Tools to facilitate working on Machine Learning in the Industry

Shreya Bali

CMU-CS-22-126
July 2022

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Dr. Chinmay Kulkarni, Dr. Sherry Tongshuan Wu

Submitted in partial fulfillment of the requirements for the degree of Master's in Computer Science.

Keywords: Machine Learning, CSCW, Toolkit, Design

For my parents

Abstract

There is an increasing interest in using machine learning(ML) across a variety of industries like finance, education and healthcare. Working on such ML product features, however, remains a point of friction for both non-technical and technical workers who without supporting infrastructure/technical knowledge struggle to fully use their skills in the context of ML. This thesis introduces novel approaches embodied in new systems to help facilitate workers better utilize the opportunities afforded by this rapidly evolving technology. At a high level, we rephrase problems encountered while ideating for ML-related features and utilizing scientific advancements as gaps in communication between the scientific community and the industry, and develop systems to help correct for this. Specifically, in the first project IdeaLens, we explore using real-world use-cases of past ML work as boundary objects while communicating technical abilities of ML work to help non-technical designers come up with new ideas within their domain. In the second project InToResearch we describe the design of a framework to spearhead an alternate ecosystem of ML research papers catered specifically to industry audience, and explore the use of TLDRs to help them navigate and find relevant information more efficiently

Acknowledgments

There are a lot of people without whom this thesis would not be possible

-

Contents

July 25, 2022
DRAFT

List of Figures

List of Tables

Chapter 1

Background and Motivation

Machine Learning (ML) models have resulted in applications proven to be useful in diverse fields such as education[1], social media[2], healthcare[3], finance[4] and disaster detection[5]. Consequently, there is a growing desire to incorporate this new technology in different industries to make products ”smarter” and augment human actions[6, 7, 8]. This seemingly ubiquitous interest in Machine Learning has, unfortunately, not been matched with corresponding support for employees working on such product development. Indeed, the machine learning application development pipeline is currently filled with friction, with both technical and non-technical workers experiencing problems while working with it [9, 10, 11, 12, 13, 14, 15]. While the goals of such features are similar to those in other software development(for eg. fulfilling the needs of users), the unique nature of this specific technology[13, 16] has made it harder to achieve these goals.

Past work interviewing real-world designers, for instance, has highlighted that designers continue to face challenges understanding the capabilities of Machine Learning[9, 10, 11, 12, 13, 14, 15, 17]. Similarly, technical workers working on these areas have been seen to observed to encounter problems using their skills and keeping up with the rapidly evolving nature of this specific technology[]. For context, in 2019, more than 3 Artificial Intelligence-related pre-prints were submitted to axXiv per hour[18]. The definitions of AI/ML in itself are continuously changing, and it has been considered difficult to come up with ontologies/define the capabilities of ML[19, 20].

This thesis introduces novel approaches embodied in new systems to help facilitate workers better utilize the opportunities afforded by this rapidly evolving technology. At a high level, we rephrase some problems encountered while ideating for ML-related features and utilizing scientific advancements as gaps in communication between the scientific community and the industry, and develop systems to help correct for this. Specifically, in the first project **IdeaLens**, we explore using real-world use-cases of past ML work as boundary objects while communicating technical abilities of ML work to help non-technical designers come up with new ideas within their domain. In the second project **InToML** we describe the design of a framework to spearhead an alternate ecosystem of ML research papers catered specifically to industry audience, and explore the use of TLDRs to help them navigate and find relevant information more efficiently.

Throughout these two projects, we first motivate the creation of the specific by analyzing past literature/self-surveys, describe and justify the design of the tool, report relevant results and then discuss the implications of the tool for workers working on Machine Learning.

Chapter 2

IdeaLens: Leveraging Analogical Reasoning to facilitate designer-led ML Ideation

Designers are trained in reframing, bricolage, and similar skills, enabling them to ideate for such features can potentially lead to innovative yet useful ideas, and yet those trying to integrate Machine Learning into their products continue to face problems brainstorming new and feasible ideas, especially when they seek to leverage recent technical advances [9, 10, 11, 12, 13, 15, 17]. The current friction in ideating for ML-related features is therefore not only problematic for the designers who are unable to effectively contribute to the ideation phase of product development which are an important part of application development[21], but also for society at large, as it is not able to adequately utilize the skills of designers. Indeed, past research suggests that companies often end up introducing designers in the Machine Learning design process too late, which results in products that do not serve the needs of end-users[22].

A large focus of past work in helping designers ideate with Machine Learning has remained on developing methods to increase their technical literacy[13]. For example, researchers have designed education material to improve designers' technical literacy [23], created tools that help designers "play" with available data [24], and developed platforms that help them use AI without a lot of coding [25].

However, not only is increasing designer's technical expertise a challenging goal that requires considerable time investment before designers can effectively contribute to brainstorming considering the broad and rapidly evolving capabilities of ML, this approach is also incompatible with designerly ways of thinking about "design materials", which emphasize a focus on design possibilities, rather than technical properties. For instance, renown industrial designer Eero Saarinen understood the design possibilities of bending plywood in three directions simultaneously, leading to iconic designs such as the Saarinen Executive Chairs, but not the technical properties of how plywood was manufactured. Machine learning is no dirrect: interviews with practicing designers suggests that designers do not want too much technical knowledge about machine learning, concerned that this might "distract" them from their real tasks[22].

Unfortunately, this lack of technical knowledge coupled with a dearth of supportive infrastructure today prevents designers from effectively contributing to ideation as they either end up

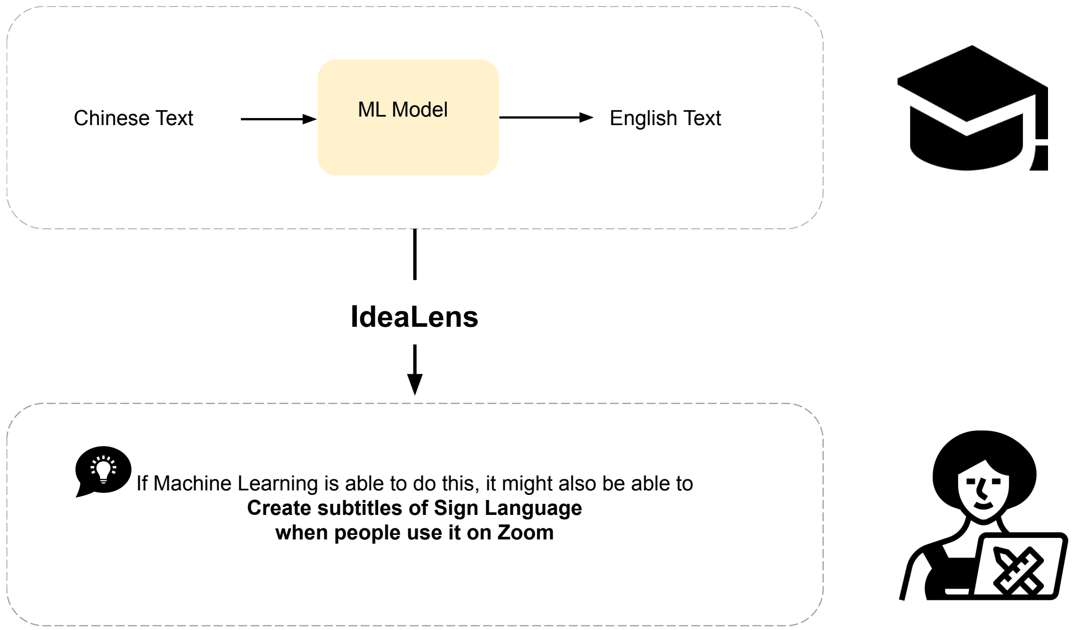

Figure 2.1: Sample Idea created using analogical reasoning in IdeaLens: Here, we show an example idea generated by one of the participants of the study who successfully used analogical reasoning to extrapolate that Creating subtitles from Sign Language might be a feasible use case after seeing the Chinese text to English text use-case

coming up with very basic ideas or consider Machine Learning as ”magic” and come up with capabilities beyond the purview of the field[9, 10, 11, 12, 13, 15, 17]. Both of these go against the adopted goals of brainstorming sessions which focus on coming up with a large number of feasible ideas[26], and as a result, designers are often not included in the ML application ideation[27]

In this chapter, we take an alternate approach to ML-related feature ideation. At a high-level, we consider designer-led ML ideation as an ongoing collaboration between the ML scientific community and designers, where the scientific community is generating technological breakthroughs that the designers can make use of for their products. To facilitate this collaboration, we construct a toolkit, IdeaLens which allows designers to leverage analogical reasoning and use these technical capabilities without having to understand the technical details.

The principles of machine learning nudge the ideas generated using this method to remain within the purview of machine learning(more details given in section 2.1.3), while the relatively open-ended framework of analogical reasoning allows for divergent and creative ideas.

We tested IdeaLens with designers(n=12) with varying levels of technical expertise in machine learning and found that the toolkit helped them come up with ideas that were within the domain of machine learning and diverse. In this chapter, we describe the design and evaluation of IdeaLens-in terms of both usage patterns and the final ideas generate, which provides further insights to this way of coming up with new ML ideas.

Through this work, we demonstrate that: use-cases form powerful boundary objects(i.e objects that can take on different roles for different people[28]) between technical experts and designers and that analogical reasoning is a powerful scaffolding technique for brainstorming. At the same time, our work has immediate implications for practice: the toolkit can be directly used as a source of inspiration while coming up with new ML-related ideas. Participants already displayed real world interest in both introducing the toolkit and the ideas they ended up generating as a result of the sessions.

2.1 Background and Related Work

This work builds on three growing areas of interest in CSCW research: developing techniques/toolkits that help aid ideation[29], helping make Machine Learning easier to work with for technical/non-technical workers[30, 31], and envisioning ways to help include designers/take a designer-ly approach to general software[32] and ML application development[33].

In this section, we underscore (1) the goals of the ideation stage of product development and some relevant methods that have been seen to (2) the current practices and problems encountered in new ML feature ideation, (3) the cognitive theory behind and past related usage of analogical reasoning, and (4) the principles of ML that allow for analogical reasoning to be used as a successful ideation technique.

2.1.1 New Product/Feature Ideation: Tools and techniques

Over the years people have started taking a more scientific approach to innovation and explored how and why certain techniques can aid in coming up with such ideas. Brainstorming and divergent ideation is now considered a core part of application development: the first step in a common design thinking methodology used in a lot of such companies ”Double Diamond” involves coming up with divergent ideas (Discover) and only later in the second stage are these ideas narrowed down[34]. This process has allowed companies to continuously evolve and innovate.

Metrics of brainstorming/ideation sessions highlight the importance of of quantity, novelty, feasibility and usefulness of the generated ideas[35, 36, 37]. However, people often find it hard to divert from their usual way of thinking[38, 39], which poses a challenge in this process. Consequently, a decent amount of past work has looked into developing toolkits to facilitate generation of good quality ideas by both (1) nudging users to think divergently and creatively and (2) steering individual ideas in a specific direction.

- *Nudging users to think divergently*: One way that past research has explored encouraging divergent ideation is by creating tools with open-ended prompts to help users generate large number of ideas. For example, in Spinneret[40], authors used Mind Maps to aid creative ideation by ”non-obvious” concept association. In another line of work, researchers have explored the use of visual metaphors, where users are shown different pictorial images and asked to come up with novel associations to generate ideas[41]. Similar techniques have also been used to help designers think about design ideas: for example, Kumar et al.[42] developed a bricolage technique to help web-based designers be inspired by existing websites.
- *Toolkits to facilitate ”quality” ideation*: In addition to generating a large quantity of creative ideas, these ideas should also ideally be useful and feasible and different toolkits

have been seen to successfully steer generated ideas in such directions. For instance, several method cards have been created to communicate byte-sized pieces of information to steer the direction of the idea[43]. One famous example of this is the the IDEO toolkit[44] which was developed to help nudge designers to consider of human-centric factors while ideating. The easy to parse structure of such cards also allow users to think about factors that they might not even be aware of beforehand. For example, in the paper "Tiles: A Card-based Ideation Toolkit for the Internet of Things"[45], authors successfully use a card-based tool to help non-experts come up with ideas related to augmented objects.

In IdeaLens, we take some inspiration from past work that has successfully nudged brainstorming in 'good' directions, while adapting to the unique problems and nature of Machine Learning to help attain this goal.

