Atlas of Human-AI Interaction (v1)

PAT PATARANUTAPORN*, MIT Media Lab, Massachusetts Institute of Technology, USA
CHAYAPATR ARCHIWARANGUPROK*, Creatorsgarten, Thailand
PATTIE MAES, MIT Media Lab, Massachusetts Institute of Technology, USA

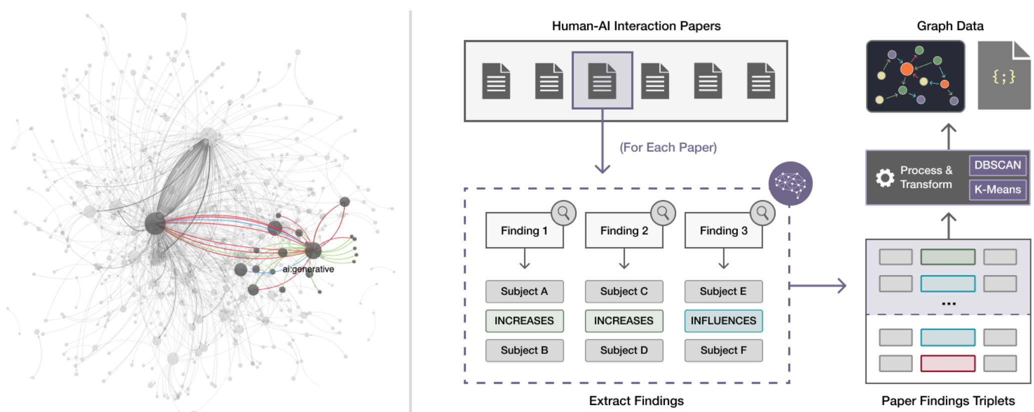

Fig. 1. (Left to Right) An Impact Graph of Human-AI Interaction Paper Focusing on the Node "AI:Generative," and a Process for Create an Impact Graph.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; • **Computing methodologies** → **Artificial intelligence**; • **Information systems** → **Information retrieval**.

Additional Key Words and Phrases: Human-AI Interaction, Research Findings, Research Network Analysis

1 Introduction

The rapid integration of artificial intelligence into everyday life has created an urgent need to understand its varied impacts on human experience and social structures. While numerous studies have examined specific instances of human-AI interaction, we lack a comprehensive understanding of how these diverse findings connect and relate to one another across the broader landscape of human-computer interaction research. This paper presents an innovative approach to synthesizing and mapping the complex web of AI's impacts on human life through an analysis of HCI papers using LLMs as analytical tools to identify and connect empirical findings.

Our work moves beyond traditional literature reviews that typically organize papers by themes or topics. Instead, we develop what we term an "impact atlas" - a novel framework that traces connections between papers based on the similarity and relationships between their empirical findings about AI's effects on human experience. This approach allows us to identify patterns in how AI systems influence people's lives across different contexts, revealing both beneficial outcomes and potential risks that might not be apparent when examining individual studies in isolation.

The contributions of this work are threefold:

*Both authors contributed equally to the paper

Authors' Contact Information: Pat Pataranutaporn, MIT Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, patpat@media.mit.edu; Chayapatr Archiwaranguprok, Creatorsgarten, Bangkok, Thailand, chayapatr.arc@gmail.com; Pattie Maes, MIT Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, pattie@media.mit.edu.

(1) A novel methodological framework for using LLMs to analyze and synthesize findings across large bodies of HCI literature
(2) An "impact atlas" that maps the landscape of AI's effects on human experience, revealing previously unidentified patterns and relationships
(3) A set of evidence-based principles for designing AI systems that promote positive outcomes while mitigating potential harms

This atlas provides researchers and practitioners with a new tool for understanding the complex dynamics of human-AI interaction, while also demonstrating the potential of LLMs as research synthesis tools. By mapping the connections between empirical findings rather than just thematic similarities, we offer a more nuanced and actionable understanding of how AI shapes human experience across different contexts.

