人工智能前沿专题
人机交互

彭振辉
中山大学人工智能学院
2023.11.13
The Fairmont, San Francisco, California USA | Oct. 29 – Nov. 1, 2023

The ACM Symposium on User Interface Software and Technology (UIST) is the premier forum for innovations in human-computer interfaces. Sponsored by ACM special interest groups on computer-human interaction (SIGCHI) and computer graphics (SIGGRAPH), UIST brings together people from diverse areas including graphical & web user interfaces, tangible & ubiquitous computing, virtual & augmented reality, multimedia, new input & output devices, Human-Centered AI, and CSCW. The intimate size and intensive program make UIST an ideal opportunity to exchange research results and ideas.
Opening Keynotes

Paper Submission Statistics

- Number of submissions: 483 (+30%)
- Accepts: 121 (8 with shepherding) + 2 from TOCHI
- Acceptance Rate: 25.1%
Opening Keynotes

MidJourney CEO

Executives

David Holz

Previously: Founder Leap Motion, Researcher at NASA, Max Planck
• Don't start a startup because you want to be in charge of something
• 用视觉元素交互进行图片生成，效果很好。
• Do you publish? — Bring something out that people can use. 每周都跟别人分享自己的想法，这可以看做我的publish... Am I helping the world? I feel good.
Opening Keynotes

GenAssist: Making Image Generation Accessible

Mina Huh
The University of Texas at Austin
Austin, TX, USA
minahuh@cs.utexas.edu

Yi-Hao Peng
Carnegie Mellon University
Pittsburgh, PA, USA
yihao@cs.cmu.edu

Amy Pavel
The University of Texas at Austin
Austin, TX, USA
apavel@cs.utexas.edu

Fig. 1: GenAssist makes image generation accessible by providing the logical structure of image generation results. Given
GenAssist is a system that enables blind or low vision creators to generate images by providing rich visual descriptions of the generation results. With the given text prompt and a set of generated images, GenAssist uses a large language model (GPT4) to generate prompt verification questions based on the text prompt, and image-based questions based on individual image captions (BLIP-2). GenAssist also extracts the visual content and style of the images using the vision-language model (CLIP, BLIP2), and object detection model (Detic). All of the information is then summarized using the GPT-4 to generate the comparison descriptions and per-image descriptions.

https://minahuh.com/GenAssist/ 用的都是现有技术, why best paper? 框架，for social good

Do the right things right
建模人类行为很难，chatgpt提供了一种解决思路，
prompt+perceive+memory;
generate plan, reflection and action by the language model。通过一个simulation，分析这些生成式agents的交互行为。通过interviews with the agents来做分析。
Generative Agents

Interactive Simulacra of Human Behavior

Joon Sung Park
Stanford University

Joseph C. O'Brien
Stanford University

Carrie J. Cal
Google Research

Meredith Ringel Morris
Google DeepMind

Percy Liang
Stanford University

Michael S. Bernstein
Stanford University
Generative Agents

- Perceive
- Plan
- Retrieve
- Reflect
- Act
Agents' actions impact the game environment

“Make and drink coffee”

- Cup -> cleaned
- Coffee machine -> turned on
- Chair -> occupied
Large language models can be prompted to generate human behavior conditioned on a variety of experiences.

"[name] is a [description]"

Social Simulacra (UIST '22)

Joon Sung Park, Lindsay Popowski, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein.
- Maria is chatting with Klaus
- The chair is empty
- Giorgio is playing the piano
Generative Agents: Interactive Simulacra of Human Behavior

Fastest growing repo on GitHub
Half a million visited the demo

We thank Stanford HAI, Google Research, Hasso Plattner Design Thinking Research Program, the Siegel Family Endowment, OpenAI, and the MSR PhD Fellowship for their funding support.

Joon Sung Park

Demo Contact Twitter:
reverie.herokuapp.com/arXiv_Demo
joonspk@stanford.edu / joonsungpark.com
joon_s_pk

帮助更好地使用大模型的交互工具研究
I plan to move to San Francisco as a recent graduate. What are some nice neighborhoods in SF?

San Francisco has several neighborhoods that are popular among recent graduates and young professionals due to their vibrant atmospheres, proximity to job opportunities, and entertainment options. Here are some nice neighborhoods in San Francisco for recent graduates:

1. **The Mission District**: The Mission is known for its artsy vibe, vibrant street art, and a wide range of dining options, from trendy restaurants to taquerias. It has a young and diverse population, making it a popular choice for recent graduates.

2. **SoMa (South of Market)**: SoMa is a bustling neighborhood with a mix of tech companies, startups, and cultural venues. It’s great for professionals looking to be close to job opportunities and enjoy a lively nightlife scene.

3. **Lower Haight**: This neighborhood offers a hip and eclectic atmosphere with unique boutiques, bars, and a strong sense of community. It’s a great option for those who want a slightly quieter neighborhood but still want access to nightlife.

... 

Remember that San Francisco’s cost of living is high, so consider your budget when choosing a neighborhood. It’s also a good idea to visit these neighborhoods in person to get a feel for their unique atmospheres and decide which one aligns best with your preferences and lifestyle. Additionally, factors like proximity to public transportation, work, and amenities should be considered in your decision-making process.
How can we support information exploration & sensemaking with LLMs?

把LLM的搜索进行层级化，应对用GPT生成时答案过长过乱的问题
Simple

How many people live in San Francisco?

As of my last knowledge update in September 2021, the population of San Francisco was estimated to be around 883,305 people.

Complex

I plan to move to SF as a new graduate. What factors should I consider?

Moving to San Francisco as a new graduate is an exciting prospect, but there are several factors to consider to ensure a successful transition. Here are some key factors to keep in mind:

1. Cost of Living: San Francisco is known for its high cost of living. Be prepared for expensive housing, groceries, transportation, and entertainment. Make sure to create a budget and have a clear understanding of your expenses.

2. Housing: Finding affordable and suitable housing can be challenging. Research neighborhoods, consider roommates to share rent costs, and be prepared to act quickly when you find a place that suits your needs.

