends user-centered design and matchmaking. Recent studies report that this hybrid process blending user-centric and tech-centric innovation is already emergent in industry best practices [86, 94, 96]. We began to rethink the role of design in this innovation process. We drew insight from research showing communication breakdowns between data scientists, domain experts and product managers [62, 70]. Instead of asking designers to envision AI concepts in isolation, we considered designers as experts in ideation who could "facilitate ideation between data scientists and domain experts" [93]. We wondered if priming teams with examples of AI capabilities where moderate performance creates value would lead to better concepts, low-risk yet high-value opportunities.

6 DESIGN EXPERIMENT 3: BLENDING UCD & MATCHMAKING

We wanted to improve the process of brainstorming AI concepts. We had three questions to investigate:

(1) How can ideation blend UCD and matchmaking?
(2) Can designers effectively scaffold data scientists and domain experts in brainstorming, in generating ideas that broadly cover the problem-opportunity space?
(3) Does priming ideation with examples of moderate model performance help to generate concepts that are lower-risk in terms of technical feasibility yet still high-value?

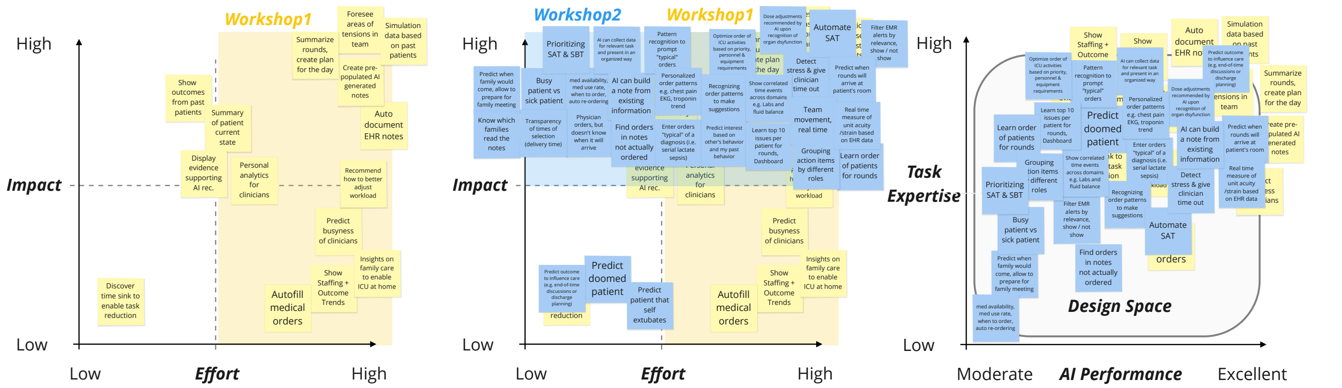
6.1 Design Process

We had an ongoing collaboration with a team of clinicians and data scientists to improve critical care medicine in the intensive care unit (ICU). The team had access to a rich dataset collected across 39 ICUs from 18 hospitals. The project goal was to broadly explore research opportunities for analytics and AI to improve critical care practice. Prior to the project, none of the clinical nor data science team members used formal, structured ideation methods nor engaged in human-centered design. In this section, we provide a brief overview of the team, research procedure and activities, and data analysis.

Our research team (n=22) consisted of 6 data scientists, 10 healthcare professionals, and 6 HCI and design experts. The data science team members had backgrounds in data analytics, healthcare analytics, and AI research. The healthcare members all had experience in critical care medicine and included 4 attending physicians, 2 fellows, 2 nurses, and 2 non-clinical healthcare experts. The HCI/design members had backgrounds in interaction design, service design, human-AI interaction, and data visualization.

We conducted two ideation workshops to generate AI concepts. A major challenge for AI innovation in healthcare is ensuring clinician acceptance [40, 92, 95]. Thus, our goal was to produce ideas that are feasible *and* clinically relevant. Each workshop had 15-17 participants involving at least one participant from each role (i.e. physician, nurse, healthcare expert, data scientist, HCI researcher, designer). Workshops were conducted in-person. The HCI/design team facilitated and participated in the brainstorming.

Workshop 1. We followed a traditional, user-centered approach to provide a baseline for measuring the impact of our modified approach. The workshop consisted of introductions (10 min), two rounds of ideation sessions (30 min each), concept assessment (40 min), and debriefing (10 min). Ideation ran parallel in two groups, where groups swapped stations at the end of the first round to build on each others' ideas. Team members first ideated individually, then shared their ideas and brainstormed as a group. We probed clinicians to draw on their lived experiences to recall pain points and envision potential AI-enabled solutions. Data science team members expanded on whether training data existed and if the concept could be built. Following ideation, we collectively assessed



a) Workshop 1 Impact-Effort b) Workshop 1&2 Impact-Effort c) Workshop 1&2 Task Expertise-AI Performance

Figure 6: Impact-effort analysis of AI concepts from workshop 1 (a) and workshop 2 (b); task expertise-AI performance analysis of workshop 1 and 2 (c).

and reflected on the concepts, mapping them on an impact-effort matrix [34].

Workshop 2. The second workshop started with a slide overview of AI examples and capabilities generated from our resource. We adapted the capability language to be less precise and more familiar. None of the examples were medical, and most involved situations where moderate model performance co-creates value. The capabilities included *observe and surface information* (contextual web search); *classify things* (email spam filter); *listen and type* (real-time meeting transcription); *read text* (text message entity recognition); *predict text* (email sentence completion); *cluster similarities* (online shopping recommender system); *discover patterns* (smartwatch activity trends). We created a slide for each capability (see Figure 5 and supplementary materials), presenting the slides and also hanging them on the wall.

We conducted two rounds of ideation. As we ideated, we asked clinicians if they could think of situations where a capability might be useful. Data science team members elaborated on what might be feasible. We pushed back on concepts that required near-perfect model performance, probing for situations where moderate model performance would still be valuable. After the second round of ideation, we collectively assessed the concepts and held a debriefing.

We recorded and transcribed both workshops (three hours of recordings per workshop with ideation sessions running in parallel, six hours in total). We documented the artifacts produced during the workshops, including ideation outcomes and impact-effort matrices from the assessment activities. We also conducted a post-workshop analysis using the Task Expertise-AI Performance matrix. We analyzed the workshop transcripts and outcomes using affinity diagramming [34] to identify key themes and gain insights into how the design activities impacted the workshop outcomes.

6.2 Findings

Both workshops were successful in facilitating ideation. All team members reported that they felt engaged and that they brainstormed successfully. However, there was a contrast between the workshop outcomes. Figure 6 shows the assessment of the concepts produced in each workshop.

Workshop 1, where we followed a traditional brainstorming approach, produced almost no concepts that were low-effort. Many ideas were actually high effort-low impact, only about half seemed relevant and useful for critical care medicine (Figure 6a). Overall, the ideation lacked breadth. A majority of ideas focused on improving clinical decision making, particularly around trust, feedback, and explainability (e.g., *AI can take feedback on why it is wrong; recommendation rationale should be clear*). The second largest theme was around automated documentation (e.g., *automatically generate notes from clinical conversations; autofill or autocomplete notes and orders; learn and document only what is most important*). Ideas often captured desired behaviors for existing AI systems (e.g., *recommendation is not intrusive; recommendation comes when ICU team is together*) or current pain points (e.g., *placing orders is a burden; I want to eliminate and delegate tasks*). Few ideas described new AI-enabled interactions (e.g., *predict sedation dose for ventilated patients; foresee areas of tension between clinicians; personal analytics for clinicians for self-reflection; recommend how to better adjust workload*).