2.1.2 Analogical Reasoning: Cognitive Theory and Past Usage

Analogical reasoning is a specific way of thinking based on the idea that because two or more things are similar in some respects, they are probably also similar in some further respect[46]. Rather than understanding the mechanics behind this transfer or coming up with universal principles, the focus of this form of reasoning remains on identifying similarities in the two domains and then applying known facts in one field to the other "similar" field. In fact, past work highlights that even users with no expertise or knowledge in the actual process that makes the carry over feasible have been able to extrapolate useful facts in new domains. For example, various people have created successful products by taking inspiration from nature-without necessarily knowing about the biological or physical reasoning behind these[47]. Similarly, executives without expertise in areas have been seen to come up with solutions "at par with experts" by using this way of thinking [48].

Analogical Reasoning has been used extensively as a problem solving mechanism where people use identified problem areas and try to think about different "similar" areas to come up with innovative solutions. While this makes for a powerful way of thinking, different factors can effect the viability of ideation and the innovativeness of the end results. For example, the distance between the 'source' and the 'destination' domain has been observed to make a difference but the exact pattern of has been a matter of contention: some researchers have claimed that examples from domains farther away from the domain have been seen to help people generate more novel and innovative solutions[49], while others have claimed that nearby examples are sometimes more helpful in creating new ideas[50]. Examples from nearby domains have also been seen as easier to think about and simpler to apply, and people often end up only considering ideas from nearby domains[51]. Recognizing the power of this approach, some past work has also explored developing toolkits to further facilitate this way thinking by developing ways to make relevant example retrieval easier[52] and techniques to effectively use it to think creatively[53]

In IdeaLens, we make use of analogical reasoning to navigate around the problem of designers not being able to ideate effectively with Machine Learning as understanding the limitations and capabilities of this technology is hard. This falls in the category of using analogical reasoning for ideation, but instead of using it for completely open-formed thinking or come up innovative solutions to well-defined problems, we explore how non-technical experts are able to extrapolate technical abilities and come up with new ways of using the technology in their own products.

Here, while the domain is fixed, the specific problems solved/new abilities added to the product are more open-ended. In doing so, IdeaLens explicitly makes use of two of areas of past analogical research: (1) non-experts in the actual solution can still extrapolate its use-case in new domains and (2) analogical reasoning has been able to generate novel ideas/solutions in the past.

2.1.3 Machine Learning Principles that allow for cross-domain transfer of ideas

Since most machine learning models first extract useful features, perform some action on these features and finally generate the result- similar actions on extract-able features should generally lead to ideas within the purview of the field. In IdeaLens, we aim to use this intuition to help designers come up with diverse ideas in their own domains. Even if the generated ideas do not use exactly the same infrastructure as the ideas they are inspired by, ideas that follow similar structure as original idea should generally fall within the wide domain of machine learning

ML researchers have used similar properties while being inspired from past work. The most explicit example of this is transfer learning[54] where a model trained on some task is used with some additional training on a different task. For example, models trained on on classifying images in the ImageNet dataset[55] have been used for tasks like medical image analysis[56] and action recognition in videos[57]. This works because once models are able to ”extracting” useful features on a task, they are able to carry over this ”knowledge” to other domains. Sometimes, however, this transfer is using similar structures. For example, while Transformers[58] were originally made for Language tasks, the same concept has been seen to be useful across diverse domains like computer vision(eg. [59, 60]) and audio(eg. [61, 62]). Other times, there are concepts that end up being useful across different domains, for example estimating and optimizing in for rewards(in reinforcement learning) has resulted in diverse use-cases like self-driving car[63], playing poker[64] and automatic real-time bidding for display advertising[65].

This cross-domain knowledge transfer is encouraging and provides further evidence that working concepts used for ML models can carry over to other domains, and we exploit this fact explicitly in the IdeaLens ideation approach.

2.2 Brainstorming Toolkit

The underlying theme of the IdeaLens toolkit is to help designers come up with ML Application ideas by using analogical reasoning on past machine learning work in a different domain. The principles of machine learning allow such ideas to still likely be within the purview of machine learning(Section 2.1.3) while the open-endedness of this prompt is expected to allow for divergent ideation. To further help users come up with new ideas using this approach, IdeaLens makes use of a three cue-system: IdeaLens card, IdeaLens card hint and product information/constraint . A detailed description of each of these features of the IdeaLens toolkit and their purpose is given in Section 2.2.3.

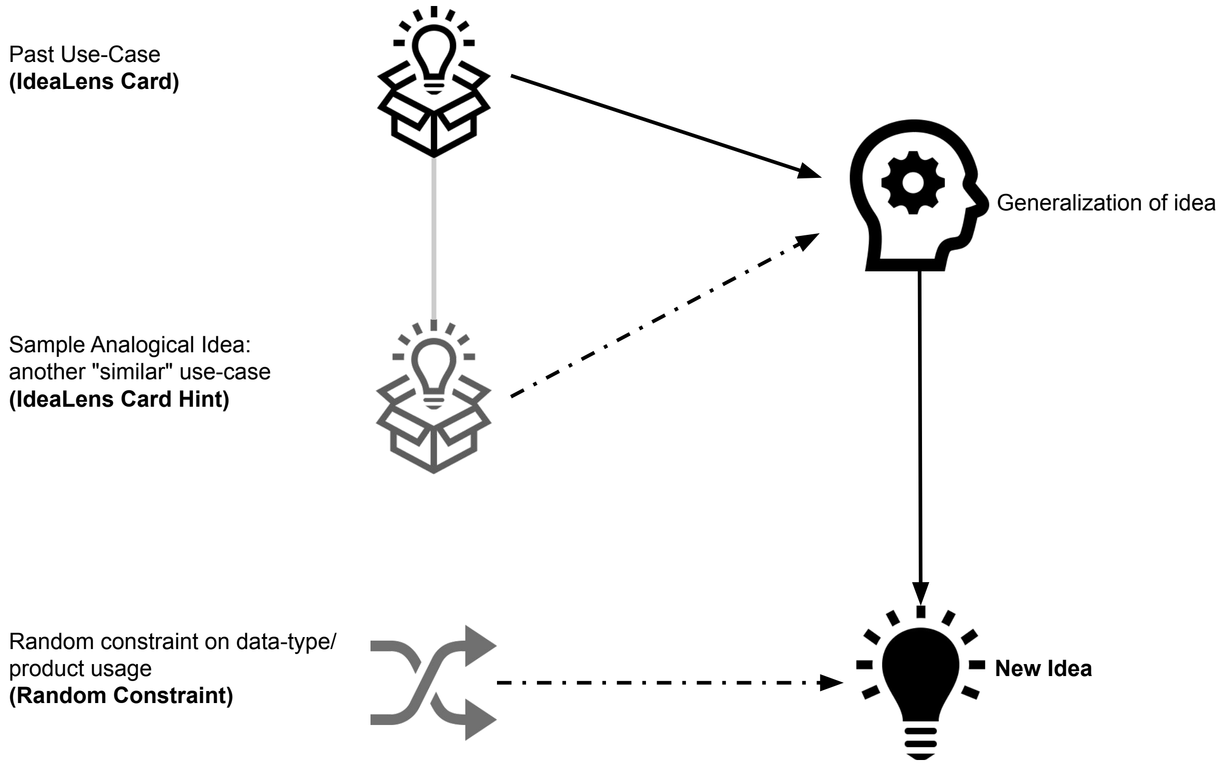

Figure 2.2: IdeaLens Ideation Procedure (dotted lines represent optional additional usage of source to help generate the destination)

2.2.1 Implementation

The IdeaLens Toolkit is available as a fully functional website at https://ml-brainstorming-toolkit.vercel.app. The frontend of the website was developed using Next.js and the backend was developed using Django. The backend and the database are hosted on Digital Ocean, and the frontend is hosted on Vercel.

While method cards are quite frequently distributed in a printed format, since Machine learning is a rapidly evolving field with a lot of use cases, we decided to host the IdeaLens Cards online which gives us the flexibility to dynamically add new cards over time. This also allows for easier distribution and usage of the card toolkit.

2.2.2 User Flow

After a user logs in to the website, they are shown an **About Tab** which contains instructions about the task and the various steps to ideate. After reading the instructions, users are asked to head over to the **Project Info Tab** where they supposed to add in some information about the product they are working with-namely the kinds of data/things they expect to encounter when users use their product, and the different use-cases users might use their product for. Finally, after completing the details of Project Info tab, users are asked to head over to the **Ideate Tab** where they are shown the Ideation interface and asked to come up with new feature ideas.

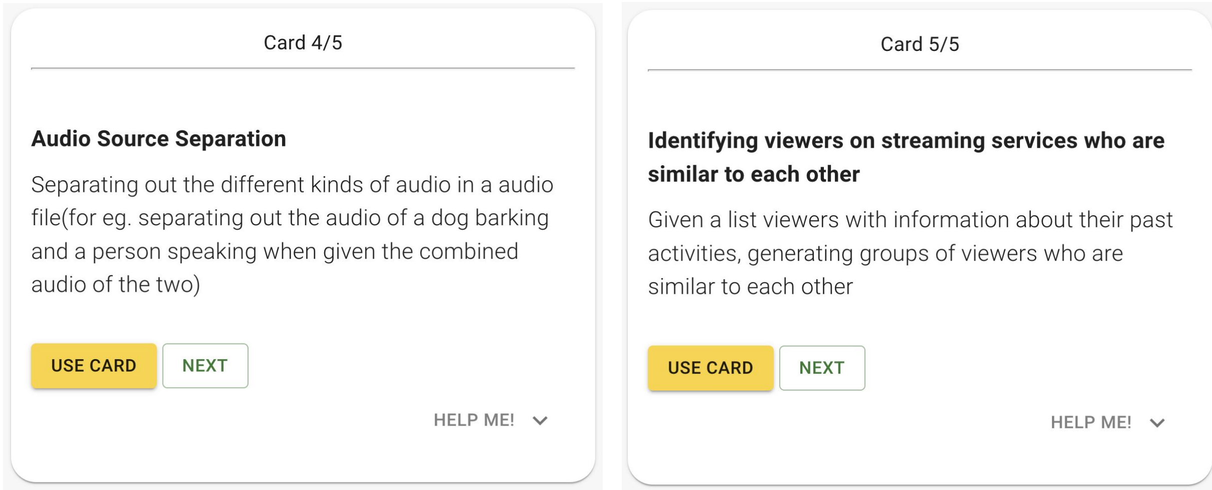

Figure 2.3: Sample IdeaLens Cards

2.2.3 Description, Purpose and Design of Individual IdeaLens Features

In this section, we outline the features and the different design decisions taken while implementing the IdeaLens toolkit. Through these features and design decisions, we focussed on the two goals of this interface : (1) helping users ideate well by allowing for divergent ideation while decreasing cognitive load when possible, and (2) making it easy to extend with additional ML abilities the platform later.

IdeaLens Cards

Each IdeaLens card contains an example use-case where machine learning has been used in the past. Users are encouraged to think about ideas by coming up with generalizations that these use-cases and then applying these generalizations in their own domains. Specifically, the prompt is phrased as follows: "If machine learning has been used to do ___ in the past, what should it also be able to do in your domain?". Two sample such cards are shown in Figure 2.3

Purpose: IdeaLens cards allow for diversity of ideas while nudging the ideas to be within the domain of machine learning. Showing past use-cases allows the generated ideas to still be within the purview of machine learning, while asking users to transform ideas from other domains to their domains allows for unique interpretations and divergent ideation. Ideas generated via analogical reasoning based on such cards are likely to be potentially feasible by the structure of machine learning pipelines(Section 2.1.3): very coarsely, most machine learning models work by first extracting useful features from input data, transforming these learned features in some manner, and then finally generating the output. Analogical ideas in other domains are expected to have similar "actions", and if we are able to generate appropriate features in the new domain, the idea is expected to be atleast potentially feasible.

Design Decisions: IdeaLens aims to use existing solved problems in machine learning as the basis of new ideas. There are multiple ways this can be done: explaining the technical details of the projects, generalizations and taxonomies of ideas etc. We decided to utilize use-cases as the main basis for these ideas because of the following reasons:

- *Examples are easy to parse*: Examples are generally considered easy to parse. In brainstorming sessions, we do not want to overwhelm end users as that defeats the purpose of the cards and examples help give ideation material succinctly.
- *Use-cases can help generate diverse ideas*: Specific use-cases can invoke different thoughts in users and elicit different and diverse ideas from the same cards. This allows use-cases to be a good source of inspiration without explicitly giving away the idea (more about this is described in the results section)
- *Using use-cases as a basis allows the framework to be easily extendable*: Some alternative methods like creating generalizations and taxonomies are tricky to develop when a field is rapidly evolving. In contrast, sample use-cases can usually be extracted easily, making the framework easily extensible in the future. Since there are already many use-cases and new use-cases of machine learning are being discovered, it might be useful to crowd-source IdeaLens cards from research chapter readers in the future. Examples seem like an easy way to extract information useful to designers and extend the number of such cards available. Creating models useful for real world applications is considered one of the main goals of machine learning research[66] and consequently most applied-ML research papers already mention some use-cases of their models or of their datasets.

