Latent Lab: Exploration Beyond Search and Synthesis

by

Kevin F. Dunnell

B.Sc., University of Michigan (2013)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Kevin F. Dunnell
Program in Media Arts and Sciences
May 19, 2023

Certified by: Andrew B. Lippman
Senior Research Scientist
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by: Tod Machover
Academic Head
Program in Media Arts and Sciences

Latent Lab: Exploration Beyond Search and Synthesis

by

Kevin F. Dunnell

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on May 19, 2023, in partial fulfillment of the
requirements for the degree of
Master of Science in Media Arts and Sciences

Abstract

This Master's thesis investigates the potential of artificial intelligence (AI) models, particularly machine learning and natural language processing techniques, to facilitate brainstorming and ideation in the invention process. The thesis centers around the iterative development of "Latent Lab," an interactive tool for exploring relationships among MIT Media Lab research projects. The work offers insights into AI systems as co-inventors by addressing the challenges of organizing, searching, and synthesizing content. Our method for interacting with the material is based on "exploration" rather than search. The primary objective was to create a human-AI co-invention system and evaluate its performance on the novelty of co-created ideas. However, the research underscored the importance of accurate data organization for meaningful data generation. Consequently, later versions of Latent Lab focused primarily on improving data organization and interactive exploration. The tool's success was measured by its effectiveness in familiarizing users with research projects at the Media Lab, ultimately laying the foundation for the future development of human-AI co-invention systems.

Thesis Supervisor: Andrew B. Lippman
Title: Senior Research Scientist, Program in Media Arts and Sciences

Latent Lab: Exploration Beyond Search and Synthesis

by

Kevin F. Dunnell

This thesis has been reviewed and approved by the following committee members

Andrew B. Lippman
Senior Research Scientist
Program in Media Arts and Sciences

Pattie Maes
Professor of Media Technology
Program in Media Arts and Sciences

Vishal Sikka
Founder and CEO
Vianai Systems Inc.

Acknowledgements

I would like to thank my advisor, Andy Lippman, for his guidance, support, and encouragement throughout my research journey.

I would also like to extend my sincere gratitude to my committee members, Pattie Maes and Vishal Sikka, whose insights and mentorship have been instrumental in shaping my research and ambitions.

I am deeply thankful to Trudy Painter and Andrew Stoddard, who made Latent Lab possible and the long hours of work enjoyable.

Thank you to the Viral lab: Alessandra Davy-Falconi, Mike Jiang, Dagmawi Haile, and Erick Oduniyi, as well as my cohort and the Media Lab community, who have been a constant source of inspiration, motivation, and assistance.

I appreciate Ziv Epstein and Hope Schroeder's vital guidance on user testing, Shayne Longpre's pioneering use of Latent Lab, and Rob Lincourt's advice and advocacy for our project within Dell.

Finally, my heartfelt thanks go to Kelly Marinick, my family—Dave, Sue, Katie, and Jimmy Dunnell—as well as my friends for their unwavering love and support throughout this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The untapped potential of collective knowledge—often obscured within silos and by tribal knowledge—has profound implications for the evolution of ideas and innovation across organizations, companies, and research groups [7]. This collective knowledge often predates most contributors and may outlive their tenures as well as those of archivists tasked with its preservation. The advent of the internet and digital technology has transformed the landscape of knowledge exploration, shifting the challenge from preserving physically scarce artifacts to navigating an abundance of digital information. However, despite the digital revolution, the underlying structure of information organization has remained strikingly similar. For instance, the list of results returned from a search engine query closely resembles knowing where to find specific documents in a filing cabinet. While items adjacent to the targeted location may be relevant—similar to the top results on a search engine page—users are unlikely to explore beyond this immediate vicinity by perusing thousands of search result pages or opening other filing cabinet drawers.

Current list view search approaches fail to fully leverage the interaction methods afforded by modern computation. This limitation restricts exploration across diverse sources and impedes the discovery of interconnected relationships. It is optimized for search, information retrieval tasks with one correct answer, rather than browsing, where results emerge through interaction. Exploration and transformation of a conceptual space are more likely to lead to "historical creativity" that produces surprising and valuable ideas that are new to humankind [2]. Additionally, the common underpinnings of modern search tools amplify this exploration impediment of complex concepts by prioritizing keywords over semantic meaning, leading to a deterioration of context. Alternatively, synthesis tools such as chatGPT offer a solution to context representation and present a paradigm shift in user interface design. The conversational alternative to traditional keyword-based search allows for a two-way dialogue between human and machine rather than a one-way query. ChatGPT's widespread acceptance validates what we started with – the interaction is valuable. While ChatGPT's interaction method improves list-based search, it obfuscates the underlying (or lack thereof) sources of information. A further drawback of ChatGPT is its limited text-based interaction, which can be overwhelming and may benefit from a more structured design. Our work expands upon the exploratory and interaction dimensions.

This thesis details the journey of building a better exploration tool that extends the capabilities of previous search and synthesis tools to encompass browsing and active visual interaction. The journey reinforces the value in the exploratory nature of the creative process rather than the solution itself. The tool, Latent Lab, has been iteratively designed to facilitate the exploration of an extensive data set with the assistance of AI, offering a glimpse of how it can be extended to produce new ideas. By leveraging leading data manipulation libraries, interactive front-end visuals, and large language models (LLMs), the latest iteration of Latent Lab transcends the limitations of keyword-centric search, enabling users to engage in semantically meaningful exploration and synthesis large data sets.

1.1 Thesis Overview

In the pursuit of developing a human-AI co-invention system, one aspect stood out as critical - the "exploration process," which refers to the interactive and iterative procedure of generating, testing, and refining ideas. In the context of invention, it involves users' engagement with the system to navigate through diverse conceptual spaces, discover novel combinations of existing ideas, and generate new concepts. Each version of Latent Lab provided a new understanding of what is essential for an optimized exploration tool. User feedback was instrumental in guiding the selection and prioritization of new features. Some early suggestions were technically unfeasible at the time. However, these were ultimately addressed in subsequent iterations thanks to the rapid pace of advancements in the field of AI, particularly LLMs, over the last year.

The latest iteration, Latent Lab V3.0, was analyzed quantitatively by collecting and examining empirical data obtained from user tests. These user tests compared the performance of Latent Lab with an existing search tool, the Media Lab's website, in facilitating the exploration of research projects at the Media Lab. The findings suggest that Latent Lab improves upon mental support and insight extraction compared to existing tools for understanding large knowledge bases.

1.2 Thesis Vision

Latent Lab aims to serve as a versatile data exploration tool designed to uncover and expose the intricate, interconnected relationships within diverse data sets. By facilitating the seamless aggregation of disparate data sets, irrespective of their structure and format, it seeks to promote a comprehensive understanding of complex data landscapes. Ultimately, the tool strives to identify high-impact ideas, ranging from research projects and patents to innovative teams, thus significantly advancing knowledge and technology.

Chapter 2

Background

2.1 Knowledge Organization

In his seminal 1945 essay "As We May Think," Vannevar Bush introduced the visionary concept of the "memex," an electromechanical device for organizing and retrieving vast amounts of information [4]. Bush's memex emphasized associative indexing, where information is organized based on associations and relationships rather than hierarchical structures, anticipating the rise of hypertext and hyperlinking systems. The memex also served as a prototype for personal knowledge management, enabling customizable, user-centered knowledge organization.

Similarly, in Feynman's Tips on Physics, a subset of the Feynman Lectures on Physics, Richard Feynman discusses the importance of understanding the relationships between concepts in physics rather than simply memorizing formulas. He uses the metaphor of "triangulation" to illustrate how a deep understanding of the interconnections between physics principles allows one to fill in gaps in knowledge, reconstruct forgotten concepts, and make discoveries [11]. Figure 2-1 illustrates this concept, from left to right: "Imaginary map of all the physics formulas," "Forgotten facts can be recreated by triangulating known facts," and "Discoveries are made by triangulating from the known to the previously unknown."

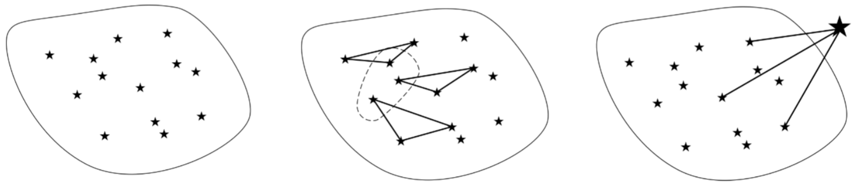


Figure 2-1: Feynman Triangulation

Decades later, Danny Hillis and his colleagues at Applied Minds founded Metaweb Technologies to develop a semantic data storage infrastructure for the Internet. This endeavor led to Freebase, an open and structured database of the world's knowledge. Freebase leveraged the idea of organizing information based on relationships and associations of user-contributed content meta tags, akin to Bush's concept of associative indexing and Feyman's Triangulation method. Freebase aimed to create a large-scale knowledge base that was accessible and shareable, and it eventually became the basis of the Google Knowledge Graph after Metaweb Technologies was acquired by Google [6]. However, the richness of associations was never exposed in the Google user interface.

The progression of knowledge organization encompasses concepts like the memex, associative indexing, and Feynman's triangulation method. The creation of Freebase and its subsequent transformation into Google Knowledge Graph highlights the importance of relationships between concepts. Information visualization's significance lies in making intricate associations comprehensible and easily accessible.

2.2 Knowledge Visualization

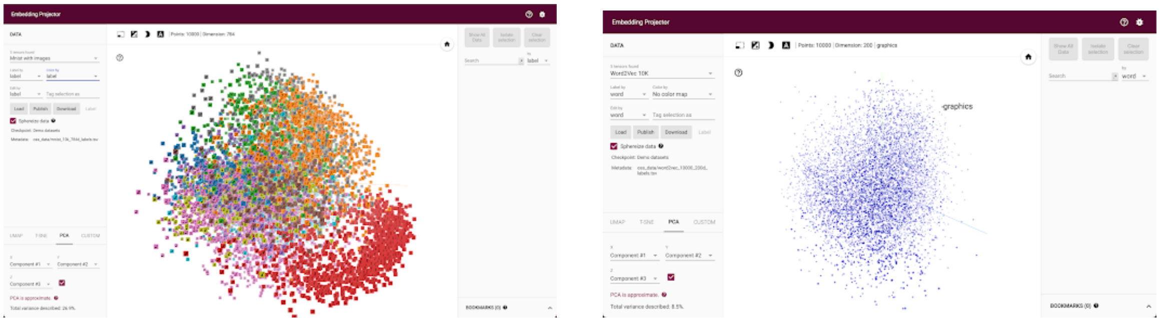
The foundation of information visualization can be traced back to Shneiderman's influential paper, "The eyes have it: A Task by Data Type Taxonomy for Information Visualizations" [24]. In this landmark work, Shneiderman proposes a taxonomy for designing visualizations based on the tasks intended to support and the types of data they represent. The visual information-seeking mantra, "overview first, zoom and filter, then details-on-demand," emerges from this paper, serving as a guiding principle for designing visual interfaces that facilitate interaction with vast knowledge bases.

A foray into interaction history is presented in "Footprints: History-rich Tools for Information Foraging," which investigates how the records of past interactions between people and information can inform and guide future interactions [32]. The authors introduce a theoretical framework of six properties characterizing interaction history systems and tools based on a navigation metaphor. The Footprints system is subsequently tested, revealing that users equipped with these tools can complete browsing tasks more efficiently, emphasizing the value of interaction history in digital environments.

In Hearst's book, "Search User Interfaces," the design of search interfaces for navigating and exploring large knowledge bases is thoroughly examined [13]. Covering topics such as query reformulation, faceted navigation, and clustering of search results, Hearst provides valuable insights into effective result presentation and interaction techniques for complex data. Shortly after, Bostock, Ogievetsky, and Heer presented D3.js, a JavaScript library for creating data-driven documents and interactive visualizations [3]. This library has significantly impacted the field of information visualization by enabling the creation of highly customizable visualizations that respond to user interactions.

Heer and Shneiderman's paper, "Interactive Dynamics for Visual Analysis," underscores the vital role of interaction in visual analysis [14]. The authors propose a framework for designing interactive visualizations and elucidate the significance of dynamic interactions in enabling users to explore and comprehend extensive data sets. Speiser's thesis, "WorldLens: Exploring World Events Through Media," builds upon Heer and Schneiderman's concepts of data exploration using Bostock et al.'s library, D3, and demonstrates the design and development of a system that provides a visual representation of global events as they unfold [26]. WorldLens adopts a data-driven approach, gathering data from various content sources and transforming it into an interactive visualization to facilitate informative discovery across multiple media types.

Following the second AI winter and the dawn of the Deep Learning revolution, the Embedding Projector is presented as a system for visualizing and understanding high-dimensional



(a) Embedding Projector - MNIST Images

(b) Embedding Projector - Word2Vec Words

Figure 2-2: Embedding Projector Visualizations

embedding spaces [25]. This interactive tool facilitates the exploration of properties of embeddings, such as local neighborhoods and global geometric structures, distinguishing it from other high-dimensional visualization systems.

Collectively, these works have established the foundation for developing effective visual interfaces for interacting with large knowledge bases. They have enhanced our understanding of user navigation and exploration of complex data while providing tools and frameworks that support the design of interactive visualizations that enable data analysis.

2.3 Information Retrieval

The evolution of search and information retrieval has seen significant advancements over the years, from keyword-based search methods to word embeddings and, more recently, cognitive or semantic search methods.

In the early days of information retrieval, keyword-based search methods were prevalent. One of the foundational works in this area was the paper by Karen Spärck Jones, titled "A Statistical Interpretation of Term Specificity and Its Application in Retrieval" (1972). In this paper, Spärck Jones introduced the concept of inverse document frequency (IDF), a vital component of the frequency-inverse document frequency (tf-idf) weighting scheme, which became widely adopted in keyword-based search systems. IDF measures the importance of a term in a collection of documents, allowing for the ranking of documents based on the specificity of their terms [27].

The field of information retrieval then saw a shift towards embedding-based search methods with the introduction of continuous vector representations of words or word embeddings. Tomas Mikolov and his colleagues proposed the Word2Vec model in their paper "Efficient Estimation of Word Representations in Vector Space" (2013). The Word2Vec model learns vector representations of words from large text corpora and has been influential in embedding-based search. Using word embeddings allows for capturing semantic relationships between words, making it possible to search for documents based on the semantic similarity of their contents [19].

More recently, advancements in language models and deep learning have led to the development of cognitive or semantic search methods, allowing systems to retain a more significant context beyond individual words. One of the most notable contributions in this area is the BERT (Bidirectional Encoder Representations from Transformers) model, introduced by Jacob Devlin and his colleagues in their paper "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" (2018). BERT is a transformer-based language model pre-trained on large text corpora and can be fine-tuned for various natural language processing tasks, including semantic search. BERT's bidirectional context-aware representations provide a powerful mechanism for understanding the context and semantics of queries, thereby improving the accuracy and relevance of search results [8].

Overall, the evolution of search and information retrieval has transitioned from simple keyword-based methods to sophisticated models that can understand the semantics and context of queries, resulting in improved search experiences.

2.4 Human-AI Co-Invention

Marvin Minsky's "Society of Mind" proposes that human intelligence results from the interaction and cooperation of many simple, individually unintelligent agents, each performing specialized functions [20]. These agents, or “mind modules,” form a society within the mind, enabling complex thought and behavior. This concept parallels human-AI collaboration, where the cooperative interaction between humans and AI systems can lead to enhanced intelligence and problem-solving capabilities. Just as individual mind modules contribute unique abilities to the overall functioning of the mind, AI systems can augment human intelligence by providing specialized skills, such as data analysis, pattern recognition, and predictive modeling. The collaboration of humans and AI can thus be viewed as a “society” in which diverse agents work synergistically to achieve outcomes neither could accomplish alone.

In Minsky’s “Symbolic Vs. Connectionist” paper, he discusses the distinction between “neat” and “scruffy” approaches to AI [21]. The “neat” approach seeks to create well-structured systems with formally defined rules and representations, whereas the “scruffy” approach is more flexible, using heuristic and analogical reasoning. TRIZ (Theory of Inventive Problem Solving) is an example of a “neat” approach. It epitomizes the systematic analysis of patterns and principles to generate innovative solutions. The problem-solving methodology initially developed by Genrich Altshuller is based on analyzing patterns in problems and their solutions to generate innovative ideas. The methodology identifies a finite set of inventive principles that can be applied to solve a wide range of problems, making it a valuable tool for co-invention with AI. The work of George Polya also played a significant role in the development of TRIZ, as he demonstrated that invention could be reduced to a finite set of combinations, further supporting the systematic approach to problem-solving that TRIZ embodies [1, 23].

Building on the foundation of systematic problem-solving methodologies such as TRIZ, more recent developments in the field have incorporated data-driven approaches powered by AI. DELPHI (Dynamic Early-warning by Learning to Predict High Impact) is an artificial intelligence framework developed by Weis and Jacobson (2021) that can provide an “early alert” signal for future high-impact technologies. The framework operates by analyzing patterns in previous scientific publications and using these patterns to predict the likely impact of new research. DELPHI identified pioneering papers on experts’ lists of critical foundational biotechnologies, sometimes as early as the first year after publication. The researchers also demonstrated the use of DELPHI to highlight scientific papers predicted to have a high impact by 2023, thereby supporting better funding allocation for scientific research [31].

"How to Generate (Almost) Anything" is a recent example of exploring the "scruffy" approach of human-AI collaboration in pushing the boundaries of creativity. The project demonstrated how AI can augment human capabilities and inspire the creation of novel and unique designs. The team trained several AI models to generate creative content in various domains, such as music, fashion, and culinary arts. For example, they generated AI-composed music, designed dresses, and concocted new perfumes, all with the assistance of AI. The initiative highlights the positive impact of AI as a complementary tool that can lead to creative and productive outcomes, allowing humans and AI to work together to generate content that may not have been possible otherwise [33].

The paper "From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People" emphasizes the distinction between human-AI interaction and collaboration. While AI and machine learning (ML) are increasingly being applied in real-world contexts, collaboration requires more than mere interaction. True collaboration encompasses mutual goal understanding, proactive task co-management, and shared progress tracking. The authors highlight the importance of incorporating the Computer-Supported Cooperative Work (CSCW) perspective into AI research to facilitate successful human-AI collaboration within complex human workflows [30].