Interestingly, on the Task Expertise-AI Performance matrix, most ideas mapped to the upper right corner. Our ideas often required near-perfect AI performance to be useful and focused on situations with high uncertainty where the task is difficult even for highly trained experts (Figure 6c). For instance, one concept was about using deep learning to help discover the right amount of sedation for a patient on a ventilator. Too little sedation and the patient suffers from pain and anxiety, which inhibits healing. Too much sedation and patients run the risk of delirium, which can cause lasting psychological harm. This is a hard problem that needs excellent model performance, and it requires very high quality healthcare data, which may not exist.

Workshop 2 produced concepts that fell across the top of the impact-effort matrix, we were able to identify high impact-low effort ideas (Figure 6b). Ideas also mapped to a broader set of themes. Examples include AI systems that would improve coordination between clinicians (e.g., *generate a schedule for nurses and respiratory therapists for extubation*); systems that improved logistics and resource allocation (e.g., *predict which medications would be needed* *based on current patients and pre-order from pharmacy*); systems that inferred workload and effort, possibly in support of dynamic staffing (e.g., *classify patients as busy or non-busy*); systems that better support attention management (e.g., *classify patients based on uncertainty*); systems that anticipate and surface needed information (e.g., *learn relevant information based on patient conditions*). Ideas seemed to follow the capability descriptions in the examples (i.e. action + inference), which resulted in way fewer non-AI ideas.

Reviewing AI capabilities and examples prior to ideation seemed to have a great impact on healthcare members. Throughout our ideation process, they repeatedly recognized situations where a capability could be useful, and then effectively transferred that capability to a healthcare opportunity. Our team quickly adopted this example-based approach and started drawing from other familiar examples. For instance, a physician brought up Amazon’s anticipatory shipping that pre-ships and stocks items when there is a high chance that customers will soon order them [75]. Clinicians shared that there are situations where it takes hours for medications to arrive from the pharmacy, especially in busier wards. They discussed how patient records could be used to predict which medications would likely be needed the following day. This is a relatively low risk idea as the worst outcome from an inference error would be that the clinicians would need to order medicine from the pharmacy – so never worse than the current state.

Overall, considering examples where an imperfect, moderate performance model could create value broadened our ideation: the Task Expertise-AI Performance matrix showed a better coverage of the larger problem-opportunity space (Figure 6c).

6.3 Reflection

Design Experiment 3 confirmed our hunch that a user-centered mindset was a hindrance for envisioning AI concepts. We were trying to address user needs that did not need to be addressed with AI. Similar to the AI-centric fallacy that views all problems as nails that can be solved with the AI hammer, we were trying to fix screws with a hammer.

Taking a complementary approach, reviewing AI capabilities and examples to probe domain experts where these could be useful, seemed to work better. Our ideation process produced low risk-high value ideas. We moved away from focusing on high-risk situations, such as clinical decision making, and identified many low risk situations where moderate performance AI could support clinical tasks. These tasks were often not difficult. They were simply too tedious in terms of volume and with respect to the need for speed in the work (e.g. *looking at all the patients and pre-order from the pharmacy; predicting cases where there might be a deviation from the standard of care*). This is a fruitful place for human-AI complementarity, marking a clear space for the design innovation of AI.

Overall, our modified approach provided a glimpse into what successful ideation might look like. Workshop debrief sessions and reflections echoed this as well. Team members expressed that the exercise was useful to inform the research agenda: *“There is a lot of inertia towards high risk-high reward projects or low risk-low reward areas that doesn’t move the needle in a meaningful way. … The exercise was really valuable to identify ideas that are worth doing, as every research portfolio should have some of these in a balanced way.” (Physician)*

7 DISCUSSION

Researchers noted a gap in AI innovation: AI products fail when they fail to address a real user need or generate enough value for the service provider to offset development and operational costs [94]. Practitioners report AI project failures due to selecting and working on the wrong problem. Current resources, such as human-AI guidelines, provide little support for discovering and selecting problems where AI might be an optimal solution [94]. We set out to close this gap by improving the ideation process of AI products, product features, and services.

Our work builds on prior observations showing that experienced innovation teams created internal resources curating AI capabilities and examples to scaffold ideation [90, 93, 94]. We engaged in a reflective design process to develop a design resource that delineates what AI can do through an analysis of AI-enabled features commonly found in the real world. We conducted design experiments to inform our understanding of how, when, and in what form this resource might be used in the design process for effective ideation.

Below, we discuss how this resource addresses the challenges of ideation, and how it might be improved and extended. We reflect on the implications of this case for (1) developing design resources to close the AI innovation gap, (2) mapping AI’s design space, and (3) exploring new innovation processes for AI.

7.1 Implications for Design Resources

Prior work suggested that AI capability abstractions and examples might support envisioning [90, 93, 94], yet little work explored how, when, and in what form these might be useful. Our work advances these efforts; having distinct capabilities and examples proved useful both in design experiment 2 and 3. In this section, we discuss how future research can operationalize, extend, and mine this resource to further its usefulness.

7.1.1 Developing Alternative Forms and Resources. We operationalized our resource as slides, one version that focused on more abstract capabilities (experiment 2) and one on more specific capabilities encapsulated in an example (experiment 3). We see several opportunities for mining from this resource to generate alternative forms and presentations (e.g. worksheets, flash cards, interactive visualizations, etc). Future forms could capitalize on the rich set of information encoded, including the multiple levels of abstraction (i.e., Levels 1-4), data types, domains, value co-creation, model performance, etc. Below, we outline a few directions for developing forms to enable designers and innovators to browse, filter, and scrutinize AI capabilities for the design task at hand:

(1) **Exploring target inferences across levels of abstraction:** Designers working on an IoT enabled smart home system could ask “What are all the different ways to notice if someone is at home?” Following *detect person* (high-level, abstracted inference) from Level 3 to Level 2 reveals that people can be detected by their voice, touch, motion, face, hand, and body (low-level inference).
(2) **Exploring capabilities related to specific data types:** Taking a data-driven design approach, designers working on a customer support application could investigate how they could make call transcripts more useful to both customers

and support staff. They might browse capabilities and inferences related to text data (e.g. *identify sentiment; identify product; identify user query; generate response to user query; generate text summary*).

(3) **Exploring examples related to domains:** Designers may search through examples within specific domains to gain a sense of what capabilities, inferences and data types have been previously used, and what seem underutilized.

(4) **Exploring examples related to value co-creation:** Designers may use different ways AI could create value for customers and service providers as lenses for perspective taking (e.g. *saving time, accelerating tasks, automating tedious work, reducing cognitive load,* etc).

7.1.2 Extending Examples. It is a challenging task to document AI capabilities, as AI technologies advance rapidly. Our goal was not capturing *everything* AI can do, instead we set out to provide a *good enough* coverage to begin the process of sensitizing designers to what AI can do. For this reason, we view this resource as *version 1.0* – a snapshot of AI capabilities that are commonly found in current products and services and that are immediately available. We intend this resource to be an open source, living resource that is stable enough for researchers to extend it. Could adding new examples expand the set of eight high-level capabilities? What additional data types and inferences should be captured in existing examples? For instance, recent developments in generative AI and its more general abilities opens the door for many new capabilities that seem just around the corner. Future research should extend, critique, and refine the framework and corpus of examples.

