Kicking Off the Flywheel of AI + XR Innovations with XR Blocks



Ruofei Du
XR Labs Lead / Mgr @ Google
March 10, 2026 @ MIT HCI seminar
www.ruofeidu.com
X: DuRuofei@
me@duruofei.com

Ruofei Du

Self Intro

www.ruofeidu.com

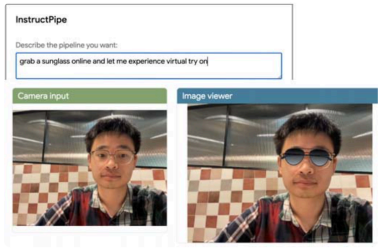


Publications
Projects
Software
Videos
Media
Talks
Artsy

XR Labs Lead

© Ruofei Du | 2025-3-28 | DuWebKit

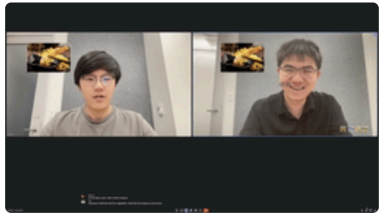
Featured Publications



InstructPipe: Building Visual Programming Pipelines in Visual Blocks With Human Instructions Using LLMs 🎖️ Honorable Mentions Award

Zhongyi Zhou, Jing Jin, Vrushank Phadnis, Xiuxiu Yuan, Jun Jiang, Xun Qian, Jingtao Zhou, Yiyi Huang, Zheng Xu, Yinda Zhang, Kristen Wright, Jason Mayes, Mark Sherwood, Johnny Lee, Alex Olwal, David Kim, Ram Iyengar, Na Li, and Ruofei Du
Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI), 2025.

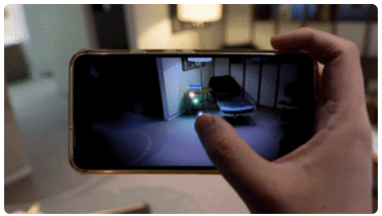
pdf, doi | project, video | cited by, cite



Visual Captions: Augmenting Verbal Communication With On-the-fly Visuals 📹 Open Source, Real-time, Live!

Xingyu Bruce Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Alex Olwal, Peggy Chi, Xiang 'Anthony' Chen, and Ruofei Du
Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI), 2023.

pdf, lowres, doi | project, video, code, data, news | cited by, cite



DepthLab: Real-time 3D Interaction With Depth Maps for Mobile Augmented Reality 🚀 100K downloads

Ruofei Du, Eric Turner, Maksym Dzitsiuk, Luca Prasso, Ivo Duarte, Jason Dourgarian, Joao Afonso, Jose Pascoal, Josh Gladstone, Nuno Cruces, Shahram Izadi, Adarsh Kowdle, Konstantine Tsotsos, and David Kim
Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST), 2020.



Geollery: A Mixed Reality Social Media Platform 🌎 Live Demo of a Metaverse of Mirrored World

Ruofei Du, David Li, and Amitabh Varshney
Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI), 2019.

pdf, doi | website, project, video, slides, demo, supp | cited by, cite

Self Intro

Ruofei Du



Ruofei Du

Ruofei Du serves as Interactive Perception & Graphics Lead / Manager at Google and devotes to creating novel interactive technologies for XR. As a Research Scientist, Ruofei's research covers a wide range of topics in technical HCI, Graphics, and Perception, including XR interactions, visual programming, augmented communication, XR social platforms, digital human, foveated rendering, accessibility, and deep learning in graphics. Du serves as an Associate Chair in program committee of CHI and UIST; an Associate Editor of IEEE TCSVT. He holds 3 US patents and has published over 30 peer-reviewed publications in top venues of HCI, Computer Graphics, and Computer Vision, including CHI, UIST, SIGGRAPH Asia, TVCG, CVPR, ICCV, ECCV, ISMAR, VR, and I3D. In his own words: I am passionate about inventing interactive technologies with graphics, perception, and HCI. See my research, artsy, projects, youtube, talks, github, and shadertoy demos for fun!

Personal Website
Google Scholar

Research Areas

- Human-Computer Interaction and Visualization →
- Machine Intelligence →
- Machine Perception →

Authored Publications

Self Intro

Ruofei Du



Ruofei Du

FOLLOW

GET MY OWN PROFILE

Interactive Perception & Graphics Lead / XR Labs Manager, Google XR
Verified email at google.com - Homepage

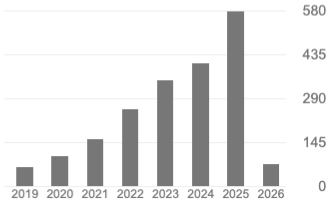
AI + XR Interactive Perception Interactive Graphics Human Computer Interaction
WebAI + WebXR

TITLE	CITED BY	YEAR
DepthLab: Real-time 3D Interaction with Depth Maps for Mobile Augmented Reality R Du, EL Turner, M Dzitsiuk, L Prasso, I Duarte, J Dourgarian, J Afonso, ... UIST '20: Proceedings of the 33rd Annual ACM Symposium on User Interface …	207	2020
Kernel Foveated Rendering X Meng, R Du, M Zwicker, A Varshney I3D '18: Proceedings of the ACM on Computer Graphics and Interactive …	117	2018
Language-based Colorization of Scene Sketches C Zou, H Mo, C Gao, R Du, H Fu SIGGRAPH Asia '19: ACM Transactions on Graphics 38 (6), 1-16	104	2019
SketchyScene: Richly-Annotated Scene Sketches C Zou, Q Yu, R Du, H Mo, YZ Song, T Xiang, C Gao, B Chen, H Zhang ECCV '18: European Conference on Computer Vision	100	2018
CollaboVR: A Reconfigurable Framework for Creative Collaboration in Virtual Reality Z He, R Du, K Perlin ISMAR '20: 2020 IEEE International Symposium on Mixed and Augmented Reality …	89	2020
Montage4D: Real-time Seamless Fusion and Stylization of Multiview Video Textures R Du, M Chuang, W Chang, H Hoppe, A Varshney I3D '18: Journal of Computer Graphics Techniques 8 (1), 1-34	81 *	2019
Geollery: A Mixed Reality Social Media Platform R Du, D Li, A Varshney CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in …	80	2019
Eye-dominance-guided Foveated Rendering X Meng, R Du, A Varshney IEEE Transactions on Visualization and Computer Graphics (TVCG) 26 (5), 1972 …	78	2020
Augmented Object Intelligence with XR-Objects MD Dogan, EJ Gonzalez, K Ahuja, R Du, A Colaço, J Lee, ... Proceedings of the 37th Annual ACM Symposium on User Interface Software and …	70	2024
PRIF: Primary Ray-based Implicit Function BY Feng, Y Zhang, R Du, A Varshney European Conference on Computer Vision (ECCV)	59	2022
Multiresolution Deep Implicit Functions for 3D Shape Representation Z Chen, Y Zhang, K Genova, S Fanello, S Bouaziz, C Haene, R Du, ... 2021 IEEE/CVF International Conference on Computer Vision (ICCV)	56	2021

Cited by

VIEW ALL

	All	Since 2021
Citations	2068	1826
h-index	26	24
i10-index	42	39



Public access

VIEW ALL

0 articles — not available
18 articles — available

Based on funding mandates

Co-authors

VIEW ALL

- Alex Olwal — Research & Engineering Leader …
- Amitabh Varshney — Dean, College of Computer, Mat…
- Danhang Tang — Research Scientist, Google
- Yinda Zhang — Google Research
- David Kim — Perceptive Experiences Lead at …

Self Intro

Ruofei Du



Follow

Ruofei Du

@DuRuofei

Interactive Perception & Graphics Lead/Mgr @Google, inventing novel #interactive #AI in #XR. PhD '18 @umdcs #VR #AR #Graphics #HCI #VisualBlocks #UIST #CHI

San Francisco, CA duruofei.com Joined April 2011

940 Following **2,018** Followers

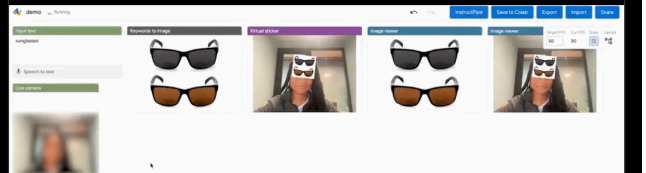
Posts Replies Media

Pinned

Ruofei Du @DuRuofei · Apr 19

Check out #InstructPipe - an initial step that turns your idea into an AI pipeline with visual programming!