Participants are shown each card separately and are unable to see all the cards in the beginning. This is expected to help them focus on individual cards without getting overwhelmed by the collection or needing to come up with a mental model of machine learning. This is also consistent with the strategy to constraint thinking used by a lot of other card based tools(eg. IDEO Method Cards[44], PLEX Cards[67]).

Generation of IdeaLens cards: These cards aim to appropriately display machine learning work in potentially different domains to designers with minimal technical ML experience and help them spearhead the ideation process. For this, we essentially display a sample use-case mentioned in past ML research papers as the seed for the new ideas generated and new cards can be easily created by following a similar process.

IdeaLens Card Hints

Each IdeaLens card is further equipped with a ”Help me!” button that provides another use-case that makes use of a ”similar” ML algorithm but in a different domain to help guide the user's thought process. Two sample such cards with their respective hints are shown in Figure 2.4.

Purpose: The IdeaLens card hints are expected to help users generalize the ideas better and also provide inspiration for sample ideas. Each such card is expected to help provide participants a sample extrapolation of the machine learning use-case to show how a generalization of the IdeaLens card can be applied to a different domain. The commonalities in these two use-cases is also expected to provide users with additional help to come up with one sample generalization.

Design Decisions: Providing sample ideas in another domain is expected to help users find commonalities in the domain easier but could also restrict thinking to the specific generalization explored. To help out the participants while still allowing for room for creative ideation, we decided to provide one additional use-case that has been solved by using a similar ML al-

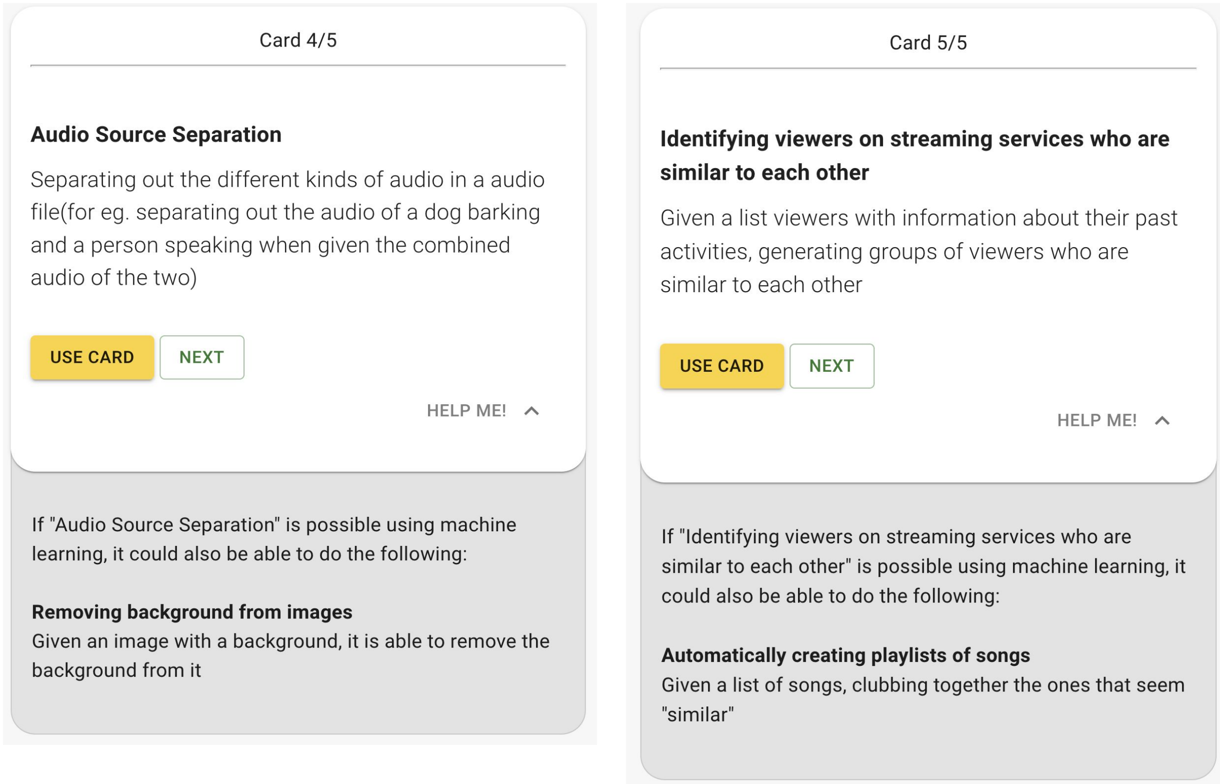

Figure 2.4: Sample IdeaLens Cards with their respective ”hints”

gorithm/model. This minimalistic approach not only lets new cards be added in later easily but is also expected to prevent users to become overly fixated in a single direction of thinking. By default the ”Help Me!” button is shown after around 10 seconds after the IdeaLens card is displayed to encourage users to think about the IdeaLens card and different possible generalizations initially and then use the additional example to help with generalization.

Generation of IdeaLens Card Hints: Each hint used in the study is essentially a machine learning use-case in a different domain whose model shares algorithmic similarities with the original IdeaLens card.

Project Information & Random Constraint

Before they can use start using the cards, users are asked to create lists of potential use-cases of the platform and kinds of data that exists in their platform in the Project Info tab. Additionally, they are given an option to randomly generate a random fact about their platform while using the use-case cards to help them restrict the scope they are working with. Screenshots of user-entered product information and a corresponding random constraint Figure 2.5.

Purpose: The Project Info tab is expected to help users be cognizant of what the actual use-cases of the platform are, and accordingly generate ideas in line with what would actually be useful in the platform. This can not only help decrease the cognition load but can also further increase the diversity of the ideas generated. In particular, once the user presses the Random

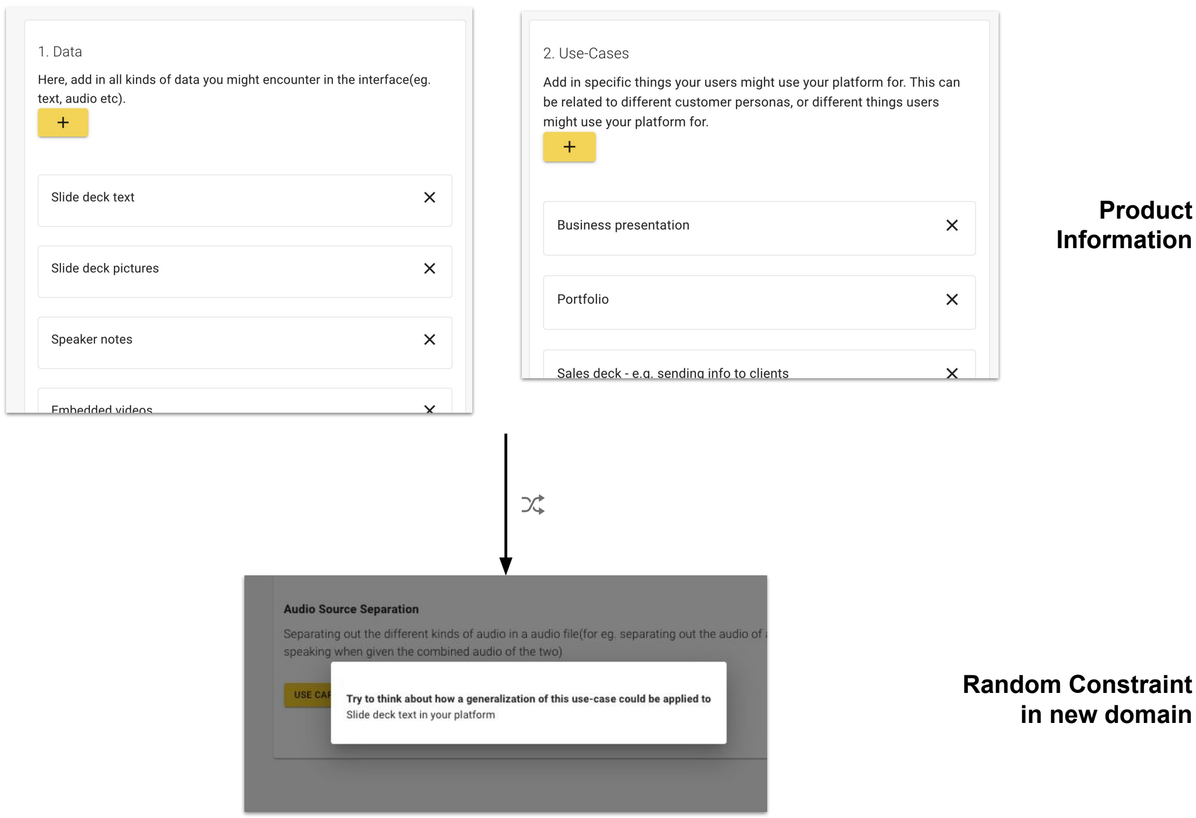

Figure 2.5: User-entered Product Information and a corresponding Sample Random Constraint

Constraint Button, they are shown either a restriction on the ”kind of data they are working with” or a ”use-case” they should consider while thinking about the idea.

Generation of Random Constraints: The random constraints in the domain are randomly generated using the information added by the user in the Project Info tab.

2.3 Evaluation: Case Studies and an analysis of ideas generated

We recruited n=12 participants to study IdeaLens in action. Each session was conducted over Zoom with Screen sharing enabled. The goal of these studies was to validate the use of this approach to help designers ideate machine learning-related features, and also observe usage patterns. For each such study, this included a two step process: asking participants to use the toolkit to come up with ideas, followed by analyzing the generated ideas.

2.3.1 Participant Recruitment and Studies Description

We recruited a combination of students and real world UX experts with different levels of self-reported ML Expertise. Participants were recruited using mass emailing and personal contacts. In total, we recruited 12 participants across the two case-studies. This size is similar to comparable past researches in ideation interfaces[68, 69], toolkits that help integrate non-technical experts in other stages of the machine learning pipeline[70, 71] and studies evaluating designer experiences with machine learning[15, 22, 72].

The table below summarizes the relevant job details and expertise levels of each of the recruited participants.

	Job Title	UX Expertise (self-reported out of 5)	ML Expertise (self-reported out of 5)	Product Case-Study
P1	MHCI Student	4	2	Powerpoint + Zoom
P2	MHCI Student	5	4	Powerpoint + Zoom
P3	UX Designer	4	3	Powerpoint + Zoom
P4	CEO, UX Consultancy	5	1	Powerpoint + Zoom
P5	UX Design Manager	5	2	Powerpoint + Zoom
P6	Head of Design	5	2	Powerpoint + Zoom
P7	Product Designer	4	1	Powerpoint + Zoom
P8	BDes Student	2		Powerpoint + Zoom
P9	BDes Student	3	1	Powerpoint + Zoom
D1	**Independent Game Designer**	**5**	**2**	**Customizable VR Video Game**
D2	**Product Designer**	**4**	**1**	**Business Chat Platform**
D3	**Product Designer (past)**	**4**	**1**	**EdTech Platform**

These case-studies helped us understand the broader applicability of IdeaLens while also noting the patterns of ideating over the same product. To appropriately achieve both these goals: we divided our evaluation study into two parts:

- The first **9** participants were asked to imagine they are working on two real world application sequentially: Powerpoint and Zoom, and use the cards to brainstorm feature ideas for the product. The order of these products was randomly selected for each participant. Making the product the same across participants allowed us to further investigate the success and failure modes of the brainstorming toolkit. Powerpoint and Zoom were specifically chosen as participants were expected to already be familiar with these products and they provided a different enough. Of these nine participants, three (P1, P3, P4) were able to get through only one of the of the two case studies due to time constraints: two(P1, P3) used all the time on one product, one(P4) was only available for 1 hour. Two of these participants(P1, P4) completed the Powerpoint task, and one(P3) completed the Zoom task. Hence, in total 8 participants tried to come up with ML Application ideas for Microsoft Powerpoint and 7 for Zoom. Each session took approximately 1.5 hours and was conducted independently over Zoom to help isolate the different interaction patterns with the interface.