7.1.3 Capturing Additional Dimensions. Initially, our focus was on capturing capabilities in each example, along with value co-creation and domain information. Discovering that capabilities alone were not enough without the consideration of model performance led us to capture and encode this emergent dimension for each example.

Building on the broader human-AI interaction literature, we outline a few missing dimensions that we intend to capture in our future work. First, our resource does not detail the relative difficulty or cost of development or maintenance. For example, generating the next word a user might type is typically easier than generating the remaining half of a sentence or a paragraph. Similarly, it does not detail the sequence or interrelationships between capabilities. For instance, forecasting demand is necessary to forecast the price of a home listing (Smart Pricing). Future work should explore how to encode the feasibility of a capability, a key consideration in assessing and prioritizing early phase design concepts [93].

Second, our resource captures AI's capabilities, yet misses its limitations. There is an ongoing extensive discussion on identifying, anticipating, and mitigating AI's potential harm across HCI, FAccT and AI research communities. *All* the examples and capabilities in our collection have risks associated around fairness, bias, and errors: *Biometric Security* applications entail disparities in race and gender [10], *Deepfakes* enable the spread of misinformation and harmful content [42]. How might innovation teams systematically and broadly explore potential harm during early phase AI design and development? Recent work proposed creating an "AI Incident Database" by cataloging real world AI failures [58]. This marks a clear space for future research to improve AI capability resources to encode both capabilities and limitations.

7.2 Mapping AI's Design Space

Prior literature deliberated on what makes AI a particularly difficult design material to work with [91]. In this work, we contribute the discovery of *model performance* as a key consideration in interplay with *task expertise*. Throughout our design process, we found ourselves asking *"how well does this AI concept need to perform to be useful?"*, which led us to create the task expertise-AI performance matrix. The matrix provides a novel and actionable perspective for describing and navigating AI's problem-opportunity space. Our initial ideas were mostly difficult to build: we were searching places that required a lot of human intelligence and expertise, and near-perfect AI performance to be useful (e.g. decision support systems). Interestingly, the heat map of 40 AI examples indicates that this might not be the richest search space. The majority of AI applications in the real world seem to have moderate model performance; these are situations where AI systems can still provide enough value with imperfect results. For instance, video captions and voicemail transcriptions are helpful for people to quickly skim information, even when one word out of ten will be incorrect.

This understanding provides a valuable lens for generating and assessing AI design concepts. There is a larger design space for leveraging AI capabilities; it suggests that innovators should look for places for making moderate performance AI systems useful to find opportunities for innovation. Does intentionally searching for moderate AI performance lower the risk of coming up with infeasible and/or low-value ideas? How can we sensitize designers and innovators to this search space, beyond sharing examples and metaphors like "the drunk island" [47]? Future research should investigate this proposed mapping of AI's design space to provide new insights into effective ideation and problem selection in early phase AI product development.

7.3 Exploring New Innovation Processes for AI

AI as a design material requires new design processes beyond user-centered design [88, 94, 96]. Insights we gained from our design process echo this. We felt imitations when following user-centered approaches for ideation: the pain points we considered were often situations where AI is not the optimal solution. We suspect that asking domain experts what would be most valuable has unintentionally led our team to focus on points of great uncertainty or edge cases where AI is not likely to work. Our modified process, starting with AI capabilities and examples, and asking domain experts to recognize situations where these would be useful, led to more effective ideation. It blended the strengths of user-centered and technology-centered innovation processes. It was user-centered in that we drew from clinicians' lived experiences when probing what is useful. It was technology-centered, as in matchmaking [7], we started with a review of technical capabilities and had the data science team lean in on what is doable.

We propose that this modified design process is *better* – it is more likely to result in a broad coverage of the problem-opportunity space. This marks a clear space for future HCI and design research:

How might we create design processes that account for AI capabilities and limitations as well as human needs? What roles can designers and domain experts play in the early phase AI design and development? AI products fail when they fail to account for human needs [94]; on the other hand, solely following a human-centered process ignores the value AI can bring. More work is needed to understand how to combine UCD and matchmaking approaches. We encourage design researchers to lean in to sketch and prototype new design processes for innovating AI products and services.

7.3.1 Integrating Design Ideation into AI Product Development. Probing the usefulness of the capabilities and examples enabled us to gain insights into how such resources might be integrated into current product development processes. We highlight four entry points for introducing design ideation resources:

(1) Envisioning new AI-driven features for an existing product or service,
(2) Envisioning application concepts for a core AI capability (e.g. large pretrained models for text or image generation),
(3) Exploring potential value in datasets to enable novel AI-based interactions,
(4) Designing a dataset with domain stakeholders with an eye for downstream applications (e.g. see [37] for a discussion of how data can and should be designed).

We view these starting points as a continuum between UCD and matchmaking. Our pilot experiment provides an example of (1), whereas our design process with clinicians and data scientists is closer to (3). Future research should further investigate how and when to integrate design resources for effective ideation.

7.3.2 Developing Better Resources and Processes for Risk Assessment. Our work provides preliminary evidence on improving the ideation process for AI concepts. However, challenges remain in assessing and selecting concepts to move forward to prototyping. How can we systematically evaluate the quality of concepts? What design processes can support a holistic assessment of risks and harm of early phase concepts *before* selecting what to build? UX practice has been evolving as practitioners need to account for many considerations to ensure the development of responsible AI systems [82, 83]. Current assessment and prioritization tools, such as the impact-effort matrix, fail to account for the complexity of assessing AI systems. Recent literature highlights practitioner-created assessment tools and processes that factor in many considerations, including risk, frequency of use, model accuracy, data quality, and cost [83, 94]. Future research should develop better approaches to holistically assess and prioritize AI concepts.

8 LIMITATIONS AND FUTURE WORK

Our work has three limitations. First, our corpus of AI examples is not exhaustive. It is limited to commonly found AI features, thus lacking new, emergent capabilities and unique commercial capabilities only available in a few systems. Second, our team was involved in both the development and evaluation of the resource as part of a real world design process. Future work should assess the use of AI capabilities and examples with design teams who are new to this resource. Third, while our resources captures AI capabilities, it does not currently capture limitations and potential harm. We note that ideating on "what AI *can* do" should always entail asking "what AI *should* or *should not* do". Because of these limitations, we describe this resource as *version 1.0*, and we intend it to address these aspects in our future work.

9 CONCLUSION

AI products fail when they fail to provide value for users and services. HCI and design thinking can play an important role in addressing AI's innovation gap. We took a step towards addressing this gap by engaging in a reflective design process. We created a resource capturing AI capabilities based on 40 features commonly found across many products and services. Our resource captured the high-level AI capabilities across all examples and low-level detailed capabilities for each example. Our pilot assessment revealed AI model performance as a critical consideration for ideation. We experienced limitations when employing a user-centered approach to AI ideation. In response, we adopted a hybrid approach blending user-centered design and matchmaking, probing domain experts on where a capability could be useful. We invite HCI and design researchers to critique, assess, and extend the resource and approach employed in this work.