Come to @Zhongyi_Zhou_'s @acm_chi talk on 5/1 9:30am at G304 and my talk at 10:30-11:10am in North 1F. #VisualBlocks #WebAI #LLM #Agent

> **Google AI** @GoogleAI · Apr 19
>
> Visual programming frameworks like Visual Blocks can speed up ML pipeline building, but novice users can still struggle with setting them up. Read how InstructPipe, a research prototype for ML pipeline design, helps users build workflows more easily →goo.gle/4lB4pav
>
> 

Self Intro

Ruofei Du





Inventor, more than a researcher

CHI / UIST Program Committee / Session Chairs / Invited Speaker

2x Best / Distinguished Paper Awards
5x Best Paper Honorable Mentions
1x Best Student Paper Awards

Self Intro

Ruofei Du



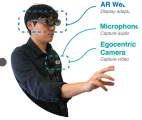
InstructPipe
CHI '25
Best Paper Honorable Mention



RetroSphere
IMWUT '23
Distinguished Paper Award



Rapsai
CHI '23
Best Paper Honorable Mention



Human I/O
CHI '24
Best Paper Honorable Mention

Geollery
CHI '19, Web3D '19

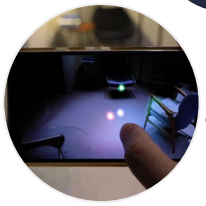
Social Street View
Web3D '16 *Best Paper Award*



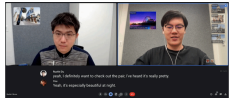


ChatDirector
CHI '24

Human-Computer Interaction



DepthLab
UIST '20
100K+ Installs



Visual Captions
CHI '23

Graphics

Perception



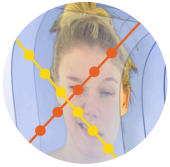
GazeChat
UIST '21



Ad hoc UI
CHI '22 Interactivity



LogRectilinear
IEEE VR '21 (TVCG)
TVCG Honorable Mention Award



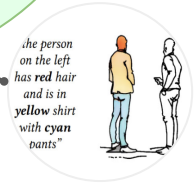
FaceFolds
I3D' 2024
Best Student Paper Award



Kernel Foveated Rendering
I3D '18, VR '20, TVCG '20
ACM CGIT 2023 Most Read Paper



Montage4D
I3D '18, JCGT '19
MSR TechFest'18

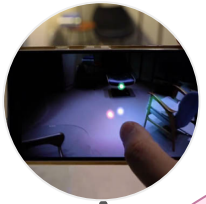


SketchyScene
TOG (SIGGRAPH Asia) '19, ECCV '18

Self Intro

Ruofei Du

DepthLab
UIST '20
100K+ Installs



Rapsai
CHI '23
Best Paper
Honorable Mention

Human I/O
CHI '24
Best Paper
Honorable Mention



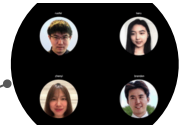
Ad hoc UI
CHI '22
Interactivity





InstructPipe
CHI '25
Best Paper
Honorable Mention

GazeChat
UIST '21



Geollery
CHI '19, Web3D '19, VR '19



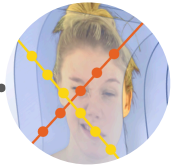
Social Street View
Web3D '16
Best Paper Award



Interaction and Communication



Visual Captions
ThingShare
CHI '23



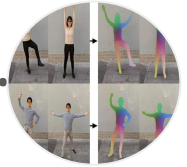
FaceFolds
I3D' 2024
Best Student Paper Award

OmniSyn
IEEE VR '22



Digital World

Digital Human



MonoAvatar
HumanGPS
CVPR' 21, '23

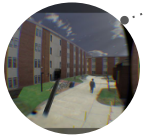
LogRectilinear
IEEE VR '21 (TVCG)
TVCG Honorable Mention