2 Background and Related Works

2.1 Research Synthesis Methods in HCI

Research synthesis methods in HCI represent a fundamental area of related work for our atlas approach. Traditional systematic literature reviews and meta-analyses have long served as primary tools for understanding the landscape of human-computer interaction research. Researchers have employed various scientometric approaches to map research landscapes, often utilizing computational methods such as topic modeling and bibliometric analysis to process large volumes of academic literature. However, these approaches typically focus on clustering papers by topics or keywords rather than analyzing the relationships between empirical findings. This limitation has made it difficult to understand how different research outcomes relate to and influence each other across the field, particularly in the rapidly evolving domain of human-AI interaction.

2.2 Knowledge Graph Construction from Academic Literature

The use of Large Language Models (LLMs) for scientific literature analysis represents a more recent but rapidly growing body of related work. Researchers have begun exploring applications of LLMs in scientific text processing, developing sophisticated methods for extracting structured information from academic papers. This includes work on prompt engineering for scientific tasks and approaches to ensuring the reliability of LLM-based analysis. While these efforts have shown promising results in tasks like paper summarization and information extraction, they have not yet been fully applied to the challenge of synthesizing research findings across large bodies of literature. Current work in this area has primarily focused on individual paper analysis rather than creating comprehensive maps of research findings and their relationships.

3 Design and Implementation

Our methodology for creating the impact atlas follows a systematic approach to extract, process, and visualize research findings from academic literature. The process consists of five main stages: (1) gathering research abstracts, (2) extracting findings triplets, (3) processing triplets, (4) semantic clustering, and (5) graph transformation. Each stage was designed to progressively refine and structure the research findings while maintaining their semantic relationships.

3.1 Gather Research Abstracts

We collected research abstracts through a targeted query of "human-ai interaction" in the ACM Digital Library. This specific query was chosen to maintain focus on direct human-AI interaction findings rather than broader human-computer interaction or artificial intelligence topics. We applied several filtering criteria to the initial results. First, we restricted our collection to papers categorized

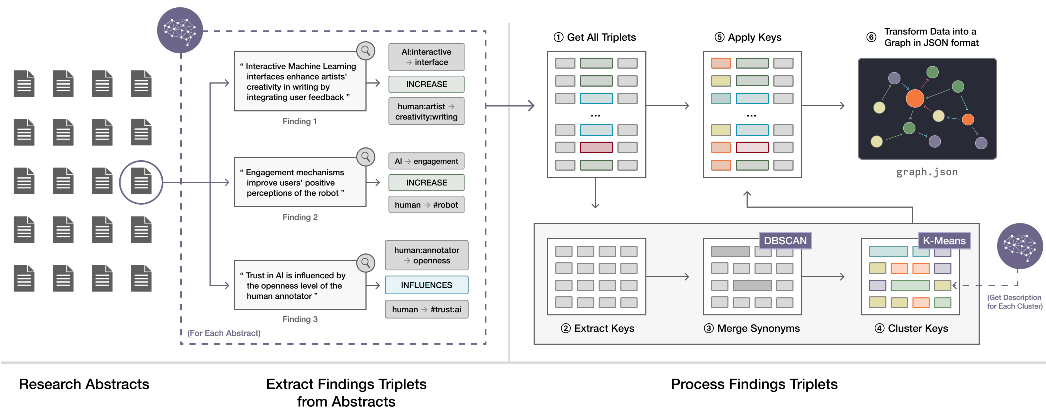

Fig. 2. Process to Create an Impact Graph based on Paper Data.

as extended abstracts, research articles, works in progress, posters, and short papers. Additionally, we only included papers that had corresponding abstracts available in Semantic Scholar. The data collection was performed on October 27, 2024, resulting in 749 papers.

3.2 Extract Findings Triplets from Abstracts

The extraction process leverages GPT-4's natural language understanding capabilities through a two-stage pipeline. First, we extract the findings of each papers from its abstract. each paper may have several findings. If no empirical findings are extracted we instructed GPT-4 to explain the reason. Then, for each extracted findings, we instruct the model to transform the natural language findings into a triplet.