ACKNOWLEDGMENTS

We would like to thank Ken Holstein, Nik Martelaro, Dominik Moritz, Adam Perer, Carolyn Rosé, Reid Simmons and Haiyi Zhu for reviewing the collection of AI examples, and Gabe Clapper, Liz Gerber, Ben Shneiderman, Fernanda Viegas, Martin Wattenberg, and Qian Yang for their feedback. We also would like to thank the students for their help at various stages of the project: Anushree Abhyankar, Shravya Bhat, Neha Deshmukh, Weiting Gao, Min Kyoung Lee, Shuyu Liu, Janet Peng, Ningjing Sun, Proud Taranat, Xuehui Zhang. Finally, we would like to thank our anonymous reviewers for their feedback.

This material is based upon work supported by the National Science Foundation under Grant No. (2007501) and work supported by the National Institutes of Health (R35HL144804). The first author was also supported by the Center for Machine Learning and Health (CMLH) Translational Fellowships in Digital Health. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.

REFERENCES

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
[2] Apple. 2019. Human Interface Guidelines: Machine Learning. https://developer.apple.com/design/human-interface-guidelines/technologies/machine-learning/introduction/
[3] Microsoft Azure. 2022. Test accuracy of a Custom Speech model. https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/how-to-custom-speech-evaluate-data?pivots=speech-studio
[4] Jesse Josua Benjamin, Arne Berger, Nick Merrill, and James Pierce. 2021. Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[5] Thomas Binder and Johan Redström. 2006. Exemplary design research. (2006).
[6] Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. *Patterns* 2, 2 (2021), 100205.

[7] Sara Bly and Elizabeth F Churchill. 1999. Design through matchmaking: technology in search of users. *interactions* 6, 2 (1999), 23–31.
[8] Kirsten Boehner and Carl DiSalvo. 2016. Data, design and civics: An exploratory study of civic tech. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2970–2981.
[9] Sander Bogers, Joep Frens, Janne Van Kollenburg, Eva Deckers, and Caroline Hummels. 2016. Connected baby bottle: A design case study towards a framework for data-enabled design. In *Proceedings of the 2016 ACM conference on designing interactive systems*. 301–311.
[10] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.
[11] Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design.* Morgan kaufmann.
[12] Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, and Alexander Chen. 2020. Teachable machine: Approachable Web-based tool for exploring machine learning classification. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–8.
[13] Michael Chromik, Florian Lachner, and Andreas Butz. 2020. ML for UX?-An Inventory and Predictions on the Use of Machine Learning Techniques for UX Research. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–11.
[14] Alan Cooper. 1999. *The inmates are running the asylum.* Springer.
[15] Scott Davidoff, Min Kyung Lee, Anind K Dey, and John Zimmerman. 2007. Rapidly exploring application design through speed dating. In *UbiComp 2007: Ubiquitous Computing: 9th International Conference, UbiComp 2007, Innsbruck, Austria, September 16-19, 2007. Proceedings 9*. Springer, 429–446.
[16] Wesley Hanwen Deng, Nur Yildirim, Monica Chang, Motahhare Eslami, Kenneth Holstein, and Michael Madaio. 2023. Uncovering the Practices and Opportunities for Cross-functional Collaboration around AI Fairness in Industry Practice. (2023).
[17] Carl DiSalvo, Jonathan Lukens, Thomas Lodato, Tom Jenkins, and Tanyoung Kim. 2014. Making public things: how HCI design can express matters of concern. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*. 2397–2406.
[18] 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*. 1–17.
[19] 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*. 278–288.
[20] Graham Dove and Sara Jones. 2014. Using data to stimulate creative thinking in the design of new products and services. In *Proceedings of the 2014 conference on Designing interactive systems*. 443–452.
[21] Steven Dow, T Scott Saponas, Yang Li, and James A Landay. 2006. External representations in ubiquitous computing design and the implications for design tools. In *Proceedings of the 6th conference on Designing Interactive systems*. 241–250.
[22] Steven P Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L Schwartz, and Scott R Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 4 (2010), 1–24.
[23] Ruofei Du, Na Li, Jing Jin, Michelle Carney, Scott Miles, Maria Kleiner, Xiuxiu Yuan, Yinda Zhang, Anuva Kulkarni, Xingyu Liu, et al. 2023. Rapsai: Accelerating Machine Learning Prototyping of Multimedia Applications through Visual Programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–23.
[24] Reaktor Education. 2018. Elements of AI. https://www.elementsofai.com/
[25] Tatiana Ermakova, Julia Blume, Benjamin Fabian, Elena Fomenko, Marcus Berlin, and Manfred Hauswirth. 2021. Beyond the hype: why do data-driven projects fail?. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. 5081.
[26] Melanie Feinberg. 2017. A design perspective on data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2952–2963.
[27] Rebecca Fiebrink, Dan Trueman, and Perry R Cook. 2009. A Meta-Instrument for Interactive, On-the-Fly Machine Learning.. In *NIME*. 280–285.
[28] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: cultural probes. *interactions* 6, 1 (1999), 21–29.
[29] William Gaver. 2012. What should we expect from research through design?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 937–946.
[30] William W Gaver, John Bowers, Andrew Boucher, Hans Gellerson, Sarah Pennington, Albrecht Schmidt, Anthony Steed, Nicholas Villars, and Brendan Walker. 2004. The drift table: designing for ludic engagement. In *CHI'04 extended abstracts on Human factors in computing systems*. 885–900.
[31] Fabien Girardin and Neal Lathia. 2017. When user experience designers partner with data scientists. In *2017 AAAI Spring Symposium Series*.
[32] Katerina Gorkovenko, Daniel J Burnett, James K Thorp, Daniel Richards, and Dave Murray-Rust. 2020. Exploring the future of data-driven product design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
[33] Nielsen Norman Group. 2018. Using Prioritization Matrices to Inform UX Decisions. Retrieved September 14, 2022 from https://www.nngroup.com/articles/prioritization-matrices/
[34] Bruce Hanington and Bella Martin. 2019. *Universal methods of design expanded and revised: 125 Ways to research complex problems, develop innovative ideas, and design effective solutions.* Rockport publishers.
[35] Patrick Hebron. 2016. Machine learning for designers. (2016).
[36] Matthew K Hong, Adam Fourney, Derek DeBellis, and Saleema Amershi. 2021. Planning for natural language failures with the ai playbook. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.
[37] Aspen Hopkins, Fred Hohman, Luca Zappella, Xavier Suau Cuadros, and Dominik Moritz. 2023. Designing Data: Proactive Data Collection and Iteration for Machine Learning. *arXiv preprint arXiv:2301.10319* (2023).
[38] Katherine Isbister and Kristina Höök. 2009. On being supple: in search of rigor without rigidity in meeting new design and evaluation challenges for HCI practitioners. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2233–2242.
[39] Anniek Jansen and Sara Colombo. 2023. Mix & Match Machine Learning: An Ideation Toolkit to Design Machine Learning-Enabled Solutions. In *Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–18.
[40] Stephanie Tulk Jesso, Aisling Kelliher, Harsh Sanghavi, Thomas Martin, and Sarah Henrickson Parker. 2022. Inclusion of clinicians in the development and evaluation of clinical artificial intelligence tools: a systematic literature review. *Frontiers in Psychology* 13 (2022).
[41] Xiaoneng Jin, Mark Evans, Hua Dong, and Anqi Yao. 2021. Design heuristics for artificial intelligence: inspirational design stimuli for supporting UX designers in generating AI-powered ideas. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–8.
[42] Deborah G Johnson and Nicholas Diakopoulos. 2021. What to do about deepfakes. *Commun. ACM* 64, 3 (2021), 33–35.
[43] Mayur P Joshi, Ning Su, Robert D Austin, and Anand K Sundaram. 2021. Why so many data science projects fail to deliver. *MIT Sloan Management Review* (2021).
[44] Claire Kayacik, Sherol Chen, Signe Noerly, Jess Holbrook, Adam Roberts, and Douglas Eck. 2019. Identifying the intersections: User experience+ research scientist collaboration in a generative machine learning interface. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.
[45] Rochelle King, Elizabeth F Churchill, and Caitlin Tan. 2017. *Designing with data: Improving the user experience with A/B testing.* " O'Reilly Media, Inc.".
[46] Rafal Kocielnik, Saleema Amershi, and Paul N Bennett. 2019. Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
[47] Cassie Kozyrkov. 2022. Making Friends with Machine Learning: Advice for finding AI use cases. https://kozyrkov.medium.com/imagine-a-drunk-island-advice-for-finding-ai-use-cases-8d47495d4c3f
[48] P Krogh and Ilpo Koskinen. 2020. *Drifting by intention.* Springer.
[49] Peter Gall Krogh, Thomas Markussen, and Anne Louise Bang. 2015. Ways of drifting—Five methods of experimentation in research through design. In *ICoRD'15–Research into Design Across Boundaries Volume 1: Theory, Research Methodology, Aesthetics, Human Factors and Education*. Springer, 39–50.
[50] 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. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.
[51] Peter Kun, Ingrid Mulder, and Gerd Kortuem. 2018. Design enquiry through data: appropriating a data science workflow for the design process. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*. 1–12.
[52] Michelle S Lam, Zixian Ma, Anne Li, Izequiel Freitas, Dakuo Wang, James A Landay, and Michael S Bernstein. 2023. Model Sketching: Centering Concepts in Early-Stage Machine Learning Model Design. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–24.
[53] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.
[54] Q Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly Understanding: Information Needs for Model Transparency to Support Design Ideation for AI-Powered User Experience. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.