Recently released GPT-4 is a large multimodal model that represents a significant advancement in the field of natural language understanding and has the potential to serve as a powerful tool in the context of human-AI interaction. The model demonstrates proficiency across various tasks, including writing, translation, and code generation, thus offering valuable support for enhancing human creativity and problem-solving efforts. Despite these strengths, GPT-4 is primarily an interactive tool rather than a true collaborator capable of mutual goal understanding or proactive task co-management. As such, the GPT-4 Technical Paper also emphasizes the need to address the model's limitations, such as hallucinations and biases, to ensure safe and effective interactions [22].

In a recent New Yorker article, “There Is No A.I.,” Jaron Lanier remarked, “Digital stuff as we have known it has a brittle quality that forces people to conform to it, rather than assess it. ... A positive spin on A.I. is that it might spell the end of this torture if we use it well” [15]. By improving upon these limitations of LLMs and carefully considering the design of the “society” encompassing humans and GPT-like systems, we strive to create a system that amplifies human capabilities, where AI is not the sole solution but rather a means of reducing the rigidity of existing search systems.

Chapter 3

Motivation

With over 4,000 projects showcased on the Media Lab's website, challenges arise both internally and externally. From an internal perspective, research groups seek to identify project overlaps and uncover collaboration opportunities. From an external standpoint, sponsors strive to discover ongoing research in the lab that aligns with their immediate interests. Crucially, both internal and external parties aim to collaboratively synthesize new research projects that cater to the mutual interests of sponsors and research groups. Latent Lab was designed to facilitate the exploration of Media Lab research projects, quickly identify associative relationships between projects, and prompt thinking at the intersection and edge of existing work.

This thesis traces the evolution of Latent Lab, as it was initially developed before the availability of LLMs fine-tuned with Reinforcement Learning from Human Feedback (RLHF), such as ChatGPT. We initially faced certain constraints but gradually made progress. The introduction of RLHF-LLMs, however, significantly enhanced our ability to analyze projects in greater depth and facilitated the creation of new ones. What began as a promising idea quickly became feasible with the advent of ChatGPT. As such, we felt it was necessary to illustrate the journey that led to our current achievements. The most recent iteration of Latent Lab was developed within weeks of the APIs becoming available. Readers should note that understanding the tool's earlier iterations is not required to comprehend the final version, and this can be achieved without reading Chapter 4, "Early Design Iterations."

3.1 Media Lab Website

The existing keyword-based search on the Media Lab website presents difficulties in identifying interesting projects due to the sparsity and creative nature of the lab's work. Consequently, the search process may not adequately reveal potential connections between projects, hindering the exploration of innovative ideas. Latent Lab addresses this issue by leveraging embedding-based search and a dynamic and interactive way to visualize associations and gaps between existing research projects. The primary objective of Latent Lab is to facilitate the ideation process by providing a starting point that surpasses a mere intelligent average of the lab's previous work, thus promoting the generation of innovative ideas and fostering collaboration.

3.2 Expansion to Additional Data Sets

To further validate the applicability and scalability of Latent Lab, additional data sets were added to the platform for users to toggle between. First, a subset of the US Patent database was considered, encompassing approximately 7,000 patents dating back to 1976, the earliest patents accessible via the USPTO API[1]. All of the 7 million patents available through the API were processed for inclusion. However, a bottleneck in front-end rendering capabilities was encountered, restricting the system to handling around 10,000 data points without compromising performance. This limitation underscores the need for further optimization and refinement of Latent Lab to accommodate larger data sets and enhance its utility in broader contexts. Second, the Center for Constructive Communication at the Media Lab provided a data set from their early "Super Spreaders" project, which explores the spread of misinformation over social media during the COVID-19 pandemic. This data set was

[1]USPTO API:`https://developer.uspto.gov/api-catalog/bulk-search-and-download`

also truncated to 10,000 data points but still provided value to the research group; more details can be found in Chapter 6.

Chapter 4

Early Design Iterations

4.1 Latent Lab V1.0: 2D Canvas, PCA, and OPTIMUS-Embedded Research Titles

The inaugural iteration of Latent Lab showcased a 2D canvas where projects, once selected from the traditional list-based sidebar, became visible. This visualization utilized 32-dimensional embeddings from OPTIMUS, a model released by Microsoft [16], and dimensionality reduction via PCA to illustrate the relationship between chosen projects. The two primary goals of Latent Lab are: 1. exploration of existing projects and 2. generation of new project ideas. In this context, users would individually select projects of interest for exploration on the latent canvas. The system then computed the average of these selected projects in the 32-dimensional space to achieve the second goal. This average, represented in 2D space by a black square, was used to synthesize the title for a newly generated project.

4.1.1 Latent Lab V1.0 Overview

The Latent Lab UI has three main components: the left sidebar, the latent canvas, and the lower-right synthesized tile. The left sidebar contains a tile grid of existing Media Lab research projects, a search bar at the top to filter projects of interest with keyword-based

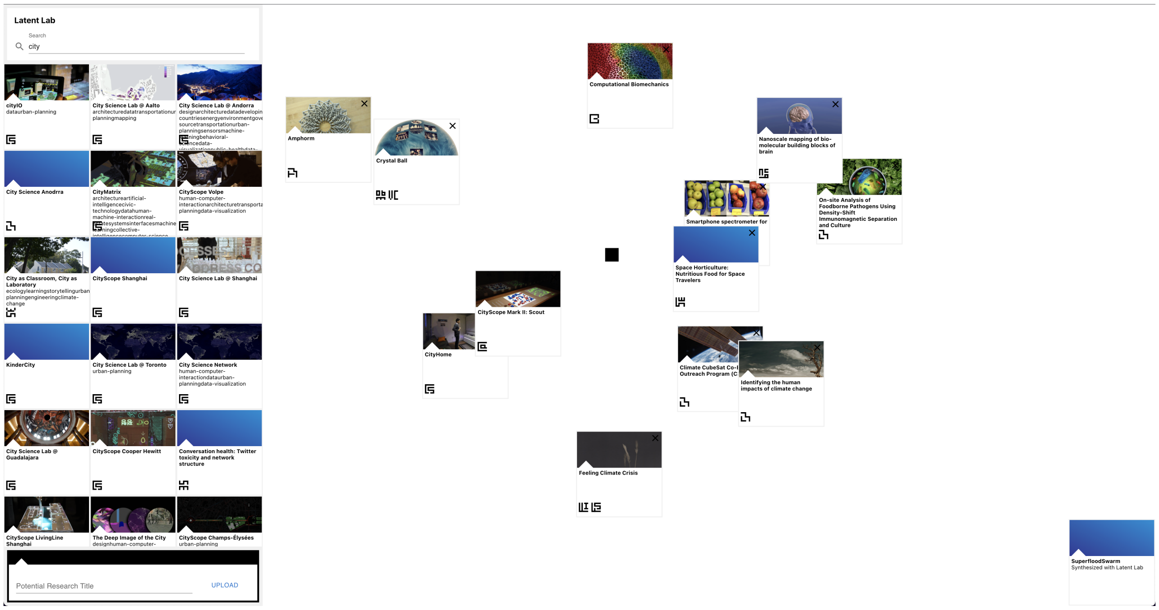


Figure 4-1: Latent Lab V1.0 User Interface

search, and a text input field at the bottom for users to upload research titles that describe ideas of their own. The latent canvas is a visualization of all selected research projects and user-uploaded ideas, represented by their corresponding 32-dimension title embedding vectors, per the default embedding size of OPTIMUS, compressed to 2 dimensions. The latent canvas is essentially an explorable representation of OPTIMUS's latent space. It is altered to produce seemingly Media Lab-ish research titles by fine-tuning OPTIMUS with the titles of Media Lab research. Titles, rather than abstracts or complete project descriptions, were used at the time because of the 512 input token limitation. The lower-right synthesized tile displays a synthesized project title, influenced by all selected research projects and user-uploaded ideas in the latent canvas. The 32-dimension embedding vector of each selected title is averaged to identify a location in the latent space of the OPTIMUS model to sample from and decode a synthesized title of the research.

4.1.2 OPTIMUS Fine-Tuning

OPTIMUS was created and pre-trained by Microsoft. The model is a composition of transformer-based models, Google's BERT as the encoder and OpenAI's GPT-2 as the

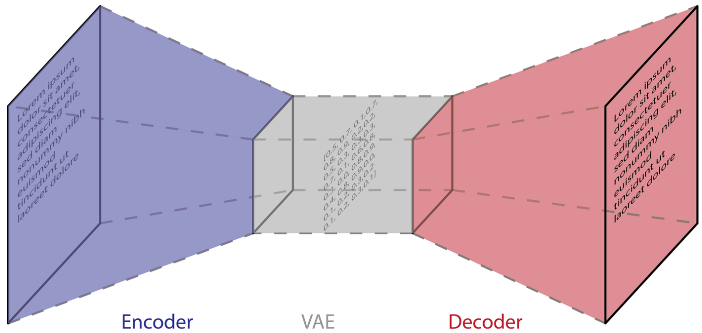


Figure 4-2: OPTIMUS Architecture Sketch

decoder. A Variational Auto-Encoder (VAE) between these two transformer models creates a somewhat smoothly explorable latent space unique for text models, given their sequential prediction method [16]. The data set used to fine-tune OPTIMUS consists of details for all 4,265 projects in the Media Lab as of January 2022. Additional details on the data set can be found in the "Media Lab Data Set" section. The titles of the 4,265 research projects were split into training and testing .txt files of 3800 and 465 strings, respectively. The VAE in the OPTIMUS architecture was trained for 100 epochs, while the weights of the encoder and decoder were held constant during training.

4.1.3 Media Lab Data Set

Projects on the Media Lab website[1] are defined by the following schema:

[1]Media Lab website: `https://www.media.mit.edu/research/?filter=projects`

Active	True or False whether the project is still active.
Description	Text description of the research project, similar to an abstract.
End On	The date the project ended on (YYYY-MM-DD).
Groups	A list of the Media Lab research groups involved in the project.
Tags	A list of keywords associated with the project, often blank.
Modified	A date-time format of when the project was last modified.
People	A list of the email addresses of the researchers involved in the project.
Published	True or False whether the project is published on the Media Lab website.
Slug	String corresponding to the URL of the project on the Media Lab website.
Start On	The date the project started on (YYYY-MM-DD).
Title	A string of the title for the research project.
Visibility	A categorical level of visibility on the Media Lab website.
Website	A string of the URL for any external website associated with the project.

The details and completeness of each component vary widely between projects. For example, of the 4,265 projects that were visible across viewing privileges of the website in January 2022, only 1,299 contained details corresponding to the "Slug" component (used to determine the header image representing the project).

4.1.4 V1.0 Findings and Feedback

During our first iteration of Latent Lab, we discovered that the title alone did not adequately describe a research project. Using titles as representations of research projects was a limitation primarily due to the technological constraints of encoding only 32 vector values and having a context window restricted to 512 tokens. Based on user feedback, we learned that the ability to view all projects within the latent canvas would be highly valued, as it would provide an immediate understanding of the data set as a whole. In subsequent iterations, this proved advantageous over requiring users to select specific projects of interest before visualizing any information on the latent canvas.

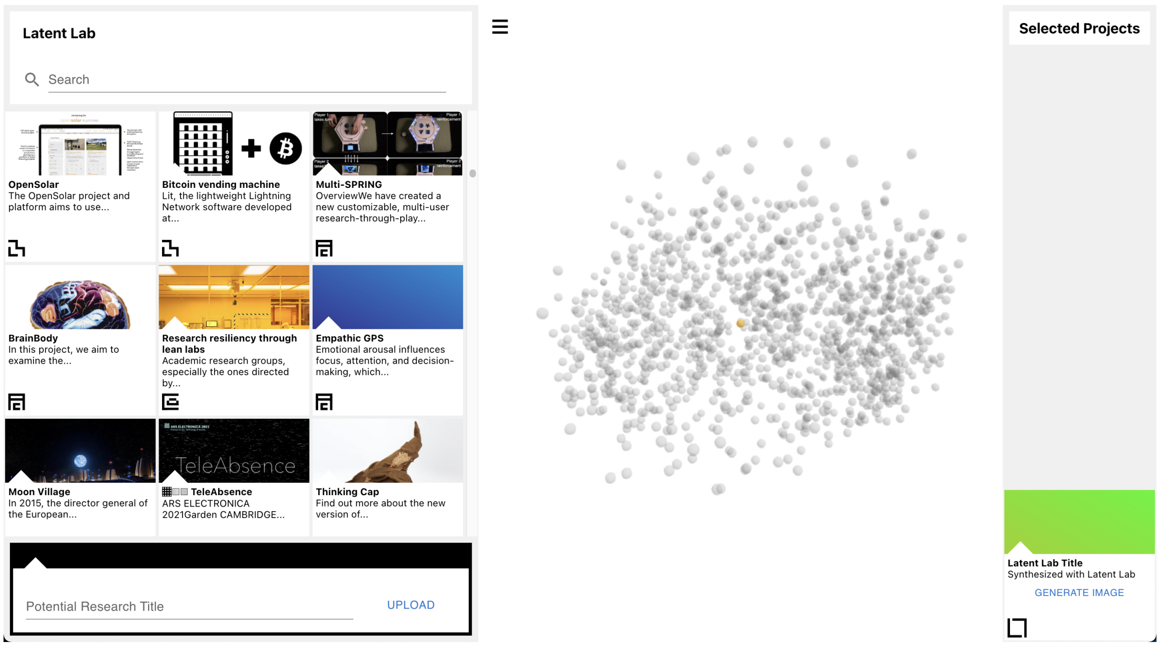


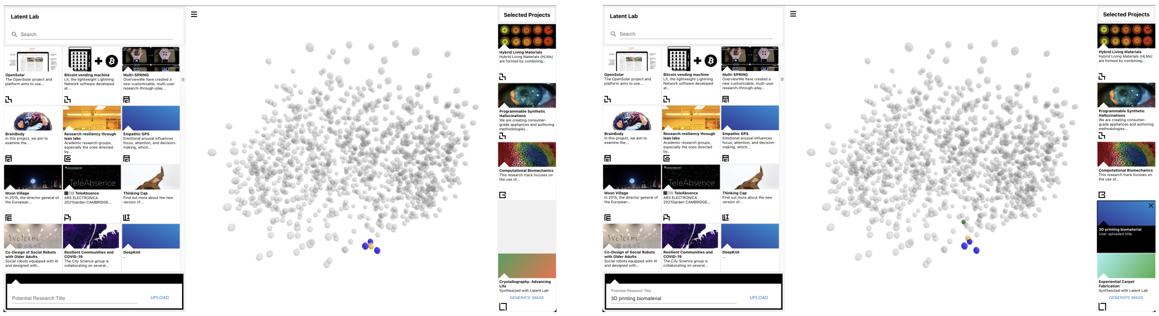
Figure 4-3: Latent Lab V2.0 UI

4.2 Latent Lab V2.0: 3D Canvas, PCA, and OPTIMUS-Embedded Research Titles

In the second iteration of Latent Lab, we explored using a 3D visualization. Again, PCA was used for dimensionality reduction, but this time with three principal components. This version aimed to provide a more immersive experience with the additional dimension and a holistic view of the data set with the ability to see all projects in the latent canvas together.

4.2.1 Latent Lab V2.0 Overview

Selecting a project from the left sidebar of Latent Lab V2.0 reveals its position as an abstract sphere in the latent canvas and adds it to the list of selected projects on the right rail. Project spheres can also be selected directly from the latent canvas for inclusion in the list of selected projects. Still bound by the 512 token constraint in V2.0, OPTIMUS considers only the titles of each project for creating 32-dimensional embeddings. The lower-right synthesized tile broadcasts the latent interpolated title, considering all project titles



(a) Existing Projects (b) User Uploaded

Figure 4-4: Project Exploration in Latent Lab V2.0

in the list of selected projects. The Three.js library was used for rendering and interacting with the 3D visualization.

The visuals expose neighboring pieces of research, highlighting opportunities for internal collaboration. The yellow orb at the center of selected research projects in the latent canvas denotes where the model synthesizes a title in the PCA-compressed latent space. Selecting additional projects or removing selected projects will influence the location of the yellow orb and the resulting generated title. The synthesized title is sampled at the 32-dimension average of all selected titles and displayed in the bottom right corner of the UI.

Additionally, users can upload their own ideas for research titles - areas of interest and potential research yet to be explored. The input text is passed to the encoder of OPTIMUS, the resulting 32-vector array is then passed through the saved PCA model, and the position of the input idea is rendered to the latent canvas. This can also be used as an implicit representation to guide the navigation through the latent space of ideas. Again, where this uploaded title is placed in the latent space signifies its relative similarity to nearby existing research, a potential collaboration opportunity. The latent space between the uploaded title and disparate existing titles represents a potential opportunity for further exploration.

4.2.2 V2.0 Findings and Feedback

In the second iteration of the Latent Lab, we acknowledged the helpfulness of displaying all projects at once, as learned from version 1.0. User feedback suggested incorporating metadata to enhance the research projects beyond gray spheres and provide additional context. The title remained inadequate for fully describing a research project. However, we remained restricted to a 512-token context window. To address these issues, we explored integrating external models and modalities for enriching the generated projects in version 2.1. We specifically investigated using CLIP-VQGAN-generated images to effectively convey research concepts and enhance the user experience and understanding of the data set.

4.3 Latent Lab V2.1: 3D Canvas, PCA, OPTIMUS-Embedded Research Titles, and Multimodality (Images - VQGAN)

4.3.1 VQGAN Image Generation Method

The method used in Latent Lab V2.1 (Figure 4-5) to generate a synthesized research title builds upon the method of steering VQGAN with CLIP discussed in the Taming Transformers paper [10]. Figure 4-6 illustrates this workflow. Given a generated research title output from OPTIMUS, the title is passed along with a random noisy image to CLIP. The image and title similarity is calculated, the loss is back-propagated through the VQGAN generator, and the pixel values of the generated image are adjusted to minimize this loss. The newly adjusted image is then passed again to CLIP, along with the synthesized research title, and this process is repeated 300 times. The resulting image is then available to view in the Latent Lab UI.