3. Transportation: San Francisco has a well-developed public transportation system, including buses, trams (Muni), and BART (Bay Area Rapid Transit). Consider the proximity of your home to work or school when choosing a neighborhood.

4. Taxes: Understand California's tax system, which...
Canvas <-> Hierarchy

**Canvas**
Space for exploring each topic

**Hierarchy**
Space for visualizing relationships between topics
Bicarbonate (HCO₃⁻) is an important molecule in the regulation of base balance in the body. Which enables users to navigate across different...

Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models

https://www.youtube.com/watch?v=qfMpUX3MqJ0
Ssensecape *motivates* exploration

“... hierarchical interface gave me clues on how to search, what to search.

*It motivated* me to search and explore more complicated information.”
Paper Presentation

Graphologue
Exploring Large Language Model Responses with Interactive Diagrams

Peiling Jiang*
Jude Rayan*
Steven Dow
Haijun Xia

Creativity Lab University of California San Diego
creativity.ucsd.edu

* equal contribution
Graphologue introduced a UIST conference with Sensecape.

Unlike linear and text responses from GPT-4:
- Appropriate levels of details
- Meaningful organization
- Flexible interactions
An earthquake releases energy from the hypocenter, which is located under the Earth's surface. The energy propagates through the Earth in waves, causing a disturbance called seismic waves. Earth is an elastic solid, and the waves propagate through it with different velocities depending on whether they are body waves or surface waves. Body waves include primary (P) waves and secondary (S) waves, which travel through the Earth's interior and are the first to arrive at the surface. Primary waves travel faster than secondary waves and do not displace particles in the direction of travel. Secondary waves displace particles perpendicular to the direction of travel and are slower than primary waves. Surface waves include Love waves and Rayleigh waves, which travel near the Earth's surface and cause more pronounced ground motion. Love waves displace particles in two directions at right angles to the wave propagation, while Rayleigh waves move in an elliptical retrograde motion.
[Earthquakes] - (due to) - [complex processes] [complex processes] - (involving) - [the Earth's crust] [complex processes] - (involving) - [mantle]

GPT-4 response to parse and render as diagrams
Relationship annotation

[AI systems ($N1$)] can be [divided into ($H$, $N1$, $N2$; $H$, $N1$, $N3$)]
Earthquakes occur due to complex processes involving the Earth’s crust, mantle, tectonic plates, and associated forces. One major process involved in earthquake formation is plate tectonics, which refers to the movement and interaction of large sections of the Earth’s lithosphere.

Plate boundaries play a crucial role in earthquake formation. Three main types of Plate Boundaries: Divergent Boundaries, convergent Boundaries, and Transform Boundaries. At these boundaries, tectonic plates interact in different ways, leading to the formation of earthquakes.

Stress and strain build-up along plate boundaries as a result of tectonic forces. When Stress and strain exceed the strength of the rocks, they break and release energy in the form of seismic waves, which is felt as an earthquake.

Finally, the magnitude and intensity of an earthquake depend on several factors, including the amount of energy released, the depth of the earthquake’s focus, and the distance from the epicenter.

**Graphologue: Exploring Large Language Model Responses with Interactive Diagrams (Preview)**

https://www.youtube.com/watch?v=fpe73mF9J6Y
Cells, Generators, and Lenses

Design Framework for Object-Oriented Interaction with Large Language Models

Tae Soo Kim, Yoonjoo Lee, Minsuk Chang, Juho Kim

llm-objects.kixlab.org
Motivation

Traditional Text Editor-Like Design

Once upon a time in a quaint coastal town, there lived a sailor named Captain James O’Malley. He was known far and wide for his tales of the sea and the adventures he had encountered on his trusty ship, the “Siren’s Call.”

Captain O’Malley had been sailing the open waters for as long as he could remember. His weathered face bore the marks of countless sunrises and sunsets over the vast expanse of the ocean. His eyes sparkled with the wisdom of a man who had seen both the beauty and the fury of the sea.

One brisk morning, as the sun painted the sky in hues of pink and orange, Captain O’Malley stood on the shore, gazing out at the horizon. The salty breeze tousled his salt-and-pepper hair, when he noticed something in the horizon.
Design Framework

Object-Oriented Interaction

Cells

He was known far and wide for his tales of the sea and the adventures he had encountered on his trusty ship, the "Siren's Call."

Captain O'Malley had been sailing the open waters for as long as he could remember.

His weathered face bore the marks of countless sunrises and sunsets over the vast expanse of the ocean.

His eyes sparkled with the wisdom of a man who had seen both the beauty and the fury of the sea.

One brisk morning, as the sun painted the sky in hues of pink and orange, Captain O'Malley stood on the shore, gazing out at the horizon.

The salty breeze tousled his salt-and-pepper hair, and he felt a longing in his heart for one last great adventure.

Generators

Lenses

With a determined gleam in his eye, he ordered his crew to prepare the ship and set sail.

As he squinted his eyes to get a better look, he realized it was a ship in distress.

As he squinted his eyes and adjusted his telescope, Captain O'Malley let out a gasp.
User Study (N=18)
Are the Objects Effective?

**Participants**
Novice writers (N=18)

**Tasks**
Writing advertisements for 2 products

**Conditions**

**Treatment**

**Control**
It all started when I accidentally picked up the wrong suitcase at the airport. It all happened because I forgot to check my yellow ribbon.
A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping

Zheng Zhang
Jie Gao
Ranjodh Singh Dhaliwal
Toby Jia-Jun Li

应用大语言模型的文章
In The Dark Knight, Batman has a villainous character.

Please select the keywords to explore:

- Actions and behavior of Batman
- Motivations of Batman
- Contrast with traditional heroic characteristics
- Batman’s use of fear and intimidation tactics
- Batman’s willingness to bend or break the law
- Batman’s disregard for personal and public safety
- Batman’s ambiguous moral code

GENERATE DISCUSSION POINTS
VISAR

1. Hierarchical writing goal recommendation
2. Argumentative sparks
3. Synchronized text and visual planning
4. Varying prototype editing strategies
思考：但这整个流程，会不会跟通常写作的流程有冲突？值得探讨。
PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions

John Joon Young Chung
SpaceCraft Inc.
Los Angeles, USA
jjyc@spacecraft.inc

Eytan Adar
University of Michigan
Ann Arbor, USA
eadar@umich.edu

Abstract

Sometimes, it is not steerable enough

“rough oil painting of cyberpunk city”

Flatter image...