[55] Youn-kyung Lim, Sang-Su Lee, and Da-jung Kim. 2011. Interactivity attributes for expression-oriented interaction design. *International Journal of Design* 5, 3 (2011).

[56] Joseph Lindley, Haider Ali Akmal, Franziska Pilling, and Paul Coulton. 2020. Researching AI Legibility through Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[57] Nirav Malsattar, Tomo Kihara, and Elisa Giaccardi. 2019. Designing and Prototyping from the Perspective of AI in the Wild. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 1083–1088.

[58] Sean McGregor. 2021. Preventing repeated real world AI failures by cataloging incidents: The AI incident database. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 15458–15463.

[59] Steven Moore, Q Vera Liao, and Hariharan Subramonyam. 2023. fAIlureNotes: Supporting Designers in Understanding the Limits of AI Models for Computer Vision Tasks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[60] Camille Moussette and Richard Banks. 2010. Designing through making: exploring the simple haptic design space. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*. 279–282.

[61] Cosmin Munteanu, Ronald Baecker, Gerald Penn, Elaine Toms, and David James. 2006. The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 493–502.

[62] Nadia Nahar, Shurui Zhou, Grace Lewis, and Christian Kästner. 2022. Collaboration Challenges in Building ML-Enabled Systems: Communication, Documentation, Engineering, and Process. *Organization* 1, 2 (2022), 3.

[63] Marianna Obrist, Rob Comber, Sriram Subramanian, Betina Piqueras-Fiszman, Carlos Velasco, and Charles Spence. 2014. Temporal, affective, and embodied characteristics of taste experiences: A framework for design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2853–2862.

[64] William Odom and Tijs Duel. 2018. On the design of OLO Radio: Investigating metadata as a design material. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–9.

[65] William Odom, John Zimmerman, and Jodi Forlizzi. 2014. Placelessness, spacelessness, and formlessness: experiential qualities of virtual possessions. In *Proceedings of the 2014 conference on Designing interactive systems*. 985–994.

[66] Fatih Kursat Ozenc, Miso Kim, John Zimmerman, Stephen Oney, and Brad Myers. 2010. How to support designers in getting hold of the immaterial material of software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2513–2522.

[67] Google PAIR. 2019. People + AI Guidebook. pair.withgoogle.com/guidebook

[68] Samir Passi and Solon Barocas. 2019. Problem formulation and fairness. In *Proceedings of the conference on fairness, accountability, and transparency*. 39–48.

[69] James Pierce, Phoebe Sengers, Tad Hirsch, Tom Jenkins, William Gaver, and Carl DiSalvo. 2015. Expanding and refining design and criticality in HCI. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2083–2092.

[70] 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. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–25.

[71] Johan Redström. 2005. On technology as material in design. *Design Philosophy Papers* 3, 2 (2005), 39–54.

[72] Donald Schön and John Bennett. 1996. Reflective conversation with materials. In *Bringing design to software*. 171–189.

[73] Ben Shneiderman. 2020. Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, 4 (2020), 1–31.

[74] Chris Speed and Jon Oberlander. 2016. Designing from, with and by Data: Introducing the ablative framework. In *Proceedings of the International Design Research Society Conference*.

[75] J Spiegel, M McKenna, G Lakshman, and P Nordstrom. 2014. Amazon us patent anticipatory shipping. *Amazon Technologies Inc* 12 (2014).

[76] 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*. 1162–1177.

[77] Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, et al. 2022. Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence. *arXiv preprint arXiv:2211.06318* (2022).

[78] Hariharan Subramonyam, Colleen Seifert, and Eytan Adar. 2021. Towards A Process Model for Co-Creating AI Experiences *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1529–1543. https://doi.org/10.1145/3461778.3462012

[79] Hariharan Subramonyom, Jane Im, Colleen Seifert, and Eytan Adar. 2022. Human-AI Guidelines in Practice: The Power of Leaky Abstractions in Cross-Disciplinary Teams. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*.

[80] Petra Sundström, Alex Taylor, Katja Grufberg, Niklas Wirström, Jordi Solsona Belenguer, and Marcus Lundén. 2011. Inspirational bits: towards a shared understanding of the digital material. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1561–1570.

[81] Philip van Allen. 2018. Prototyping ways of prototyping AI. *Interactions* 25, 6 (2018), 46–51.

[82] Rama Adithya Varanasi and Nitesh Goyal. 2023. "It is currently hodgepodge": Examining AI/ML Practitioners' Challenges during Co-production of Responsible AI Values. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[83] Qiaosi Wang, Michael Madaio, Shaun Kane, Shivani Kapania, Michael Terry, and Lauren Wilcox. 2023. Designing Responsible AI: Adaptations of UX Practice to Meet Responsible AI Challenges. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[84] Joyce Weiner. 2020. Why AI/data science projects fail: how to avoid project pitfalls. *Synthesis Lectures on Computation and Analytics* 1, 1 (2020), i–77.

[85] Mikael Wiberg. 2014. Methodology for materiality: interaction design research through a material lens. *Personal and ubiquitous computing* 18, 3 (2014), 625–636.

[86] Maximiliane Windl, Sebastian S Feger, Lara Zijlstra, Albrecht Schmidt, and Pawel W Wozniak. 2022. 'It Is Not Always Discovery Time': Four Pragmatic Approaches in Designing AI Systems. In *CHI Conference on Human Factors in Computing Systems*. 1–12.