SketchyScene
TOG (SIGGRAPH Asia) '19, ECCV '18

VideoFields
Web3D '16

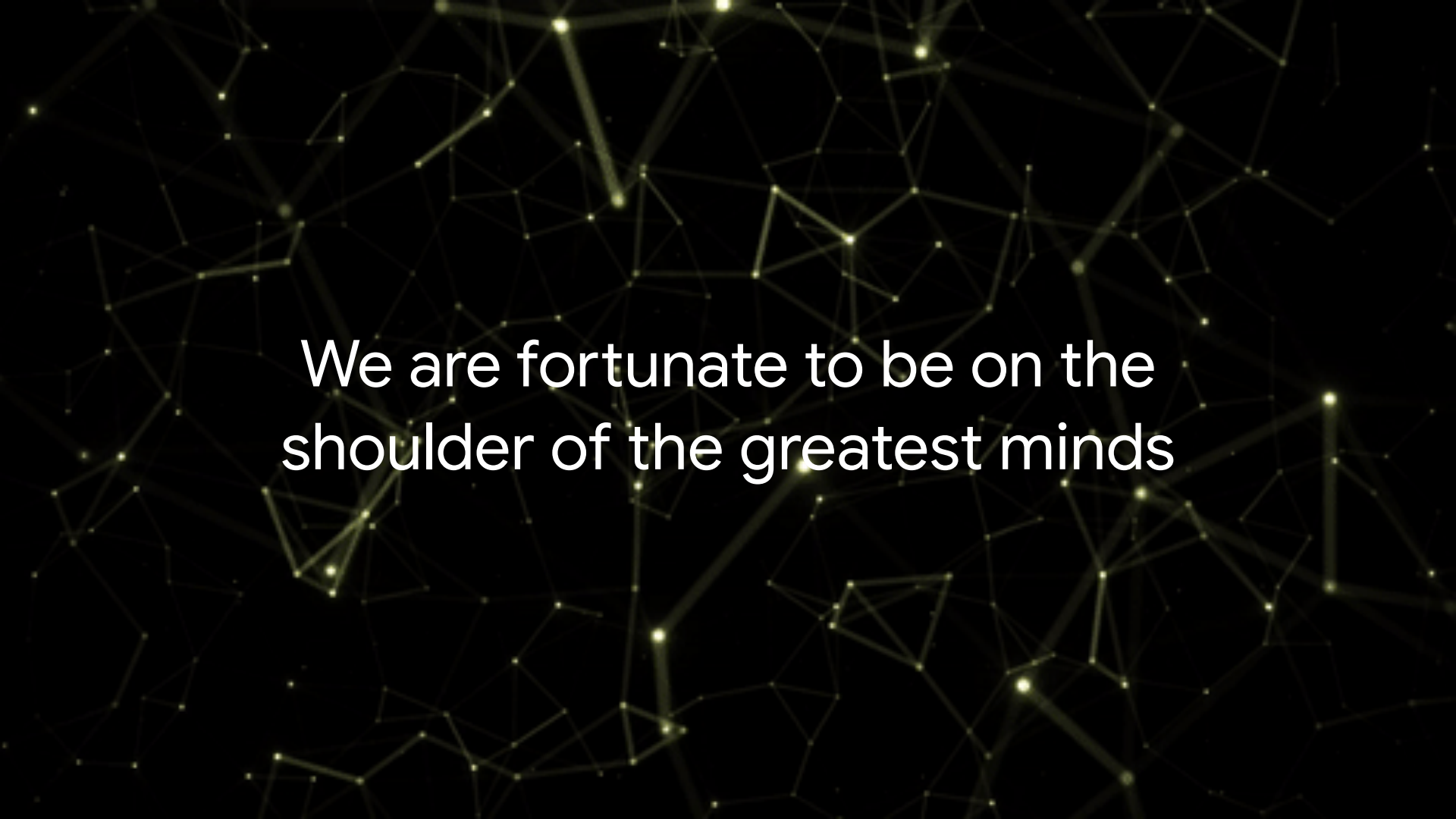


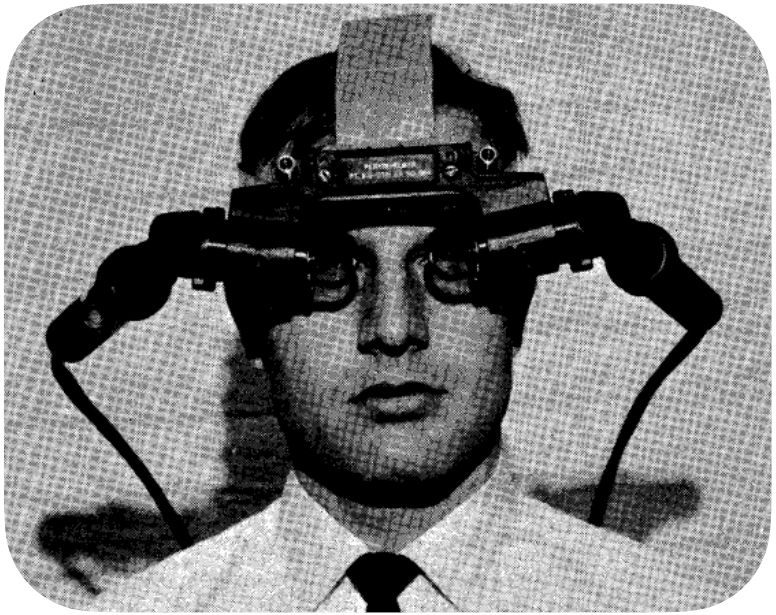
Montage4D
I3D '18
JCGT '19

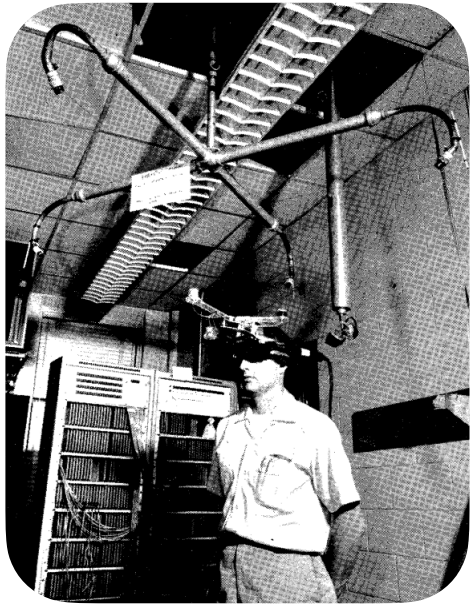




ChatDirector
CHI '24


We are fortunate to be on the
shoulder of the greatest minds















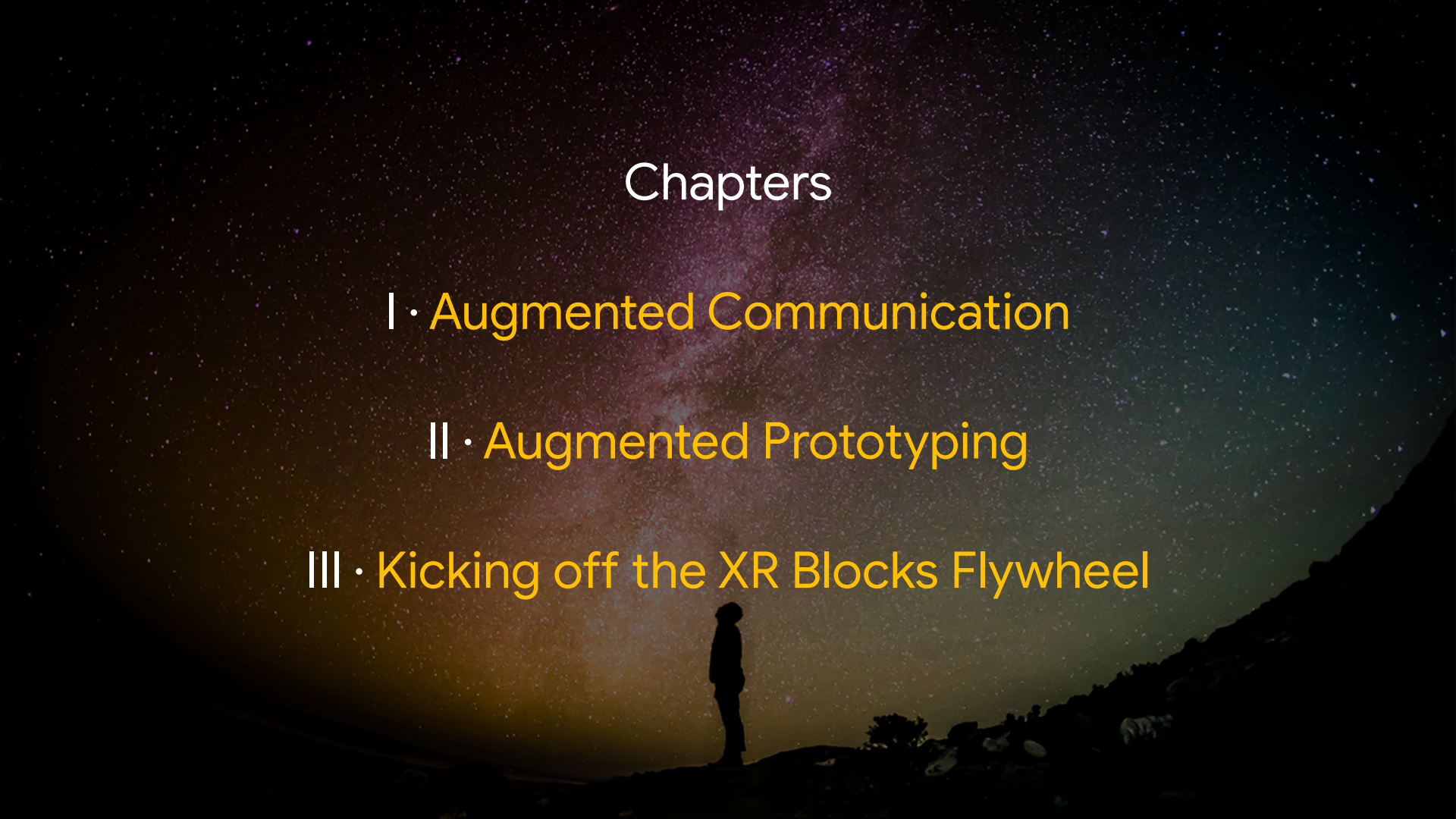





As a teenager...


Snow Cras
Neal Stephenson,
1992.


XR
a medium to make information more useful and accessible and help people to better communicate in their physical life


Chapters
I · Augmented Communication
II · Augmented Prototyping
III · Kicking off the XR Blocks Flywheel

Chapter I · Augmented Communication




SOCIAL
MEDIA
BLOG
SMS
BLOG
SMS








Social Street View
Mirrored world with
geo-tagged social media
Social Street View


Social Street View
Immersive Education
L'esprit de Paris
est au 29 rue de Rivoli.

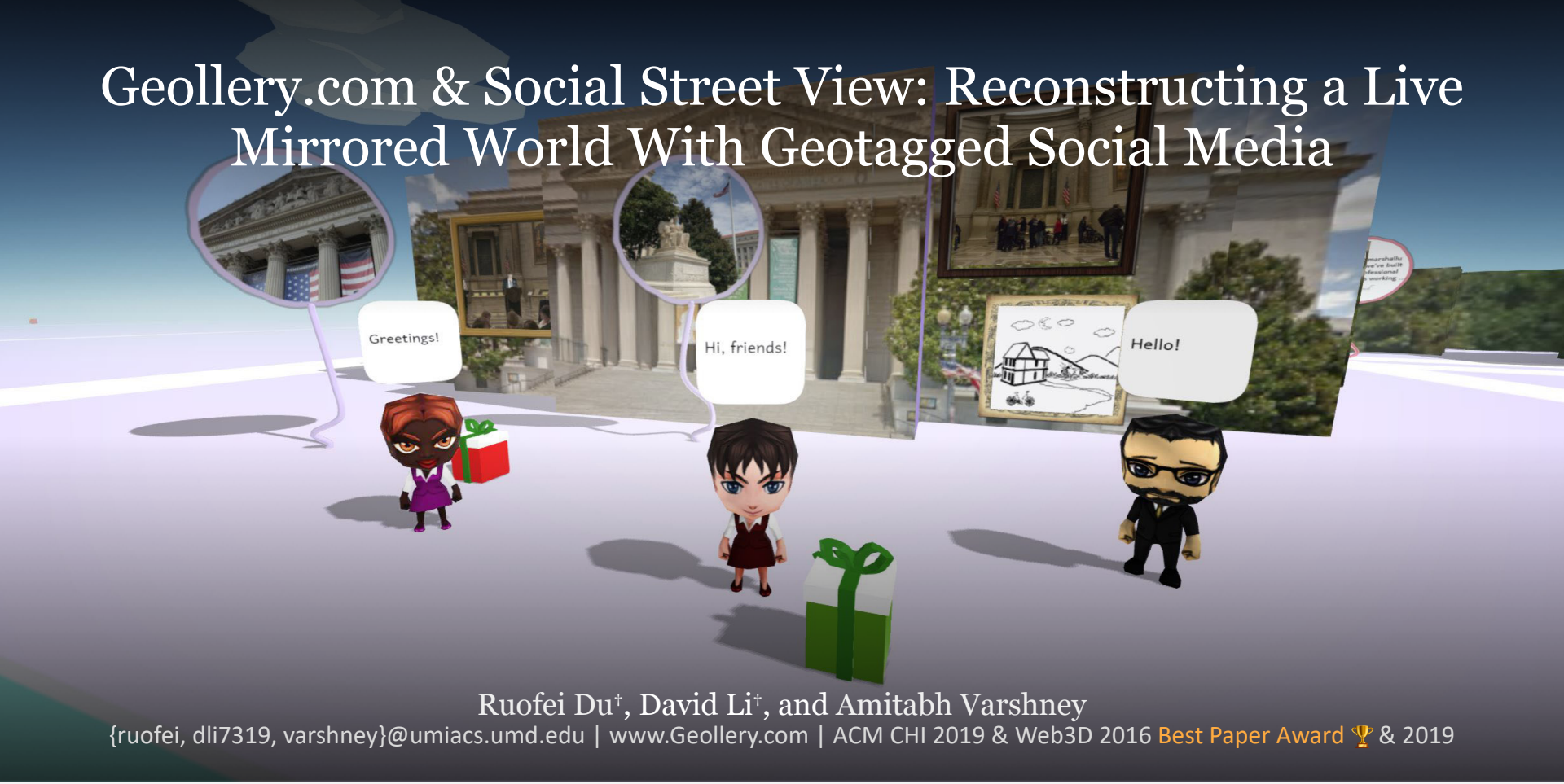

Social Street View
Immersive Tourism



Towards Augmented World...


Greetings!
Hi, friends!
Hello!