How much flat though?
A photo of a rabbit ⇒ A ceramic cup

An action painting piece ⇒ A lego dragon
PromptPaint
Steering Text-to-Image Generation Through Paint Medium-like Interactions
John Joon Young Chung and Eytan Adar

https://www.youtube.com/watch?v=3SPRsm213_I
Automated Conversion of Music Videos into Lyric Videos

Jiaju Ma**, Anyi Rao*, Li-Yi Wei†, Rubaiat Habib Kazi†, Hijung Valentina Shin†, and Maneesh Agrawala**

Stanford University*, Adobe Research†, Roblox‡
UIST 2023
Kinda dream that can't be sold
Stage 1
Lyrics Phrasing and Line Breaking

Stage 2
Lyrics Placement via Optimization

Stage 3
Lyrics Rendering with Animation

I Ain't Worried by OneRepublic
1. Lyrics Phrasing and Line Breaking

- DG1: Follow song phrasing
- DG2: Break long lines
- DG3: Highlight word being sung
- DG4: Maintain color contrast

2. Lyrics Placement via Optimization

- DG5: Synchronize to music
- DG6: Place near focus of attention
- DG7: Place consistently

Lorem ipsum dolor sit amet

consectetur adipiscing elit
DG6: Place near focus of attention

Convolution

Distance

Darker means lower energy

Foreground Cost Map

Focus of Attention Cost Map

2. Lyrics Placement via Optimization
Automated Conversion of Music Videos into Lyric Videos

https://www.youtube.com/watch?app=desktop&v=nZ9kOwvWgIU   https://vimeo.com/858468154
Mirrorverse

LIVE TAILORING OF VIDEO CONFERENCING INTERFACES

Jens Emil Grønbæk,1 Marcel Borowski,1 Eve Hoggan,1 Wendy E. Mackay,2
Michel Beaudouin-Lafon,2 Clemens N. Klokmose1

1 Aarhus University | 2 Université Paris-Saclay, CNRS, Inria

UIST 2023
REPLICATING MEETING EXPERIENCES
IMPROMPTU REPROGRAMMING OF TOOLS
Mirrorverse: Live Tailoring of Video Conferencing Interfaces

https://www.youtube.com/watch?v=jBiG48rW6Ck
Papeos
Augmenting Research Papers with Talk Videos

Tae Soo Kim  
KAIST

Matt Latzke  
AI2

Jonathan Bragg  
AI2

Amy X. Zhang  
University of Washington

Joseph Chee Chang  
AI2
Design Goals

DG1
Fluid Switching

DG2
Surface Visuals

DG3
Coarse & Fine
3 FORMATIVE AND CO-DESIGN STUDY

To explore the design space for combining research papers and talk videos, we conducted a formative study where participants explored the opportunities and challenges in combining the two formats from the perspectives of both readers and authors.

3.1 Participants

We invited 14 researchers who had previously published at least one paper and created accompanying talk videos. 10 were doctoral students, 2 were mentors’ students, and the remaining 2 were a postdoc and an undergraduate student. 10 of the 14 participants identified their discipline as human-computer interaction (HCI) or related sub-fields (e.g., visualization, AI Barrett). 3 as natural language processing (NLP), 3 as machine learning (ML), and 1 as computer vision (CV).

3.2 Apparatus

Consistent scholarly papers and talk videos as the same time is a new experience that may be hard for participants to imagine. In a preliminary version of this formative study, we gave participants (n=6) a paper and talk video pair side-by-side and instructed them to "understand the content of the paper based on your real-life habits." Although participants could freely choose how they wished to consume the paper and video, they all watched the whole video first and then delved into the paper. Participants expressed how this was not due to a lack of desire but due to the whole high-cost of cross-referencing between formats. This preliminary study revealed that unannotated papers and videos were inadequate to explore how readers wanted to leverage both formats together.

Thus, we developed a technology probe (Fig. 2) where we could pre-link segments of a talk video to relevant passages in the paper (e.g., paragraphs, figures) and color-code them so that participants could switch between the two formats with lower effort. Before the study, one of the authors manually created the links between the papers and videos for three papers in each of the assigned participants’ research fields (e.g., empirical HCI, systems HCI, NLP, CV). To create these links, the author followed criteria that were based on insights from the preliminary study: segment the video on stable transitions, and link segments to paragraphs based on content similarity (e.g., phrases, figures) while following the paper’s reading order.

3.3 Study Procedure

The study consisted of two consecutive sessions. First, there was a formative session where participants took the perspective of paper readers and used the technology probe (Fig. 2) to read a paper where several passages were pre-linked to relevant segments of the talk video. Then, in a co-design session, participants took the perspective of paper authors and considered designs for combining their own research papers and talk videos.

For the formative session, participants chose their preferred paper from the set of pre-linked paper-video pairs and, while thinking aloud, read the paper using the technology probe for 30 minutes. In the probe, linked passages in the paper were highlighted, and participants could click on a linked passage to automatically navigate to the corresponding segment in the video. The video segments were also displayed under the video timeline, and participants could click on a video segment to scroll to the corresponding passage in the paper. After the reading period, participants were asked about the benefits and drawbacks of using the probe and the talk video during paper reading.

For the co-design session, participants took the perspective of
respectively, visually represent importance. By skimming based reader could prioritize reading interface and a few important be included in the conference implementation details and remembering the thumbnails per, the user can also develop paper to help them return to if the thumbnail or transcript about the video segment, the highlight bar to peek into Fig. 4a).

and Video. As the user is read-struggle to understand certain particular sections. For exam-

in the implementation details already familiar with the back-
d/ or summarized explanation highlight bar or video note it-
ote “activates” the video note to a video player that starts.