Baseline images were captured for the control title, “Tools for Super-Human Time Perception”. The image in column 1 of Table 4.1 is the header image associated with the control title of research on the Media Lab website. The image in column 2 results from passing the control title to a pre-trained implementation of VQGAN and CLIP, as defined in the workflow in Figure 4-6. The VQGAN model was pre-trained with Imagenet, so it has not seen many images like the baseline image and therefore cannot reconstruct it accurately. The CLIP model evaluated, ViT-B/32, was trained on 12 image+text pairs data sets, none of which included the control title and baseline image combination. This means it is improb-

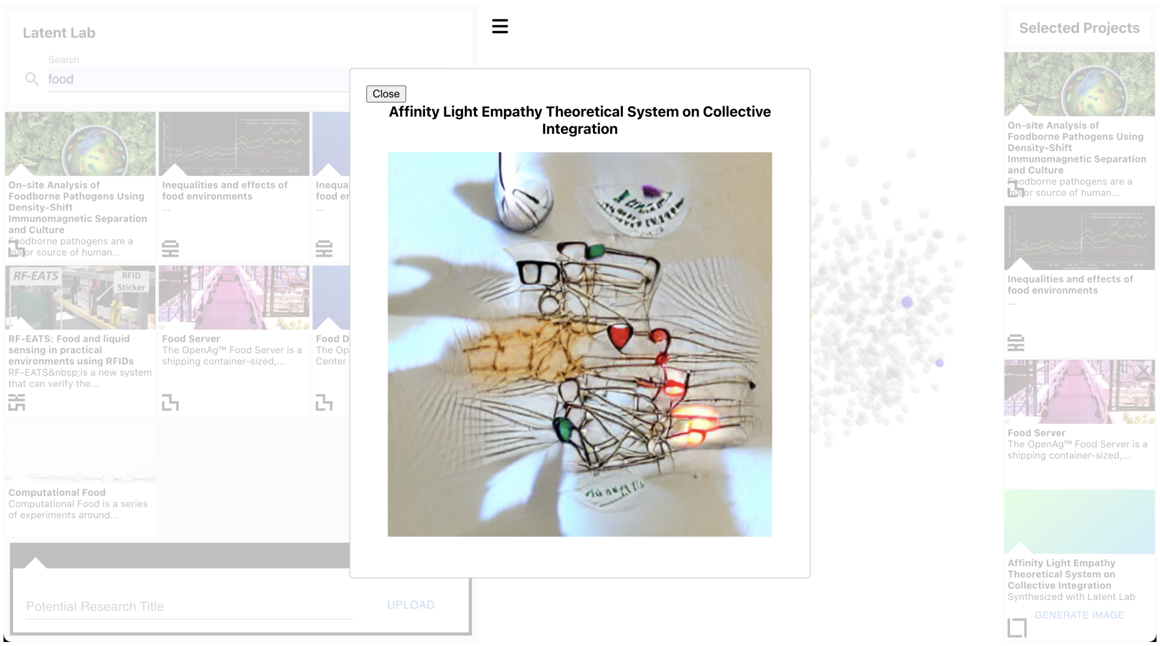


Figure 4-5: Latent Lab V2.1 UI - VQGAN Image Generation

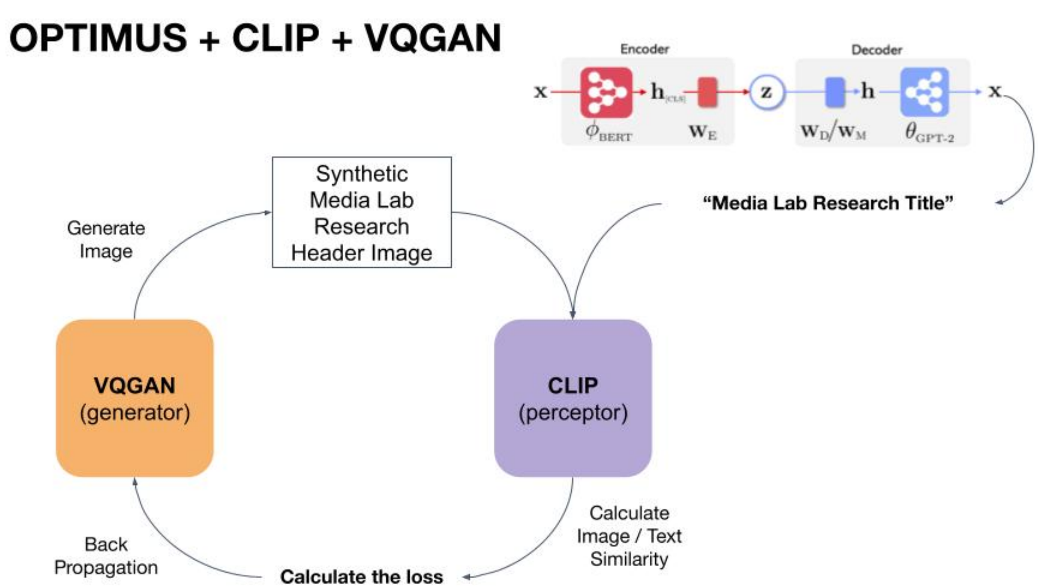
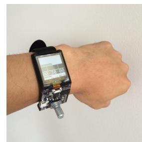
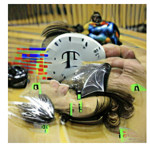


Figure 4-6: VQGAN Image Generation Workflow

Table 4.1: Baseline Image and Standard Generation

1	2
	
Baseline	Pre-Trained CLIP + VQGAN

able that the input text to CLIP would yield a location in latent space near the embedding vector of the baseline image. Therefore, minimizing this image/text similarity loss would not help produce an image similar to the baseline. In an attempt to solve these issues, VQGAN and CLIP were fine-tuned with the Media Lab data set.

4.3.2 Data Preparation

As noted in section 4.1.3, only 1,299 of the 4,625 research projects include header images and are considered for the data set to fine-tune VQGAN and CLIP. The Slug of each project is used to scrape the header image from the media lab website. The image is then resized to 256x256 pixels with a white background if the image is not already square and then saved locally.

4.3.3 VQGAN Fine-Tuning

VQGAN was fine-tuned as suggested in the Taming-Transformers GitHub repository. First, weights and config files were downloaded for the VQGAN model fine-tuned on the Imagenet data set. The config file was then modified to use the downloaded weight file as the starting checkpoint, and the path to the Media Lab training and testing .txt files were added as the corresponding data sets. The training and testing .txt files contained a list of paths to the header images stored locally. An 80/20 train/test split was used, where the training set included the first 1039 images and the test set included the remaining 260. The model was fine-tuned on a Google Colab GPU for 336 epochs with a batch size of 4 until the validation reconstruction loss converged.

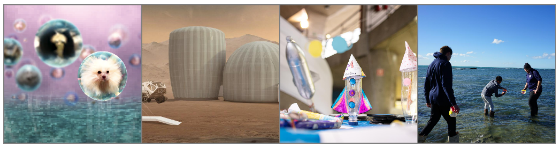
4.3.4 CLIP Fine-Tuning

CLIP was fine-tuned by adapting code found in the CLIP GitHub repository. The state dictionary and preprocessing method were downloaded for the pre-trained ViT-B/32 CLIP model. Images and titles of the 1,299 projects were loaded into the memory of a Google Colab session to create separate arrays of each. Images were converted to pillow (PIL) images, and a data loader was created to zip the titles and images. Additionally, the data loader was configured to call the preprocessing method downloaded with the pre-trained model. This ensured a similar normalization scheme and resizing (224 x 244) of the pre-training images was applied to all fine-tuning images.

CLIP is trained by minimizing the similarity between different images and title pairs while maximizing the similarity of paired images and titles. This means batch sizing is an important parameter. The training code referenced above suggests using a batch size of 20, which requires dropping 19 images from the data set for a divisible quantity of training instances. Instead, a batch size of 10 was used, and only nine instances were dropped initially. During the training process, it was noticed that additional image transformations might be beneficial. An additional preprocessing step was added to support this. Due to the training process, the stacking method was important; each transformation set (image and text pairs) was stacked after the prior transformation set. This ensured there were never batches with multiple transformations of the same image. In these scenarios, all 1,299 pairs were used, each with a number transformation that was a multiple of the batch size.

4.3.5 Latent Lab V2.1 Implementation

A Python Flask API was created for generating an image given a synthesized research title. The text generation back end of Latent Lab runs as a container application on one of the Media Lab servers. Given the size of the fine-tuned VQGAN and CLIP models and the



(a) Target Validation Images

(b) Reconstructed Images

Figure 4-7: Visual Comparison of Generated Images

supporting frameworks required, a separate container app was created to host the image generation portion of Latent Lab. The image generation requires 13GB of GPU RAM, and the individual GPUs on the Media Lab servers are 10 GB. As an alternative, a Google Cloud Virtual Machine with one GPU was provisioned to support the image generation service.

4.3.6 V2.1 Image Generation Results and Discussion

The methodology described above enabled the generation of a supporting image for a synthetic research title in the Latent Lab application. The individual results of fine-tuning VQGAN and CLIP, and the implementation into the web application, all contribute to this method's success and room for improvement.

VQGAN Fine-Tuning Results

The Python package Test Tube was used to monitor the training of VQGAN per the Taming Transformers GitHub repository setup. The validation reconstructed images were monitored, but the .csv files Test Tube created for the validation reconstruction loss were overlooked and not plotted until this section was written. Figure 4-7 is a sample of the visual inspection.

The left batch of images (Figure 4-7 (a)) are the baseline input images, which were passed to the encoder of the fine-tuned VQGAN to create a corresponding embedding vector. The images on the right (Figure 4-7 (b)) are the output images, reconstructed from the embedding vectors by the decoder of the fine-tuned VQGAN. Visual inspection confirms

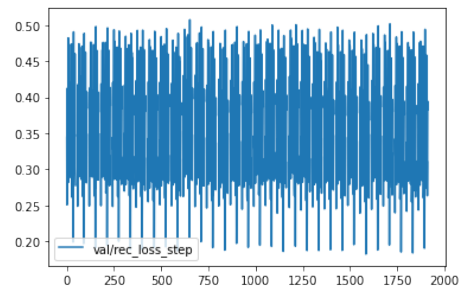


Figure 4-8: VQGAN Validation Reconstruction Loss Plot

the fine-tuned VQGAN's ability to represent the Media Lab research images as embedding vectors in the model's latent space.

Figure 4-8 plots the validation reconstruction loss at each step of gradient descent throughout the training process. The validation reconstruction loss plot oscillates sharply between 0.50 and 0.20, without any strong observable trend. This suggests that the fine-tuning process may have not had a large impact on the ability of the pre-trained VQGAN to represent the Media Lab research images as embedding vectors in the model's latent space. A test to confirm this would be to try reconstructing the same input images in Figure 4-6 with the pre-trained VQGAN and comparing the results with the output images of Figure 4-6.

CLIP Fine-Tuning Results

The fine-tuned CLIP model demonstrates significant improvement in the ability to associate a Media Lab research title with an image that more closely resembles the baseline image than possible with the ViT-B/32 pre-trained CLIP model. However, even considering this improvement, the resulting image is still significantly visually different compared to the target baseline image. This is largely due to the size and sparsity of the fine-tuning data set.

The initial attempts at fine-tuning CLIP are captured in Table 4.2 and expose three major findings; a very low learning rate, evaluation mode, and 32-bit floating point values for weights should be used when fine-tuning CLIP. Turning on evaluation mode, meaning that the normalization layer of CLIP is frozen and not adjusted during gradient descent, appears

Table 4.2: Generated Output for "Tools for Super-Human Time Perception": Fine-Tuned VQGAN, Fine-Tuned CLIP: Batch Size = 10

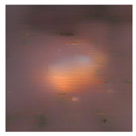
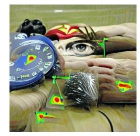
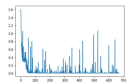
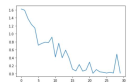
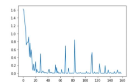
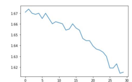
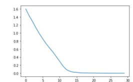
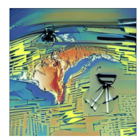
	1	2	3	4	5
Output					
Loss Plot					
Learning Rate	1e–5	1e–5	1e–5	1e–7	1e–6
Eval. Mode	No	No	No	Yes	Yes
Img. Dtype	int x 255	int	fp32	fp32	fp32

Table 4.3: Generated Output for "Tools for Super-Human Time Perception": Fine-Tuned VQGAN, Fine-Tuned CLIP: Lr = 1e-6, Eval. Mode, fp32

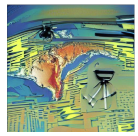
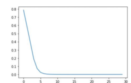
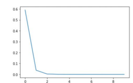
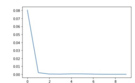
	1	2	3
			
Batch Size	10	2	20

essential in fine-tuning CLIP. A learning rate of 1e–6 seems to be an ideal combination of quick and smooth convergence that yields an expressive output image.

The batch size was adjusted to further align the generated output image with the baseline image in Table 4.1, column 1. It was hypothesized that increasing batch size would force a stronger differentiation between image/text pairs from other image/text pairs in the data set. A smaller batch size of 2 was also tested. Table 3 suggests that larger batch sizes produce a more cohesive and homogeneous output image than smaller batch sizes, which makes sense considering the training process of CLIP discussed prior. The subsequent tests were run with a batch size of 20.

After optimizing hyperparameters, the focus shifted to data augmentation for further improvement of image generation. As aforementioned, the Media Lab data set is small and sparse. To help improve the size, random image augmentation was used to generate a larger set of images. Table 4.4 showcases the resulting images from CLIP models fine-tuned

Table 4.4: Generated Output for "Tools for Super-Human Time Perception": Fine-Tuned VQGAN, Fine-Tuned CLIP: Lr = 1e-6, Eval. Mode, fp32, Batch Size = 20

	1	2	3	4
Output				
Loss Plot	Not Available			
Transformations	1x	10x	50x	80x

with an augmented data set that was 1, 10, 50, and 80 times larger than the original data set. With 80x transformations, the loss plot mostly converges after a single epoch. This suggests that additional transformations of these types will offer significantly diminished improvement.

Tuning hyperparameters and augmenting the data helped improve image generation, but the data set is far too sparse to generate images identical to those representing existing projects. The sparsity results from Media Lab research images covering such a wide range of content, spanning abstract images to images of physical inventions to images of screen-based applications. This could potentially be remedied by expanding the data set with image/text pairs from sources similar to the Media Lab. Examples could be header images and titles from publications like Wired and Tech Crunch or research projects from other similar institutional research programs, like the University of Colorado Boulder's ATLAS or Aalto's Media Lab in Helsinki, Finland.

Implementation Results

Clicking the research title of an existing research project within Latent Lab will open the corresponding web URL in a new browser tab. In prior versions of Latent Lab, clicking the title of a synthesized title of research would result in a 404 Page Not Found error. Now, in V2.1, after a user chooses to generate an image for a synthesized title of the research, a modal will open, displaying the generated image when the research title is clicked.

This multi-step user interaction is not ideal, but results from the current latency in the image generation process. It takes about 3 minutes to generate an image with this method, so the front end is designed to require an explicit generation request from the user. One solution to the latency could be to optimize the generated code to be distributed across multiple GPUs. Additionally, this version has no back-end scaling or queuing in place to prevent a server request overload, which can occur from spamming or multiple concurrent user sessions. This could be fixed by using a container orchestration service.

4.3.7 V2.1 Findings and Feedback

Again, using research titles as proxies for entire projects was noted to be a poor means of organizing the relationships between projects and even worse for a new-project generation. While the image generation approach is somewhat intriguing, it offers limited utility in validating the meaning or feasibility of generated idea titles. Furthermore, during the course of this iteration, new open-source alternatives for image generation, such as Stable Diffusion, emerged.

4.4 Latent Lab V2.2: 3D Canvas, PCA, OPTIMUS-Embedded Research Abstracts, and Multimodality (Images - Stable Diffusion)

To improve project visualization, Latent Lab V2.2 employed 3D visualization and higher-dimensional embeddings (762 dimensions). Again, PCA was used for dimensionality reduction, and a single representative sentence of the project description was considered. GPT-3 was used for preprocessing to help extract this representative sentence, which was then used to fine-tune the OPTIMUS model. This version also used GPT-3 downstream to expand upon a generated sentence describing a research project to create an abstract of the project and condense the representative sentence down to a title for the research project. Additionally, an instance of Stable Diffusion V1.4 was provisioned on a Media Lab server to

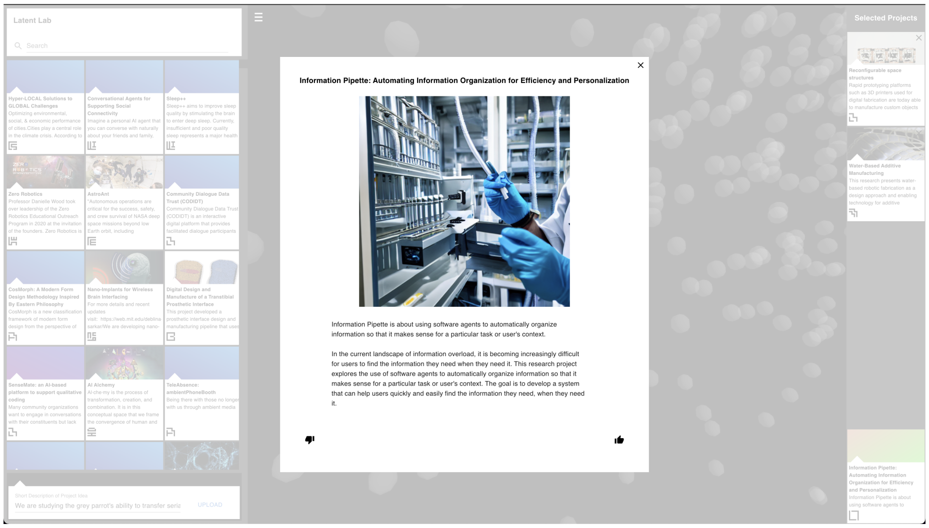


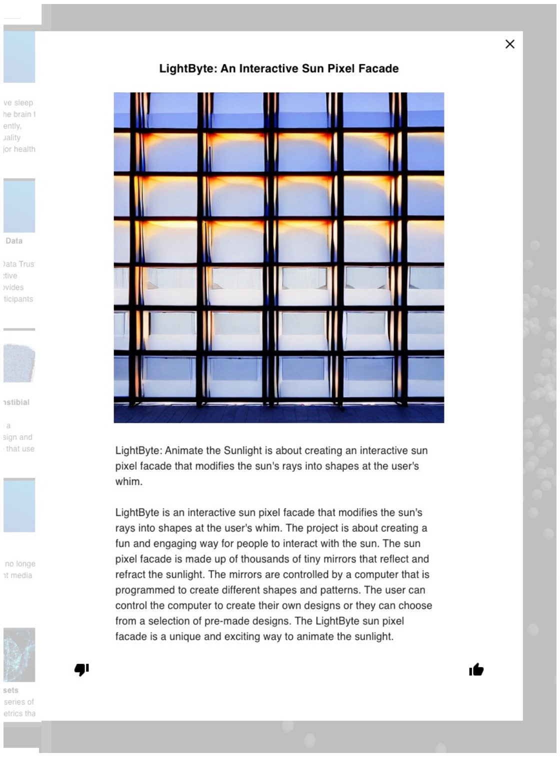
Figure 4-9: Latent Lab V2.2 UI

support image generation instead of the VQGAN method. The resulting containerized version of Stable Diffusion was open-sourced on GitHub[2]. A voting system was incorporated to gather feedback on generated projects, allowing users to vote up or down. Figure 4-9 shows a generated project within the V2.2 UI.