Geollery.com & Social Street View: Reconstructing a Live Mirrored World With Geotagged Social Media



Ruofei Du[†], David Li[†], and Amitabh Varshney
{ruofei, dli7319, varshney}@umiacs.umd.edu | www.Geollery.com | ACM CHI 2019 & Web3D 2016 Best Paper Award 🏆 & 2019

UNIVERSITY OF MARYLAND
UMIACS
THE AUGMENTARIUM
VIRTUAL AND AUGMENTED REALITY LAB
AT THE UNIVERSITY OF MARYLAND
COMPUTER SCIENCE
UNIVERSITY OF MARYLAND, COLLEGE PARK

System Overview

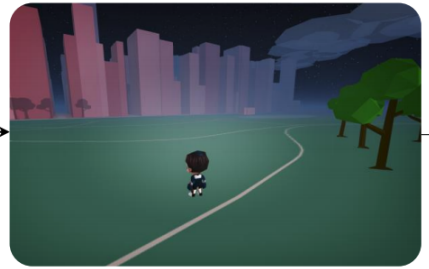
Geollery Workflow



2D polygons and metadata from OpenStreetMap



shaded 3D buildings with 2D ground tiles



added avatars, clouds, trees, and day/night effects



Geollery fuses the mirrored world with geotagged data, street view 360° images, and virtual avatars.



internal and external geotagged social media

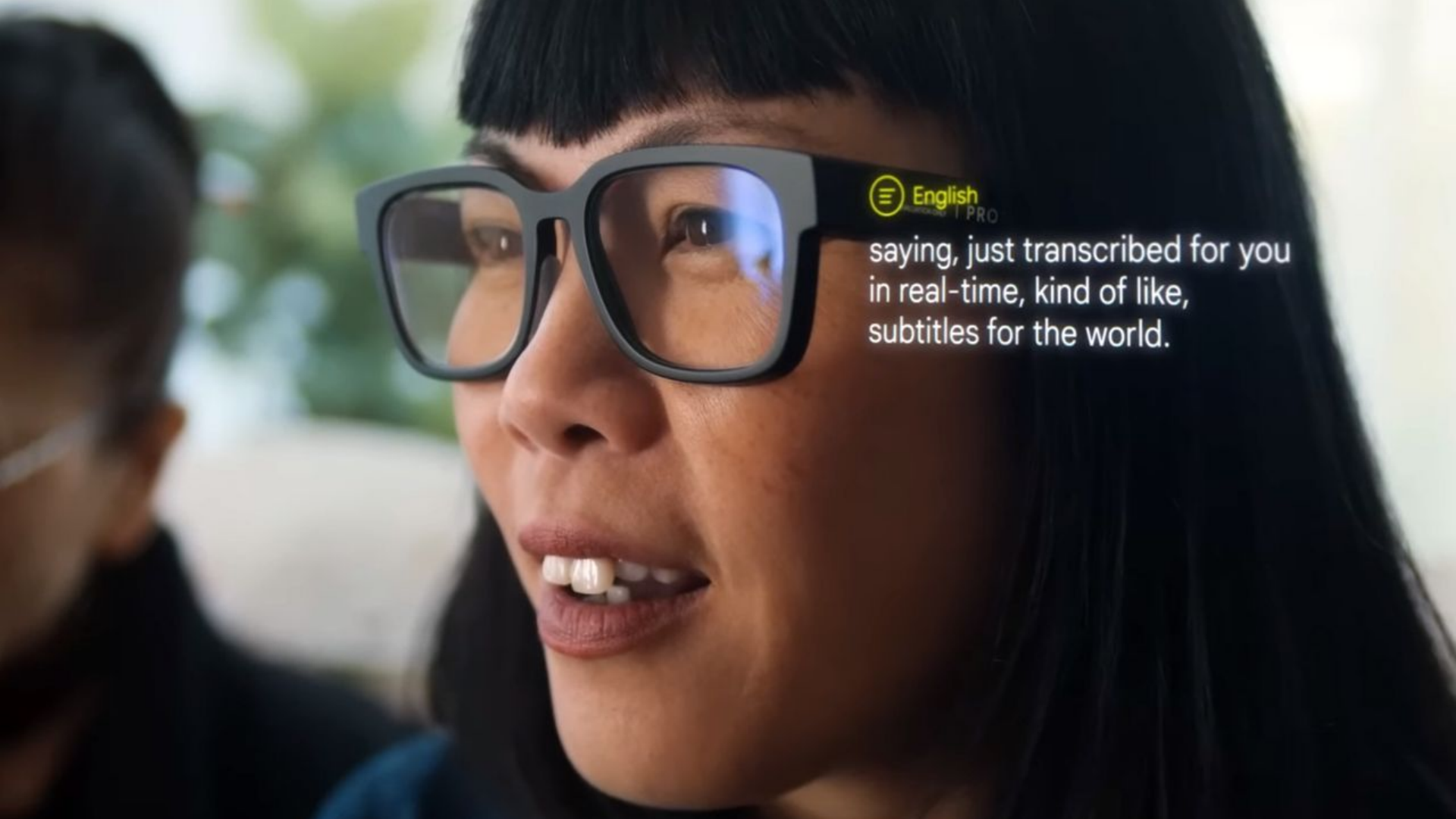






virtual forms of social media: balloons, billboards, and gifts


Shahram Izadi
Partner Research Manager


English
PRO
saying, just transcribed for you
in real-time, kind of like,
subtitles for the world.


Sasha

Total Relighting: Learning to Relight Portraits for Background Replacement

ROHIT PANDEY*, SERGIO ORTS ESCOLANO*, CHLOE LEGENDRE*, CHRISTIAN HÄNE, SOFIEN BOUAZIZ, CHRISTOPH RHEMANN, PAUL DEBEVEC, and SEAN FANELLO, Google Research

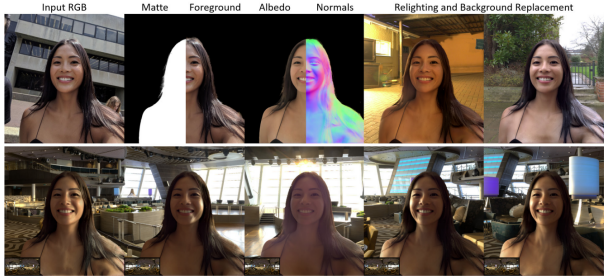


Fig. 1. Given a portrait and an arbitrary high dynamic range lighting environment, our framework uses machine learning to composite the subject into a new scene, while accurately modeling their appearance in the target illumination condition. We estimate a high quality alpha matte, foreground element, albedo map, and surface normals, and we propose a novel, per-pixel lighting representation within a deep learning framework.

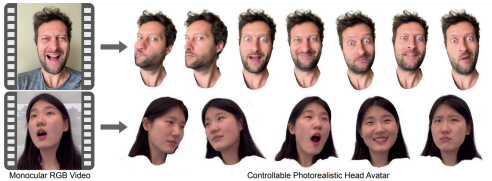
Pandey et al. ACM Trans. Graph., SIGGRAPH 2021

MonoAvatar: Learning Personalized High Quality Volumetric Head Avatars from Monocular RGB Videos

CVPR 2023

Ziqian Bai[1,2], Feitong Tan[1], Zeng Huang[1], Kripasindhu Sarkar[1], Danhang Tang[1], Di Qiu[1], Abhimitra Meka[1], Ruofei Du[1], Mingsong Dou[1], Sergio Orts-Escolano[1], Rohit Pandey[1], Ping Tan[2], Thabo Beeler[1], Sean Fanello[1], Yinda Zhang[1],
[1]Google, [2]Simon Fraser University

Paper | Supplementary | arXiv



MonoAvatar builds a 3D avatar representation of a person using just a single short monocular RGB video (e.g., 1-2 minutes). We leverage a 3DMM to track the user's expressions, and generate a volumetric photorealistic 3D avatar that can be rendered with user-defined expression and viewpoint.

MonoAvatar, Bai et al. CVPR 2023, CVPR 2024

FaceFolds: Meshed Radiance Manifolds for Efficient Volumetric Rendering of Dynamic Faces

SAFA C. MEDIN, MIT and Google, USA
GENGYAN LI, ETH Zürich and Google, Switzerland
RUOFEI DU, Google, USA
STEPHAN GARBIN, Google, United Kingdom
PHILIP DAVIDSON, Google, USA
GREGORY W. WORNELL, MIT, USA
THABO BEELER, Google, Switzerland
ABHIMITRA MEKA, Google, USA

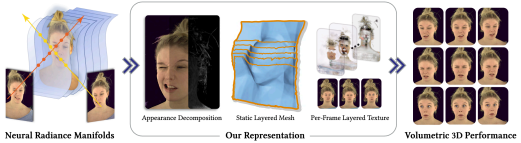


Fig. 1. We model a dynamic face sequence as a set of radiance manifolds, which are exported as a static layered mesh and an animated texture. This allows for smoothly controlling the quality vs. memory and compute footprints, while achieving efficient and photorealistic rendering of volumetric scenes using established graphics pipelines without any neural network integration.

Medin et al. SIGGRAPH I3D 2024
Best Student Paper Award

IM-Portrait: Learning 3D-aware Video Diffusion for Photorealistic Talking Heads from Monocular Videos

Yuan Li[1,2*§] Ziqian Bai[2*] Feitong Tan[2] Zhaopeng Cui[1†] Sean Fanello[2]
Yinda Zhang[2†]
[1]State Key Lab of CAD&CG, Zhejiang University [2]Google

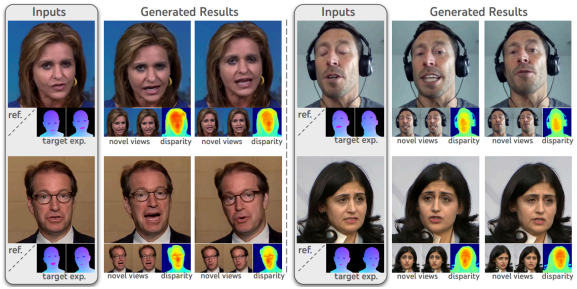


Figure 1. We propose a 3D-aware video diffusion model for talking head synthesis. Given an image as identity and a sequence of tracking signals (as shown on the left for each example), our model directly generates videos in Multiplane Images (MPIs) in a single denoising process, which is ready for efficient novel-view rendering. This enables immersive viewing experience, *e.g.* rendering binocular stereo or perspective distortion in VR headset. Please see our website for more results: https://y-u-a-n-l-i.github.io/projects/IM-Portrait/.