- The last **3** participants were asked to use the toolkit for any one of the products they have actually worked on in designer-roles. The participants in this study were expected to be more familiar with the product they were brainstorming on, including the goals of the products and the desires of their customers. Hence, in addition to using designer-ly skills, this study was expected to help us further evaluate and analyze the usage of the toolkit in a more real world context. This also helped us further evaluate the usage of the platform across more industries. Each session took approximately 1 hour and was conducted independently over Zoom.

Our case studies, thus, spanned 5 different products and included 12 participants. We conducted a total of 21 ideation activities (first nine participants being included in two each, and the last three in one each). A combination of simulating ideation on fixed products, and letting participants ideate on their own products helped us better understand the differences in patterns of thinking across participants and across different domains.

2.3.2 IdeaLens Cards and Hints used

For the course of the study, we populated the toolkit with five sample IdeaLens cards and hints. To maintain consistency across the different case-studies and participants, these are kept constant throughout the study but they were presented in a random order to the different participants. This is obviously not expected to be an exhaustive list representing all possible machine learning use-cases but represents some relatively different domains, scopes and algorithms where ML has been used in the past. Specifically, we used the following ML algorithms/models: Seq2Seq, Recurrent Neural Networks, Convolutional Auto-Encoders, KMeans Clustering and Reinforcement Learning in the IdeaLens cards. The specific use-cases mentioned in the IdeaLens cards worked in domains across audio/songs, text, images, streaming services, playing cards, and driving.

The two real world use-cases(IdeaLens card, IdeaLens card hints respectively) used in each of these cards is given in the table below. We also mention the specific algorithms/models corresponding to these cards in the table for further reference.

	IdeaLens Card Use-Case	IdeaLens Card Hint (Another ”similar” Use-Case)	Sample ”similar” model/algorithm used across these tasks
1	*Translating Chinese Text to English Text*	*Generating the song lyrics of a song* : Given the audio version of a song, generating the text version of it	Seq2Seq models like Transformers ([73] & [74])
2	*Stock price prediction*: Predicting future stock price using past trends and market information	*Text Auto-complete*: Generating the next letter given some text	Recurrent Neural Networks ([75] & [76])
3	*Audio Source Separation*: Separating out the different kinds of audio in a audio file (for eg. separating out the audio of a dog barking and a person speaking when given the combined audio of the two)	*Removing background from images*: Given an image, removing its background	Convolutional Auto-Encoders ([77] & [78])
4	*Identifying viewers on streaming services who are similar to each other*: Given a list viewers with information about their past activities, generating groups of viewers who are similar to each other	*Automatically creating playlists of songs*: Given a list of songs, clubbing together the ones that seem ”similar”	KMeans Clustering ([79] & [80])
5	*Playing poker*: Suggesting the strategy while playing poker when given a hand of cards	*Auto-driving*: Given the road conditions, suggesting the directions	Reinforcement Learning ([63] & [64])

2.3.3 Metrics of Evaluation

Three different metrics were used to evaluate the effectiveness of the brainstorming toolkit: participant feedback, machine learning expert feedback on the feasibility and innovativeness of ideas, and feedback on the usefulness of ideas.

Participant Activity and Post-Completion Survey

Each ideation session was conducted independently over Zoom to help isolate the usage across different levels of ML/UX expertise. Screen sharing was enabled and participants were encouraged to think aloud while ideating. At the end of the brainstorming activity, participants were asked to respond to a survey about the toolkit and reflect on its usage. Quantitavely, this included rating the understandability of each of the cards on a 7-point Likert scale and whether they thought the toolkit helped them come up with ideas they would not have had otherwise.

Technical review of generated ideas

We recruited 6 independent evaluaters with technical ML expertise to rate the feasibility of the ideas generated and each idea was judged by two unique evaluators. We reached out to past students of grad-level ML courses/TAs of similar courses in a large private university in Pittsburgh. Each evaluator was allocated a random case-study and a shuffled list of generated ideas. They were then asked to fill out a Google Form judging the feasibility of each allocated idea.

The measure of feasibility used was inspired from the Microsoft Playbook for ML Feasibility Studies[81]. Essentially, evaluators were asked to judge whether the task cannot be completed without ML(i.e manual coding would not work), and then how feasible they believe completing the ones using ML would be. This gave us a measure of whether the generated ideas were indeed feasible and mark ideas according to the following categories: Do not require ML, Definitely Feasible using ML, Seems hard but probably feasible, and nearly impossible/would require SOTA work.

Usefulness review of generated ideas

In addition to ideas being within the scope of machine learning, the ideas should be useful for the toolkit to be considered helpful. For the first part of the study, we recruited users of the applications (i.e. Zoom and PowerPoint) to judge the usefulness of the suggested features, such that each idea was evaluated by at least four users. We recruited a total of 20 users for Zoom, and 20 users for PowerPoint using the Prolific online platform, and paid them $1.00 for their ratings. This rating task took each rater approximately five minutes.

For the second part of the study, since we had recruited designers who were working with specific products with often niche clienteles, we could not evaluate the ideas by users viewpoints directly but asked the participants themselves whether the toolkit helped them come up with useful ideas for their platform. This is predicted to be a good proxy as these designers had been working on the product and are expected to understand client needs.

2.4 Results

In this section, we report the results from the case-studies in light of the various goals of IdeaLens: utilizing use-cases as effective boundary object, inspiring designers to come up with effective ideas for their products and exploring the various patterns used by the participants.

While we conducted different case-studies, the results and patterns of usage were largely the same across the sessions. Hence, we report the results of these studies together and highlight the slight differences in the observations within these results.

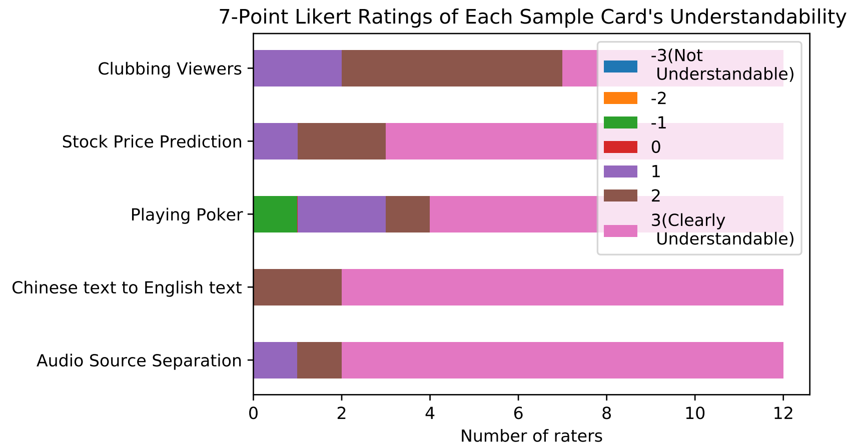

Figure 2.6: Likert Ratings of Card Understandability: Participants were generally able to understand the different cards and what they were trying to communicate

2.4.1 Participants usage of IdeaLens cards as boundary objects allowed them to extrapolate ML capabilities and generate new ideas

Most participants(10/12) were able to come up with atleast one idea per card during the course of the study. The few who were not able to use them, mentioned that they were having trouble using the cards in their context, and not because they were unable to understand the use-case or what it was trying to communicate. In fact, on the 7-point Likert scale, most participants rated most of cards as ”very understandable” (Figure 2.6).

Moreover, most of these ideas generated extrapolating these various interpretations were in fact valid ML use-cases according to the technical evaluators hired as a part of the study(Figure 2.7), despite the participants being asked to focus solely on the quantity and not worry about the usefulness/feasibility of the ideas. Each generated idea was evaluated by 2 unique technical experts and the inter-rater reliability across the tasks was 83.8%.

At the same time, the differences in the sample cards themselves helped us further understand how to phrase future use-cases to show the users. One participant for example mentioned: [wrt to the *”Poker Strategy”* card] *”This one was a bit confusing.. the strategy part helped me but I guess I’m not that familiar with Poker so that threw me off a little bit” (P8)*. Another participant (D2) raised the concern that while the *”Clubbing Viewers”* use-case was understandable in that it adequately communicated the capabilities, it would be easier to extrapolate it to new use-cases if it also mentioned what the algorithm did after clubbing the users.

2.4.2 Divergent interpretations of IdeaLens Cards spearheaded ideation and encouraged creative thinking

Overall, all participants came up with multiple new ”ML” ideas in different domains: at the end of the first study, there were a total of **92** unique ideas for Microsoft Powerpoint generated by the 8 participants ideating for this feature, and **52** for Zoom across the 7 participants ideating for it; at the end the second case study, D1 came up with **8** ideas for their virtual reality video game, D2 with **22** for their business chat platform, and D3 with **15** for their edTech platform. Thus, in total, the toolkit helped users come up with **189** ”ML” ideas across 5 product areas. In fact, a large majority(11/12) of participants felt that the IdeaLens framework helped them come up with ideas they would not have had otherwise.

Further, consistent with the hypothesis that this framework allows for divergent ideation by encouraging diverse but valid interpretations, each card was able to generate multiple ideas across the different ideas. Figure 2.7 shows a distribution of these along with feasibility ratings: Each card was seen to create multiple feasible ideas across the different use-cases.

Participants were able to carry over the ideas to different domains

: The IdeaLens toolkit was tested on 5 products and each of these products contained multiple different modalities, which helped us further understand the use of these five cards across different possible domains. After organizing ideas according to the different kinds of data they work on, we observed that participants had used these cards across diverse areas like audio, images, video, text, slide design/format, multimodal input and numerical data.

IdeaLens cards provided different pathways to help users come up with ideas

Analogical reasoning allows thinking at different levels of distance[82] and we observed a similar pattern used for ideating by participants using our toolkit, with most participants thinking at ”different distance levels” to help them ideate. Carrying over the ideas from different domains helped them view their products and the problems experienced through different perspectives.

In particular, we observed new ideas being created by extending mentioned use-cases and sometimes previous ideas in three different ways. We describe these below to provide a flavor of how IdeaLens helps ideation, rather than as clearly distinct categories of ideas.

- **New ideas by focusing on explicitly mentioned ’action’ words alone:** In this way of generating ideas, participants seemed to either focus on the specific ”action” key words or similarities between the use-cases in the IdeaLens card and the IdeaLens card hint alone, without necessarily thinking about additional details of these respective use-cases. This allowed them to make valid assumptions about the varying abilities of machine learning algorithms and come up with relevant ways to use these in their domain of interest. For example, one participant(P3) came up with an idea to *”Consolidate design feedback: Grouping similar comments together”* in Powerpoint after seeing the *”Clubbing Similar Viewers”* card and the corresponding hint to *”Automatically create playlists of similar songs”*. The generated idea did not have a lot in common with either of these two use-cases outside of the actual task performed by the ML algorithm.

  Reflection and Implications: In a way, this form of ideation is similar to generating ideas after being explained the corresponding capabilities of machine learning. However, we did not need to explicitly tell participants these capabilities or explain the mechanism behind it to help participants generate these new ideas.

- **New ideas by exploring specific themes explored in the IdeaLens card examples + generalizing** Participants also thought about additional details of the specific use-cases in the IdeaLens cards and IdeaLens card hints beyond what was explicitly mentioned in the cards to help guide their ideation process, and often look at their own from unique perspectives. For example, upon reading the auto-driving hint corresponding to Poker IdeaLens card, P1 thought about the potential of machine learning to make real-time changes, and came up with an idea that allows the *flow of the presentations to change according to the amount of time left*. Same card reminded another participant(D3) of money and strategy instead and they came up with an idea to *"Suggest users the best way to spend a fixed amount money based on past activities"* on their EdTech platform. This pattern was largely observed across different participants who focussed on not just the generalization but some specific facts about the ideas to help them generate more ideas and think of their problem area creatively.

Reflection and Implications: While focussing on surface level details of the source domains has been considered to be harmful for problem solving using analogical reasoning and we had not included a lot of additional details for the mentioned use-cases, in this case trying to find corresponding similarities helped nudge users in more divergent ways of thinking, look beyond the exact umbrella category of machine learning models. This allowed them to see their domains in creative ways and come up with novel ideas. At the same time, this observation can also inform the design of future cards as less familiarity with domains could potentially impact the ideation process.

- **Minor modifications on existing ideas:** Finally, some examples mentioned in the cards could also be directly applied to the domains explored and participants were able to generate new ideas by modifying these a bit. For example, most participants came up with a translation-related idea after seeing the *Chinese to English translation* IdeaLens card almost immediately.