Transcript To combine papers and videos, we propose Papeos, a novel reading experience that augments papers with segments from talk videos. The Papeo reader places segments as
Styloette: Styling the Web with Natural Language

Daihan Choi
School of Computing, KAIST
Daejeon, Republic of Korea
daihan.choi@kaist.ac.kr

John Kim
School of Computing, KAIST
Daejeon, Republic of Korea
johnkim@kaist.ac.kr

ABSTRACT

Styloette is a new tool that enables users to style the Web using natural language. Users can write instructions for styling websites in their own language, and the tool automatically generates the corresponding CSS code.

CCS CONCEPTS
- Human-centered computing — Human language interfaces
- Web-based metaphors, interfaces, and tools — Experimental studies (in UI)

KEYWORDS
Web Design, Natural-Langauge Interface, End-User Programming, Machine Learning

Figure 1: Styloette enables users to style the Web using natural language. Users can write instructions in their own language, and the tool automatically generates the corresponding CSS code.

References

Within-Subjects User Study

Participants: Early-stage HCI researchers (N=16)

Materials: 3 HCl systems papers

Task: Summarize the systems section

3 Conditions:
- Paper
- Paper + Video
- Papeos

- Introduction
- Interface Tutorial
- Read Abstract
- Read and Summarise Systems Section
- Post-Survey
- Post-Interviews
Augment Static Papers...

... with Existing Content
Papeos  Augmenting Research Papers with Talk Videos

Play with the Papeo of Papeo
papeo.app/demo

https://papeo.app/demo
Video2Action: Reducing Human Interactions in Action Annotation of App Tutorial Videos

Sidong Feng¹, Chunyang Chen¹, Zhenchang Xing²

¹ Monash University
² CSIRO
However

App Tutorial Videos may lack of **accessibility**.

- Video may play too **fast** to watch, and the actions performed sometimes are **not obvious** to be realized

- Auditory information may not accessible to **non-native-language users** or **hearing-impaired users**
Video2Action, a lightweight non-intrusive approach to automatically acquire the actions from the video, reducing the burdens of video creators in action annotation.
Video2Action: Reducing Human Interactions in Action Annotation of App Tutorial Videos

https://www.youtube.com/watch?v=EGD4BgABIbw
UIST 2023

PEANUT: A Human-AI Collaborative Tool for Annotating Audio-visual Data

Zheng Zhang *  Zheng Ning *  Chenliang Xu  Yapeng Tian  Toby Jia-Jun Li
Baseline Annotation Method
Challenge of Fully Automation

- Existing multi-modal models are not robust for achieving accurate annotation
- Single-modality models are hard to automatically correspond annotations in visual and auditory modality
PEANUT: A Human-AI Collaborative Tool for Annotating Audio-Visual Data

https://www.youtube.com/watch?v=PpPaxbxV5Bl
## Study Result

<table>
<thead>
<tr>
<th></th>
<th>Avg. Seconds of Completion/Frame</th>
<th>Number of Frames</th>
<th>Accuracy (clIoU)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline (Human only)</strong></td>
<td>7.73</td>
<td>149.45</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>PEANUT (Human+AI)</strong></td>
<td>5.12</td>
<td>488.85</td>
<td>0.93</td>
</tr>
</tbody>
</table>

- **Baseline**:
  - Avg. Seconds of Completion/Frame: 7.73
  - Number of Frames: 149.45
  - Accuracy (clIoU): 0.72

- **PEANUT**:
  - Avg. Seconds of Completion/Frame: 5.12
  - Number of Frames: 488.85
  - Accuracy (clIoU): 0.93

- **Comparisons**:
  - Avg. Seconds: 33.8% decrease
  - Number of Frames: 227% increase
  - Accuracy (clIoU): 29.2% increase
Soundify: Matching Sound Effects to Video

David Chuan-En Lin\textsuperscript{1}  Anastasis Germanidis\textsuperscript{2}
Cristóbal Valenzuela\textsuperscript{2}  Yining Shi\textsuperscript{2}  Nikolas Martelaro\textsuperscript{1}

\textsuperscript{1}Carnegie Mellon University, \textsuperscript{2}Runway
Design Principles

1. Surface suitable sounds
2. Synchronize sounds to video
3. Tune spatial parameters (pan and volume)
4. Stack multiple soundtracks
This Work

Studio quality sound effects libraries + CLIP (open vocab classification) + Activation map

Soundify Demo

https://www.youtube.com/watch?v=VhsD699osIM
User Study

Task
Create a layered sound environment for a video

Baseline
Manual editing with Adobe Premiere Pro

Participants
12 professional video editors

Measures
- Workload (NASA Task Load Index, 5-point Likert)
- Task Completion Time (seconds)
- Usability (System Usability Scale, 5-point Likert)
Live Demo
https://soundify.cc

Project Page
https://chuanenlin.com/soundify

Robust Finger Interactions with COTS Smartwatches via Unsupervised Siamese Adaptation

Wenqiang Chen, Ziqi Wang, Pengrui Quan, Zhencan Peng, Shupei Lin, Mani Srivastava, Wojciech Matusik, John Stankovic.

Massachusetts Institute of Technology
University of California, Los Angeles
VibInt AI
University of Virginia
Play games
Finger/hand gesture recognition using IMU


Domain Adaptation

- Target new users produce tapping samples in daily usage.
- There is no label in the target new users’ data.
Robust Finger Interactions with COTS Smartwatches via Unsupervised Siamese Adaptation

https://www.youtube.com/watch?v=TW3D19p6rqs
Color Field

- Color Field is a novel visualization of color filters
- Professional Vision as a framework for expertise
- Design CSTs to help users understand their domain

Prototype:
https://creativity.ucsd.edu/colorfield

https://creativity.ucsd.edu/colorfield
Raise your hand if you have used ...
Color Filters on your phone?