[87] Allison Woodruff, Sarah E Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A qualitative exploration of perceptions of algorithmic fairness. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–14.

[88] Qian Yang, Nikola Banovic, and John Zimmerman. 2018. Mapping machine learning advances from hci research to reveal starting places for design innovation. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–11.

[89] Qian Yang, Justin Cranshaw, Saleema Amershi, Shamsi T Iqbal, and Jaime Teevan. 2019. Sketching nlp: A case study of exploring the right things to design with language intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[90] 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 designing interactive systems conference*. 585–596.

[91] 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*. 1–13.

[92] Qian Yang, John Zimmerman, Aaron Steinfeld, Lisa Carey, and James F Antaki. 2016. Investigating the heart pump implant decision process: opportunities for decision support tools to help. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4477–4488.

[93] Nur Yildirim, Alex Kass, Teresa Tung, Connor Upton, Donnacha Costello, Robert Giusti, Sinem Lacin, Sara Lovic, James M O'Neill, Rudi O'Reilly Meehan, et al. 2022. How Experienced Designers of Enterprise Applications Engage AI as a Design Material. In *CHI Conference on Human Factors in Computing Systems*. 1–13.

[94] Nur Yildirim, Mahima Pushkarna, Nitesh Goyal, Martin Wattenberg, and Fernanda Viégas. 2023. Investigating How Practitioners Use Human-AI Guidelines: A Case Study on the People+ AI Guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.

[95] Nur Yildirim, John Zimmerman, and Sarah Preum. 2021. Technical Feasibility, Financial Viability, and Clinician Acceptance: On the many challenges to AI in Clinical Practice. In *Proceedings of the AAAI 2021 Fall Symposium: Human Partnership with Medical Artificial Intelligence*.

[96] Sabah Zdanowska and Alex S Taylor. 2022. A study of UX practitioners roles in designing real-world, enterprise ML systems. In *CHI Conference on Human Factors in Computing Systems*. 1–15.

[97] Angie Zhang, Alexander Boltz, Jonathan Lynn, Chun-Wei Wang, and Min Kyung Lee. 2023. Stakeholder-Centered AI Design: Co-Designing Worker Tools with Gig Workers through Data Probes. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[98] John Zimmerman and Jodi Forlizzi. 2014. Research through design in HCI. In *Ways of Knowing in HCI*. Springer, 167–189.

[99] John Zimmerman, Changhoon Oh, Nur Yildirim, Alex Kass, Teresa Tung, and Jodi Forlizzi. 2020. UX designers pushing AI in the enterprise: a case for adaptive UIs. *Interactions* 28, 1 (2020), 72–77.

[100] John Zimmerman, Aaron Steinfeld, Anthony Tomasic, and Oscar J. Romero. 2022. Recentering Reframing as an RtD Contribution: The Case of Pivoting from Accessible Web Tables to a Conversational Internet. In *CHI Conference on Human Factors in Computing Systems*. 1–14.

	AI Feature | Domain	**Capability Level 1**	**Capability Level 2**	**Capability Level 3**	**Capability Level 4**
1	**Biometric Security**	Detect face in image	Detect face	Detect person	Detect
	Risk Mitigation & Security	Identify face in image	Identify face	Identify person	Identify
		Detect fingerprint in image	Detect fingerprint	Detect person	Detect
		Identify fingerprint in image	Identify fingerprint	Identify person	Identify
		Detect voice in audio	Detect voice	Detect person	Detect
		Identify voice in audio	Identify voice	Identify person	Identify
		Detect face in depth map	Detect face	Detect person	Detect
		Identify face in depth map	Identify face	Identify person	Identify
2	**Fraudulent Transaction Detection**	Identify fraud in transaction	Identify transaction anomaly	Identify anomaly	Identify
	Risk Mitigation & Security	Estimate fraud likelihood of transaction	Estimate entity risk	Estimate risk	Estimate
		Discover fraud in transactions	Discover transaction anomaly	Discover anomaly	Discover
3	**Smartwatch Workout Detection**	Detect motion in sensor stream	Detect motion	Detect activity	Detect
	Healthcare	Identify workout in sensor stream	Identify workout	Identify activity	Identify
		Detect step in sensor stream	Detect motion	Detect activity	Detect
		Detect hard fall in sensor stream	Detect motion	Detect activity	Detect
		Estimate energy expenditure of user	Estimate consumption	Estimate outcome	Estimate
4	**Drug Discovery**	Discover relationships between drugs and treatment outcomes	Discover correlations	Discover relationship	Discover
	Healthcare	Generate protein structure of drug	Generate chemical attribute	Generate attribute	Generate
		Generate protein interaction of drug	Generate chemical attribute	Generate attribute	Generate
		Generate physio-chemical reaction of drug	Generate chemical attribute	Generate attribute	Generate
		Generate bioactivity of drug	Generate chemical attribute	Generate attribute	Generate
		Estimate toxicity of drug	Estimate chemical attribute	Estimate attribute	Estimate
		Estimate promise of drug	Estimate success	Estimate outcome	Estimate
		Discover new uses of drug in drug-treatment relationships	Discover correlations	Discover relationship	Discover
5	**Medical Imaging Analysis**	Detect medical anomaly in image	Detect visual anomaly	Detect anomaly	Detect
	Healthcare	Identify anomaly as tumor in image	Identify visual anomaly	Identify anomaly	Identify
		Identify malignant tumor in image	Identify class	Identify attribute	Identify
		Estimate size of tumor	Estimate object size	Estimate world	Estimate
		Identify tumor type in image	Identify class	Identify attribute	Identify
		Discover medical anomaly in image	Discover visual anomaly	Discover anomaly	Discover
6	**Synthetic Health Data Generation**	Generate new patient data from patient data	Generate new data	Generate numeric data	Generate
	Healthcare	Generate missing elements of patient data	Generate new data	Generate numeric data	Generate
		Generate new medical images from medical image	Generate new image	Generate image	Generate
		Generate high-res detail for low-res medical image	Generate image detail	Generate image	Generate
		Generate detail for occulded area of medical image	Generate image detail	Generate image	Generate
		Generate missing view for medical image	Generate new image	Generate image	Generate
7	**Crop Monitoring**	Detect crop stress in image	Detect visual anomaly	Detect anomaly	Detect
	Manufacturing & Agriculture	Identify crop stress type in image	Identify class	Identify attribute	Identify
		Estimate growth of crop	Estimate world activity	Estimate activity	Estimate
		Forecast yield of crops	Forecast financial outcome	Forecast outcome	Forecast
		Forecast yield impact of resource plans	Forecast financial impact	Forecast impact	Forecast
8	**Defect Detection**	Detect product defect in image	Detect visual anomaly	Detect anomaly	Detect
	Manufacturing & Agriculture	Identify defect cause in image	Identify class	Identify attribute	Identify
		Identify defect type in image	Identify class	Identify attribute	Identify
		Discover product defect in image	Discover visual anomaly	Discover anomaly	Discover
		Estimate defect likelihood in product	Estimate entity risk	Estimate risk	Estimate
9	**Robotic Pick and Place**	Detect objects in image	Detect object	Detect world	Detect
	Manufacturing & Agriculture	Identify object in image	Identify object	Identify world	Identify
		Estimate location and orientation of object	Estimate object orientation	Estimate world	Estimate
		Generate motion and grasping path to object	Generate motion plan	Generate plan	Generate
		Act motion and grasping path to pick by minimum moves	Act motion plan	Act plan	Act
10	**Predictive Maintenance**	Detect machine sound in audio	Detect object	Detect world	Detect
	Energy & Infrastructure	Identify machine breakdown in audio	Identify audio anomaly	Identify anomaly	Identify
		Estimate breakdown likelihood of machine	Estimate system risk	Estimate risk	Estimate
		Forecast breakdown point of machine	Forecast failure point	Forecast time	Forecast
		Estimate breakdown duration of machine	Estimate event duration	Estimate duration	Estimate
		Estimate breakdown cost of machine	Estimate financial impact	Estimate impact	Estimate