3.2.1 Extract Findings from Each Abstract. : The prompt instructs the model to filter abstracts for concrete findings with the involvement of the concerned parties, that is, the interaction between human and AI system and their results. Each finding extracted must be in the form of a clear subject-predicate-object structure. For instance, "Interactive Machine Learning interfaces enhance artists' creativity" would be a key finding.

As a result, 75.83% (568 papers) contained extractable empirical findings, while 24.17% (181 papers) did not present direct findings due to their nature. The papers without explicit findings primarily consisted of conceptual frameworks (99 papers), workshop announcements (47 papers), systematic and scoping reviews (14 papers), and other types, such as design methodology papers (9 papers) and position papers (3 papers). With one paper may have multiple findings, the extraction results in 1077 findings.

3.2.2 Convert Each Finding into a Triplet. : To construct more structural information, GPT-4 is then utilized to transform each finding into a structured triplet in the form `[cause, relationship, effect]`. The procedure is described below.

- **Subject (Cause/Effect) Classification**: To avoid scope ambiguity, elements are categorized into Human (e.g., `human:student`), AI (e.g., `ai:generative`), or Concepts/Objects (CO) for abstract ideas that can be encapsulated within the first two types (e.g., `co:collaboration`, `co:justice`). In addition, several causes and effects do not occur by the agency of the subject but rather their characteristic or features; we further categorize them using the pipe ("|") symbol. For example, Anthropomorphic AI explainability become `ai:anthropomorphic|explainability`.

- **Relationship Standardization**: Each relationship between subjects is normalized into three types:

```
- INCREASES: Direct positive impact on measurable attributes
- DECREASES: Direct negative impact on measurable attributes
- INFLUENCES: Complex or indirect effects on behavior/perception
```

- **Feature Normalization**: The prompt handles special cases through standardized notation:

```
- Perception markers: In several cases, there are nuances between the feature and the
    subject's idea or perception of the feature (e.g. the differences between learner's
    efficiency and learner's perception of efficiency). Thus, we use a "\#" prefix as our
    standard notation (e.g., human's perception of trust -> human:#trust)
```

For instance, "Engagement mechanisms improve users' positive perceptions of the robot" becomes `[ai|engagement, INCREASES, human:\#robot]`, capturing both the causal relationship and the perceptual nature of the effect.

3.3 Process Findings Triplets

For each triplet being generated through separated prompts, there are several differences in the wording. We employ several methods to merge and cluster the subjects to improve interoperability. We first construct a set of unique keys from the triplets. Each key represents a subject-feature pair that appears in either cause or effect positions. For instance, from the triplet `[human:expert|knowledge, INFLUENCES, ai|performance]`, we extract `"human:expert|knowledge"` and `"ai|performance"` as distinct keys. This approach accounts for the bidirectional nature of relationships where elements can serve as causes and effects across different findings. We then generate an embedding of each key using OpenAI's `text-embedding-3-large`, resulting in a 3096-dimensional vector for each key.

3.4 Merge Synonyms

The synonym merging phase addresses semantic redundancy in the extracted keys through density-based clustering. The synonym merging phase addresses semantic redundancy in the extracted keys through density-based clustering. We employ HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) on the embeddings, configured with a high epsilon value to capture only a cluster with very high similarity. Using cosine similarity as the distance metric enables the identification of semantically equivalent terms while accounting for variations in terminology. For each identified cluster, we select a canonical representative by first computing the cluster's centroid through mean embedding aggregation. The representative term is then selected as the one whose embedding has the highest cosine similarity to this centroid. In the actual process, we use epsilon=0.2, resulting in the detection of 59 clusters as illustrated in Figure 3. Two samples of the merged clusters are given below.