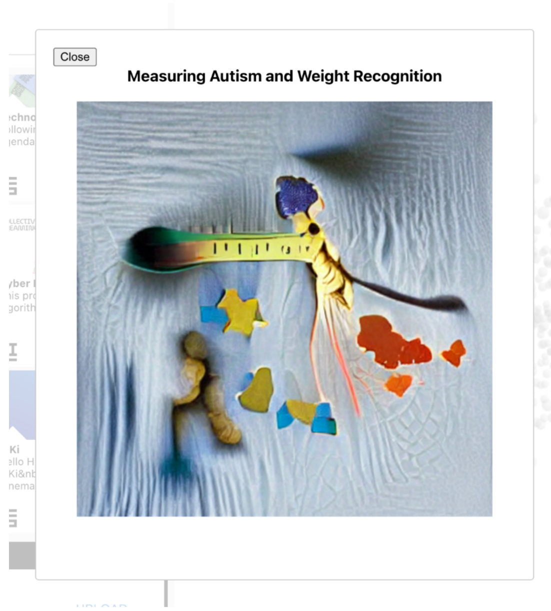
A comparison of output images from V2.1 and V2.2 is presented below. These examples, though not representing identical project combinations, emphasize each version's strengths and weaknesses. V2.2's Stable Diffusion method generates less abstract, more convincing supporting images in significantly less time than V2.1's VQGAN method. For instance, in Figure 4-10 (a), V2.2's "LightByte: An Interactive Sun Pixel Facade" on the left provides a plausible depiction of such an art installation. In contrast, V2.1 abstractly visualizes the less cohesive title "Measuring Autism and Weight Recognition" on the right (Figure 4-10 (b)). It's important to remind readers that both project text and supporting image generation methods have been enhanced in V2.2 since V2.1.

However, the explicit images' lack of abstractness may limit their ability to produce "happy accidents"—a term Epstein et al. use to describe unexpected deviations in generated images from their prompts, leading to new user insights [9]. Moreover, Stable Diffusion occasionally

[2] Containerized Stable Diffusion Server: `https://github.com/viral-medialab/stable_diffusion_server`



(a) V2.2 - Combination of "Whispers of the Mountain"[3], "Sociomedia Garden"[4], and "Attentive Electronic Skins For Low Power, Multi-functional Operation"[5]



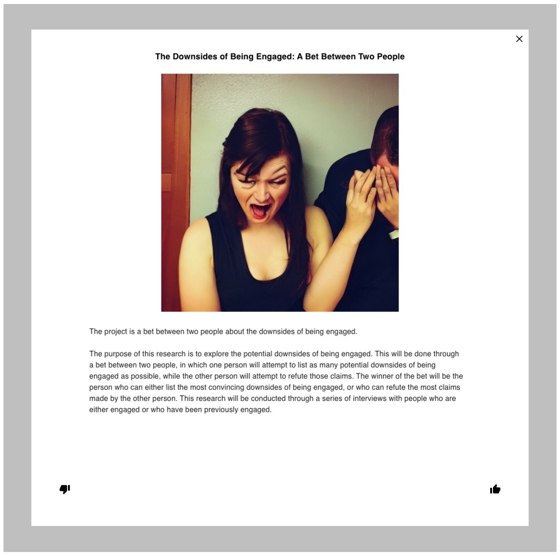
(b) V2.1 - Combination of "Freedom Radio"[6], "Black Forest"[7], and "Lifelong Personalized Models"[8]

Figure 4-10: Latent Lab V2.2 and V2.1 Image Comparison 1

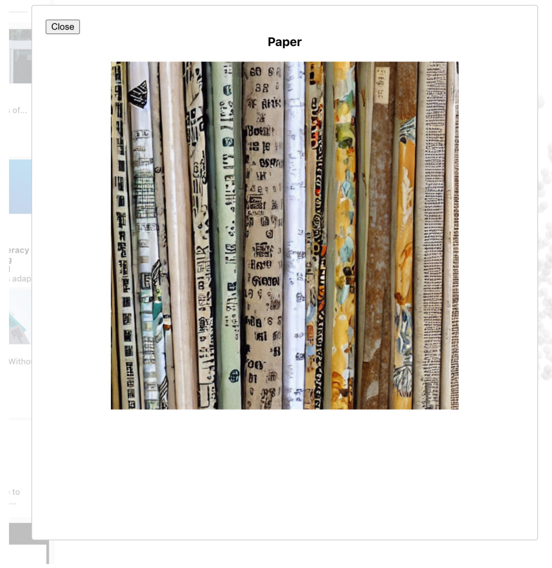
falters with highly theoretical and abstract research project ideas, seemingly attempting to visualize the text description rather than the concept itself. In contrast, VQGAN excels at such abstract visualization.

Figure 4-11 (a), generated by V2.2, represents "The Downside of Being Engaged: A Bet

[3]https://www.media.mit.edu/projects/whispers-of-the-mountain/overview/
[4]https://www.media.mit.edu/projects/sociomedia-garden/overview/
[5]https://www.media.mit.edu/projects/attentive-skins/overview/
[6]https://www.media.mit.edu/projects/freedom-radio/overview/
[7]https://www.media.mit.edu/projects/blackforest/overview/
[8]https://www.media.mit.edu/projects/lifelong-personalized-models/overview/
[9]https://www.media.mit.edu/projects/child-driven-machine-guided-literacy-activity/overview/
[10]https://www.media.mit.edu/projects/project-octopus/overview/
[11]https://www.media.mit.edu/projects/honest-crowds/overview/
[12]https://www.media.mit.edu/projects/are-you-still-here/overview/
[13]https://www.media.mit.edu/projects/inspire/overview/



(a) V2.2 - Combination of "Child-Driven, Robot-Guided Literacy Activity"[9] and "Project Octopus"[10]



(b) V2.1 - Combination of "Honest Crowds"[11], "Are you still here?"[12], and "INSPIRE"[13]

Figure 4-11: Latent Lab V2.2 and V2.1 Image Comparison 2

Between Two People" through an image from 'The Uncanny Valley.' Though the peculiar and unsettling project depiction is somewhat faithful to the title, it's more disconcerting than inspiring. Similarly, "Paper," produced by V2.1 and shown in Figure 4-11 (b), lacks any discernible ideation insight. Despite this, its vibrant illustration adds a touch of excitement to the otherwise ordinary generated research title.

Figure 4-12 (a), generated by V2.2 for "The LEV Project: A New Way to Manipulate Physical Matter," appears to be a compelling research title interpolation between the selected projects. However, it clearly struggles in its attempt to visualize the concept rather than the text. Conversely, the image on the right, created by V2.1 for the project idea "Affinity Light Empathy Theoretical System on Collective Integration," shows inspiring elements. The eye-

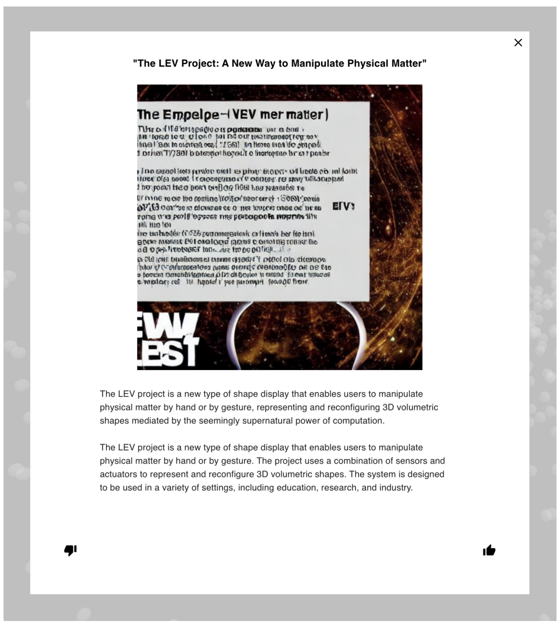
[14]`https://www.media.mit.edu/projects/knittedkeyboard-ii/overview/`

[15]`https://www.media.mit.edu/projects/navajo/overview/`

[16]`https://www.media.mit.edu/projects/on-site-analysis-of-foodborne-pathogens-using-density-shift-immunomagnet`
`overview/`

[17]`https://www.media.mit.edu/projects/inequalities-and-effects-of-food-environments/`
`overview/`

[18]`https://www.media.mit.edu/projects/food-server/overview/`



(a) V2.2 - Combination of “KnittedKeyboard II”[14] and “Navajo Star Traveler”[15]

(b) V2.1 - Combination of “On-site Analysis of Foodborne Pathogens Using Density-Shift Immunomagnetic Separation and Culture”[16], “Inequalities and effects of food environments”[17], and “Food Server”[18]

Figure 4-12: Latent Lab V2.2 and V2.1 Image Comparison 3

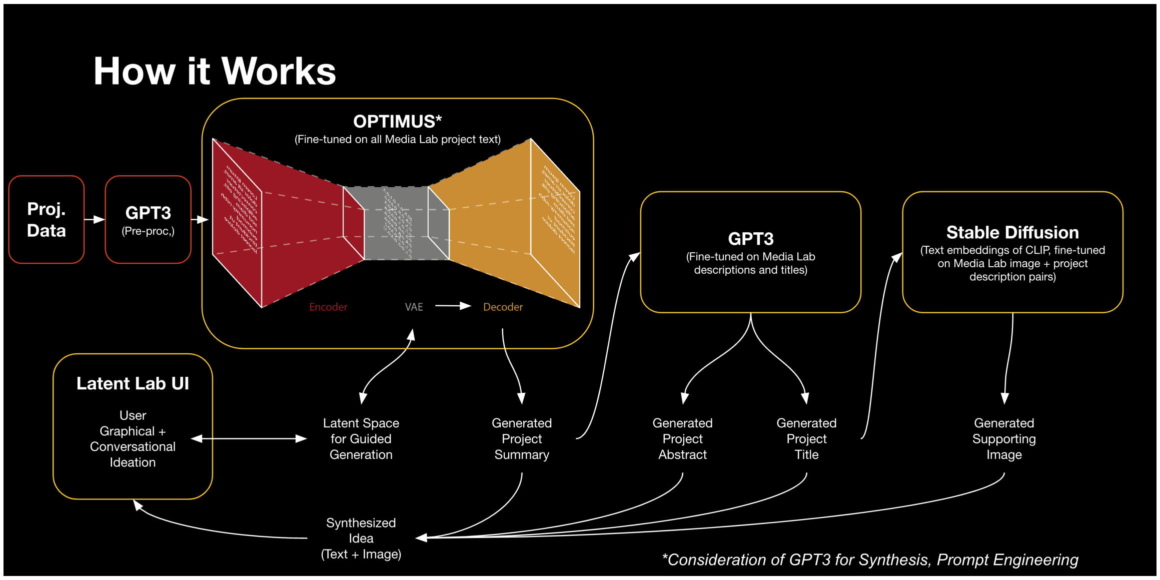


Figure 4-13: Latent Lab V2.2 Architecture

glasses, light, heart, and connecting veins collaborate to conjure a dream-like visualization of the perplexing project title.

4.4.1 V2.2 Larger Context Architecture

The system architecture in V2.2 remains centered around OPTIMUS to enable guided language generation from an abstract level using latent vectors, overcoming challenges faced with large language models like GPT-3. Differing from previous iterations, the OPTIMUS VAE was fine-tuned with Media Lab project data using a larger vector space (768 dimensions) with a similar-sized context window (512 tokens). GPT-3 is employed to both expand descriptions into project abstracts and condense them into titles. Along with Stable Diffusion V1.4, Latent Lab V2.2 uses multiple AI models to create a multi-modal "synthesized idea" for user feedback.

4.4.2 V2.2 Findings and Feedback

In Latent Lab V2.2, newly generated project ideas displayed convincing results, with cohesive project descriptions, representative titles, and improved output images compared to

V2.1 with VQGAN. However, the organization of existing Media Lab projects seemed to lack explainable relative placement in the latent canvas. Projects seemingly unrelated were placed near each other and vice versa, countering our goal of facilitating exploration. User feedback suggested that the correct, meaningful organization of existing project ideas should be a higher priority than further improving the quality of generated project ideas. These findings, coupled with OpenAI's recent announcement of the Ada embeddings service offering larger token input and increased embedding vector length, prompted us to shift focus from project generation to project organization and exploration. A preliminary test with OpenAI's embedding service produced a more compelling latent canvas organization than we achieved using OPTIMUS. Despite its advancements, version 2.2 lacks visual quality and high-level metadata to provide immediate insights. Although we partially adhered to Shneiderman's mantra of offering details on demand, the zoomed-out view does not effectively guide users to find the information they seek. Shifting the focus to organization and access in the next iteration helped to organically address this issue.

Chapter 5

Methods: Latent Lab Today

5.1 Latent Lab V3.0: UMAP, Topic Extraction, Density Contours, and GPT-Embedded Research Descriptions

Version 3.0 of Latent Lab[1] prioritizes enhanced data set organization and employs Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction, replacing the previously used PCA method. This iteration incorporates GPT-3 Ada embeddings for a comprehensive organization that considers full project descriptions and introduces an automated pipeline for topic extraction and organization with unstructured data sets. Moreover, the front end has been reimagined to give users a more captivating and insightful exploration experience.

5.1.1 Architecture

The Latent Lab V3.0 architecture is designed to prioritize performance and scalability. The backend is built using Fast API and Python and is hosted on one of the Media Lab servers. The front end, on the other hand, employs Vercel, Next.js, React, and Typescript. Vercel, a cloud platform, is utilized for static frontends and serverless functions. The initial

[1] Latent Lab:`https://latentlab.ai/`

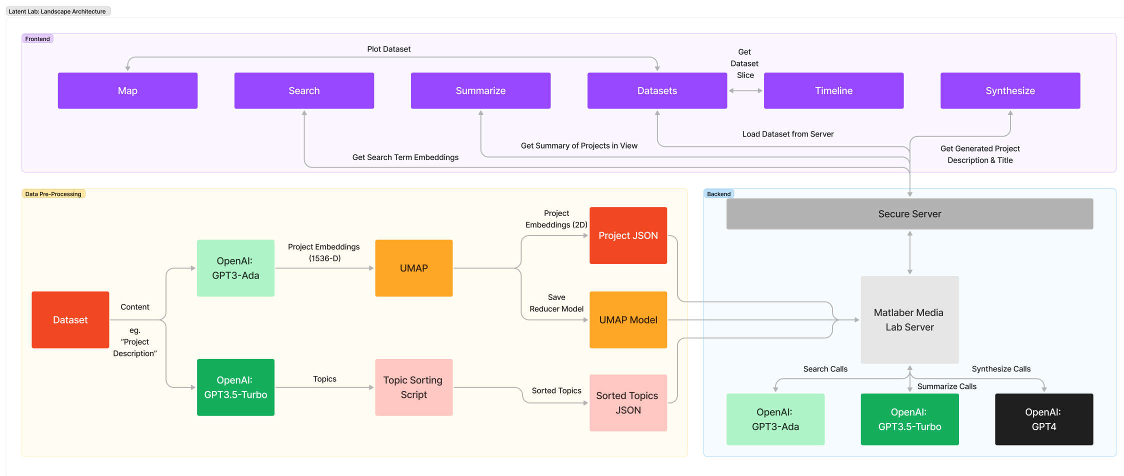


Figure 5-1: Latent Lab V3.0 Architecture

approach was to execute all operations on the front end. However, due to the absence of a fully JavaScript-ported version of UMAP, the decision was made to incorporate a server. This also facilitated server-side rendering to expedite the loading of various data sets.

The data processing pipeline is mostly automated and runs independently of the back end for each new data set. The pipeline generates three primary artifacts: a project JSON containing the unstructured data and embedding data for every project, a pickled UMAP model trained on the data set embeddings to reduce dimensionality to 2D, and a sorted Topics JSON containing all topics produced by the pipeline, organized by the topics with the most associated projects. These artifacts are supplied to the Media Lab server, which, in response to various frontend interactions, either accesses or manipulates the data through one of the GPT API endpoints. Figure 5-1 illustrates this workflow of the various components.