  In addition to this, however, participants also started associating existing knowledge with new ideas. For example, P5 came up with an idea they titled *"Ripping of from Copy.ai"* for Powerpoint with the following description *"Suggest alternative titles given a specific sentiment (funny, instructional, serious, academic)"* after seeing the *Chinese to English translation* IdeaLens card, and another idea to *"automatically generate Video Filters... similar to Snapchat"* after seeing the *Audio Source Separation* IdeaLens card for Zoom.

  Reflection and Implications: Facilitating or nudging relavent example retrieval or providing direct but diverse examples is helpful in its own right as remembering relevant examples while ideating as it helps brings relevant examples to working memory and using existing close-by ideas have also been seen to create useful and creative ideas[50]. However, such examples could also make it harder to ideate divergently/away from ideas that have already been explored-as users could start strongly associating some cards with the already explored idea in the domain. *"I was able to come up with more new ideas for the Audio Source Separation card in the Powerpoint case.... and not for Zoom... because I wasn't able to think beyond what has already been implemented" (P6).* Interestingly, this same card actually had the highest total number of ideas of all the IdeaLens cards in Zoom in total in the end-suggesting different individual patterns on such fixations. This initial fixation was not just limited to the exact example used in the toolkit but also with respect to ideas that participants were reminded of. For example, D2 mentioned that they were having trouble coming up with an idea for the Playing poker card as this was reminding them of an idea they had already explored in the team and had to cycle back to this card before they were able to come up with a new idea. In such situations, random constraints prove to be specially helpful as they nudged users in a different direction. This finding also supports further benefits of adding in multiple cards because fixation could not only be driven by the use-case itself but also by the analogical ideas these remind users of-which is harder to predict/optimize for.

2.4.3 Participants' and Real-world users' interest in the generated ideas

We evaluated the actual usefulness of the ideas using a combination of participant's feedback and the opinions of real-world users on the generated ideas.

Participants thought the toolkit helped them come up with creative ideas and displayed real-world interest in the generated ideas

Multiple participants also seemed to be genuinely interested in and excited by the ideas they ended up creating and mentioned this often unprompted. While talking aloud during the interview, P3 mentioned that while they spent more time coming up with the idea related to Stock price prediction, they ended up really liking one of the ideas they generated(using Zoom to give metrics of performance): *"... actually if you can create this, that would be great! I would use it!" (P3).* Another one of the participants from the first study(P5) reached out to us to explore the usage of the toolkit in helping their team ideate machine learning related features for their product.

A similar pattern was observed in the second study as well. In this case, the participants had already been involved in multiple brainstorming sessions, including thinking about ways to include machine learning in their products. However, the toolkit was still able to help them generate ideas that they thought were useful for their application. For example, one participant (D1) asked the interviewer if they could get a copy of their ideas during their session as they thought that the ideas were useful to the actual brainstorming of their product. Other participants also displayed comparable interest in the ideas they ended up creating and the toolkit in general. *"When I first started [working at my company and using machine learning], I had trouble understanding what's possible. I knew what we had done previously when I joined the team, ... [the ideation procedure] was more based on what the customers said ... like scheduling an appointment so I kind of tried to find how that could apply to other parts of our product... and making more obvious connections... something like [the IdeaLens toolkit] where you've given those different examples and trying to like extrapolate from there is definitely very useful" (D2).* Other participants suggested support that would be useful beyond brainstorming, such as validating ideas: *" I felt pretty good about the ideas, because I think most of them would be helpful, but you [also] need to topic validate the ideas" (D3)*

Real world users of Zoom/Powerpoint are interested in the new feature ideas generated as a result of the first case study

Finally, we also conducted a preliminary study to evaluate the usefulness of the ideas generated in the first case study as a sanity check that the ideas generated using the toolkit where in fact useful. The second case study involved ideating for use-cases for the companies directly and we were not able to test these ideas because of (1) very specific clientele and more importantly (2) the vested interests of the participants in the actual ideas. However, the first study provided a good opportunity to explore the potential use of the generated ideas as they were generated as a result of a simulation. We asked real-world users of Zoom and Powerpoint to evaluate the usefulness of the ideas on a 7-point Likert scale. Each idea was evaluated by atleast 5 independent such users, and here we report the summary of these results. Overall, **73/90** of the ideas created for Powerpoint had strictly positive mean score and **39/51** for Zoom. The split across the different cards is given in Figure 2.8, and most of the ideas even within each of the card categories are observed to have strictly positive usefulness score.

2.4.4 Additional Usage Patterns

Focus on structural similarities between tasks, rather than categorical differences

One problem in machine learning related ideation is that without explicit technical education about machine learning, users often have certain incomplete pre-conceived mental models. This restricts them to think in linear manners and potentially only come up with limited ideas. While going through the set of cards, P4 asked the interviewer why stock price prediction and auto-complete were mentioned together, as *"Stock price prediction uses past trends to predict values"* while *"auto-complete is NLP"*. Both of these are in fact similar use-cases of machine learning and often use very similar models(for eg. Recurrent Neural Networks have been used for both Stock Price Prediction[76] and for Text Auto-Complete[75]). The cross-domain hint helped them break away from the over dependence on incomplete technical details, and in the post-survey interview, P4 specifically mentioned *"The hints helped me a lot"* . Similarly, while ideating about a use-case inspired from Audio Source Separation, P5 mentioned that their idea *"might not be completely related as it uses Computer Vision"* but later became more comfortable coming up with new ideas in domains different from the one mentioned on the cards.

The participants mentioning technical terms without understanding the theoretical concepts behind machine learning or how it is able to learn from different kinds of data was surprising, and suggests that designers might have been looking for different ways to understand the concepts behind this new technology. The incomplete mental models could however restrict their thought process, and result in very limited sets of ideas- thinking about ideas they were already aware about in the "same" domain.

As with "Mind-Map"-based tools, participants started focussing on similarities instead of categorical differences in ideas. This is beneficial as it leads to more divergent thinking and was seen to still keep a majority of the generated ideas within the domain of machine learning. They were sometimes also excited by how ML ideas in relatively far off domains could help them view features of their own products differently, for eg. D3 thought that the Poker strategy card was *"surprisingly useful"* for coming up with new ML Application ideas for the EdTech platform they had worked on.

Differing usage of the additional optional features: Hints and Random Constraints

- "Hints" used as additional analogies, and potentially avoided fixation While optional, most (10/12) participants used the "Help Me!" button across the IdeaLens cards. Many participants seemed to use them as additional analogies, P1 for instance, specifically saw the Hint as a separate use-case to further help them ideate.

  Some participants regulated their use of the button to avoid fixation: (6/12) used the button only after trying to generate atleast one idea; others (6/12) used the button immediately.
- The use of the Random Constraint button was much more scattered and different users seemed to have different opinions on the usefulness of this button. Most(8/12), however, still used it in more than one isolated incident to help inspire new ideas, and the randomness of the cards seemed to encourage more divergent thinking when used. For example, one participant mentioned *"I probably wouldn't have thought about the Breakout Room idea[to automatically generate groups of people based on how active they are in the Zoom meeting] if it was not for the Random Constraint"(P2)*. In contrast, however, a different participant said *"I didn't use the Random Constraint a lot.... but I think it's better to have it in rather than not have it... I liked that you had mechanisms to help get unstuck"(P8)*. While the usage of the Random constraint differed across the different participants, the product information indeed seemed to help the participants be in the right frame of mind while ideating as most participants ended up generating ideas related to the information they added in in the Product Info tab. Some participants(P5, P6, D3) also went back to the Project Info tab to see this information again to come up with more ideas.

2.5 Discussion

While one goal of this work was to develop a system and generate a usable artefact that in itself helps designers come up with new ML-related feature ideas, more broadly this work aims to contribute to the continuing discussions on incorporating designers in ML-feature development and supporting creative ideation with technologies.

Through the IdeaLens approach, we advocate for viewing use-cases(both existing and conceptual) as a unit of knowledge transfer between technical experts and designers. Instead of seeing real-world use-cases as the end-results of technical work or products of ideation sessions, we believe they are a rich representation of knowledge about technical capabilities that are easy to understand, and thus a powerful boundary object. As such they served as a useful as a communication tool beyond simply communicating the actual use-case to the respective audiences.

2.5.1 Real-world use-cases of technical advances in machine learning as a way to ”communicate” technical knowledge

With the constant technical advances in ML on one hand and a corresponding desire of companies to be able to make use of ML for their own use-cases on the other, there is a definite desire/need to find good ways to adequately communicate ML capabilities to employees working on such products. The current knowledge gap encountered especially for non-technical workers including designers trying to work on this technology has thus naturally been seen as problematic.

In this toolkit, we target a specific portion of this population-”designers” and explore use-cases mentioned in ML research papers as a way to roughly communicate the various capabilities of machine learning while hiding the technical details. Through the power of analogical reasoning, designers can then make use of these as ”seeds” to come up with new use-cases of ML in their own domains, while the principles of ML help keep them within the purview of the field.

In this way, IdeaLens views use-cases as a unit of abstraction of ML capabilities and provides a sample pathway to successfully use them as a boundary objects while coming up with new ideas. Participants in our case-studies were able to effectively make use of this structure to help them generate new ideas and were even able to extrapolate the more nuanced details of some capabilities not explicitly mentioned to guide their ideas: for example, participants were able to infer that some machine learning algorithms can work in real-time by seeing the auto-driving example even though this fact was not explicitly mentioned. This is especially helpful as then these specific use-cases can then help users get inspiration from different ML capabilities without having to exhaustively go through literature/technical details behind the algorithms that make this possible, while still taking advantage of such advances to help guide their own ideas.

2.5.2 Designer-led Product Feature Ideation as a means to communicate needs/desires of customers to Technical ML Teams

Another implicit theme explored in IdeaLens is the use of ideation to communicate user-needs and desires to the technical workers. Past research has shown that UX is often an afterthought for ML applications[17]-not only resulting in under-utilization of designer skills while coming up with such ideas but also leading to creation of products that are often not useful for end-users and one reason behind this is the fact that designers are often not included in ideation sessions because of the friction they experience while ideating and the consequent inability contribute properly contribute [27].

In IdeaLens, we help designers use their skills in bricolage, knowledge about users, and their focus on creating ”useful” ideas while thinking about new ML ideas. We view the resulting ideas to be a culmination of this way of thinking, and through IdeaLens allow designers to package their information about the users/problem areas into valid ML ideas, which could eventually guide technical teams. In the initial evaluation study, participants were already able to come up with new and divergent ideas several of which were considered as creative yet feasible uses of the technology by the hired ML technical experts, and useful by end-users.

2.5.3 An extensible and flexible framework to inspire ideas in a rapidly evolving field

Past work has highlighted that non-technical experts like designers have found it hard to come up with machine learning-related features as they often lack. Previous approaches have attempted to solve for this problem by finding ways to impart technical knowledge to designers, but even if we forget the how this deviates from designerly methods of working-fully understanding/defining the capabilities of this field has been considered hard even for technical experts-the field is simply too broad and diverse, and moreover rapidly evolving.

Indeed, one overarching theme explored through the design of this toolkit is the specially broad and rapidly changing nature of the ML field. We argue that tool that aims to help users ideate with this design materials should be flexible enough to not only be able to communicate current sources of information, but also be able to extend as the field inevitably discovers more possibilities and opportunities.

The card-based interface used in IdeaLens, for example, allows for an easy transfer of knowledge of these broad capabilities by splitting up the broad capabilities into different algorithms and providing examples of similar algorithms as abstractions to communicate this information without delving into technical details. Moreover, the IdeaLens approach also provides an easy and scalable method to allow new cards to be be added to the platform to create more seeds and as the field continues to evolve: each IdeaLens card is simply a use-case of a past example where ML has been used accompanied by another use-case whose machine learning algorithm makes use of a similar structure as the original use-case(ideally operating on some other domain to encourage more divergent ideation). This can often be easily extracted from such papers. Hosting the platform online also lets us utilize affordances awarded by the internet to help make this approach scalable in terms of ease of update and a broadeer distribution.

This approach is also flexible to allow for the varying motives of companies trying to use machine learning. At its current level, for example, the IdeaLens toolkit makes no assumptions about the goals of the ideation-the framework is flexible enough that it can be used for different goals-such as thinking about ideas that might provide companies a competitive edge, ideas that might require additional data collection but are worth the effort to do so and those for which off-the-shelf algorithms are applicable. In fact, we were able to observe a distribution over these in the studies also. At the same time, while the toolkit in its current state is aimed to help ”general” ML Application ideation, there are also clear ways to add on some restrictions in the current interface itself. For instance, the successful use of ’Random Constraints’ by participants provides some initial evidence of this, and provides one simple way of how such constraints can be added in in the future. For example, in the current version of IdeaLens, one kind of the random constraint asks designers to think about how a generalization of the given use-case could be applied to a specific data-type in their platform. This could easily be updated with only those data types that are already available.