- **human:ux_designer** (All members: human:ux_practitioner, human:designer, human:ux-practitioner, human:ux_designer)
- **ai:ml-assisted** (All members: ai:llm-based, ai:ml-assisted, ai:assistance, ai:ml, ai:ml-based, ai:llm, ai:assisted, ai:llm-powered)

3.5 Cluster Keys

After synonym merging, we perform semantic grouping through type-specific clustering. Keys are first segregated by their subject type (human, ai, or co) to ensure that subsequent clustering respects the fundamental categorical differences in these domains. For each type, we apply k-means clustering with parameters optimized through silhouette analysis. The number of clusters is tailored

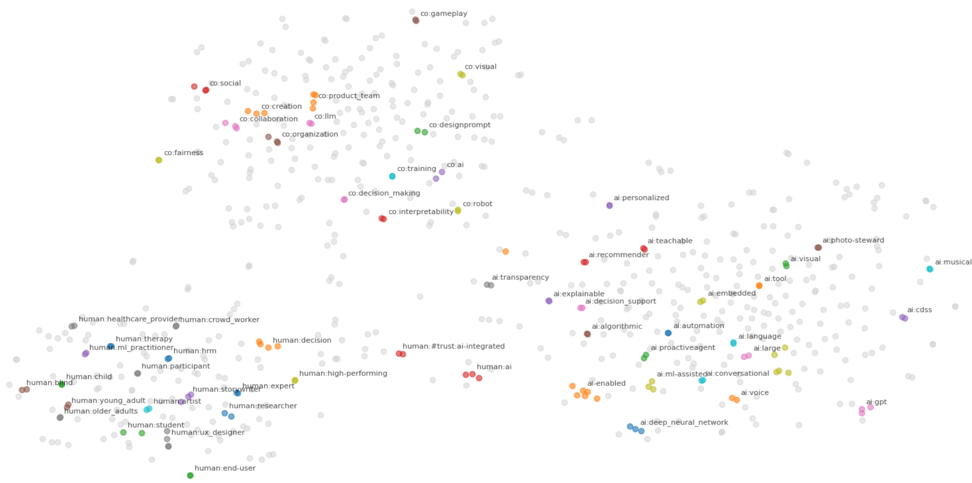

Fig. 3. Result of the HDBSCAN, each clustered is illustrated through colors. The representative word for each cluster is provided next to the cluster.

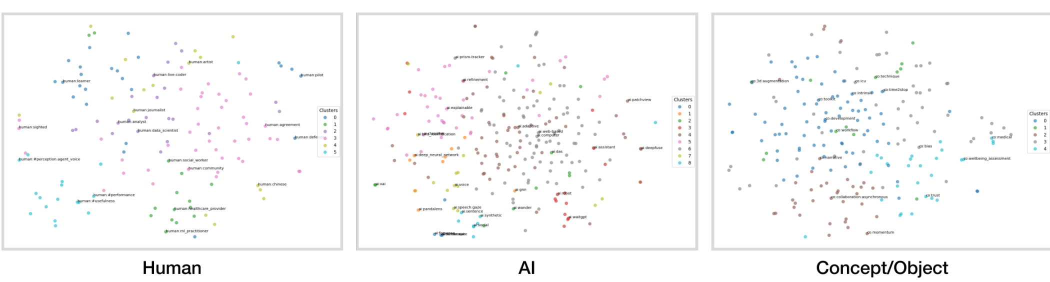

Fig. 4. t-SNE visualization of the clustered keywords in three types (Human/AI/Concept and Object).

to each category's size and semantic diversity: 5 clusters for human-related terms, 8 for AI-related terms, and 4 for concept/object terms. The results are illustrated in Figure 4.

We then generate a description of each cluster. First, we identify the 20 most representative terms from each cluster based on their proximity to the cluster centroid. These terms serve as exemplars that capture the cluster's semantic range. Then, the representative terms, along with their subject type context, are provided to GPT-4 to generate concise descriptions. These descriptions capture common themes, distinctive characteristics, and relationships to human-AI interaction.

3.6 Triplet Processing Result

Considering the subject relationships (causes and effects), the five most frequently occurring pairs are provided in the table below.