5.1.2 Components

The front end of Latent Lab V3.0 is built using a modern web stack of D3.js, React, and Next.js and is hosted on Vercel. The D3.js library is used for rendering individual data points, label placement, and contour lines in the map visualization. Next.js is used for handling routing and viewing different data sets. This allows users to alternate between

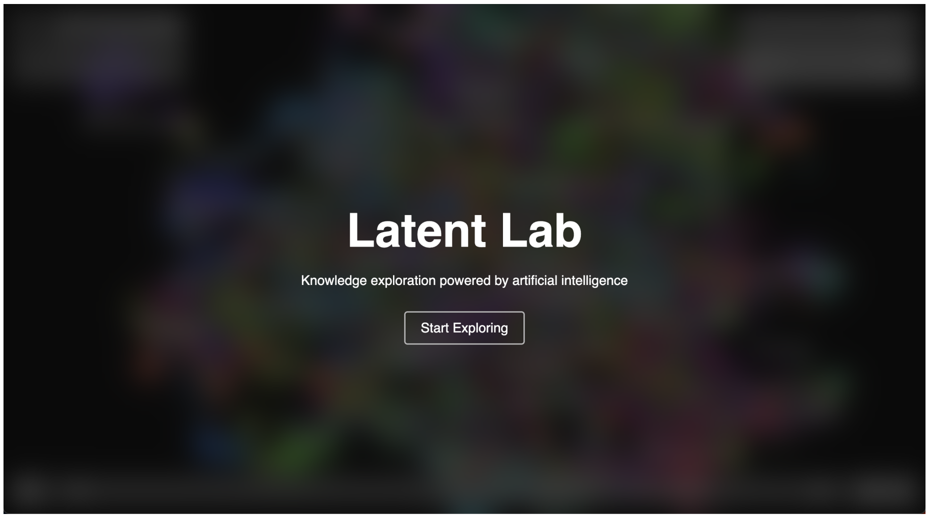


Figure 5-2: Latent Lab Splash Screen

different available datasets. These currently include the MIT Media Lab, USPTO, and "Pandemic Super Spreaders," data sets.

Map

The main UI component of Latent Lab V3.0 is the map, illustrated in Figure 5-3. Latent Lab V3.0 utilizes GPT-Ada embeddings with 1,536 dimensions to represent project descriptions. These high-dimensional representations are then transformed using the UMAP dimensionality reduction technique, resulting in two-dimensional x and y coordinates suitable for visualization. The tool presents all publicly visible projects from the MIT Media Lab as of April 12, 2023, as colored dots on the map, with each color signifying the responsible research group. Contour lines, generated with the D3.contourDensity function, indicate data density within clusters, a concept borrowed from topographic maps where they represent elevation. Paired with the timeline, the changing contour lines reveal the evolution of research concentration. The centroid coordinates for topic and subtopic labels are derived from the aggregate of projects encompassing these themes.

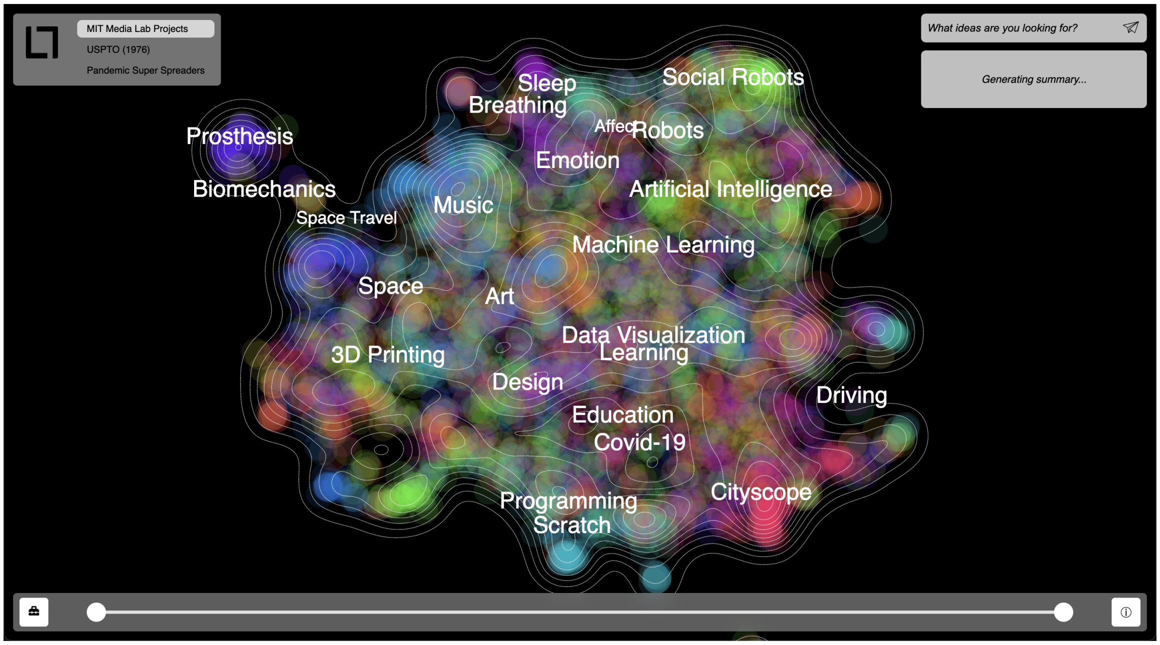


Figure 5-3: Latent Lab Map

Search

Latent Lab V3.0 employs semantic search, which compares the embedding values of search terms with those of research projects to identify semantically similar results. A search request is passed to the back-end server, and the GPT-Ada API is called to return the corresponding 1,536 value embedding for the search term. This embedding is passed to the UMAP reducer, corresponding to the data set saved on the Media Lab server, to return x and y coordinates. These x and y coordinates are returned to the front end, which dynamically zooms to the coordinates on the map and highlights the surrounding region for the user to explore. Figure 5-4 demonstrates a search for "quadratic voting" and the blockchain area of the map being highlighted.

When a project is selected in Latent Lab, the workbench expands from the left side, presenting users with the project's description. Figure 5-4 shows the project "Election ERG" being inspected. This description is utilized to create the embeddings and determine the project's 2D coordinates on the map. Users can access the project page on the Media Lab website by clicking on the project title. Furthermore, the research group responsible for the project is

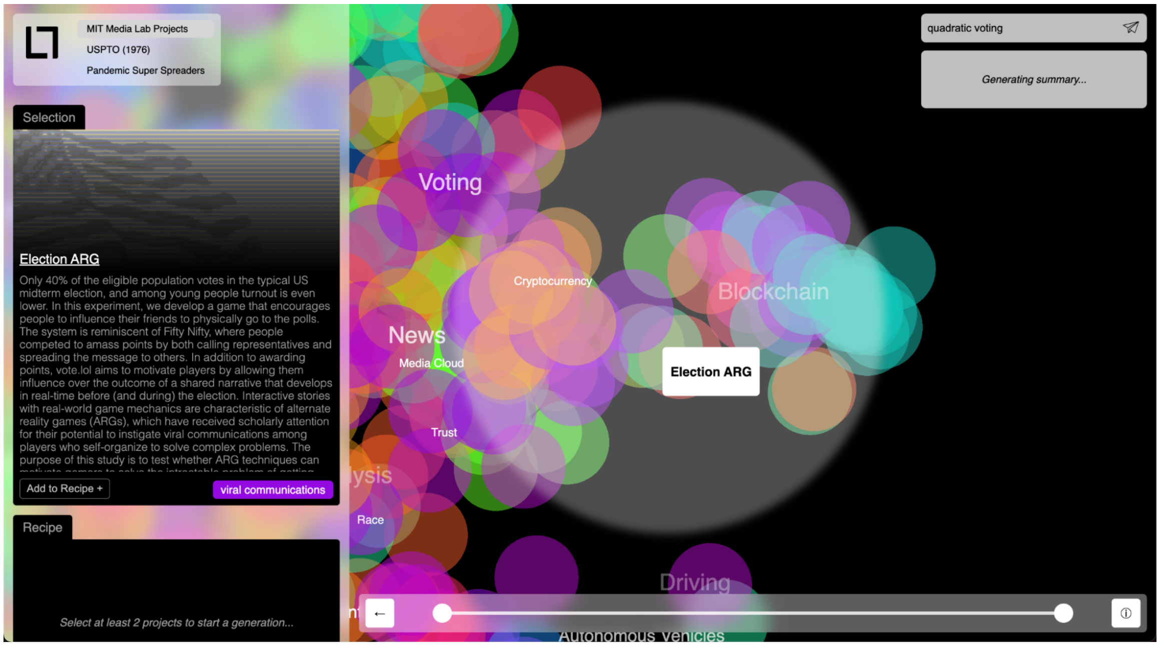


Figure 5-4: Latent Lab Search

indicated using the same color that represents the project on the map, establishing a visual connection between the workbench and the map.

Synthesis

Within the workbench, Latent Lab enables users to create a "recipe" that explores the intersection among existing projects, facilitating the discovery of potential future research directions. The intent is to leverage AI and construct interpolations between known edges of research projects, filling holes or "missing spots of goo" in the map of existing projects, as Feynman would describe it. Figure 5-5 shows the generation of new a project idea considering the entirety of two existing research projects. As suggested in the feedback on previous iterations, users can either include the entirety or a subset, such as the technology, problem statement, or community of the projects being considered (Figure 5-6).

Upon formulating a recipe, users can click "generate" to input the project descriptions into a predefined prompt, which is then modified to focus on specific aspects of the selected projects, as indicated by the user. This prompt is fed via the OpenAI API to GPT-4,

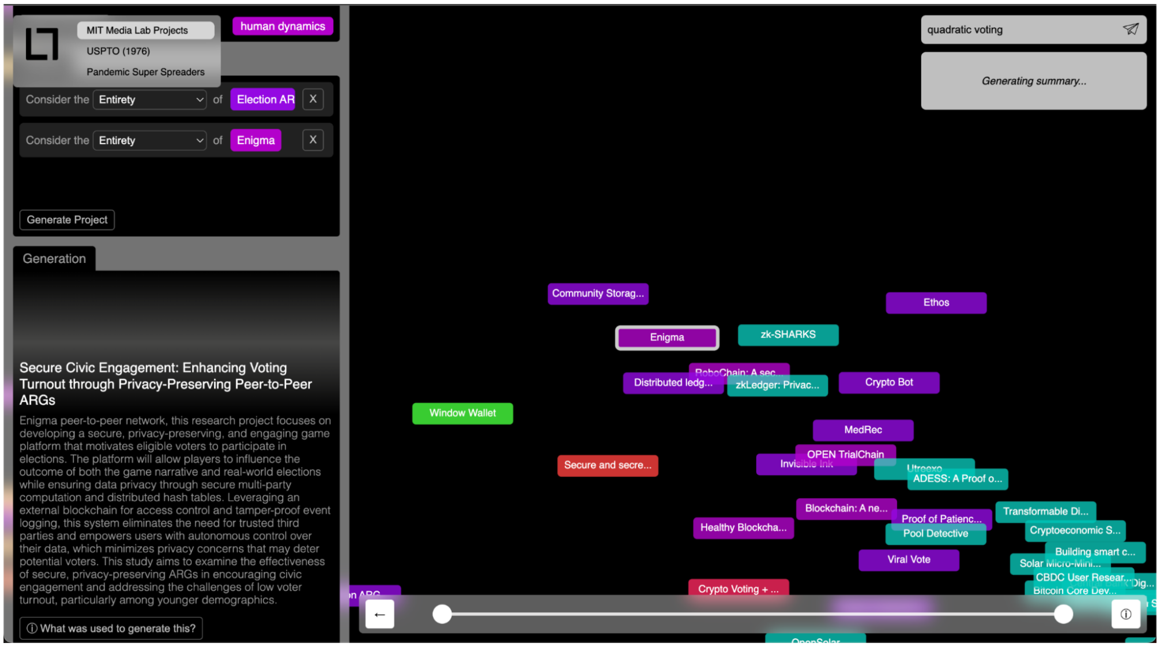


Figure 5-5: Latent Lab Synthesis: Project Entirety

which generates a synthesized project description and title. Prompting has also been used to define the JSON response structure of the generated project idea. Users can explore the exact prompt sent to GPT by selecting the button with the information icon and "What was used to generate this?" at the bottom of the generated project.

Prompt-tuning has been shown to be a more effective approach than fine-tuning LLMs in certain scenarios [28]. Unlike fine-tuning, which adjusts a subset or sometimes the entire model's parameters, prompt-tuning involves minimal changes by tuning only task-specific prompts. This method preserves the original language representations while effectively guiding the model to generate creative and contextually relevant project ideas. The advantage of prompt-tuning lies in its high performance, with fewer adjustments to the model's parameters, making it efficient and resource-friendly. Consequently, prompt-tuning is an acceptable and potentially superior method for generating new research project ideas using GPT models without requiring extensive language-tuning.

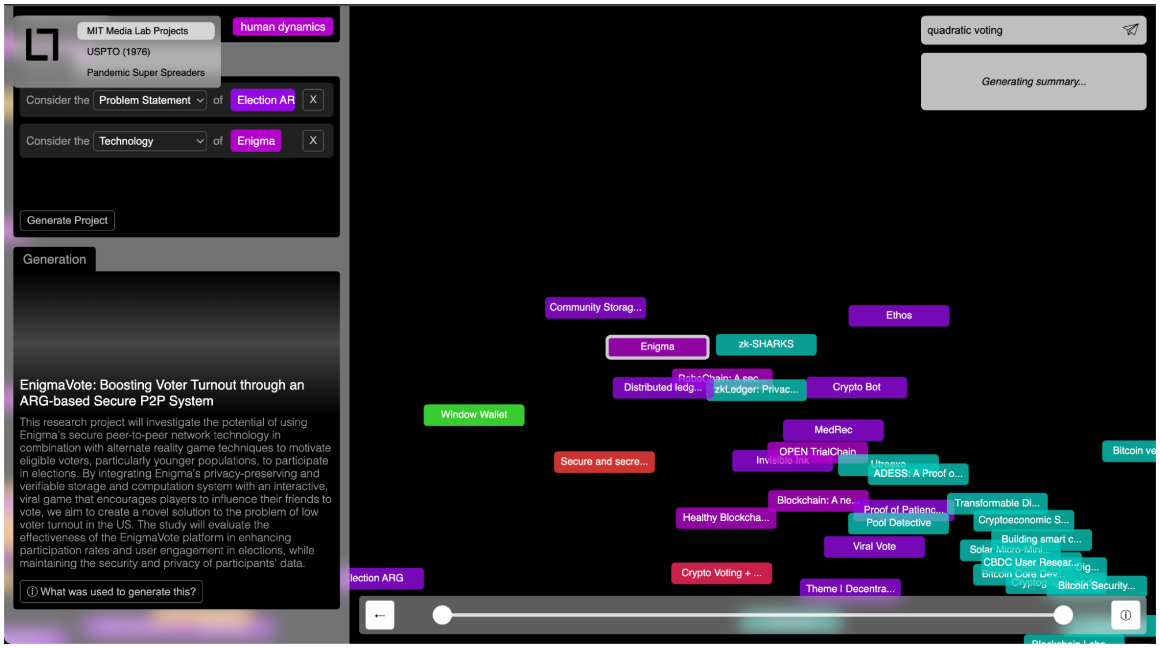


Figure 5-6: Latent Lab Synthesis: Project Components

Summarization

Below the top-right search bar is a summary of what the user is currently viewing. While topic labels provide waypoints for further discovery, the summarization feature aims to provide a digestible synthesis of all data points on the screen. Summarization is run whenever the user views the same portion of the screen for 3 seconds. This allows reduces the number of calls to the summarizer and enables the user to focus on an area of interest before generating a summary. The summary is generated by passing a random selection of 10 project descriptions currently in view to GPT-3 DaVinci with prompting that requests a 2-sentence high-level summary of the project descriptions. We envision this as the foundation for dynamic storytelling of a data set at both large-scale and fine-grained views. Figures 5-7 and 5-8 demonstrate summarization at both a data set and cluster levels.

Timeline

The timeline component within Latent Lab allows for the analysis of data sets with temporal aspects, enabling users to observe the evolution of data over a specified period. The

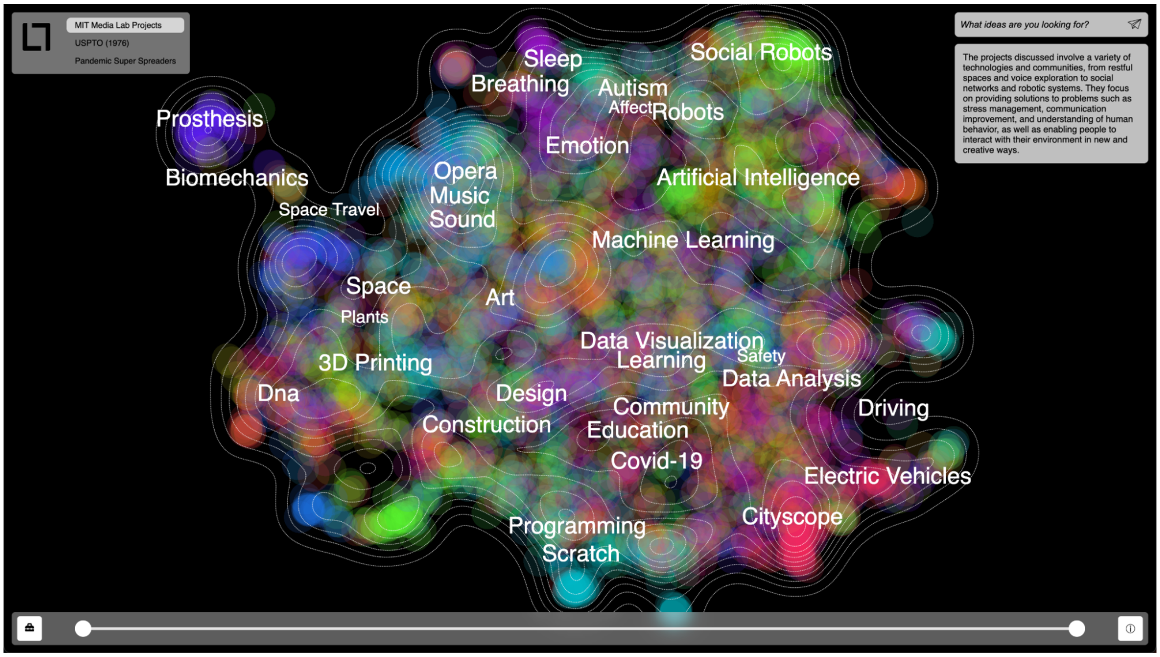


Figure 5-7: Latent Lab Summarizer: Entire Data Set

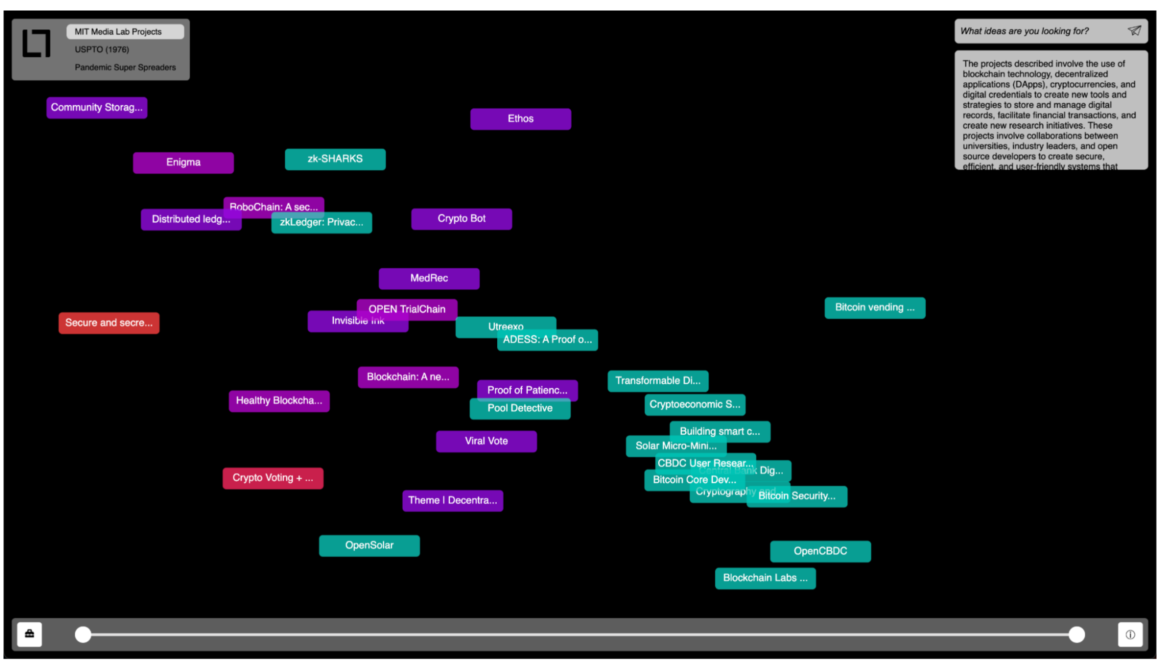


Figure 5-8: Latent Lab Summarizer: Cluster

implementation of start and end sliders facilitates the selection of a subset of data within the bounds of these markers, streamlining the visualization process. This feature proves particularly advantageous when combined with the search bar, as it aids in the identification of current or ongoing projects within a specific area of interest. Overall, the timeline component contributes to a more concise and efficient exploration of time-sensitive data sets. Figure 5-9 illustrates the timeline's capacity to visualize the evolution of the Media Lab data set over time. Additionally, Figure 5-10 shows a specific slice of the data, research conducted since the onset of the COVID-19 pandemic.