2.5.4 Study Limitations

Lack of an experimental control

The work is motivated by the fact that designers in the real world are not keen on and should be expected to learn the technical details of machine learning to be able to work with it. Our work provides an initial evidence for the use case of an analogy-based toolkit to define machine learning features. Due to the lack of an established comparison baseline for the toolkit, especially for UX researchers without prior knowledge of AI, the evaluation was made compared to the prior experiences and how the participants 'felt' while using the toolkit and no control was used. For example, while one such approach could include asking participants to come up with ML-related features without the toolkit, some participants simply did not know what machine learning is. Testing the toolkit on such users was still important as it provided us a better understanding about its usage.

Individual evaluation

Brainstorming sessions are often conducted in groups. However, in this study we chose to test out the toolkit on individual subjects. This was an intentional decision to help evaluate the usage of the toolkit across different participants with different prior ML experience. Some participants, however, did mention that having someone to *"bounce idea off from"* would have been helpful to help them further create new ideas and this can be explored in future work.

2.6 Conclusion

This chapter introduces a system, IdeaLens that leverages analogical reasoning as a way to help designers come up with valid but divergent machine learning ideas. By only showing designers the sample real-world use-cases of past ML work, we hide technical details of the algorithms while still letting designers come utilize these as sources of inspiration to guide ideation in their own domains. This work thus provides a pathway to use past use-cases as effective boundary objects to help inspire designers. More broadly, we hope that by equipping designers with a tool that helps them come up with feasible and divergent ideas, we help them be be included in ML application pipeline earlier, and thus let companies take a more user-centred approach to ideation of such features/products.

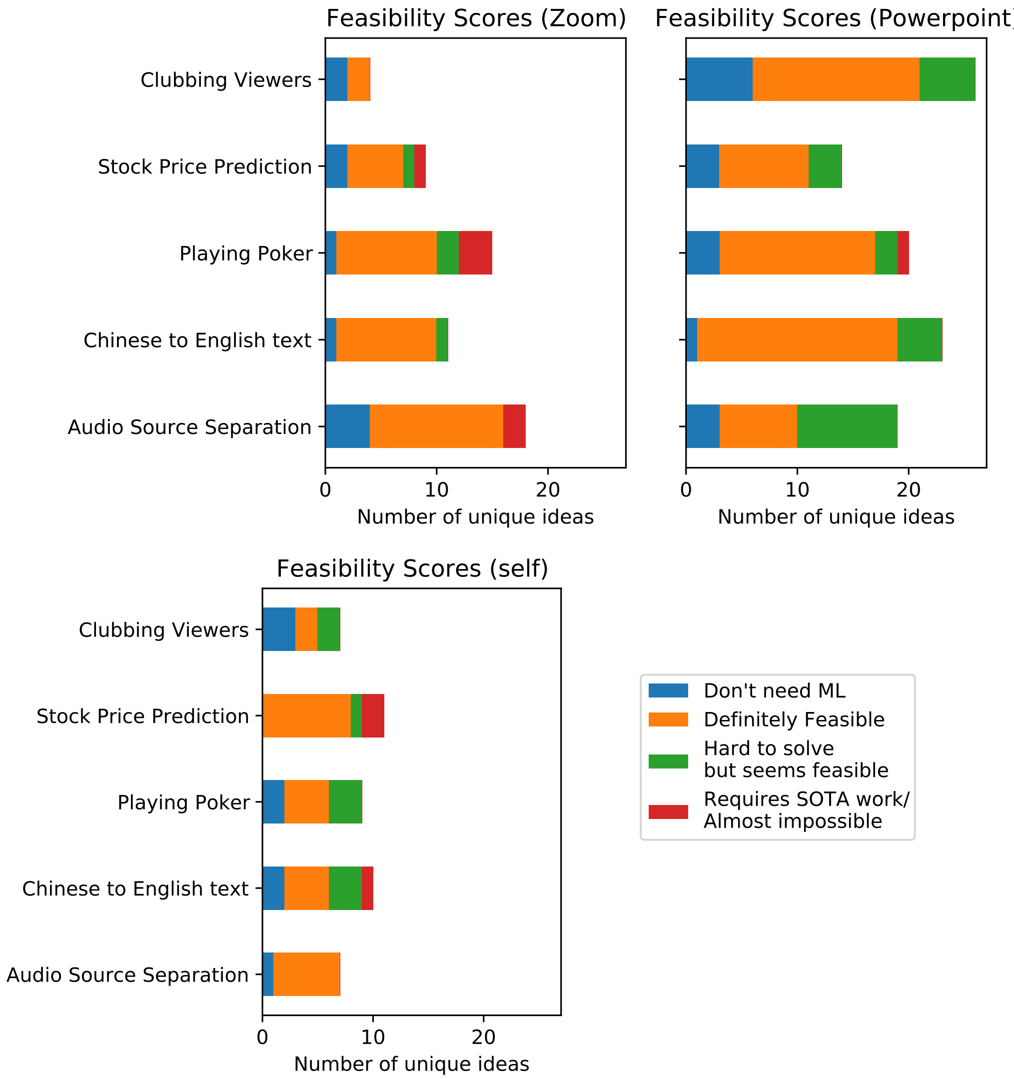

Figure 2.7: Mean Feasibility Ratings of the Generated Ideas across the different use-cases: all cards were able to create multiple ideas within the purview of machine learning. In fact, a large majority of the ideas were rated as feasible(yellow/green) despite participants being encouraged to think divergently and not care about feasibility of the ideas.

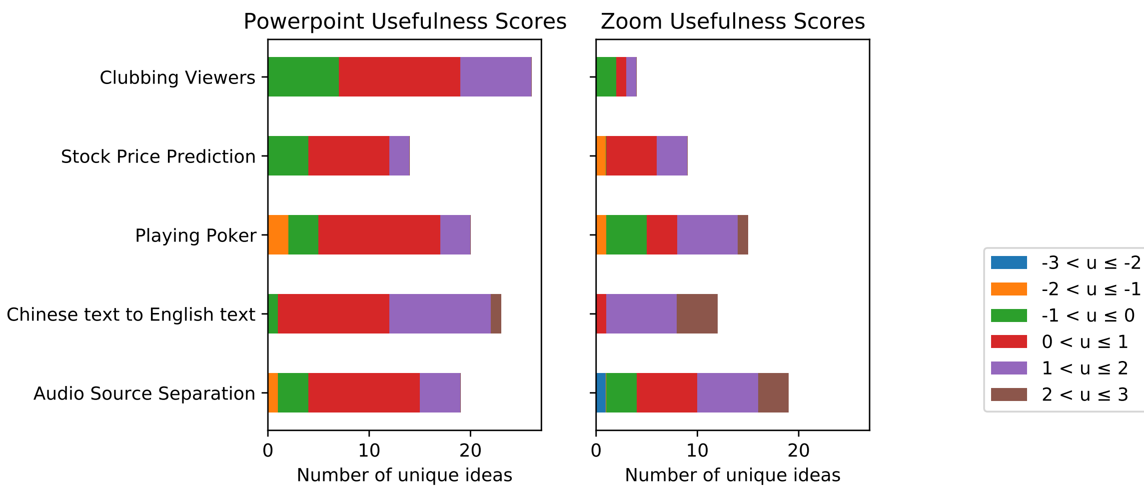

Figure 2.8: Mean usefulness scores provided by real world-users across the different cards from -3(not useful) to 3(very useful): All cards were seen to be helpful in generating multiple useful ideas

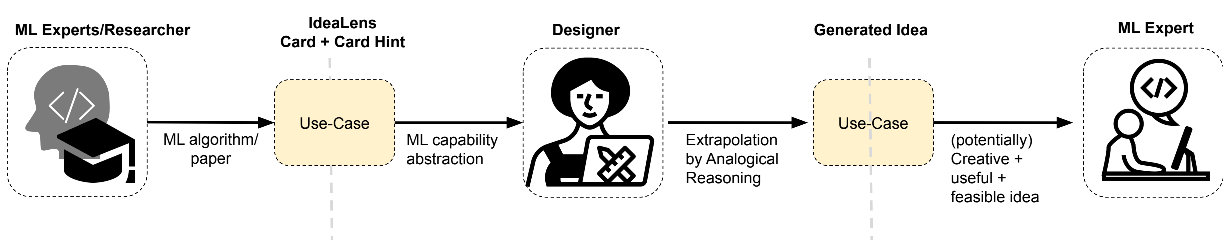

Figure 2.9: Use-case as a boundary object approach in IdeaLens: This figure shows the various steps in the pipeline where use-cases are being used as boundary objects and units of knowledge transfer in our framework

Chapter 3

InToML: An Alternate Repository of ML Research Papers

Not only are groups in both academia and industry working on machine learning-related problems, the line between academia and industry work in Machine Learning is blurry at best. Many of the models, datasets, toolkits and even theories suggested in research papers can be useful for real-world applications either directly or indirectly, and generating research useful for real world applications has been considered as one of the important goals of machine learning research([66]). Consequently, keeping up with research is useful for employees in machine learning teams. However, as in other areas of research, this transfer of knowledge is filled with friction([83]). Combining this with the rapid rate of publications, there is a huge obstacle faced by industry professionals while trying to access this source of knowledge.

This divide is especially challenging considering the probable differences in the expertise between academic and industry professionals working on the field: Machine Learning teams in the industry are generally multi-disciplinary with not everyone having technical expertise of the field. For example, teams often include UX researchers and designers who work in close collaboration with data scientists and engineers to come up with useful use-cases, design the product, deal with fairness issues and generally bring a user-voice to the product. Similarly, some teams also include domain experts who provide useful information about the domain. Both these groups have faced difficulties working in Machine Learning domain due to a lack of understanding of the capabilities or ways they can navigate the shortcomings-which in itself are evolving as the field matures. Thus while such users are expected to lack the expertise to actually read the research paper, they can still benefit from some of knowledge contained in such research papers. For example, in the case of designers this could include real world use-cases mentioned in papers which can then inspire them to come up with use-cases in their own domain, as also seen in the previous chapter. While some domain experts(for eg. law and policy makers) might be concerned more about the ethical concerns mentioned in some papers. Hence, not only is the navigation of Machine Learning papers sub-optimal for ML experts, the knowledge generated in such research papers remains elusive to the non-experts.

Friction in navigating research papers is in fact a well-documented problem in past research both for academic audiences and industry professionals. However, while some tools have been developed to help academics navigate the growing number of research papers better(eg.

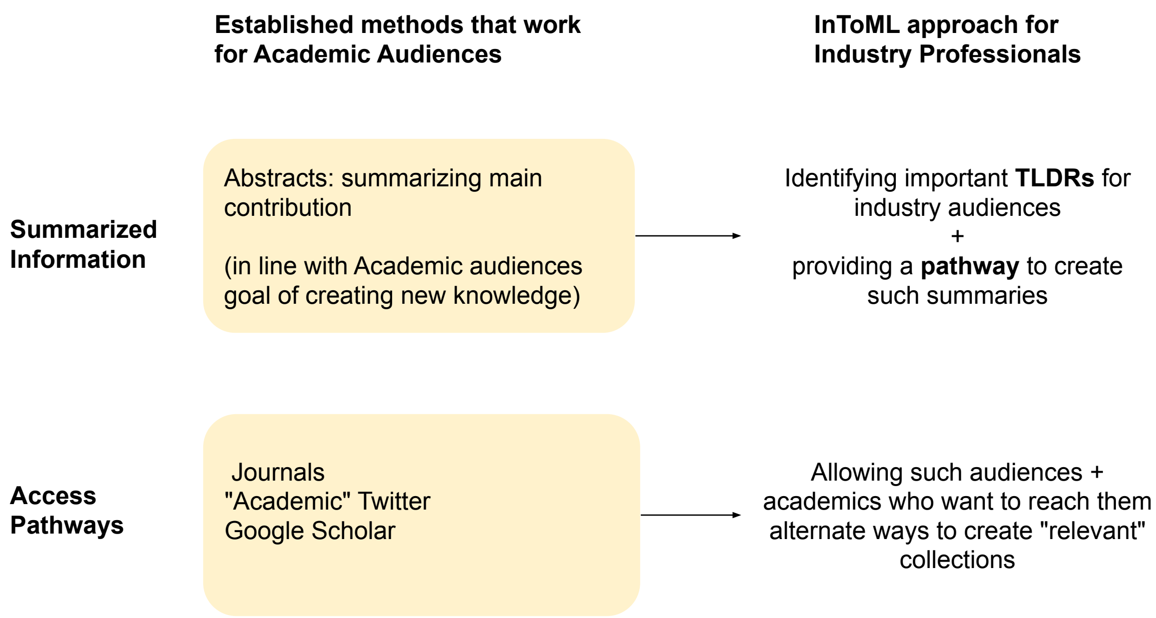

Figure 3.1: InToML suggested approach for spearheading an alternate eco-system of research papers for industry audiences

[84, 85, 86]), they fail to accommodate the differing needs of this growing faction of industry professionals who could also benefit from this source of knowledge: despite the research itself being a useful source of knowledge for both industry professionals and fellow academics, readers across these two groups often have very different goals. Academic readers focus on creating new knowledge by solving new problems or finding new solutions to existing problems. In contrast, the main goal for industry professionals is usually to create business value.