Apple Clips.
Color Field: Developing Professional Vision by Visualizing the Effects of Color Filters

https://www.youtube.com/watch?v=p08FlV1utFQ
Living Papers
A Language Toolkit for Augmented Scholarly Communication

Jeffrey Heer, Matthew Conlen, Vishal Devireddy, Tu Nguyen, and Joshua Horowitz

University of Washington
To support dynamic reading aids and explorable explanations [9,12], Living Papers produces web-based articles with a reactive runtime and extensible component system. We use Markdown as a default input file format, with syntax extensions for custom components. Articles may include executable code in languages such as JavaScript, R, and Python to generate static or interactive content. To support “backwards compatibility” with current publishing practices, the Living Papers compiler automatically converts interactive and web-based material to static content, and can generate LaTeX projects or compiled PDFs using extensible journal and conference templates. To assist not only people but also computers to more easily interpret papers, Living Papers can compile article content into accessible data structures, APIs, and software modules.

We present our design objectives for Living Papers, honed in conversations with publishing tool developers and researchers of both augmented reading aids and information extraction from academic articles. We seek to balance tensions among dynamic content, accessible authoring, participation in existing publishing workflows, and research into novel techniques. We evaluate the system by demonstration, sharing articles by ourselves and others that span formal research papers (including this one), explorable explanations, and enhanced content extraction and reuse. These examples highlight augmentations such as enhanced previews for citations and cross-referenced material, equations with interactive term definitions, and articles with dynamic content such as explorable multiverse analyses [14]. Living Papers is available as open source software, and intended to support both article authors and researchers exploring augmented forms of scholarly communication.
R Knitr Integration

Living Papers Team

This article demonstrates compile-time evaluation of R code, with results stitched back into the Living Papers article. JSON-serialized values can also be bound to variables in the reactive runtime.

R version: R version 4.1.3 (2022-03-10)

1 Plots

Let's make some plots.

https://www.youtube.com/watch?v=5-4wd3dVtEk  https://www.youtube.com/watch?v=YoCDW5FZePM
The Future of Scholarly Communication

A "Standard Toolkit" of Reading Augmentations
Definitions, equations, figures, and more...
Helpful to readers, easy for authors to apply
Further design space exploration & evaluation

Collaborative Editing & Reviewing
Novel reviewing tools, pre- and post-publication

Improved Conversion & Accessibility
How might we effectively convert rich, dynamic content to similarly effective static content?
Living Papers
A Language Toolkit for Augmented Scholarly Communication

github.com/uwdata/living-papers

https://uwdata.github.io/living-papers-template/
Artificial General Intelligence

AGI is Coming… Is HCI Ready?

Meredith Ringel Morris
Google DeepMind
Typically, HCI and other sociotechnical domains are not considered as core to the ML research community as areas like model building. However, I argue that research on Human-AI Interaction and the societal impacts of AI is vital and central to this moment in computing history. HCI must not become a “second class citizen” to AI, but rather be recognized as fundamental to ensuring the path to AGI and beyond is a beneficial one.
## Performance x Generality: AGI Level Estimates

<table>
<thead>
<tr>
<th></th>
<th>Narrow</th>
<th>General</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emerging ≈ unskilled</td>
<td>“Narrow Emerging AI” GOFAI (e.g., SHRDLU)</td>
<td>“Emerging AGI” SOTA LLMs (ChatGPT, Bard)</td>
</tr>
<tr>
<td>Competent ≥50th % skilled</td>
<td>“Narrow Competent AI” Siri, Watson, PaLI, Jigsaw</td>
<td>“Competent AGI” &lt;nothing yet&gt;</td>
</tr>
<tr>
<td>Expert ≥90th % skilled</td>
<td>“Narrow Expert AI” Grammarly, Dall-E 2</td>
<td>“Expert AGI” &lt;nothing yet&gt;</td>
</tr>
<tr>
<td>Virtuoso ≥99th % skilled</td>
<td>“Narrow Virtuoso AI” AlphaGo, Deep Blue</td>
<td>“Virtuoso AGI” &lt;nothing yet&gt;</td>
</tr>
<tr>
<td>Superhuman additional skills</td>
<td>“Narrow Superhuman AI” AlphaFold</td>
<td>“ASI” &lt;nothing yet&gt;</td>
</tr>
</tbody>
</table>
## Model Capabilities Unlock Interaction Paradigms

<table>
<thead>
<tr>
<th>Autonomy Level</th>
<th>Unlocked At...</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI as a Tool</td>
<td>Emerging Narrow AI</td>
<td>Running a grammar-checker over text</td>
</tr>
<tr>
<td>AI as a Consultant</td>
<td>Competent Narrow AI; Emerging AGI</td>
<td>Using an LLM to summarize a set of documents</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Programming with a code-generating model</td>
</tr>
<tr>
<td>AI as a Collaborator</td>
<td>Expert Narrow AI; Emerging or Competent AGI</td>
<td>Training as a chess player via a chess AI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Socializing with AI bots</td>
</tr>
<tr>
<td>AI as an Expert</td>
<td>Narrow AI Virtuoso; Expert AGI</td>
<td>Using an AI system to advance scientific discovery (e.g., protein-folding)</td>
</tr>
<tr>
<td>AI as an Agent</td>
<td>Virtuoso AGI; ASI</td>
<td>Autonomous AI-powered personal assistants</td>
</tr>
</tbody>
</table>
My AGI Timelines

Timeline 1: ≥ Expert AGI in 5-10 yrs
- ASI
- Virtuoso AGI
- Expert AGI
- Competent AGI
- Emerging AGI

Timeline 2: Competent AGI Plateau
- ASI
- Virtuoso AGI
- Expert AGI
- Competent AGI
- Emerging AGI
Back to Basics: Norman’s Gulfs 鸿沟

Gulf of Evaluation
What’s the current system state?

Gulf of Execution
How do I use this system?

The User

The World

nngroup.com NN/g
Image credit: Nielsen Norman Group
A PERSON MIGHT THINK THAT BECAUSE
HE HAS PRESSED A BUTTON ON AS REMOTE CONTROL
THAT THE TV WILL TURN OFF
IF THE REMOTE CONTROL DOES NOT WORK BECAUSE OF TECHNICAL PROBLEMS
THERE IS STILL A GULF OF EXECUTION

Gulf of execution VS Gulf of evaluation | Project Management Professional | PMP | CT Academy

https://www.youtube.com/watch?v=isAAu8s3OtM
Updating Norman’s Gulfs for the AGI Era

**Gulf of Evaluation**
What’s the current system state?

**Gulf of Execution**
How do I use this system?

**Process Gulf**
How is the system accomplishing the task?

Image credit: modification of figure from nngroup.com
Human-AGI Interaction: Novel Interaction Techniques

BCI* for “ESP” Interactions?