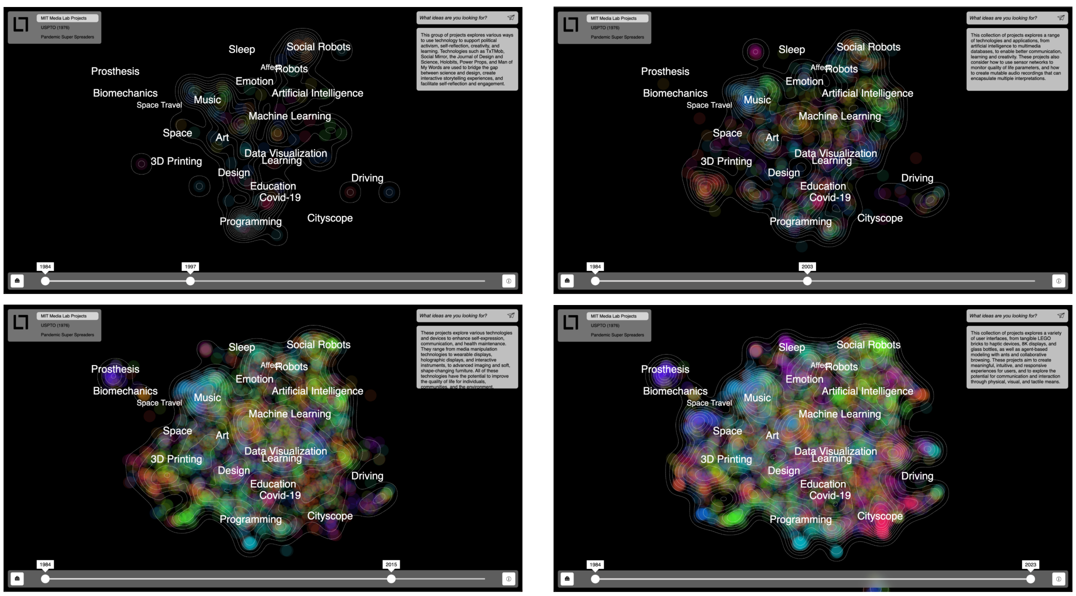


Figure 5-9: Latent Lab Timeline: Evolution

Data Set Toggling

The data processing and visualization engines of Latent Lab V3.0 have been abstracted from the underlying data set, ensuring adaptability and ease of integration with various data sources. We used a subset of the US patent data set, "USPTO (1976)," and a collection of 10,000 social media posts regarding the COVID-19 pandemic, "Pandemic Super Spreaders," as additional data sets to showcase this capability. Figure 5-11 shows the USPTO data set viewed within in Latent Lab. Users can toggle between data sets in the

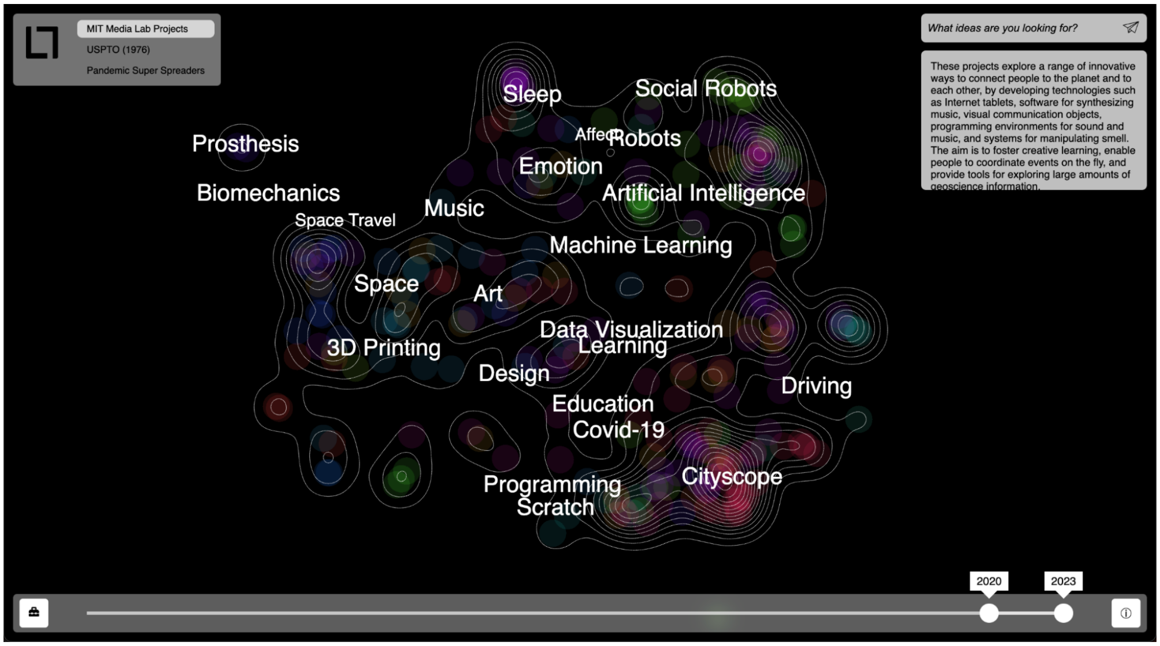


Figure 5-10: Latent Lab Timeline: Recent Slice

top left corner, enabling the seamless exploration of alternative data sets within the Latent Lab environment.

Information

Lastly, basic information about the project, including details about the individuals involved, links to relevant project pages, and contact emails, can be found in the bottom corner of the page (Figure 5-12).

5.2 Preprocessing Pipeline

An automated preprocessing pipeline was developed to process JSON data sets, obtain embedding vectors for content sections, and perform topic and subtopic extraction. Currently, Latent Lab requires users to upload project data in JSON file format, with consistent formatting across all projects (e.g., "title," "year," "description"). The preprocessing pipeline

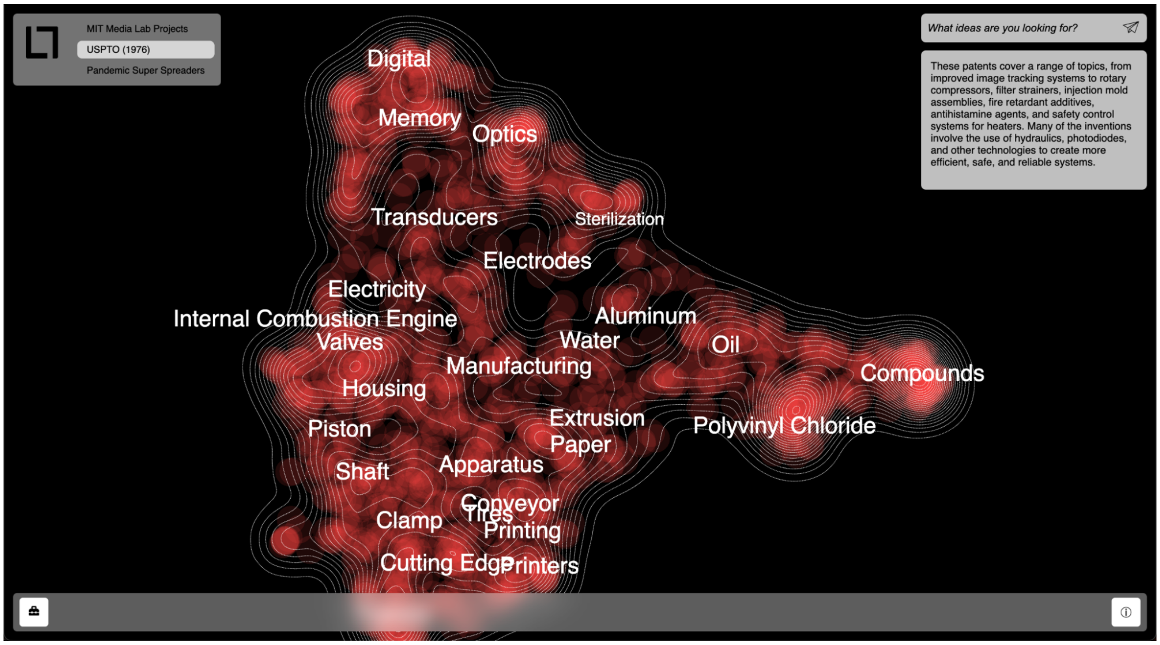


Figure 5-11: Latent Lab Patent Data Set

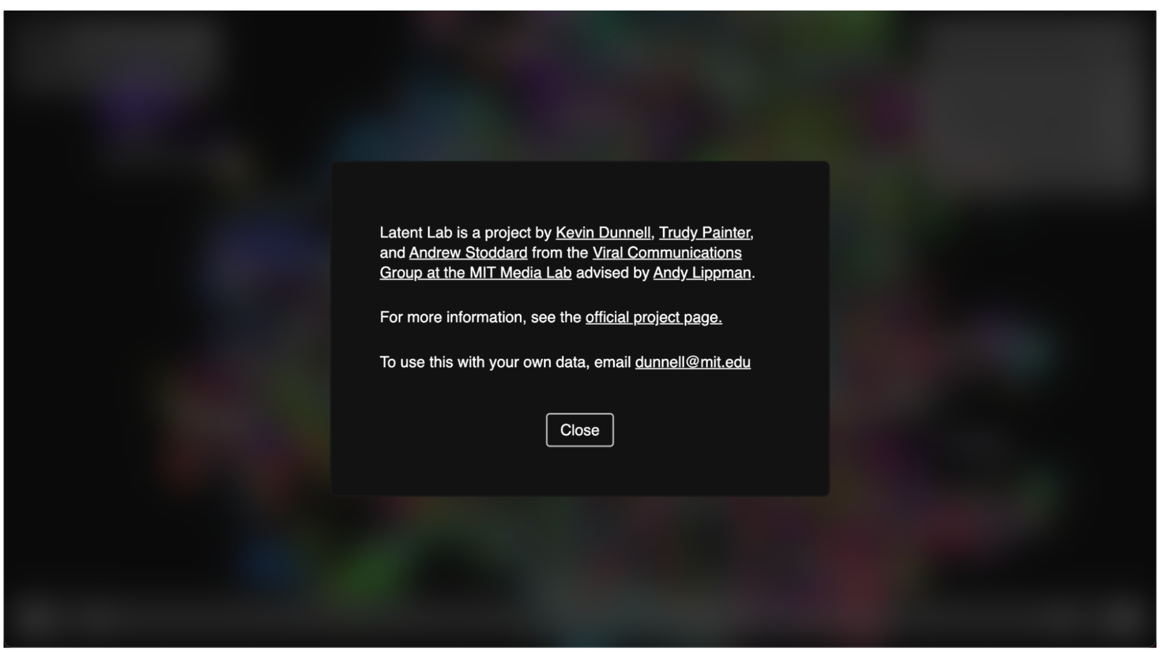


Figure 5-12: Latent Lab Information

has three main steps: vector embeddings computation, UMAP dimensionality reduction, and topic extraction.

Vector Embeddings

Embeddings are n-dimensional vector representations machine learning models use to understand input meaning. Similar inputs have closer embeddings in the latent space. OpenAI's large language model "GPT-3-Ada" provides an API endpoint to obtain embeddings from input text. Each project is assigned a 1536-value vector that corresponds to the project description.

UMAP Dimensionality Reduction

To explore a two-dimensional similarity map of project information, UMAP dimensionality reduction is applied to the 1536-dimensional vector embeddings. UMAP is preferred over techniques like PCA, as it preserves project clusters. It is also preferred to TSNE because it offers the ability to save a model, whereas TSNE is a resulting probabilistic relationship between data points with no returned model [18]. This saved model is necessary for transforming the embeddings of search requests at runtime. After UMAP, each project is assigned two values (x, y) for representation on the topic map.

Topic Extraction

Automated topic extraction differentiates Latent Lab from other embedding visualizations that lack cluster meaning insights. Instead of manually examining elements in a cluster, Latent Lab automates this process. Topics are extracted by running each project through a GPT-3 prompt, generating a list of topics for each project. For each unique topic label, the number of occurrences and associated projects are determined for each unique topic label. The label's location is then calculated by finding the centroid of the (x, y) UMAP-reduced coordinates for each associated topic.

Chapter 6

Evaluation

6.1 Latent Lab User Study

A pilot user study was initially conducted with a small group (n=8) of Media Lab students. The purpose was to evaluate the effectiveness of the survey structure and estimate the time participants might need to complete it. Upon completion and a few tweaks discussed in section 6.2, we scaled the survey to a larger pool of participants (n=94) using Prolific. The study compares Latent Lab's search functionality with existing methods for searching the Media Lab database of research projects. Metrics similar to those used in prior HCI studies of Human-Centered AI systems were considered in evaluating the system's ability to support participants' understanding of and discovery within an unfamiliar large knowledge base and the participants' attitudes towards the system [5].

6.1.1 User Study Design

Ninety-four US-based self-identified researchers participated in the user study. Each user was first asked if they had any affiliation with the MIT Media Lab, to which only 2 participants responded with a "Yes". They were also asked to respond on a 5-choice Likert scale to the question. "How familiar are you with ongoing and past research at the MIT Media

Lab?" A large majority of the participants (71) responded with a 1 (Not at all familiar), and none responded with a 5 (Very familiar). Participants were told they would test two versions of an MIT Media Lab search tool to search for an idea in their own area of interest. They were asked to search for the same idea with each tool and answer a set of questions after using each tool, as well as a final question comparing the tools at the end of the survey. Participants recorded the idea they were attempting to search for before using the existing project search page of the Media Lab website and Latent Lab in a randomized order.

6.1.2 Measurements

All participants were asked to respond to the questions below on a 5-point Likert scale.

- Clarity: Participants answered, *"How well do you understand how the system arrived at these results?"*
- Workload: The following dimensions of the NASA-TLX [12] were considered:
  - Effort: Participants answered the question, *"How much effort did you have to exert to achieve these search results?"*
  - Frustration: Participants answered a series of questions:
    - *"On a scale of irritated to content, how did you feel during the task?"*
    - *"On a scale of stressed to relaxed, how did you feel during the task?"*
    - *"On a scale of annoyed to complacent, how did you feel during the task?"*
- Engagement: Participants answered, *"How engaged were you during this task?"*
- Future Use: Participants were asked to rate the level to which they agreed with the statement, *"I would continue using this system if I could use it to access the knowledge base of my own institution/company/organization."*
- Insight: Participants were asked to rate the level to which they agreed with the statement, *"The system helped me gain a deeper understanding of the research at the MIT Media Lab."*

- Mental Support: Participants were asked to rate the level to which they agreed with the statement, *"The search tool helped me think about related ideas and search terms."*

- Trust: Mayer's dimensions of trust [17] were considered in the trust measurement. Particularly, capability, benevolence, and reliability were examined as they are key dimensions of trust.

- Overall Preference: Participants were asked, *"Which search tool did you prefer overall?"* and were offered two options, Version A and Version B.

In the study, we referred to the Media Lab website as Version A and Latent Lab as Version B to avoid biasing participants. Users tested the Media Lab website on the Project Search Page[1], focusing on research projects for a balanced comparison with Latent Lab.

6.2 Pilot Study Results

The pilot user study provided valuable insights into Latent Lab's user experience and effectiveness compared to existing search methods for the Media Lab database. The preliminary study involved eight participants in a between-subject design, where each participant tested only one system (four tested the Media Lab website, and four tested Latent Lab) to refine the survey questions and overall format.

The pilot study results indicate that participants had varying experiences when using Latent Lab and the Media Lab website to search for their ideas of interest (Figure 6-1). In general, participants found search results to be related or somewhat related to their search terms for both systems. Notably, participants reported a significantly higher degree of understanding regarding how Latent Lab arrived at search results compared to how they understood the Media Lab website returned search results. Additionally, participants experienced a significantly reduced workload when browsing with Latent Lab compared to using the Media Lab website.