In this chapter, we introduce and describe the design of a platform called that leverages Pinterest-like ”Boards” to facilitate such knowledge transfer. In particular, InToML allows readers of research papers(who might be industry professionals or academics) to curate collections of papers accessible to other users. Each board gradually builds industry-specific TLDRs of research paper that it hosts. In doing so, allows an alternate pathway to access knowledge generated in research papers by both providing curated collections and presenting papers in more accessible ways. Facilitating this knowledge transfer can be mutually beneficial with the increasing number of companies interested in adopting machine learning, and ML and HCI researchers who want to aid in the development of real-world use-cases.

In summary, this study aimed to achieve the following goals: (1) Understanding what aspects of research papers can be beneficial to different employees working on such products, (2) Designing a platform that allows specific TLDRs for ML industry professionals to help them access this information more easily, and (3) Spearheading an alternate ecosystem of research papers directed towards industry professionals. In this chapter, we describe the design of

3.1 Related Work

3.1.1 Industry-Academia knowledge exchange in Machine Learning

Different past researchers have identified fundamental problems in the knowledge exchange between ML in the industry and ML in academia. This has been look at in different perspectives- (1) direct interactions: where people from the industry and academia come together for a common goal, (2) indirect interactions: where either the industry professionals are using the resources/knowledge built by academia or vice versa for example when academics are inspired by problem areas faced by industry professionals. The problems observed in such communications are however very consistently present across both these cases and below we summarize some past work that highlights this.

In one such work, Rahman et al [83] used a case-study based approach of using Machine Learning in the industry and among other problems highlighted industry-academia knowledge transfer as one of the obstacles faced by researchers trying to . In addition, some past research has highlighted the divide between the actual needs of the industry professionals and the direction some researchers trying to help them are taking. For example, in their paper ”Improving Fairness in Machine Learning Systems:What Do Industry Practitioners Need?”[87], authors have highlighted that Fairness Toolkits fail to accomodate the real needs of industry professionals. Similarly, after interviewing designers working in machine learning, researchers discovered that they do not want to learn technical knowledge of machine learning but rather have been seen to learn from part exemplars-despite past research focussing on helping them gain technical expertise. This is also a problem witnessed in identifying areas. Past work like this seems to highlight that even though researchers are interested in research that is useful in the industry, there is a gap in the feedback loop once a research is published. This gap doesn't always necessarily have to be as large as the cases mentioned above where an entire direction of research is

In this project, we view the problems encountered by industry professionals while trying to read research papers in a similar light, i.e as a result of a gap in objectives, knowledge, and means to communicate while this does not require direct interactions as mentioned in some other work. We view research papers as sources of knowledge that could be beneficia

3.1.2 Augmenting research papers for increased accessibility

This research also builds upon a growing area of work looking into making research papers accessible to audiences. In one such work, researchers developed a TL;DR model to help generate extreme summaries of the contributions made in research papers to help easier parsing of such papers[86]. Similarly, in other line of work,

In this project, while taking inspiration from such work augmenting information we take an alternate approach and consider

Even within Machine Learning in specific, some tools have been developed to help reach their audiences better-for example, PapersWithCode[]b is a popular website that allows users to access code related to respected papers. In another such work, researchers suggested Model papers to be reported with Model Cards [] to help easy access of model's ethical considerations.

3.1.3 Sensemaking and boundary objects

This research also builds on a growing area of HCI research involving using the power of the crowds to make sense of the large amount of knowledge that exists. One such Crowdsourcing has been used to accelerate knowledge transfer in different areas in the past, including

Non-experts working in the field

Unlike other design materials, non-experts working with Machine Learning need to be more involved with the actual technology to be able to contribute well. For example, UX researchers have been seen to

3.1.4 Co-designing using flexible Platforms that help identify importance

3.2 Formative Study

To validate the need of such a toolkit and identify what employees could find useful through a survey responded to by 14 such workers.

Through a combination of asking them to give a usefulness score to some initial set of things that we thought could be useful and a an open response field, we identified the following measures useful to be displayed before they actually read the papers (outside the actual technical contribution): **(1) real-world use-cases, (2) resources used for training, (3) software/tools used during development, (4) model parameters and (5) code/dataset** .

In fact, all **14** participants ranked atleast one of the other factors the atleast as important as a summarization of the actual contribution made in the paper, while **8** ranked atleast one to be strictly more important-even though this is the focus of abstracts/past work making this information easier to access. Through these responses, we realize that industry professionals want a combination of implicit and explicit knowledge generated in such papers be communicated through such fields of metadata.

Further, at the end of the survey, we asked participants if they thought it would be useful for them to find this information across ”similar” papers being presented together in pinterest like board useful, and found an overwhelming majority (**12**) of the responders thought that this would be very useful (≥ 2 on a -3 to 3 scale).

3.3 InToML: An Alternate Repository of ML Research Papers

Based on this initial needfinding survey and after identifying the information useful for industry professionals, we developed an inital prototype of an alternate repository of research papers to make it easier for industry professionals to access this source of knowledge.

Below, we describe the design of the system across the its two goals: (1) providing an entity that allows for easier access of this information for industry professionals, and (2) providing a platform that allows users to crowdsource this information.

3.3.1 Paper-specific TLDRs

Each paper in InToML is augmented with summarized information of things important for industry professionals. In a way, these form the industry-counterparts for abstracts. In the same way as abstracts lay out summarized information of the main contribution(which is expected to be the most important takeaway from the paper for academic audiences), TLDRs summarize industry-specific important information for such audiences to parse, understand and select papers to read.

Purpose

The purpose of such TLDRs is not only to augment abstracts with information specific to industry audiences but also to allow them to take away relevant information without reading the research papers. For instance, as we mentioned real world use-cases in such research paper could inspire designers for their own ideas, and this could be extracted even without

Design Decisions

3.3.2 Paper Collections and Views

Purpose

Design Decisions

3.4 Discussion

3.4.1 An easier way to parse/select relevant research papers for industry professionals

Given the interest of our initial survey responders in getting relevant information displayed together across ”similar” papers and the expected differing needs of the papers, IdeaLens allows users to crowdsource different papers into collections. There are similar to Pinterest boards([88]) but directed towards research papers. From a viewer’s perspective, this allows users to first select boards they might be interested in, and get relevant information across these different papers together to select, compare and maybe even directly use them in their own work.

3.4.2 A pathway to create relevant metadata for industry professionals

Each research board has a specific topic and hosts papers related to that topic. Other users can then either suggest or add new research papers to these board depending on the board settings. By default, each board has the factors identified in the formative study as the default input fields for new papers to fill out, but board creators are also allowed to add in additional fields that any added papers can fill out.

Different users can add in/suggest relevant paper with its relevant metadata: each new paper added has these fields pre-filled out by what the last user(if any) filled out for the same fields.

Eventually, through these iterative changes, this information would form the metadata that we expect industry professionals to find useful.

Such boards thus aim to serve the following purposes in : (1) provide an entity to directly view papers, (2) allow ownership over crowd-sourced information to help encourage better quality control the results, (3) providing a flexible entity to allow for the different needs of different industries/industry professionals be satified.

3.4.3 Stakeholders

1. Industry Professionals:
   - Machine Learning Experts: These include data-scientists, machine learning engineers and other employees who have prior machine learning experience. These are usually people who have some past Machine Learning experience. These users benefit from both actually reading research papers and from some metadata that could help recommend papers. While reading research papers, they usually care about different aspects of the research papers from academics- making current methods of information retrieval suboptimal at best.
   - Non-Machine Learning Experts: Machine Learning teams in the industry, like other software engineering teams, are often multidisciplinary-they include members like domain experts, UX researchers and product managers. Non-experts in such teams report problems while doing their tasks as they are often unable to understand what machine learning can and cannot do. Given that the field itself is rapidly evolving, this is has no simple solution and making such research papers or the knowledge added by such papers more accessible could help provide a solution to this problem.
2. Academic Research: In addition to directly supporting industry professionals, this toolkit also aims to uncover needs of industry professionals, and provide helpful feedback to academics who are interested in facilitating industry-academia transfer. Further, creating such TLDRs of research papers can increase visibility of the said research in industry. In the article, Machine Learning that Matters([66]), Wagstaff claims that the three main steps of machine learning research include preparing data, creating a machine learning contribution and finally creating impact by persuading users to adopt technique. However, current incentives only focus on the second objective. A toolkit directed towards industry professionals, who are often the end users of the results, could eventually help place greater emphasis on the last step.

3.5 Ongoing and Future Work

Chapter 4

Conclusion and Future Work

Bibliography

[1] Danijel Kučak, Vedran Juričić, and Goran ambić. Machine learning in education-a survey of current research trends. *Annals of DAAAM & Proceedings*, 29, 2018. 1

[2] TK Balaji, Chandra Sekhara Rao Annavarapu, and Annushree Bablani. Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, 40:100395, 2021. 1

[3] Rohan Bhardwaj, Ankita R Nambiar, and Debojyoti Dutta. A study of machine learning in healthcare. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, volume 2, pages 236–241. IEEE, 2017. 1

[4] Matthew F Dixon, Igor Halperin, and Paul Bilokon. *Machine learning in Finance*, volume 1170. Springer, 2020. 1

[5] Rania Rizki Arinta and Emanuel Andi WR. Natural disaster application on big data and machine learning: A review. In *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pages 249–254. IEEE, 2019. 1

[6] Thomas H Davenport and Rajeev Ronanki. Artificial intelligence for the real world. *Harvard business review*, 96(1):108–116, 2018. 1

[7] In Lee and Yong Jae Shin. Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2):157–170, 2020. 1

[8] Peter Hofmann. *Designing and Managing Artificial Intelligence-Enabled Information Systems*. PhD thesis, 2021. 1

[9] Qian Yang, Justin Cranshaw, Saleema Amershi, Shamsi T Iqbal, and Jaime Teevan. 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*, pages 1–12, 2019. 1, 2

[10] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 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*, CHI '17, page 278–288, New York, NY, USA, 2017. Association for Computing Machinery. 1, 2

[11] Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. Investigating how experienced ux designers effectively work with machine learning. 06 2018. 1, 2

[12] Zhibin Zhou, Lingyun Sun, Yuyang Zhang, Xuanhui Liu, and Qing Gong. Ml lifecycle

canvas: Designing machine learning-empowered ux with material lifecycle thinking. *HumanâComputer Interaction*, 35(5-6):362–386, 2020. 1, 2

[13] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. *Re-Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design*, page 1–13. Association for Computing Machinery, New York, NY, USA, 2020. 1, 2

[14] Soya Park, April Yi Wang, Ban Kawas, Q. Vera Liao, David Piorkowski, and Marina Danilevsky. Facilitating knowledge sharing from domain experts to data scientists for building NLP models. *CoRR*, abs/2102.00036, 2021. 1

[15] Maximiliane Windl, Sebastian S. Feger, Lara Zijlstra, Albrecht Schmidt, and Pawel W. Wozniak. 'it is not always discovery time': Four pragmatic approaches in designing ai systems. In *CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. 1, 2, 2.3.1

[16] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pages 291–300. IEEE, 2019. 1

[17] Qian Yang. Machine learning as a ux design material: how can we imagine beyond automation, recommenders, and reminders? In *AAAI Spring Symposia*, 2018. 1, 2, 2.5.2

[18] Xuli Tang, Xin Li, Ying Ding, Min Song, and Yi Bu. The pace of artificial intelligence innovations: Speed, talent, and trial-and-error. *Journal of Informetrics*, 14(4):101094, 2020. 1