Cross-Species Interactions?


What are Merrie’s chickens saying?
AGI Design Challenge: Anticipating Societal Impacts

Merrie’s kids interact with her grandfather?  Hire Merrie’s generative ghost to edit your UIST 2100 paper?

"Joint "wild idea" w/ Jed Brubaker from CU Boulder... possible CHI 2024 Workshop"
GPT3 \rightarrow GPT3.5 也是HCI的一个大事件：1）它有一个简单好用的interface；2）它的训练改进过程是采用了human的feedback去加强的。
Reframing AGI Challenges as HCI Challenges

THE ALIGNMENT PROBLEM
Machine Learning and Human Values
BRIAN CHRISTIAN
Best-Selling Author, Algorithms to Alive By
PROMPTIFY

Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models

Stephen Brade*, Bryan Wang*, Mauricio Sousa*, Sageev Oore†, Tovi Grossman*
University of Toronto*, Dalhousie University†
Writing prompts is difficult
Prompt:
An ancient castle perched atop a mountain peak in the breathtaking Dolomites, impressionism style, with soft and dreamy brushstrokes, capturing the beautiful landscape and the vast open sky above, by John Singer Sargent, oil painting on canvas.

Suggested modifiers:
- by André Charles B伊er
- by Wilhelm Trübner
- by Eduard von Grützner
- by Hermann Bodzisch

Suggested modifiers for cluster:

Suggested modifiers unique to cluster:
Style Extension

Atomic Style

Studio Ghibli

+ Steered Subject Ideation

Lion standing majestically by a cherry blossom tree with Mount Fuji in the background, Studio Ghibli

Append

Lion standing majestically by a cherry blossom tree with Mount Fuji in the background, Studio Ghibli

ChatGPT

Few-shot Prompting

Style Extension

Lion standing majestically by a cherry blossom tree with Mount Fuji in the background, Studio Ghibli style, soft lighting, pastel colors, anime-inspired, intricate details, in the style of Hayao Miyazaki and Isao Takahata, breathtaking scenery, trending on artstation.
UIT 2023 Promptify: Text-to-Image Generation through Interactive Prompt Exploration with LLMs

https://www.youtube.com/watch?v=0u22sv8uB-I
Comparative Evaluation

14 – Participant Study

Within subjects comparison with Automatic1111
Solicited Subjective Ratings and NASA-TLX
Concluded with free usage observation + interview

Next step: 对话式生成图片，or direct manipulation
PoseVEC: Authoring Adaptive Pose-aware Effects using Visual Programming and Demonstrations

Yongqi Zhang
Cuong Nguyen
Rubaiat Habib Kazi
Lap-Fai Yu

George Mason University
Adobe Research
Adobe Research
George Mason University
Pose-aware Visual Effects

Entertainment ©JP Pirie
At-home workout ©Jeremy Ethier
Education ©Grant VanderHayden
Programming-based approach is expressive but too technical

<table>
<thead>
<tr>
<th>Authoring Workflow</th>
<th>Example</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyframe-based approach</td>
<td>Adobe Premiere, After Effects</td>
<td>Output effect isn’t reusable</td>
</tr>
<tr>
<td>Programming-based approach</td>
<td>Lens studio, Spark AR</td>
<td>Require programming and configuration</td>
</tr>
</tbody>
</table>
Previous research on Programming by Demonstration

GesturAR
Gesture Studio

MadRiddles

Gesture Knitter

Rapido
PoseVEC (Pose-aware Visual Effect Creator)

Video Canvas
Direct manipulation on video canvas to streamline authoring process.

Node Canvas
Reduce programming complexity using node UI.

Load videos, Manipulate graphics,
Create pose recognizer, Create nodes,
PoseVEC: Authoring Adaptive Pose-aware Effects Using Visual Programming and Demonstrations

https://www.youtube.com/watch?v=6_miO0YIRVM
Summary

PoseVEC
- programming by demonstration and visual programming
- direct manipulation with video canvas
- node-based programming UI

Project website
Wakey-Wakey: Animate Text by Mimicking Characters in a GIF

Liwenhan Xie*, Zhaoyu Zhou*, Kerun Yu, Yun Wang, Huamin Qu, and Siming Chen
*: equal contribution

Presenter: Qingyu Guo
讨论了generalizability，motion transfer model的局限
Our Mental Face Reconstruction System (MFRS)

We propose a fast interactive system with high usability for mental face reconstruction.

https://www.youtube.com/watch?v=qP-ncu7r4T4
Usable and Fast Interactive Mental Face Reconstruction

MGFS – Quantitative Evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>Mental Rating ↑</th>
<th>Visual Rating ↑</th>
<th>SUS ↑</th>
<th>NASA-TLX ↓</th>
<th>Time (mins) ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>4.0* ± .82</td>
<td>4.1 ± .95</td>
<td>84.6* ± 12.9</td>
<td>27.2* ± 18.4</td>
<td>10.0* ± 4.1</td>
</tr>
<tr>
<td>CG-GAN</td>
<td>4.8* ± .80</td>
<td>3.9 ± .95</td>
<td>59.0* ± 12.7</td>
<td>43.4* ± 10.9</td>
<td>17.0* ± 5.6</td>
</tr>
</tbody>
</table>

*mental and visual quality的打分分析，有意思*
Spellburst
A Node-based Interface for Exploratory Creative Coding with Natural Language Prompts

spellburst-paper.repl.it.app

Tyler Angart
Replit, Inc.
tyler@replit.com

Miroslav Suzara
Stanford University
msuzara@stanford.edu

Jenny Han
Stanford University
jennyhan@cs.stanford.edu

Christopher Pondoc
Stanford University
cipondoc@stanford.edu

Harharan Subramonyam
Stanford University
harhara@stanford.edu
Semantic sliders

1. Idea → code
2. Variations
3. Combinations
Spellburst: A Node-based Interface for Exploratory Creative Coding with Natural Language Prompts