Interestingly, AI explainability and user trust emerge as the dominant factors in their respective roles - causal and effect. While explainability leads to technical aspects, trust appears as the primary human response factor, suggesting these elements' crucial roles in human-AI relationships. Notably, these factors are not just individually significant but form the most frequent direct causal pair in the dataset. This strong explainability-trust pathway suggests a fundamental principle: transparent

Keyword (Cause)	Occurrence	Keyword (Effect)	Occurrence
ai|explanation	24	human|#trust	18
ai|system	12	human|experience	6
ai:voice|assistant	6	human|reliance:ai	6
ai:large|language_model	6	human|#trust:ai	6
ai:conversational|agent	5	human|engagement	4

Table 1. Top occurring keywords in cause and effect relationships

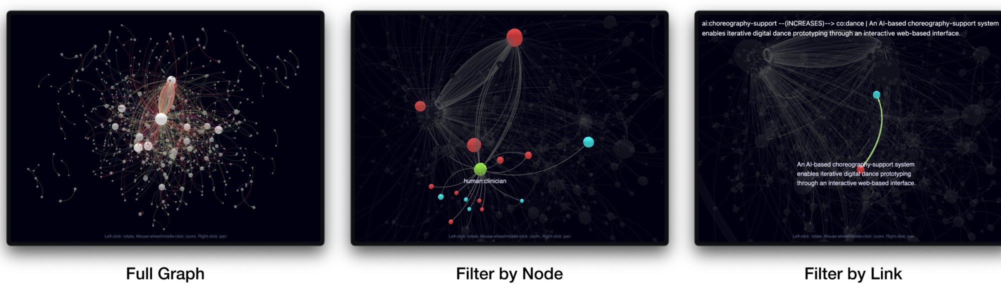

Fig. 5. Sample interface of the web visualization. (From left to right) The full graph display of the dataset; the filtered view highlighting connected nodes and edges of a selected node; and the link filter view providing details about the findings.

AI systems that can explain their behaviors may be the key mechanism for building user trust, transcending specific implementations like voice assistants or conversational agents. We depict one of the ai explainability to human's perception of trust pair in Figure 6.

3.7 Transform Data into a Graph

After keyword processing, we merged the keywords back into the original triplets and converted them into a graph-based representation. Each node represents a subject, and to reduce graph complexity, features are generally not separated into different nodes. However, features that have many linkages beyond the threshold (e.g., `ai|explainability`, `human|#trust`) are separated into different nodes to reduce cluttering and improve interpretability. We encode nodes' cluster membership into their metadata. Meanwhile, each edge represents a relationship between nodes, i.e., `INCREASES`, `DECREASES`, `INFLUENCES`. Edge metadata contains the original finding statements corresponding to the relationship and details regarding the source paper. In our case, the process results in 639 nodes and 2442 edges. The data, originally created using `networkx` package in Python, are then exported to JSON format.

3.8 Web Platform Development

We developed an interactive web-based platform to display the data as a 3D graph. The platform is implemented using Svelte as the core framework and 3d-force-graph for graph visualization. The graph library utilizes Three.js for rendering and D3.js to implement force-directed graph drawing. The interface of the platform is illustrated in Figure 5.