IM-Portrait, Li et al. CVPR 2025


How can we make real-time Avatars more accessible for non-XR users?

GazeChat

Enhancing Virtual Conferences With Gaze-Aware 3D Photos

Zhenyi He†, Keru Wang†, Brandon Yushan Feng‡,
Ruofei Du§*, Ken Perlin†*

† New York University
‡ University of Maryland, College Park
§ Google Research

* co-senior leadership

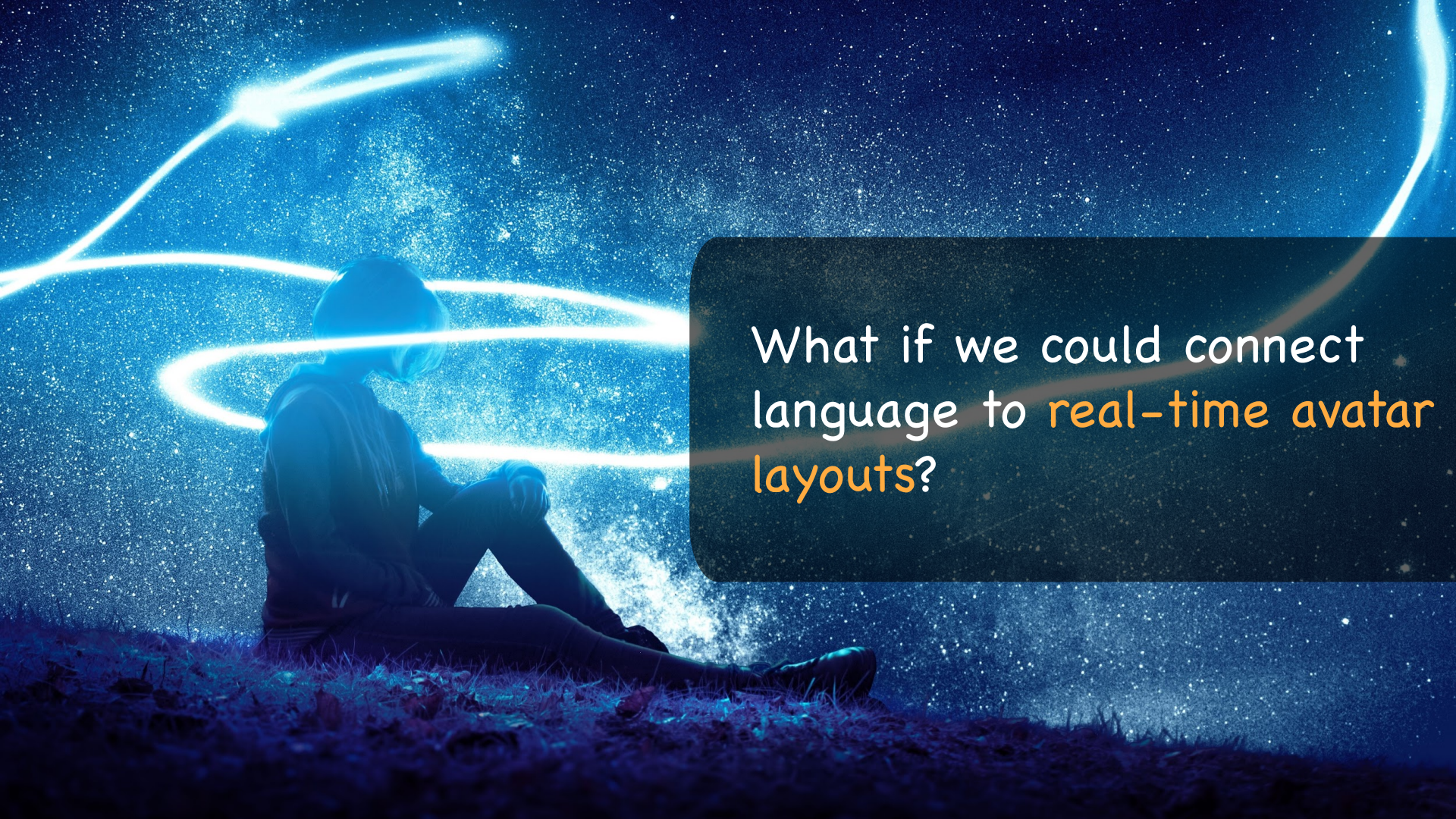


NYU FUTURE REALITY LAB

UNIVERSITY OF MARYLAND

Google

uist 2021


What if we could connect language to real-time avatar layouts?

CHI 2024
Surfing the World
11-16 May 2024



ChatDirector: Enhancing Video Conferencing with Space-Aware Scene Rendering and Speech-Driven Layout Transition



Xun Qian, Feitong Tan, Yinda Zhang, Brian Collins,
Alex Olwal, David Kim, Karthik Ramani, and Ruofei Du

Google

PURDUE UNIVERSITY



Likeness (beta)

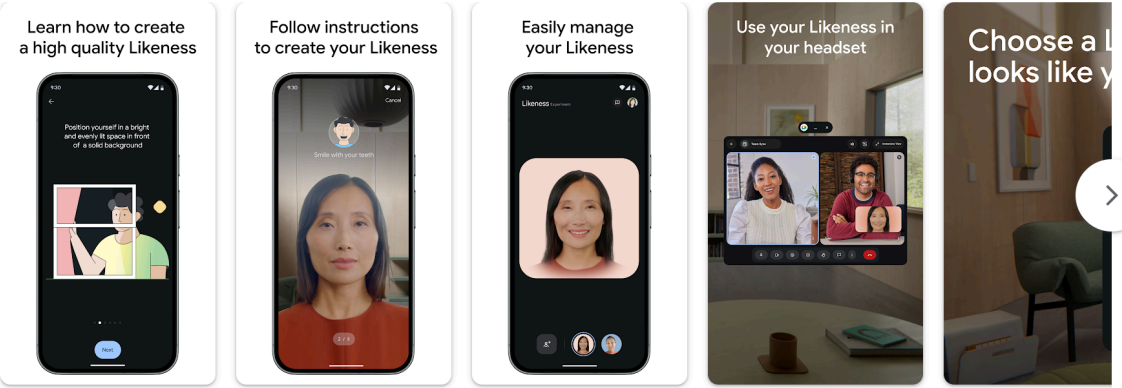
Google LLC

2.5★ 6 reviews | 10K+ Downloads | Everyone

Install on more devices | Share

This app is available for some of your devices

You can share this with your family. Learn more about Family Library



App support

More apps to try

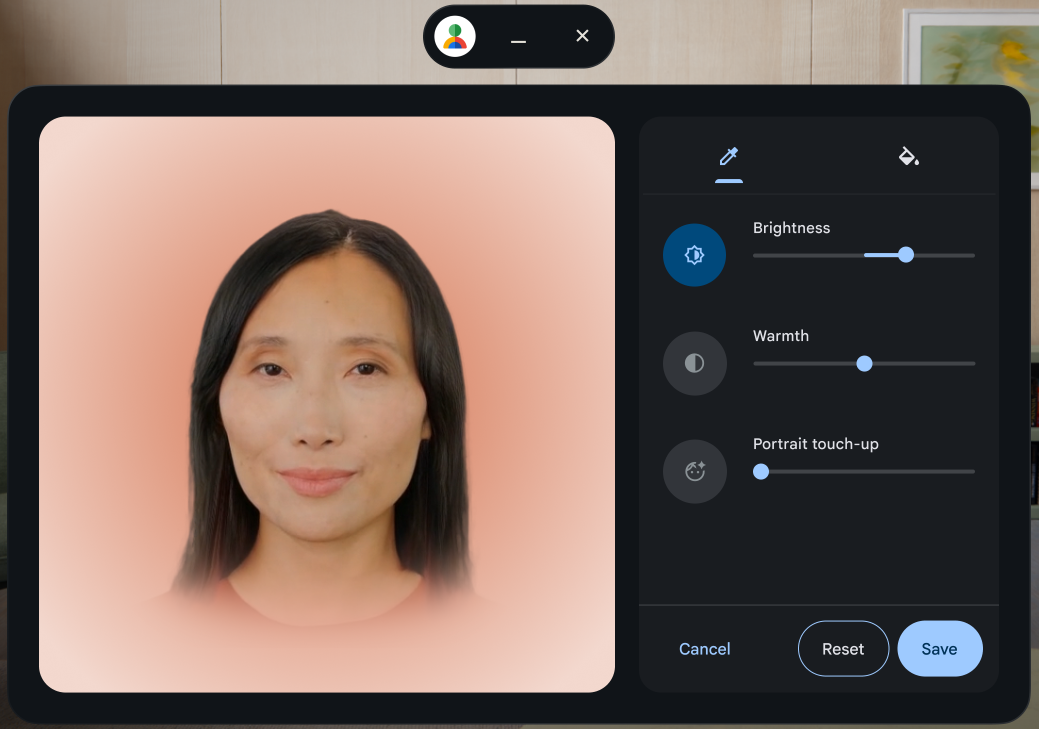
Japan Travel–Trip,Transit&Rail
NAVITIME JAPAN CO., LTD.
4.6★

Amazon Music: Songs & Podcasts
Amazon Mobile LLC
3.8★

Choose a Likeness that looks like you



Enhance your appearance




How can we further augment verbal communication in videoconferencing, today?