	AI Feature | Domain	**Capability Level 1**	**Capability Level 2**	**Capability Level 3**	**Capability Level 4**
		Forecast repair impact of machine	Forecast financial impact	Forecast impact	Forecast
11	**Home Energy Optimization**	Detect human presence in sensor stream	Detect human presence	Detect person	Detect
	Energy & Infrastructure	Estimate preferred temperature of user	Estimate preference	Estimate human attribute	Estimate
		Discover person's routine in temperatures	Discover routine	Discover activity	Discover
		Discover group routines in temperatures	Discover routine	Discover activity	Discover
		Identify person's routine in temperatures	Identify routine	Identify activity	Identify
		Forecast peak usage of energy	Forecast peak point	Forecast time	Forecast
		Generate temperature plan for user	Generate action plan	Generate plan	Generate
12	**Text Generation**	Compare phrases by partial sentence fit	Compare phrases	Compare entities	Compare
	Office Productivity & Business Workflow	Generate next word of sentence	Generate word	Generate text	Generate
		Generate ending of sentence	Generate sentence	Generate text	Generate
13	**Spam Filter**	Estimate spam likelihood of email	Estimate entity risk	Estimate risk	Estimate
	Office Productivity & Business Workflow	Identify spam in email	Identify document anomaly	Identify anomaly	Identify
		Identify spam words in text	Identify word	Identify entity	Identify
14	**Language Translation**	Detect text in image	Detect text	Detect entity	Detect
	Office Productivity & Business Workflow	Identify language in text	Identify language	Identify attribute	Identify
		Identify word translation in text	Identify word	Identify entity	Identify
		Identify phrase translation in text	Identify phrase	Identify entity	Identify
		Compare phrases by partial sentence fit	Compare phrases	Compare entities	Compare
		Generate translation of sentence	Generate translation	Generate text	Generate
15	**Meeting Summarization**	Detect voice in audio	Detect voice	Detect person	Detect
	Office Productivity & Business Workflow	Identify words in audio	Identify word	Identify entity	Identify
		Identify phrase in text	Identify phrase	Identify entity	Identify
		Identify sentence in text	Identify sentence	Identify entity	Identify
		Compare words by partial sentence fit	Compare words	Compare entities	Compare
		Generate summary of transcript	Generate text summary	Generate text	Generate
16	**AR Item Viewer**	Detect room in depth map	Detect space	Detect world	Detect
	Marketing & Sales	Detect room objects in depth map	Detect object	Detect world	Detect
		Estimate size of room	Estimate spatial size	Estimate world	Estimate
		Estimate object location in room	Estimate object location	Estimate world	Estimate
		Estimate object size in room	Estimate object size	Estimate world	Estimate
		Detect virtual-physical collision in AR	Detect object	Detect world	Detect
		Generate room with virtual and physical objects	Generate space	Generate world	Generate
		Identify room objects in depth map	Identify object	Identify world	Identify
17	**Personalized Advertisements**	Discover user similarities from user behavior	Discover similarities	Discover relationship	Discover
	Marketing & Sales	Discover ad similarities from user behavior	Discover similarities	Discover relationship	Discover
		Compare ads by will-user-click	Compare documents	Compare entities	Compare
		Compare users to ad fit	Compare consumers	Compare people	Compare
18	**Web Usage Analytics**	Compare products by will-user-click	Compare items	Compare entities	Compare
	Marketing & Sales	Compare users to product fit	Compare consumers	Compare people	Compare
		Discover navigation patterns from user behavior	Discover human behavior	Discover activity	Discover
		Discover user interests from user behavior	Discover user interests	Discover human attribute	Discover
		Discover user similarities from user behavior	Discover similarities	Discover relationship	Discover
		Identify content in web page	Identify content	Identify entity	Identify
19	**Review Analytics**	Identify subject in text	Identify text attribute	Identify attribute	Identify
	Hospitality	Identify sentiment in text	Identify text attribute	Identify attribute	Identify
		Identify user intent in text	Identify user intent	Identify human attribute	Identify
		Identify context in text	Identify text attribute	Identify attribute	Identify
		Discover topics in documents	Discover document attribute	Discover attribute	Discover
20	**Smart Pricing**	Forecast demand for listing	Forecast demand	Forecast attribute	Forecast
	Hospitality	Forecast maximum price for listing	Forecast financial attribute	Forecast attribute	Forecast
21	**Child Welfare Risk Assessment**	Forecast maltreatment risk for child	Forecast human risk	Forecast risk	Forecast
	Governance & Policy	Forecast likelihood of repeated maltreatment	Forecast human risk	Forecast risk	Forecast
		Discover relationships between child maltreatment and locations	Discover correlations	Discover relationship	Discover
22	**Infectious Disease Forecasting**	Identify words in user query	Identify word	Identify entity	Identify
	Governance & Policy	Estimate symptom-relevance of user query	Estimate document attribute	Estimate attribute	Estimate
		Discover relationship between symptom searches and infections	Discover correlations	Discover relationship	Discover
		Estimate number of infected people	Estimate infections	Estimate outcome	Estimate