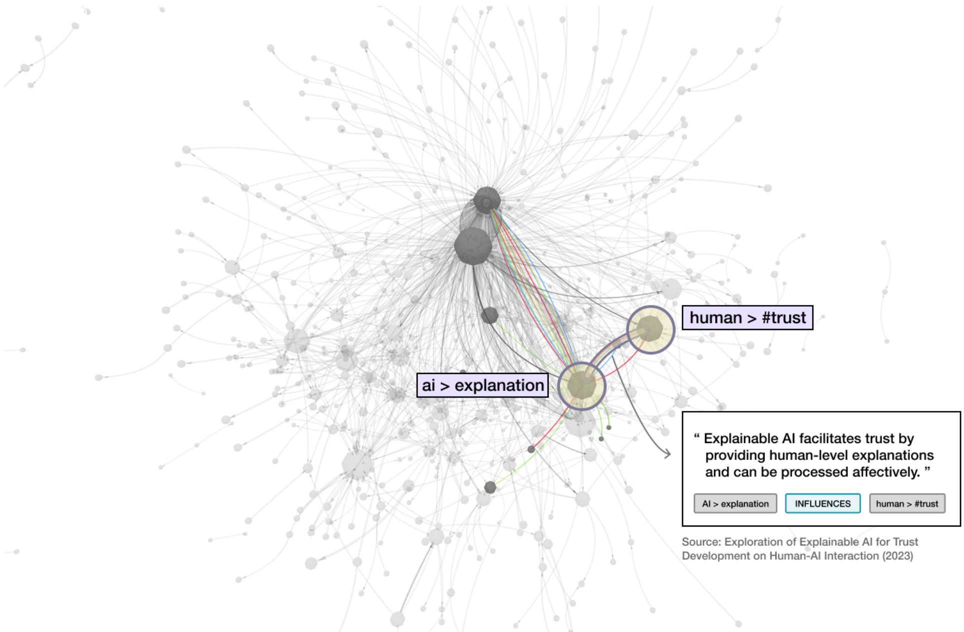

Fig. 6. Sample graph data illustrating two nodes: `ai > explanation` and `human > #trust`, connected by an edge annotated with the original finding and source paper reference.

4 Discussion

4.1 Implications

The methodological framework and findings presented in this atlas have significant implications for both research and practice in human-AI interaction. Our approach to research synthesis demonstrates the potential for using large language models to uncover patterns that might be missed through traditional literature review methods. By extracting and analyzing cause-effect relationships rather than merely grouping papers by themes, researchers can better understand how different aspects of human-AI interaction intersect and influence each other across various contexts and applications. The impact atlas approach is particularly valuable given the increasingly interconnected nature of AI systems in human activities. By mapping connections between empirical findings rather than just topical similarities, researchers can identify crucial gaps in our current understanding of how AI affects human behavior and experience. This comprehensive view reveals potential synergies between different interaction patterns and design approaches while also highlighting areas where careful consideration is needed to avoid negative outcomes. Such insights become increasingly valuable as organizations and individuals integrate AI systems into their daily workflows and practices.

A key opportunity lies in adapting and extending this framework to better capture the multifaceted nature of human-AI interaction. While our current implementation focuses primarily on direct interactions with AI systems, the methodology could be enhanced to better account for social dynamics, institutional contexts, and the broader implications of technology adoption. Such extensions would provide a more comprehensive understanding of how AI can best support human activities while avoiding potential pitfalls, ultimately contributing to the development of more effective and socially aware AI systems.

4.2 Future Works

While our current implementation of the Atlas of Human-AI Interaction provides valuable insights, several promising directions for future work could enhance its utility and scope. First, we envision continuous data addition and updates to maintain the atlas's relevance and comprehensiveness. This includes not only incorporating new research findings as they emerge but also developing automated mechanisms for detecting and integrating relevant publications. Such a dynamic approach would help track the evolution of human-AI interaction patterns over time and ensure that the atlas remains a current reflection of the field's understanding.

The current framework could be extended to accommodate more types of connections between findings. While our present implementation focuses on direct causal relationships through INCREASES, DECREASES, and INFLUENCES connections, future work could incorporate more nuanced relationship types such as mediation effects, contextual dependencies, and temporal sequences. This expansion would enable more sophisticated analysis of how different aspects of human-AI interaction relate to and influence each other, potentially revealing more complex patterns and interdependencies.

Refinement of the extraction process presents another crucial area for future development. Our current LLM-based extractors, while effective, could be enhanced through several approaches. These include developing more sophisticated prompting strategies, implementing multi-stage verification processes, and incorporating domain-specific knowledge to improve the accuracy and granularity of extracted findings. Additionally, we plan to explore methods for automatically identifying and resolving contradictory findings across different studies, potentially through probabilistic reasoning or meta-analytical approaches.

Finally, we see significant potential in developing more intensive connectivity analysis methods. This involves not only examining direct relationships between findings but also identifying higher-order patterns and clusters of interaction effects. Advanced network analysis techniques could be applied to uncover hidden communities of related findings, identify key bridge concepts that connect different areas of human-AI interaction, and predict potential emerging patterns based on existing relationships. This enhanced connectivity analysis could provide deeper insights into the complex web of relationships that characterize human-AI interaction.