[1] https://www.media.mit.edu/search/?filter=project

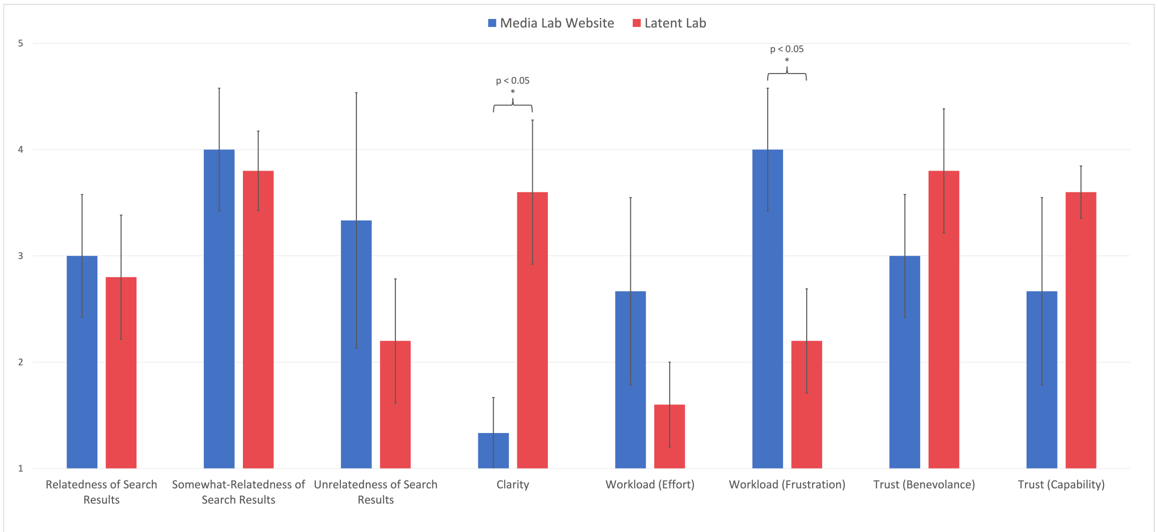


Figure 6-1: Pilot Results

During the study, it was observed that participants frequently encountered confusion regarding three questions assessing the relatedness of results to the search term, which had a 1-5 rating scale (strongly disagree to strongly agree) for unrelated, somewhat related, and related results. This led to consolidating these questions into a single question evaluating the degree of relatedness. Additionally, the framing of the benevolence question was confusing for participants, prompting an alteration and the inclusion of the "reliability" dimension of trust [17].

Based on these findings, the survey was revised, and the sample size was increased, with Prolific used as the recruitment platform. The study design was also changed from a between-subject to a within-subject study, allowing participants to test and compare both systems. This modification was made to control individual differences among participants, increase statistical power, and better assess the two systems' relative effectiveness and user experience of the two systems.

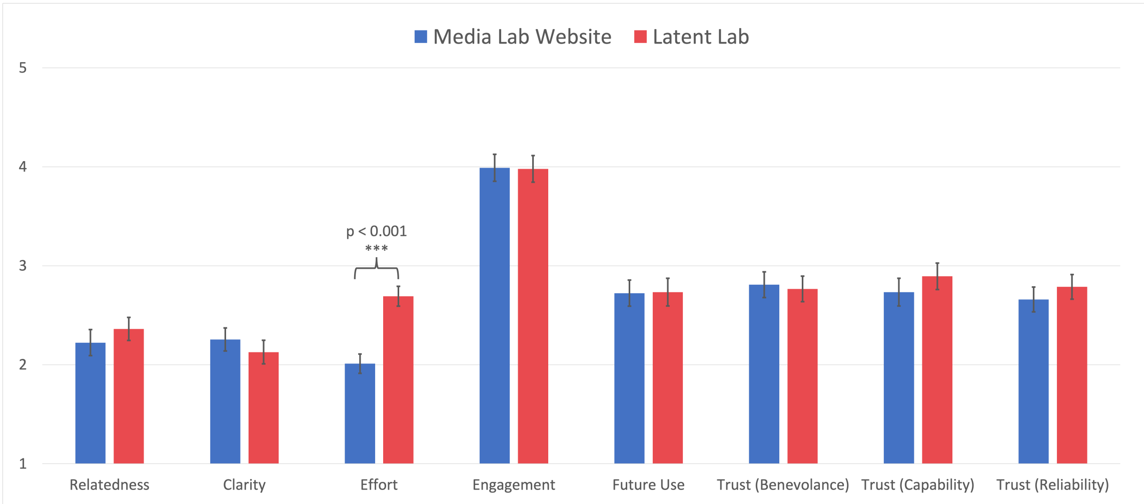


Figure 6-2: Latent Lab v. Media Lab Website: Clarity, Effort, Engagement, Future Use, and Trust

6.3 User Study Results

The scaled-up user study with Prolific users provided valuable insights into the performance of Latent Lab in comparison with the current Media Lab website project search page. In general, both tools received similar ratings regarding how related search results were to the provided search terms (Figure 6-2). However, some differences were observed in specific aspects of the user experience.

Regarding mental support and insight, Latent Lab outperformed the Media Lab website with statistical significance (Figure 6-3). Users reported that Latent Lab helped them think about related ideas and search terms more effectively and provided a deeper understanding of the research at the MIT Media Lab. This shows that Latent Lab's approach to search facilitates greater exploration and discovery for users.

When considering clarity, effort, engagement, future use, and trust (benevolence, capability, and reliability), both tools mostly performed similarly. Despite a statistically significant difference in the effort required for Latent Lab, there were no statistically significant differences in any of these metrics between Latent Lab and the Media Lab website. This suggests that Latent Lab, an unfamiliar exploration tool, is essentially on par with the existing search

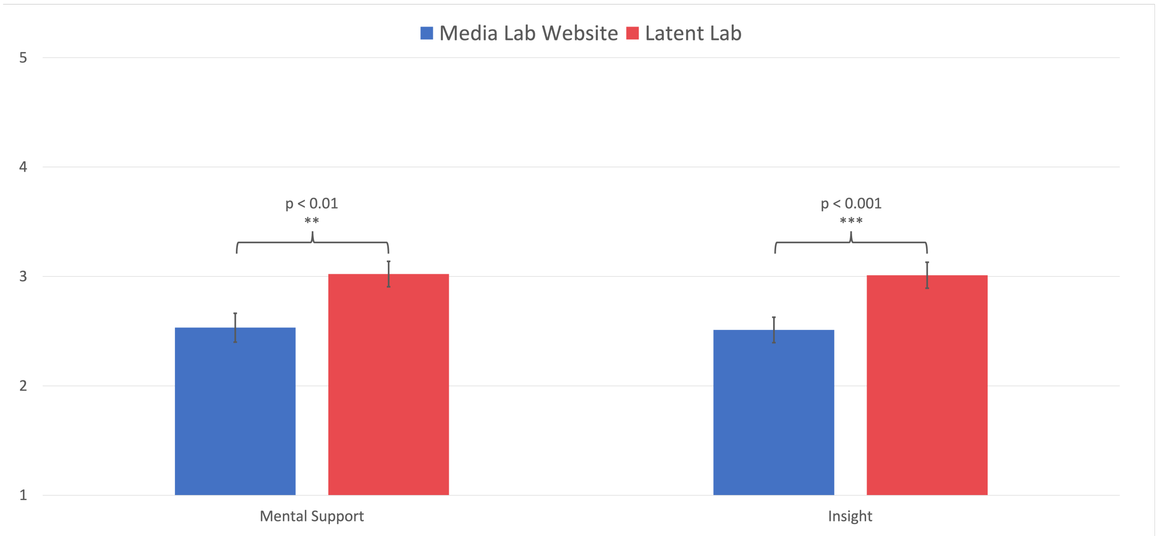


Figure 6-3: Latent Lab v. Media Lab Website: Mental Support and Insight

tool, which we found encouraging.

Regarding overall user preference, the results were nearly evenly split between the Media Lab Website (45 votes) and Latent Lab (49 votes) (Figure 6-4). This indicates a roughly 10% greater preference for Latent Lab. While Latent Lab may offer some advantages regarding mental support and insight, users may still be divided in their preference for one tool over the other. Further refinements to Latent Lab may help improve its appeal to a broader user base and solidify its position as the preferred tool.

Overall, users found both tools to be useful, but there were differences in their preferences and experiences. For the Media Lab website, users appreciated its ease of use and responsiveness, with some users noting that it was quick and provided relevant results, such as the user who searched for "AI" and found a project on "Creative AI: A curriculum around creativity, generative AI, and ethics." However, there were instances where users received no results for their queries, such as the user who searched for "The history of calendars" and received no results. Some users felt that the tool was limited in its ability to explore related topics, as seen in the feedback for "Historical archival efforts in video games," where the user noted that the tool only gave one search result.

For Latent Lab, users were intrigued by its visual interface and ability to group related

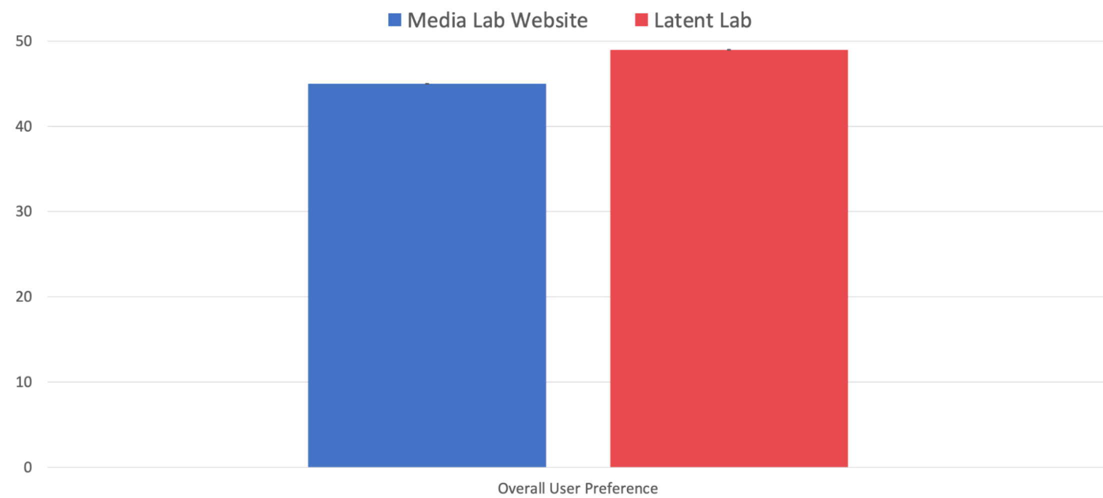


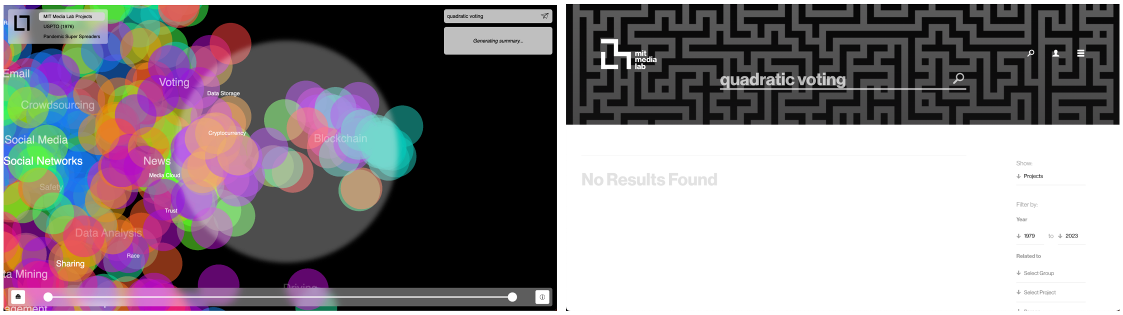
Figure 6-4: Latent Lab v. Media Lab Website: Overall User Preference

projects. Some users found it helpful for exploring related fields, as exemplified by the user who searched for "prosocial behavior" and found projects in a related field, leading them to think of new ideas. However, the tool was also described as slow, laggy, and less intuitive, with some users expressing frustration with the need to hover over individual nodes to see project details, such as the user who searched for "Does fare-free transit boost ridership?" and found the tool "slow and clunky." Additionally, some users found the interface confusing and challenging to understand, as seen in the feedback for "AI Psychology," where the user noted that the system was "harder to utilize."

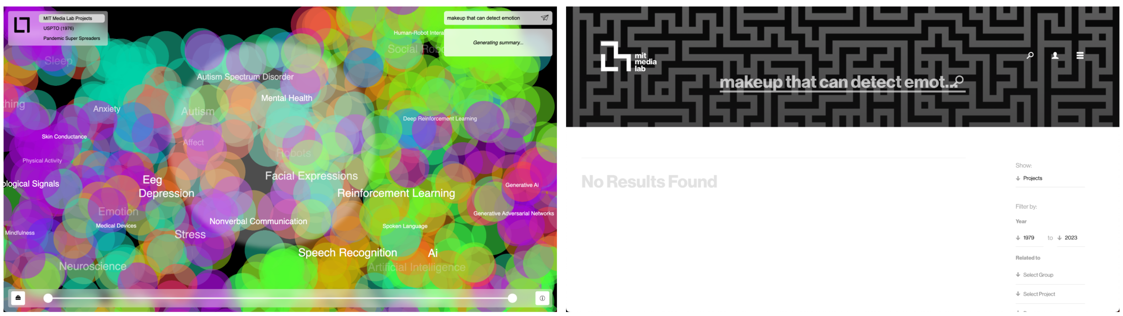
6.4 Searching Complex Ideas

Latent Lab V3.0's embedding-based search outperforms the keyword-based search method used in the existing Media Lab website when searching for complex concepts with multiple words. The two examples in Figure 6-5 illustrate this with the search terms "quadratic voting" and "makeup that can detect emotion."

Both Latent Lab and the Media Lab website can handle simpler, one or two-word search terms quite well. The example in Figure 6-6 compares searching "Artificial Intelligence" in



(a) Quadratic Voting



(b) Makeup That Can Detect Emotion

Figure 6-5: Latent Lab v. Media Lab Website: Complex Search Terms

both tools.

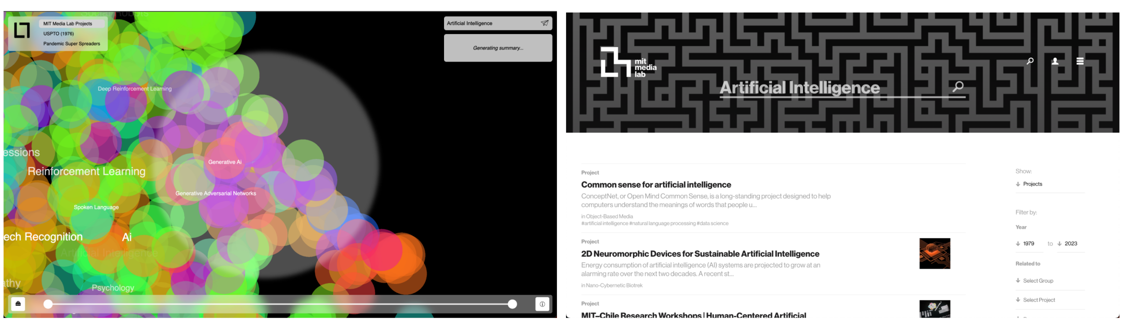


Figure 6-6: Latent Lab v. Media Lab Website: Keyword Search (Artificial Intelligence)

However, when users know the exact project they are looking for, Latent Lab breaks down, and the Media Lab website outperforms our tool. Searching "Latent Lab" on the Media Lab website will identify and return the project pages corresponding to this work. While Latent Lab cannot "find itself," it poetically takes us to an area of the knowledge landscape that centers around "MIT Media Lab" (Figure 6-7).

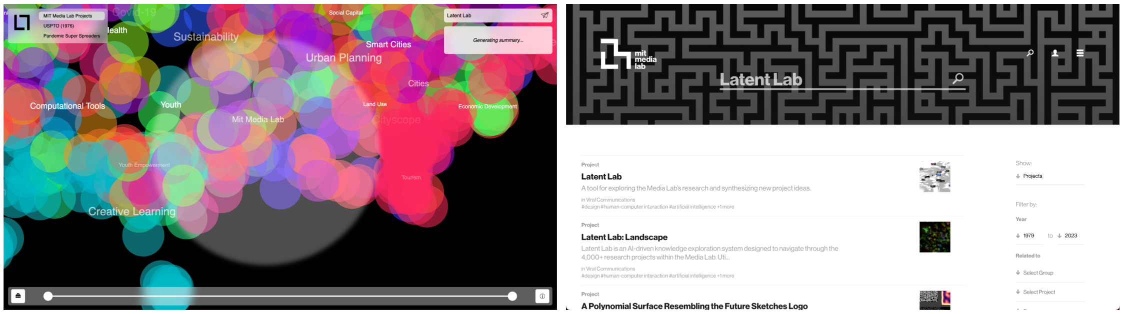


Figure 6-7: Latent Lab v. Media Lab Website: Project-Specific Search (Latent Lab)

The user study also supports this variance depending on the complexity of the search terms. Specifically, Latent Lab appears to perform well when users input complex search terms or longer phrases, while it may struggle with more exact project names or single-word keyword-based searches. For example, a user who searched for the multi-word phrase “Cognitive Decline and Speech Perception Interactions” in Latent Lab provided favorable feedback, stating, “I love this tool,” despite noting some areas for improvement in the interface. In contrast, a user who searched for the single-word term “Dermatology” found a relevant project, “Smartphone dermatoscope,” on the Media Lab Website but did not find a specific relevant result in Latent Lab. The user’s feedback for Latent Lab suggested that the tool could show unrelated information due to its map-like format. These observations indicate that Latent Lab may excel at handling complex queries where the relationships between topics are essential. However, it may benefit from improvements in handling more precise keyword-based searches. One potential improvement to Latent Lab could be to include a drop-down section in the search interface that provides keyword matches, similar to the functionality observed on the Media Lab website. This enhancement could improve the user experience, particularly for those searching for specific projects or topics using exact keywords.