Visual Captions

Augmenting Verbal Communication with On-the-fly Visuals





Xingyu “Bruce” Liu, Vladimir Kirilyuk, Xiuxiu Yuan,
Alex Olwal, Peggy Chi, Xiang "Anthony" Chen, **Ruofei Du**

github.com/google/archat

CHI23
Hamburg, Germany | Hybrid
April 23-28, 2023
reCHInnecting

Google

UCLA HCI RESEARCH

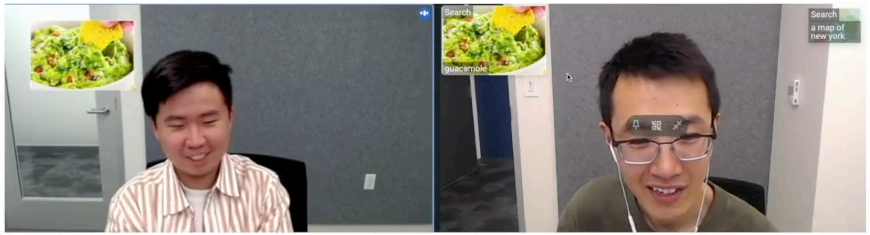
ARChat Chrome Plugin

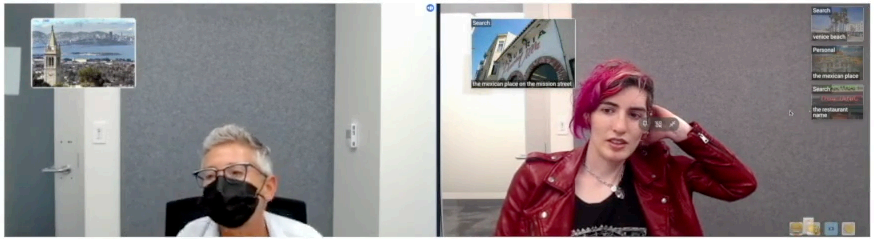


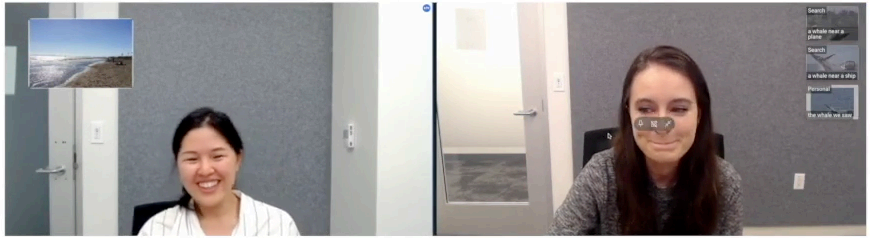
You

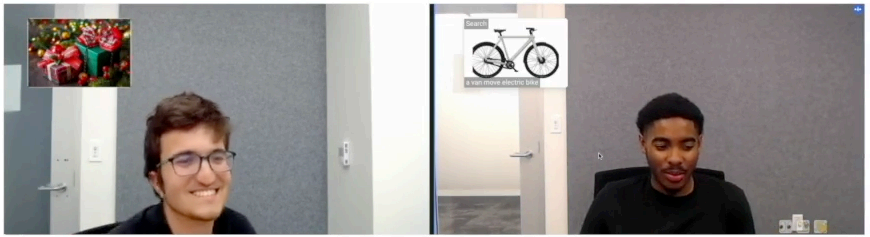
2:03 PM | uys-pdyp-bie

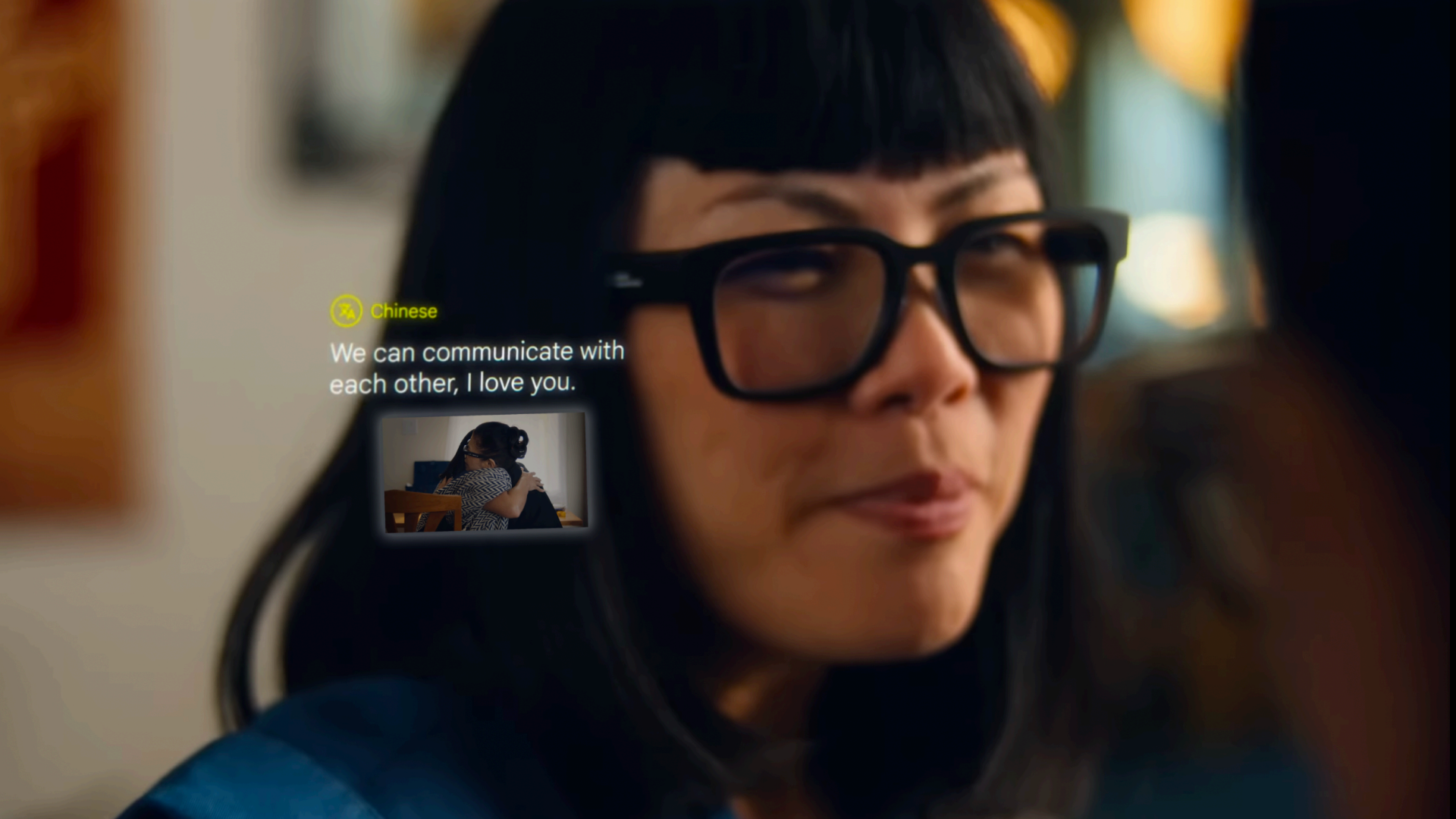
User Studies
N=26



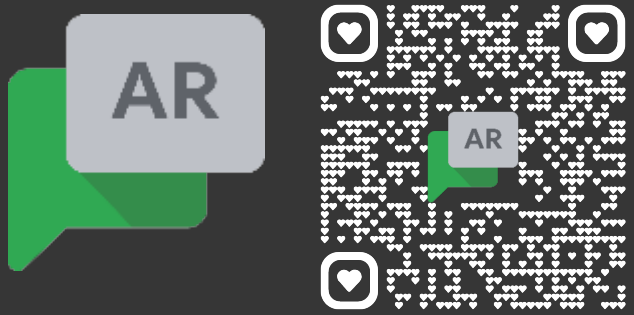







Chinese
We can communicate with each other, I love you.