[19] Scott H Hawley. Challenges for an ontology of artificial intelligence. *arXiv preprint arXiv:1903.03171*, 2019. 1

[20] Pei Wang. What do you mean by" ai"? In *AGI*, volume 171, pages 362–373, 2008. 1

[21] David Piorkowski, Soya Park, April Yi Wang, Dakuo Wang, Michael Muller, and Felix Portnoy. How ai developers overcome communication challenges in a multidisciplinary team: A case study. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1), apr 2021. 2

[22] Hariharan Subramonyam, Jane Im, Colleen Seifert, and Eytan Adar. Solving separation-of-concerns problems in collaborative design of human-ai systems through leaky abstractions. In *CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. 2, 2.3.1

[23] Shan Carter and Michael Nielsen. Using artificial intelligence to augment human intelligence. *Distill*, 2(12):e9, 2017. 2

[24] Philip van Allen. Prototyping ways of prototyping ai. *Interactions*, 25(6):46–51, oct 2018. 2

[25] Rebecca Fiebrink and Perry R Cook. The wekinator: a system for real-time, interactive machine learning in music. In *Proceedings of The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010)(Utrecht)*, volume 3, 2010. 2

[26] Eric Fulco Rietzschel. *From quantity to quality: cognitive, motivational and social aspects*

of creative idea generation and selection. PhD thesis, Utrecht University, 2005. 2

[27] Hariharan Subramonyam. *Designing AI Experiences: Boundary Representations, Collaborative Processes, and Data Tools*. PhD thesis, 2021. 2, 2.5.2

[28] Susan Leigh Star and James R Griesemer. Institutional ecology,translations' and boundary objects: Amateurs and professionals in berkeley's museum of vertebrate zoology, 1907-39. *Social studies of science*, 19(3):387–420, 1989. 2

[29] Joel Chan, Steven Dang, and Steven P. Dow. Ideagens: Enabling expert facilitation of crowd brainstorming. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, CSCW '16 Companion, page 13–16, New York, NY, USA, 2016. Association for Computing Machinery. 2.1

[30] Michael F. Clarke, Joseph Gonzales, Richard Harper, David Randall, Thomas Ludwig, and Nozomi Ikeya. Better supporting workers in ml workplaces. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, CSCW '19, page 443–448, New York, NY, USA, 2019. Association for Computing Machinery. 2.1

[31] Karen L. Boyd. Datasheets for datasets help ml engineers notice and understand ethical issues in training data. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct 2021. 2.1

[32] Melanie Feinberg. Material vision. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, page 604–617, New York, NY, USA, 2017. Association for Computing Machinery. 2.1

[33] Cecilia Aragon, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing, and Joseph Bayer. Developing a research agenda for human-centered data science. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, CSCW '16 Companion, page 529–535, New York, NY, USA, 2016. Association for Computing Machinery. 2.1

[34] Design Council. The 'double diamond'design process model. *Design Council*, 2005. 2.1.1

[35] Barry Matthew Kudrowitz and David Wallace. Assessing the quality of ideas from prolific, early-stage product ideation. *Journal of Engineering Design*, 24(2):120–139, 2013. 2.1.1

[36] Bruce A Reinig, Robert O Briggs, and Jay F Nunamaker. On the measurement of ideation quality. *Journal of Management Information Systems*, 23(4):143–161, 2007. 2.1.1

[37] E Paul Torrance. *Guiding creative talent*. Pickle Partners Publishing, 2018. 2.1.1

[38] John H Flavell, Allan Cooper, and Robert H Loiselle. Effect of the number of pre-utilization functions on functional fixedness in problem solving. *Psychological Reports*, 4(3):343–350, 1958. 2.1.1

[39] Abraham S Luchins. Mechanization in problem solving: The effect of einstellung. *Psychological monographs*, 54(6):i, 1942. 2.1.1

[40] Suyun Sandra Bae, Oh-Hyun Kwon, Senthil Chandrasegaran, and Kwan-Liu Ma. *Spinneret: Aiding Creative Ideation through Non-Obvious Concept Associations*, page 1–13. Association for Computing Machinery, New York, NY, USA, 2020. 2.1.1

[41] Noël Carroll. *Visual Metaphor*, page 347–368. Cambridge University Press, 2001. 2.1.1

[42] Ranjitha Kumar, Jerry O. Talton, Salman Ahmad, and Scott R. Klemmer. Bricolage: Example-based retargeting for web design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 2197–2206, New York, NY, USA, 2011. Association for Computing Machinery. 2.1.1

[43] Robin Roy and James P. Warren. Card-based design tools: a review and analysis of 155 card decks for designers and designing. *Design Studies*, 63:125–154, 2019. 2.1.1

[44] Method cards. 2.1.1, 2.2.3

[45] Simone Mora, Francesco Gianni, and Monica Divitini. Tiles: A card-based ideation toolkit for the internet of things. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, page 587–598, New York, NY, USA, 2017. Association for Computing Machinery. 2.1.1

[46] Wendelin Küpers. *Analogical Reasoning*, pages 222–225. Springer US, Boston, MA, 2012. 2.1.2

[47] Katherine Fu, Diana Moreno, Maria Yang, and Kristin L Wood. Bio-inspired design: an overview investigating open questions from the broader field of design-by-analogy. *Journal of Mechanical Design*, 136(11):111102, 2014. 2.1.2

[48] Daniel M Eveleth. Analogical reasoning: When a non-expert reasons like an expert. *Journal of Behavioral and Applied Management*, 1(1):795, 2016. 2.1.2

[49] Darren W Dahl and Page Moreau. The influence and value of analogical thinking during new product ideation. *Journal of marketing research*, 39(1):47–60, 2002. 2.1.2

[50] Joel Chan, Steven P Dow, and Christian D Schunn. Do the best design ideas (really) come from conceptually distant sources of inspiration? In *Engineering a Better Future*, pages 111–139. Springer, Cham, 2018. 2.1.2, 2.4.2

[51] Dedre Gentner. Analogical reasoning, psychology of. *Encyclopedia of cognitive science*, 1:106–112, 2003. 2.1.2

[52] Hyeonsu B Kang, Xin Qian, Tom Hope, Dafna Shahaf, Joel Chan, and Aniket Kittur. Augmenting scientific creativity with an analogical search engine. *ACM Transactions on Computer-Human Interaction*, 2022. 2.1.2

[53] Lixiu Yu, Aniket Kittur, and Robert E Kraut. Distributed analogical idea generation: inventing with crowds. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 1245–1254, 2014. 2.1.2

[54] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010. 2.1.3

[55] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2.1.3

[56] Mohammad Amin Morid, Alireza Borjali, and Guilherme Del Fiol. A scoping review of

transfer learning research on medical image analysis using imagenet. *Computers in Biology and Medicine*, 128:104115, 2021. 2.1.3

[57] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems*, 27, 2014. 2.1.3

[58] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2.1.3

[59] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 2.1.3

[60] Lu He, Qianyu Zhou, Xiangtai Li, Li Niu, Guangliang Cheng, Xiao Li, Wenxuan Liu, Yunhai Tong, Lizhuang Ma, and Liqing Zhang. End-to-end video object detection with spatial-temporal transformers. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1507–1516, 2021. 2.1.3

[61] Prateek Verma and Jonathan Berger. Audio transformers: Transformer architectures for large scale audio understanding. adieu convolutions. *arXiv preprint arXiv:2105.00335*, 2021. 2.1.3

[62] David Elliott, Carlos E Otero, Steven Wyatt, and Evan Martino. Tiny transformers for environmental sound classification at the edge. *arXiv preprint arXiv:2103.12157*, 2021. 2.1.3

[63] Fredrik A Dahl. A reinforcement learning algorithm applied to simplified two-player texas hold'em poker. In *European Conference on Machine Learning*, pages 85–96. Springer, 2001. 2.1.3, 2.3.2

[64] B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A Al Sallab, Senthil Yogamani, and Patrick Pérez. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 2021. 2.1.3, 2.3.2

[65] Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. Real-time bidding by reinforcement learning in display advertising. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, page 661–670, New York, NY, USA, 2017. Association for Computing Machinery. 2.1.3

[66] Kiri Wagstaff. Machine learning that matters. *arXiv preprint arXiv:1206.4656*, 2012. 2.2.3, 3, 2

[67] AndrÃ©s Lucero and Juha Arrasvuori. Plex cards: A source of inspiration when designing for playfulness. pages 28–37, 01 2010. 2.2.3

[68] Lena Tørresdal, Francesco Gianni, and Monica Divitini. The digital transformation of card-based design toolkits: From tiles to digitiles. In *CHI Greece 2021: 1st International Conference of the ACM Greek SIGCHI Chapter*, pages 1–5, 2021. 2.3.1

[69] Anton Fedosov, Masako Kitazaki, William Odom, and Marc Langheinrich. Sharing economy design cards. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–14, New York, NY, USA, 2019. Association for Comput-

ing Machinery. 2.3.1

[70] Hong Shen, Leijie Wang, Wesley H Deng, Ciell Brusse, Ronald Velgersdijk, and Haiyi Zhu. The model card authoring toolkit: Toward community-centered, deliberation-driven ai design. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 440–451, 2022. 2.3.1

[71] Hariharan Subramonyam, Colleen Seifert, and Eytan Adar. Protoai: Model-informed prototyping for ai-powered interfaces. In *26th International Conference on Intelligent User Interfaces*, pages 48–58, 2021. 2.3.1

[72] Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. Investigating how experienced ux designers effectively work with machine learning. In *Proceedings of the 2018 designing interactive systems conference*, pages 585–596, 2018. 2.3.1

[73] Wesley Lau. *Application of Machine Learning Model in Generating Song Lyrics*. University of California, Los Angeles, 2021. 2.3.2

[74] Jiacheng Zhang, Huanbo Luan, Maosong Sun, Feifei Zhai, Jingfang Xu, Min Zhang, and Yang Liu. Improving the transformer translation model with document-level context. *arXiv preprint arXiv:1810.03581*, 2018. 2.3.2

[75] Andreas Madsen. Visualizing memorization in rnns. *Distill*, 4(3):e16, 2019. 2.3.2, 2.4.4

[76] Adil Moghar and Mhamed Hamiche. Stock market prediction using lstm recurrent neural network. *Procedia Computer Science*, 170:1168–1173, 2020. 2.3.2, 2.4.4

[77] Daniel Stoller, Sebastian Ewert, and Simon Dixon. Wave-u-net: A multi-scale neural network for end-to-end audio source separation. *arXiv preprint arXiv:1806.03185*, 2018. 2.3.2

[78] Bernhard Vennemann and Thomas Rösgen. A dynamic masking technique for particle image velocimetry using convolutional autoencoders. *Experiments in Fluids*, 61(7):1–11, 2020. 2.3.2

[79] Eugene Seo and Ho-Jin Choi. Movie recommendation with k-means clustering and self-organizing map methods. In *ICAART (1)*, pages 385–390, 2010. 2.3.2

[80] Beth Logan. Music recommendation from song sets. In *ISMIR*, pages 425–428, 2004. 2.3.2

[81] Microsoft CSE. Ml feasibility studies, May 2022. 2.3.3

[82] Bo T Christensen and Christian D Schunn. The relationship of analogical distance to analogical function and preinventive structure: The case of engineering design. *Memory & cognition*, 35(1):29–38, 2007. 2.4.2

[83] Md Saidur Rahman, Emilio Rivera, Foutse Khomh, Yann-Gaël Guéhéneuc, and Bernd Lehnert. Machine learning software engineering in practice: An industrial case study. *arXiv preprint arXiv:1906.07154*, 2019. 3, 3.1.1

[84] Napol Rachatasumrit, Jonathan Bragg, Amy X Zhang, and Daniel S Weld. Citeread: Integrating localized citation contexts into scientific paper reading. In *27th International Conference on Intelligent User Interfaces*, pages 707–719, 2022. 3

[85] Raymond Fok, Andrew Head, Jonathan Bragg, Kyle Lo, Marti A Hearst, and Daniel S Weld. Scim: Intelligent faceted highlights for interactive, multi-pass skimming of scientific

papers. *arXiv preprint arXiv:2205.04561*, 2022. 3

[86] Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel S Weld. Tldr: Extreme summarization of scientific documents. *arXiv preprint arXiv:2004.15011*, 2020. 3, 3.1.2

[87] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–16, New York, NY, USA, 2019. Association for Computing Machinery. 3.1.1

[88] 3.4.1