	AI Feature | Domain	**Capability Level 1**	**Capability Level 2**	**Capability Level 3**	**Capability Level 4**
		Forecast rate of infection	Forecast rate	Forecast attribute	Forecast
		Forecast peak of infection	Forecast peak point	Forecast time	Forecast
23	**Robotic Invoice Processing**	Identify document type in image	Identify class	Identify attribute	Identify
	Finance	Detect document structure in image	Detect document attribute	Detect attribute	Detect
		Identify document structure in image	Identify document attribute	Identify attribute	Identify
		Identify paired content in image	Identify content	Identify entity	Identify
		Identify handwritten words in image	Identify word	Identify entity	Identify
24	**Stock Trading Recommendations**	Discover relationships between news and stock prices	Discover correlations	Discover relationship	Discover
	Finance	Forecast peak price of stock	Forecast peak point	Forecast time	Forecast
		Forecast price of stocks	Forecast financial attribute	Forecast attribute	Forecast
25	**Smart Speaker Question Answering**	Detect voice in audio	Detect voice	Detect person	Detect
	Leisure, Content & Media	Identify voice in audio	Identify voice	Identify person	Identify
		Identify words in audio	Identify word	Identify entity	Identify
		Identify user query in text	Identify user query	Identify entity	Identify
		Identify subject in text	Identify text attribute	Identify attribute	Identify
		Compare responses to query fit	Compare responses	Compare entities	Compare
		Generate human speech from response	Generate human speech	Generate audio	Generate
26	**Media Feed**	Identify person's name in text	Identify person's name	Identify person	Identify
	Leisure, Content & Media	Detect face in image	Detect face	Detect person	Detect
		Identify face in image	Identify face	Identify person	Identify
		Identify content in image	Identify content	Identify entity	Identify
		Generate description of image content	Generate description	Generate text	Generate
		Identify company, organization, or product in text	Identify business entity	Identify entity	Identify
		Identify place-of-interest in text	Identify place	Identify entity	Identify
		Identify sentiment in text	Identify text attribute	Identify attribute	Identify
		Estimate user engagement of media post	Estimate internal state	Estimate human attribute	Estimate
		Compare media posts by engagement	Compare documents	Compare entities	Compare
		Identify bullying in text	Identify text attribute	Identify attribute	Identify
		Identify inappropriate content in text	Identify content	Identify entity	Identify
		Identify inappropriate content in audio	Identify content	Identify entity	Identify
		Identify inappropriate content in image	Identify content	Identify entity	Identify
		Identify copyrighted content in audio	Identify content	Identify entity	Identify
		Identify copyrighted content in text	Identify content	Identify entity	Identify
27	**Game Player**	Compare game moves by game impact	Compare action plans	Compare plans	Compare
	Leisure, Content & Media	Generate game strategy for game	Generate action plan	Generate plan	Generate
		Act game moves to win by minimum moves	Act action plan	Act plan	Act
28	**Image Style Transfer**	Identify content in target image	Identify content	Identify entity	Identify
	Leisure, Content & Media	Identify style in reference image	Identify visual attribute	Identify attribute	Identify
		Estimate content similarity of images	Estimate similarity	Estimate relationship	Estimate
		Estimate style similarity of images	Estimate similarity	Estimate relationship	Estimate
		Generate stylized version of target image	Generate new image	Generate image	Generate
29	**Mobile App Face Filter**	Detect background in image	Detect background	Detect entity	Detect
	Leisure, Content & Media	Detect face in image	Detect face	Detect person	Detect
		Detect eye, mouth, and face landmarks in image	Detect face landmarks	Detect person	Detect
		Detect hair in image	Detect hair	Detect person	Detect
		Identify face gesture and expression in image	Identify gesture	Identify human attribute	Identify
		Detect human body in image	Detect human body	Detect person	Detect
		Identify body pose in image	Identify body pose	Identify human attribute	Identify
		Detect hand in image	Detect hand	Detect person	Detect
		Identify hand gesture in image	Identify gesture	Identify human attribute	Identify
		Generate virtual effects on user face and body	Generate image detail	Generate image	Generate
30	**Deepfakes**	Detect face in reference image	Detect face	Detect person	Detect
	Leisure, Content & Media	Detect eye, mouth, and face landmarks in reference image	Detect face landmarks	Detect person	Detect
		Detect human body in reference image	Detect human body	Detect person	Detect
		Identify body pose in reference image	Identify body pose	Identify human attribute	Identify
		Detect face in target image	Detect face	Detect person	Detect
		Identify face in target image	Identify face	Identify person	Identify

	AI Feature | Domain	**Capability Level 1**	**Capability Level 2**	**Capability Level 3**	**Capability Level 4**
		Estimate similarity of images	Estimate similarity	Estimate relationship	Estimate
		Generate image of target person	Generate new image	Generate image	Generate
		Identify voice in target audio	Identify voice	Identify person	Identify
		Estimate similarity of audio	Estimate similarity	Estimate relationship	Estimate
		Generate voice of target person	Generate human speech	Generate audio	Generate
31	**Lane Departure Prediction**	Detect lane in image	Detect lane	Detect world	Detect
	Transportation	Estimate lane position of vehicle	Estimate object position	Estimate world	Estimate
		Estimate lane departure likelihood of vehicle	Estimate system activity	Estimate activity	Estimate
		Identify driver's intent to depart in vehicle telemetry	Identify user intent	Identify activity	Identify
		Detect objects in sensor stream	Detect object	Detect world	Detect
		Estimate collision likelihood of lane departure	Estimate action risk	Estimate risk	Estimate
32	**Navigation Route Planner**	Estimate street and direction of vehicle	Estimate object orientation	Estimate world	Estimate
	Transportation	Forecast traffic impact of route	Forecast time impact	Forecast impact	Forecast
		Estimate travel time of route	Estimate activity duration	Estimate duration	Estimate
		Compare routes by driver preferences	Compare action plans	Compare plans	Compare
33	**Autonomous Parking**	Identify driver's intent to park in vehicle telemetry	Identify user intent	Identify activity	Identify
	Transportation	Detect objects in sensor stream	Detect object	Detect world	Detect
		Detect parking space in image	Detect space	Detect world	Detect
		Identify objects in sensor stream	Identify object	Identify world	Identify
		Estimate size of parking space	Estimate spatial size	Estimate world	Estimate
		Generate motion path to parking space	Generate motion plan	Generate plan	Generate
		Act motion path to park by minimum moves	Act motion plan	Act plan	Act
34	**Resume Screening**	Identify skills in text	Identify skills	Identify human attribute	Identify
	Human Resources & Management	Identify competence in text	Identify competence	Identify human attribute	Identify
		Identify specialization in text	Identify specialization	Identify human attribute	Identify
		Compare resumes by job fit	Compare documents	Compare entities	Compare
35	**HR Chatbot**	Identify subject in user query	Identify text attribute	Identify attribute	Identify
	Human Resources & Management	Generate response to user query	Generate sentence	Generate text	Generate
		Compare responses to query fit	Compare responses	Compare entities	Compare
36	**Workforce Scheduling**	Forecast demand for staffing	Forecast demand	Forecast attribute	Forecast
	Human Resources & Management	Estimate priority of cases	Estimate priority	Estimate attribute	Estimate
		Generate schedule for employees	Generate schedule	Generate plan	Generate
37	**Automated Essay Scoring**	Identify style in text	Identify text attribute	Identify attribute	Identify
	Education	Identify organization in text	Identify text attribute	Identify attribute	Identify
		Identify coherence in text	Identify text attribute	Identify attribute	Identify
		Estimate grade of essay	Estimate document attribute	Estimate attribute	Estimate
38	**Personalized Lesson Plans**	Estimate skill acquisition of student	Estimate learning	Estimate human attribute	Estimate
	Education	Estimate skill level of student	Estimate competence	Estimate human attribute	Estimate
		Compare math problems by skill acquisition	Compare items	Compare entities	Compare
		Compare unknown skills by learning impact	Compare learning plans	Compare plans	Compare
		Generate learning plan for student	Generate learning plan	Generate plan	Generate
		Discover student stereotypes from student behavior	Discover similarities	Discover relationship	Discover
39	**Aerial Wildlife Monitoring**	Detect animal in image	Detect animal	Detect world	Detect
	Science	Identify animals in image	Identify animal	Identify entity	Identify
		Estimate number of animals	Estimate quantity	Estimate attribute	Estimate
		Discover animal movement patterns in image	Discover animal behavior	Discover activity	Discover
		Discover animal habitats in image	Discover habitat	Discover attribute	Discover
40	**Weather Prediction**	Estimate weather condition for location	Estimate condition	Estimate attribute	Estimate
	Science	Estimate intensity of weather condition	Estimate intensity	Estimate attribute	Estimate
		Forecast temperature for location	Forecast temperature	Forecast world	Forecast