ARChat



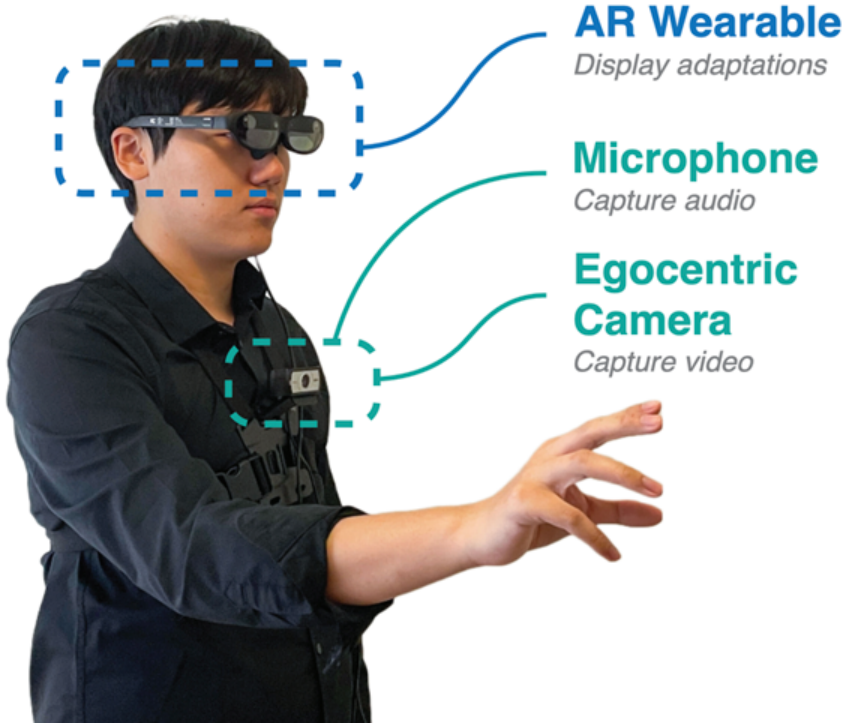
github.com/google/archat

With ARChat, the HCI / AR / VR / NLP community can make communication more interactive, effective, and accessible with real world impact.

CHI 2024 Best Paper
Honorable Mention

Human I/O

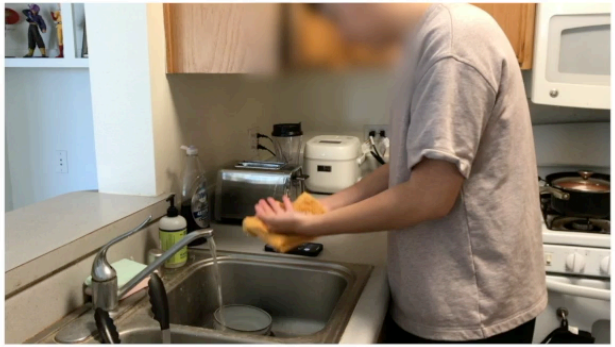
Towards a Unified Approach to Detecting Situational Impairments



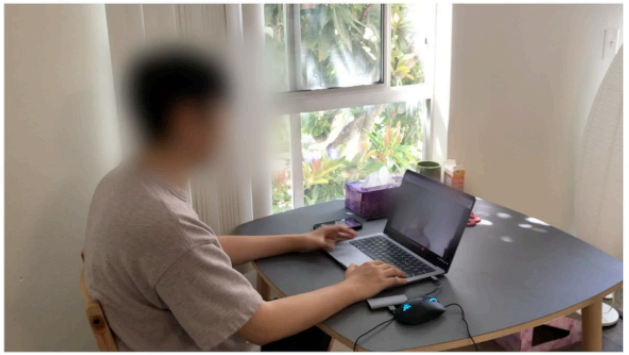
Xingyu Bruce Liu, Jiahao Nick Li, David Kim,
Xiang 'Anthony' Chen, and Ruofei Du

Google
UCLA HCI RESEARCH
CHI 2024
Surfing the World

SIIDs as Human I/O Channel Availability



Hands / Fingers: *hands are wet*



Vision: *need to make eye contact*



Hearing: *noisy hairdryer*

SIIDs as Human I/O Channel Availability

Human Input	Example use cases
Vision	*Read text message, watch video, etc.*
Hearing	*Hear notification, phone call, etc.*
Tactile	*Haptic feedback, feel temperature, etc.*
Taste	*Drink, eat, etc.*
Smell	*Smell, etc.*

Human Output	
Eyes / Gaze	*Face ID, gaze-based interaction, etc.*
Vocal System	*Conversation, voice assistants, etc.*
Hands / Fingers	*Touch screen, gesture control, etc.*
Limbs / Movement	*Walk, reach, etc.*
Head / Face	*Nodding, facial expression, etc.*



Sensible Agent:

A Framework for Unobtrusive Interactions with Proactive AR Agents



Geonsun Lee, Min Xia, Nels Numan, Xun Qian, David Li, Yanhe Chen,
Achin Kulshrestha, Ishan Chatterjee, Yinda Zhang, Dinesh Manocha, David Kim, Ruofei Du

Google

UNIVERSITY OF MARYLAND 18 56 DEPARTMENT OF COMPUTER SCIENCE

UIST 25





Google Project Astra

The next era of XR agents is here — but still **reactive**

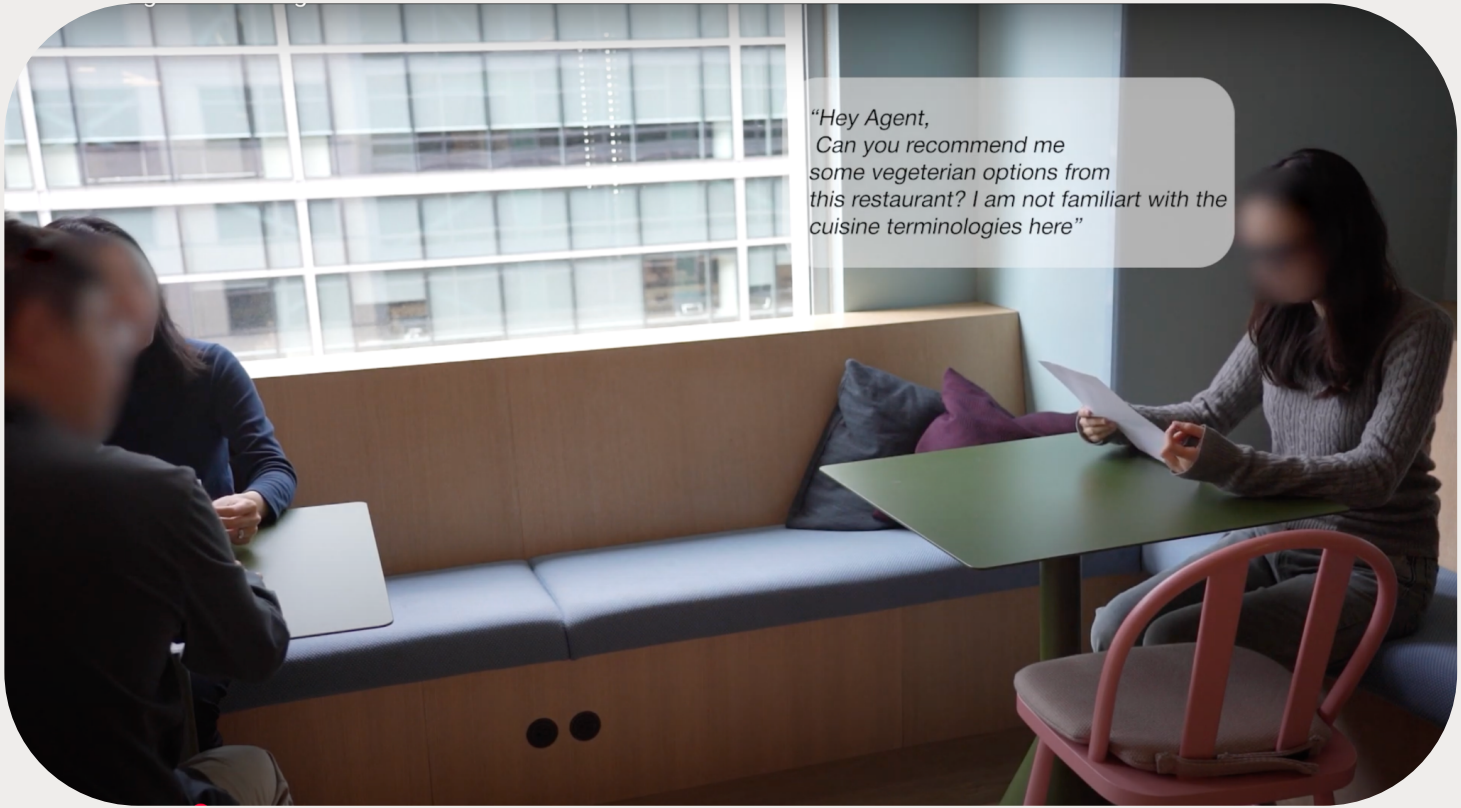
But.. voice input is not always possible or comfortable







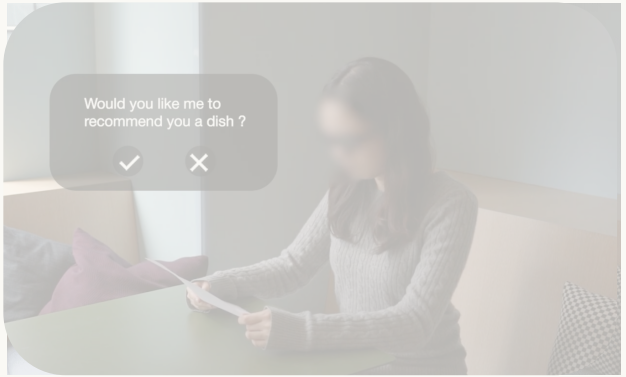
When users can't talk
How should agents talk to them?