Hariharan Subramonyam

https://www.youtube.com/watch?v=5ww0jsS9hTI
Other Scenarios in UIST

Banquet at Holloween
Poster session
Demo session
Human-Cat Interaction
Coffee break
Coffee break
People I Met
I am an incoming Assistant Professor of Computer Science at University of Rochester, also as a member of ROC-HCI group. I am currently a postdoctoral researcher in the Augmented Perception Lab, Human-Computer Interaction Institute at Carnegie Mellon University. Before that, I earned my Ph.D. degree and Bachelor’s degree from Tsinghua University. My research focus lies in Human-Computer Interaction and Mixed Reality.

My research is focused on understanding, predicting and enhancing user behavior in Mixed Reality, which follows three connected threads: 1) capture the user’s behavioral and perceptual patterns with computational methods, 2) develop input techniques and adaptive user interfaces to facilitate human-computer two-way communication, 3) explore unique behavioral enhancements enabled by Mixed Reality. I publish at ACM CHI, UIST, IMWUT and IEEE VR.

I will join UR in Janurary 2024, I am looking for students (including Ph.D. students in 2024 Fall) to join my lab, please reach out if you are interested.

Selected Publications

HandAvatar: Embodying Non-Humanoid Virtual Avatars through Hands
Yu Jiang, Zhipeng Li, Mufei He, David Lindlbauer, Yukang Yan

https://yukangyan.info/
HandAvatar: Embodying Non-Humanoid Virtual Avatars through Hands

https://drive.google.com/file/d/1gdMUFRgal0YLJY2wsiiv8v-4lyQ-RgQR/view
I am recruiting Ph.D. students, undergraduate researchers, and visiting researchers. See the Students page for details.

Toby Jia-Jun Li is an Assistant Professor in the Department of Computer Science and Engineering at the University of Notre Dame where he directs the SaNDwich Lab. Toby received a Ph.D. degree in Human-Computer Interaction at Carnegie Mellon University, where he was advised by Brad A. Myers. Toby also worked closely with Tom M. Mitchell. Prior to Ph.D., Toby received his Bachelor's degree in Computer Science from the University of Minnesota where he worked with Brent Hecht as a member of GroupLens Research.

Toby works at the intersection of Human-Computer Interaction (HCI), End-User Software Engineering, Machine Learning (ML), and Natural Language Processing (NLP) applications, where he uses human-centered methods to design, build, and study interactive systems to empower individuals to create, configure, and extend AI-powered computing systems. His recent work seeks to address the societal challenges in the future of work through a bottom-up human-AI collaborative approach that helps individual workers automate and augment their tasks with AI systems.

https://toby.li/
StoryBuddy: A Human-AI Collaborative Agent for Parent-Child Interactive Storytelling with...

https://www.youtube.com/watch?v=wvRthBqd-io
Shumin Zhai

Shumin Zhai is a Principal Scientist and Director of Gboard and Interaction Methods at Google. His research has advanced foundational models and understandings of human-computer interaction (HCI), as well as practical user interface designs, inventions, and flagship products. His publications have won numerous awards, including the ACM UIST Lasting Impact Award, the ACM Mobile HCI Impact Award, and the IEEE Computer Society Best Paper Award. He was the Editor-in-Chief of the *ACM Transactions on Computer-Human Interaction* from 2009 to 2015. He received his Ph.D. degree from the University of Toronto in 1995. In 2006, he was named to ACM's inaugural class of Distinguished Scientists. In 2010, he was named a Member of the ACM CHI Academy and a Fellow of the ACM.

https://www.shuminzhai.com/

https://scholar.google.com/citations?user=jJLvY8AAAAAJ&hl=en
https://blog.google/products/search/gboard-now-on-android/
Juho Kim is an Associate Professor in the School of Computing at KAIST, and directs KIXLAB (the KAIST Interaction Lab). He is also affiliated faculty in the Kim Jaechul Graduate School of AI at KAIST. His research in human–computer interaction and human–AI interaction focuses on building interactive and intelligent systems that support interaction at scale, aiming to improve the ways people learn, collaborate, discuss, make decisions, and take action online.

He earned his Ph.D. from MIT, M.S. from Stanford University, and B.S. from Seoul National University. In 2015–2016, he was a Visiting Assistant Professor and a Brown Fellow at Stanford University. He is a recipient of a KIISE/IEEE–CS Young Computer Reseacher Award, KAIST’s Songam Distinguished Research Award, Grand Prize in Creative Teaching, Q–Day Creative Education Award, and Excellence in Teaching Award, as well as 14 paper awards from ACM CHI, ACM CSCW, ACM Learning at Scale, ACM UII, ACM DIS, and AAAI HCOMP. He recently spent his sabbatical year as a chief scientist at Ringle Inc. to transfer his research on AI-powered analysis and diagnosis of English learners’ proficiency into a real product. He gave a keynote at NeurIPS 2022 titled “Interaction–Centric AI”.

If you're interested in working with me at KAIST, please read this page. Also, in this interview video for HCI Korea, I summarize KIXLAB's representative projects and share my thoughts on research and mentoring. English transcription is available.

Video of my recent talks on "Interaction–Centric AI": (1) NeurIPS 2022 keynote (targeted at AI audience), (2) Stanford HCI Seminar (targeted at HCI audience)

https://juhokim.com/
People I Met

Interaction-Centric AI

Juho Kim
juhokim.com | kixlab.org | @imjuhokim | juhokim@kaist.ac.kr

https://slideslive.com/38996064/interactioncentric-ai
Long-form Answer

The handshake has existed in some form or another for thousands of years, but its origins are somewhat murky. One popular theory is that the gesture began as a way of conveying peaceful intentions. By extending their empty right hands, strangers could show that they were not holding weapons and bore no ill will toward one another. . . .