6.5 Content Synthesis Results

As outlined in Chapters 3 and 4, Latent Lab was initially designed to assist in the co-creation and synthesis of high-impact research ideas. Due to time constraints and user

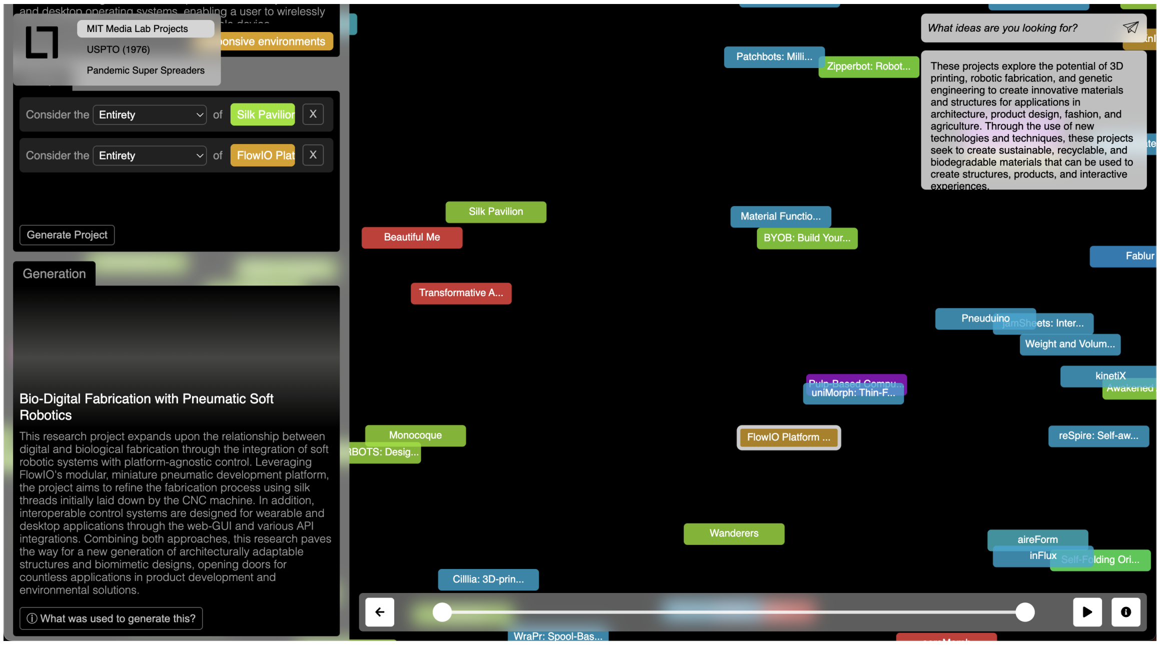


Figure 6-8: Entirety of "Silk Pavilion"[2] + "FlowIO Platform for Soft Robotics and Programmable Materials"[3]

feedback indicating a preference for exploring project relationships, our testing focus shifted. Despite this, we have qualitatively noted that the latest version of Latent Lab excels at generating new project ideas over its predecessors, adeptly integrating complete projects or specific elements such as problem statements, technologies, and associated communities from existing projects. Below are a few examples that showcase this capability in Latent Lab V3.0.

Figure 6-8 demonstrates Latent Lab's ability to combine the entirety of two distinct projects clustered in the bio-focused materials region of the Media Lab research project map. The proposed project, "Bio-Digital Hybrid Pneumatic Soft Robotics," adeptly blends the Silk Pavilion's bio-digital fabrication with FlowIO's soft robotics and programmable materials. This combination harnesses the strengths of both projects, synthesizing a potentially interesting launchpad for ideation at the intersection of soft robotics, bio-inspired structures, and programmable materials. Unlike previous iterations, Latent Lab V3.0 utilizes LLM prompting for project interpolation, moving beyond direct embedding manipulation. Syn-

[2] https://www.media.mit.edu/projects/silk-pavilion/overview/
[3] https://www.media.mit.edu/projects/flowio/overview/

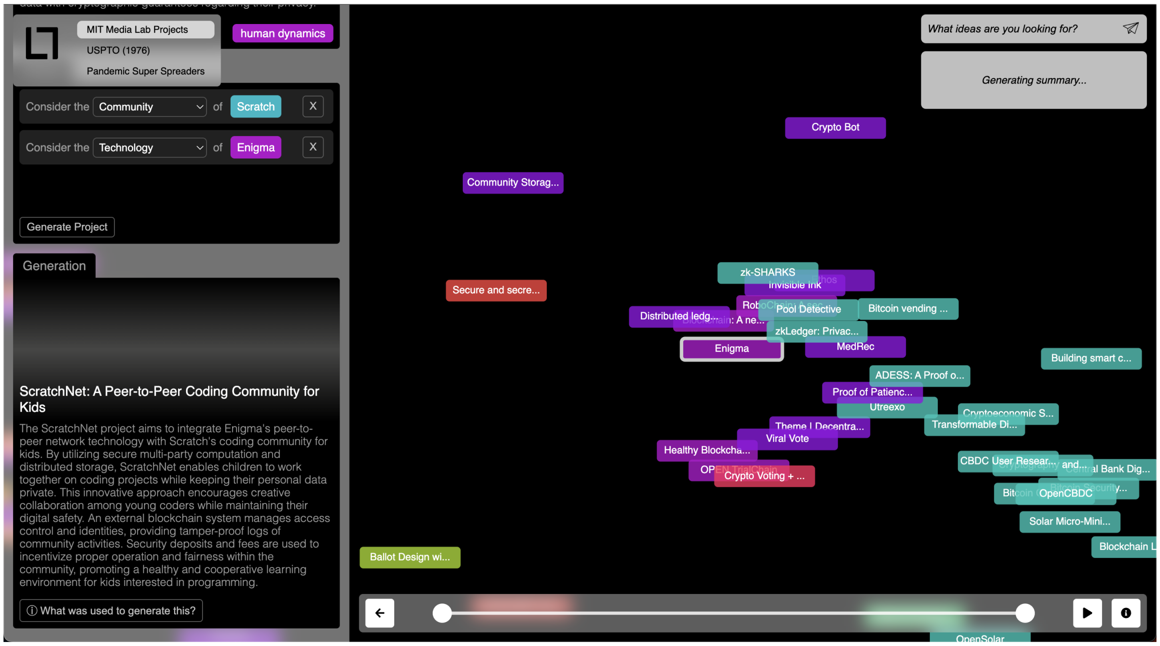
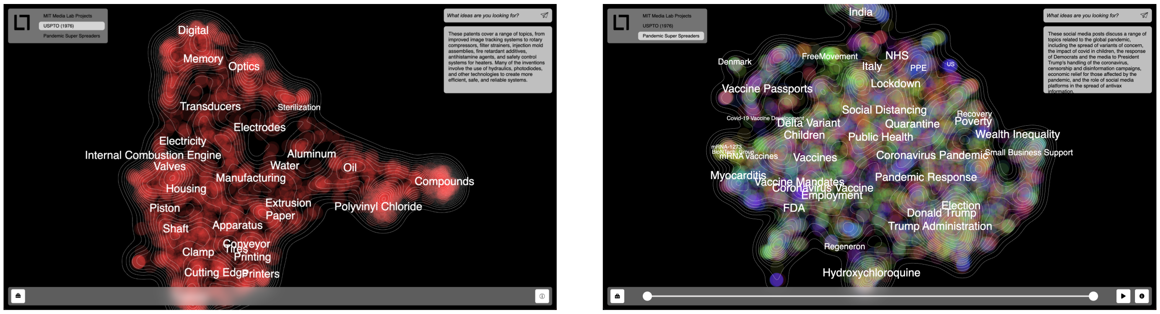


Figure 6-9: Community of "Scratch"[4] + Technology of "Engima"[5]

thesis at this higher level of abstraction enables consideration of specific project elements, as opposed to the exclusive whole-project focus of past versions.

As illustrated in Figure 6-9, the successful synthesis of Scratch and Enigma into the novel "ScratchNet" project underscores the promise of Latent Lab in facilitating the ideation process. By effectively merging Scratch's children-centric coding community and Enigma's secure peer-to-peer network technology, Latent Lab showcases its potential in generating new, innovative, and potentially impactful ideas. However, challenges, such as the child-friendly implementation of advanced technology and balancing security with accessibility, highlight areas for future research. The example of ScratchNet supports the assertion that Latent Lab can serve as an effective tool for concept development and ideation.

[4] https://www.media.mit.edu/projects/scratch/overview/
[5] https://www.media.mit.edu/projects/enigma/overview/



(a) U.S. Patents (1976) (b) COVID-19 Social Media Posts

Figure 6-10: Latent Lab Supporting Additional Data Sets

6.6 Exploring Beyond Media Lab Data

Latent Lab has proven its proficiency as a general-purpose exploration tool by effectively extending its capabilities to accommodate data sets beyond the scope of the Media Lab's collection. Demonstrating its adaptability, Latent Lab was seamlessly applied to additional data sets, including a subset of the USPTO data set and "Pandemic Super Spreaders," a collection of social media posts across Twitter, Facebook, and Instagram, to understand the spread of misinformation during the COVID-19 pandemic.

An intriguing finding from this effort is Latent Lab's capacity to reveal outlier data, which the researchers who aggregated the data had not identified beforehand. For example, several posts about "Dwayne Johnson" were included due to the team filtering by "Johnson & Johnson" to find posts regarding the company's COVID-19 vaccine. This unanticipated aspect of Latent Lab highlights its ability to detect irregularities within a data set, thus broadening its range of utility. A similar observation was made in the Media Lab data set context. A distinct cluster, referred to by us as the "island of misfits," materialized at the bottom right of the central cluster. This group was comprised of test projects and project pages that were either empty or minimally completed, further emphasizing Latent Lab's potential as an all-encompassing exploration tool.

Chapter 7

Discussion

7.1 Collaboration with Media Lab Member Companies

Several Media Lab member companies, including Dell, Deloitte, Comcast, L'Oréal, MITRE, IDEO, 3M, USGA, Walmart, and Hyundai, expressed interest in using Latent Lab for visualizing their own data sets. Conversations with these companies revealed that a common challenge is aggregating and homogenizing data from different sources is a common challenge. Latent Lab's underlying large language models may offer a solution by dynamically bridging varying data sets and standardizing formats, which we are currently investigating.

7.2 Limitations

Latent Lab, while promising, has areas for improvement, namely in content synthesis and user experience enhancement. As detailed in Chapter 4, user feedback emphasized the need for improved content organization. The current Latent Lab iteration supports idea generation through GUI-based LLM prompting, but its effectiveness in promoting knowledge exploration or improvement over text-based prompting hasn't been quantitatively assessed. Time constraints precluded a study on the user experience for generating and evaluating novel research projects. Our findings indicate that Latent Lab V3.0 may require more effort than the Media Lab website. The lack of a tutorial in our user study survey may have hindered participant understanding and increased reported effort. Regrettably, no user session logging data was captured for Latent Lab, a missed opportunity for further usage insight and workflow improvement.

7.3 Future Work

Our work on Latent Lab has only scratched the surface of defining an ideal future human-AI co-invention system. Further research could investigate Media Lab researchers' experiences with Latent Lab, specifically examining how domain experts interact with the platform. Whereas our user study focused on participants unfamiliar with the data set being explored, these future tests could explore how domain experts might see a familiar data set in a new way. This could involve students, researchers, and faculty in assessing whether Latent Lab is more effective as an internal tool than the Media Lab website for discovering related projects, fostering collaborations, and understanding research trends over time.

Optimizing the front and back end is essential, as the current system supports only approximately 10,000 data points before experiencing lag. Implementing sampling techniques and rendering only a subset of the data could scale rendering capabilities to support tens of millions of data points. This would accommodate numerous companies and additional research groups interested in importing their data sets for exploration. Using LLMs to handle arbitrary data, which we've referred to as a "driverless driver," could facilitate data ingestion from various structures and formats.

The release of GPT-4 and other models with multi-modal support and external data stream plugins makes developing a co-inventor system more achievable. Future research could focus on Latent Lab's generation capabilities, enabling them to connect with outside tools like Wolfram Alpha or the open internet to further ground generated ideas in the physical constraints of reality. Ultimately, we aim to further evolve this work towards Brett Victor's vision of a "Seeing Space," a physical environment designed to deepen our understanding of the complex behaviors of our inventions [29]. To gauge the system's effectiveness as a tool for human-AI co-invention, we plan to engage with research students from the Media Lab, across MIT, and from other institutions in user studies.

We also aim to enhance support for visualizing disparate data sets within the same view to expose insights from new juxtapositions. Several companies share a similar interest in team formation for new projects, considering talent exposed externally or in a different context of their organizations. For example, professional resume data could be imported from a separate repository and combined with project data, allowing for a joint visualization aid in identifying the ideal team for a newly generated project idea. Similarly, product usage reviews and relevant scholarly research could supplement patent information, providing a holistic view of patent impact. This integrated perspective could inform the creation of high-impact patent applications.

In the course of our work with Latent Lab, we have been inspired to explore a derivative tool for visualizing individual data in a digestible and meaningful way, fostering consumer awareness of data privacy and the value of personal data. This aspiration aligns with our larger goal of human-AI co-invention and recognizes that the democratization of AI is inseparable from data ownership. We envisage a tool reminiscent of Spotify's year-end review but applied to disparate personal data sources capable of generating regular insights. Such a tool could underscore the value of the information we routinely create, promoting data ownership. Looking ahead, these possibilities guide our journey toward empowering individuals to become active stewards of their own data.

Chapter 8

Conclusion

This thesis demonstrates the development and potential of Latent Lab as a powerful and innovative exploration tool, enabling users to delve deeper into the interconnected relationships within large data sets. By combining state-of-the-art AI technologies, such as LLMs, with visually engaging and interactive interfaces, Latent Lab transcends the limitations of conventional search methods, offering a more semantically meaningful and context-aware exploration experience. Compared to existing tools, Latent Lab's positive reception and performance in demonstrations and user studies validate the significance of the interaction itself in fostering creativity and understanding within large knowledge bases.

The Latent Lab development journey has emphasized the importance of exploration and iterative design that the tool embodies. Latent Lab represents a return to what experts in information technology have suggested and strived for but has never quite materialized at scale - a benevolent richness of interconnected information that is intuitively accessible. By focusing on the exploration process with the assistance of AI, Latent Lab empowers users to break free from the constraints of the traditional rigidity of digital tools. The success of Latent Lab provides a solid foundation for future research and development of human-AI co-invention systems, ultimately paving the way for more intuitive and effective collaborations between humans and AI that yield novel and impactful creations.

Bibliography

[1] G. Altshuller. *The Innovation Algorithm: TRIZ, Systematic Innovation, and Technical Creativity*. Technical Innovation Center, Inc., 1999.

[2] M. A. Boden. Creativity in a nutshell. *Think*, 5(15):83–96, 2007.

[3] M. Bostock, V. Ogievetsky, and J. Heer. D^3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[4] Vannevar Bush. As we may think. *Atlantic Monthly*, 176:101–108, July 1945.

[5] C.J. Cai, E. Reif, N. Hegde, J.D. Hipp, B. Kim, D. Smilkov, M. Wattenberg, F.B. Viégas, G.S. Corrado, M.C. Stumpe, and M. Terry. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.

[6] N. Carr. Freebase: the web 3.0 machine, March 2007.

[7] Martin Curley and Bror Salmelin. Open innovation 2.0: a new paradigm. *OISPG White Paper*, pages 1–12, 2013.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[9] Ziv Epstein, Hope Schroeder, and Dava Newman. When happy accidents spark creativity: Bringing collaborative speculation to life with generative ai, 2022.

[10] Patrick Esser et al. Taming transformers for high-resolution image synthesis. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12868–12878, 2020.

[11] R. P. 1. Feynman, M. A. Gottlieb, R. Leighton, and M. L. Sands. *Feynman's tips on physics: reflections, advice, insights, practice : a problem-solving supplement to the Feynman lectures on physics*. Basic Books, 2013.

[12] Sandra G Hart and Lowe E Staveland. *Development of NASA TLX (Task Load Index): Results of empirical and theoretical research*, volume 52, pages 139–183. Elsevier, 1988.

[13] M. A. Hearst. *Search user interfaces*. Cambridge University Press, 2009.

[14] J. Heer and B. Shneiderman. Interactive dynamics for visual analysis. *Queue*, 10(2):30:30–30:55, 2012.

[15] Jaron Lanier. There is no a.i. *The New Yorker*, April 20 2023.

[16] Chunyuan Li et al. Optimus: Organizing sentences via pre-trained modeling of a latent space. *arXiv preprint arXiv:2004.04092*, 2020.

[17] Roger C. Mayer, James H. Davis, and F. David Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995.

[18] L. McInnes and J. Healy. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[19] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[20] M. Minsky. *The Society of Mind*. Simon Schuster, 1988.

[21] M. Minsky. Logical vs. analogical or symbolic vs. connectionist or neat vs. scruffy. *AI Magazine*, Reprinted in AI Magazine, Summer 1991, 1990.

[22] OpenAI. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[23] G. Polya. *How to Solve It: A New Aspect of Mathematical Method*. Princeton University Press, 1945.

[24] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages, VL '96*, pages 336–343. IEEE Computer Society, 1996.

[25] D. Smilkov, N. Thorat, C. Nicholson, E. Reif, F.B. Viégas, and M. Wattenberg. Embedding projector: Interactive visualization and interpretation of embeddings. *arXiv preprint arXiv:1611.05469*, 2016.

[26] J. Speiser. Worldlens: Exploring world events through media. Technical report, MIT Media Lab, 2014.

[27] K. Spärck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28:11–21, 1972.

[28] L. Tu, C. Xiong, and Y. Zhou. Prompt-tuning can be much better than fine-tuning on cross-lingual understanding with multilingual language models. *arXiv preprint arXiv:2210.12360*, 2021.

[29] Bret Victor. Seeing spaces. `https://www.youtube.com/watch?v=klTjiXjqHrQ`, 2014.

[30] D. Wang, E. Churchill, P. Maes, X. Fan, B. Shneiderman, Y. Shi, and Q. Wang. From human-human collaboration to human-ai collaboration: Designing ai systems that can work together with people. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, 2020.

[31] J. W. Weis and J. Jacobson. Using machine learning to predict high-impact research. *Nature Biotechnology*, 2021.

[32] A. Wexelblat and P. Maes. Footprints: History-rich tools for information foraging. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 270–277, May 1999.

[33] P. Yanardag et al. How to generate (almost) anything. n.d.