Why do we shake hands?

Transform

Dialogue Tree

People have been shaking hands for a really long time, but we don't know why.

Why do we shake hands?

Let me ask you a question. How long have people been shaking hands?

That's correct! Do you want me to tell you some of the theories about why people shake hands?

One popular idea is that the gesture was started to show that the person was being friendly.

For a really long time

Okay!

DAPIE: Interactive Step-by-Step Explanatory Dialogues to Answer Children's Why and How Questions

https://www.youtube.com/watch?v=6HuH-DskjiM
Xiang 'Anthony' Chen

xac @ ucla.edu

Assistant Professor
ECE Department, UCLA
UCLA HCI Research

Fall 2023 Office Hours

My research employs human-centered design methods to build systems that catalyze advances in AI to augment human activities, supported by NSF CAREER Award, ONR YIP Award, Google Research Scholar Award, Intel Rising Star Award, Hellman Fellowship, and NSF CRII Award.

Vitae  Papers  Tweets  Reads

https://hci.prof/
I was also awarded a Google PhD Fellowship in Human Computer Interaction in 2022! Northwestern wrote a nice article featuring me as a fellowship recipient.

I am a Postdoctoral Researcher at Stanford University CS, where I collaborate with Diyi Yang and Emma Brunskill to advance AIs potential to assist in Mental Health and Psychotherapy. Mental health providers—from trained therapists to volunteer peer supporters—remain crucial, and we hypothesize that AI-assisted training and practice can improve their efficacy, confidence, and growth. Human–AI collaboration and interfaces for effectively imbuing AIs with domain-specific and personal-knowledge of this area will be important.

I received my PhD from Northwestern's Technology and Social Behavior program, where I worked closely with Prof. Haoqi Zhang and Darren Gergle. My research aimed to enhance people’s communication of intent and goals when using AI systems to assist in personally-meaningful pursuits, such as creating music compositions and sharing meaningful social experiences at distance. To do this, I advanced Human–AI Interface Layers, a technology sitting on top of existing AIs that help people in communicating their intent and expectations to their AI assistants and collaborators.

http://youralien.github.io/
Anyi Rao
Postdoctoral Scholar
Computer Science Department
Stanford University
Email: anyirao [at] stanford.edu

Bio
Anyi Rao is a Postdoctoral Scholar at Stanford with Maneesh Agrawala. He has research experiences at Meta Reality Lab, Vector Institute, University of Toronto, Hong Kong University. He received the Ph.D. at MMLab, Chinese University of Hong Kong in 2022, advised by Dahua Lin and Bolei Zhou. He studies human-centered AI for creativity and multimodality, with focuses on content generation, intelligent video editing and creation, video semantic and cinematic analysis, aiming to build connections between AI and humans for collaborative intelligence. His works include ControlNet, AnimateDiff, MovieNet, CityNeRF, Shoot360, and Virtual Dynamic Storyboard, with a Marr Prize (ICCV best paper award).

If you also have some exciting ideas and insights on the aforementioned research, please drop me an email. Let’s push it forward together.

https://anyirao.com/
Adding Conditional Control to Text-to-Image Diffusion Models

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala
Stanford University
{lvmin, anyirao, maneesh}@cs.stanford.edu

Figure 1: Controlling Stable Diffusion with learned conditions. ControlNet allows users to add conditions like Canny edges (top), human pose (bottom), etc., to control the image generation of large pretrained diffusion models. The default results use the prompt “a high-quality, detailed, and professional image”. Users can optionally give prompts like the “chef in kitchen”.

https://github.com/Mikubill/sd-webui-controlnet
Prof. Antti Oulasvirta

Antti Oulasvirta leads the Computational Behavior Lab (cbl.aalto.fi) at Aalto University and the Interactive AI research program at FCAI (Finnish Center for Artificial Intelligence). Prior to joining Aalto, he was a Senior Researcher at the Max Planck Institute for Informatics and the Cluster of Excellence on Multimodal Computing and Interaction at Saarland university. He received his doctorate in Cognitive Science from the University of Helsinki in 2006, after which he was a Fulbright Scholar at the School of Information in University of California-Berkeley in 2007-2008 and a Senior Researcher at Helsinki Institute for Information Technology HIIT in 2008-2011. During his postgraduate studies in 2002-2003, he was an exchange student at UC Berkeley's Neuropsychology Lab. He was awarded the ERC Starting Grant (2015-2020) for research on computational design of user interactions. Dr. Oulasvirta serves as an associate editor for ACM TOCHI and has previously served International Journal of Human-Computer Studies, as well as served as a column editor for IEEE Computer. He frequently participates in the paper committees of HCI conferences, including the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI). His work has been awarded the Best Paper Award and Best Paper Honorable Mention at CHI fifteen times between 2008 and 2022. He has held keynote talks at NordiCHI’14, CoDiT’14, EICS’16, IHCI’17, ICWE’19, Chinese CHI’19, and IS-EUD’23. He is a member of ELLIS (European Laboratory for Learning and Intelligent Systems). In 2019, he was invited to the Finnish Academy of Science and Letters. He was a SICSA Distinguished Visiting Fellow in 2011 and in 2022.

Research Lead

- Computational Behavior Lab at Aalto University
- Interactive AI research programme at FCAI

https://users.aalto.fi/~oulasvir/
Trip to Stanford University

胡佛塔
博士/博后办公室
校园长廊
各个研究主题的午餐会
Fei-Fei Li的办公室
周日集市

下雨了怎么办？
Trip to UC Berkeley

萨瑟塔
（钟楼）
肠粉店...
Trip in SF
Trip in SF
Trip in SF
NBA Game
HCI elements in NBA Game
Learn from a chat with Tommy (Taewook Kim)
• From Prof. Juho Kim
  • It is an important thing that let supervisors not worry about you.
  • Expectation management is important in the student-supervisor relationship.
  • You should also get into the HCI community. The community is a family that helps students grow up. Everyone should also volunteer to help the community.
• From Prof. Xiaojuan Ma
  • Don’t easily give up. Think of your co-authors. Everyone is helping you.