“Hey Agent,
Can you recommend me
some vegeterian options from
this restaurant? I am not familiart with the
cuisine terminologies here”

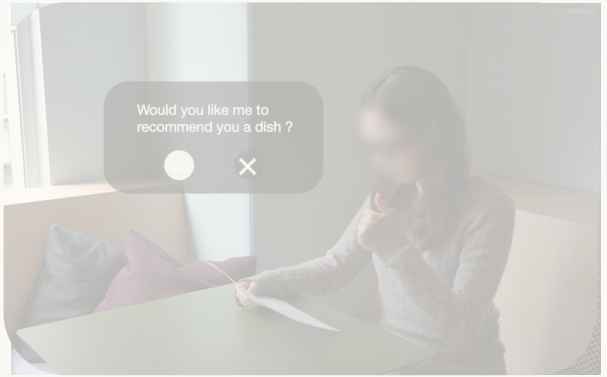
Sensible Agent Overview



Context Detection



Query Presentation



Unobtrusive User Response

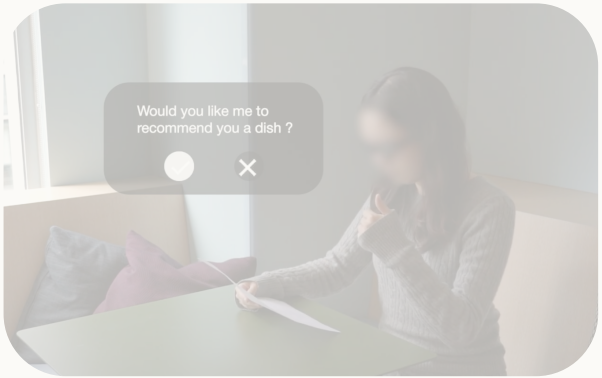
Sensible Agent Overview



Context Detection



Query Presentation

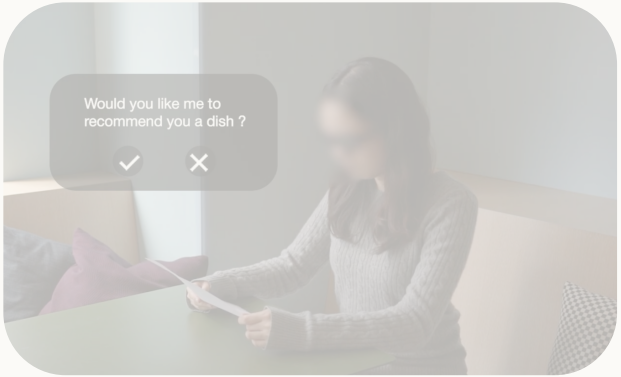


Unobtrusive User Response

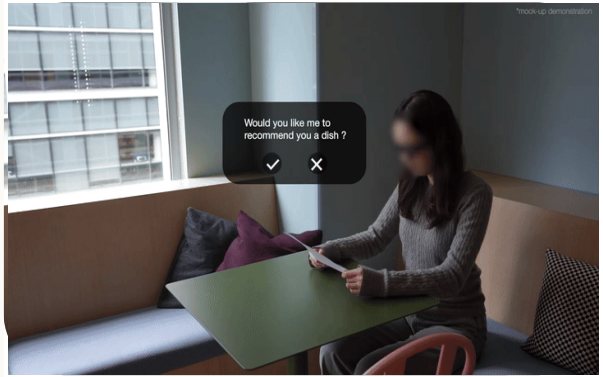
Sensible Agent Overview



Context Detection



Query Presentation

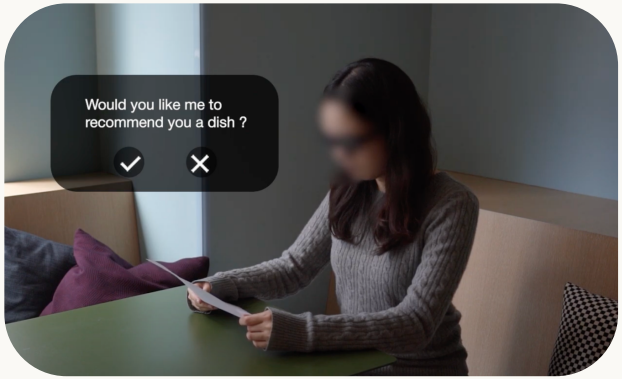


Unobtrusive User Response

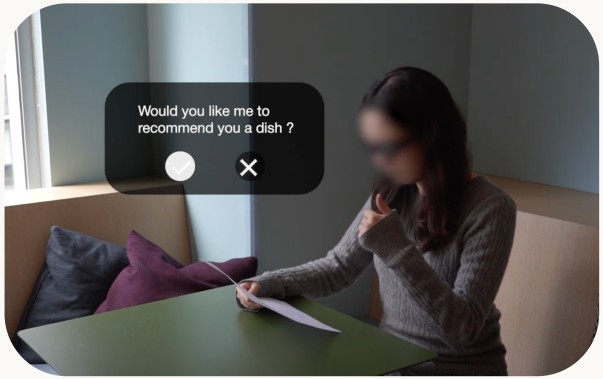
Sensible Agent Overview



Context Detection



Query Presentation

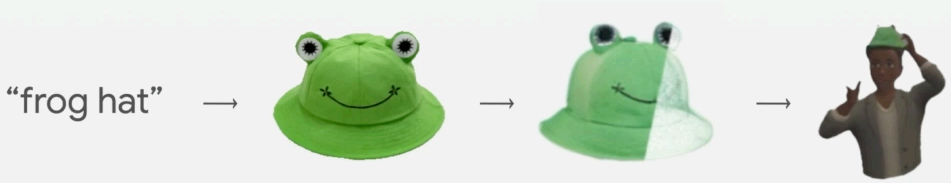


Unobtrusive User Response


How can we augment communication with everyday objects, today?

Thing2Reality

Transforming 2D Content into Conditioned Multiviews and 3D Gaussian Objects for XR Communication

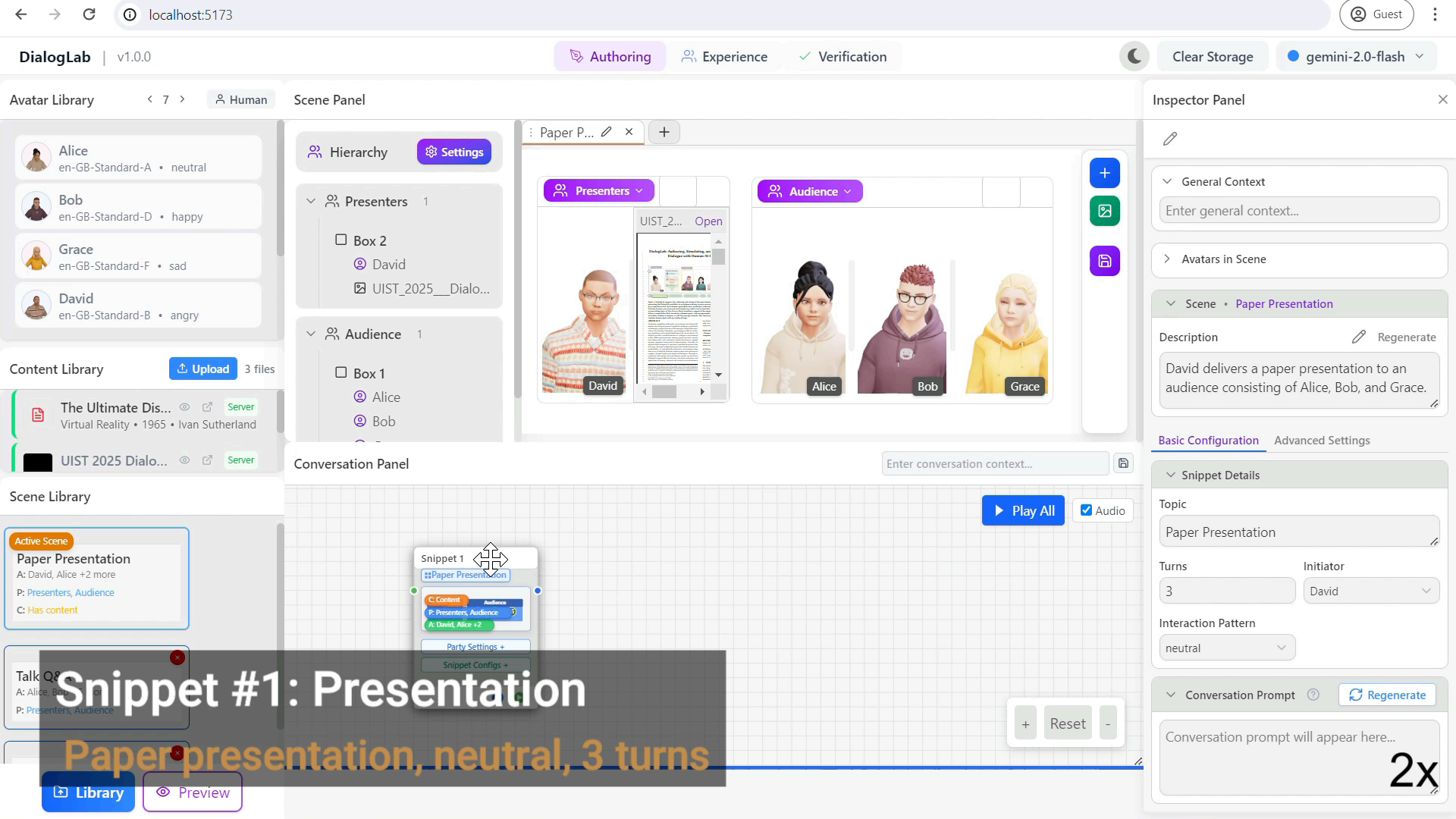


Erzhen Hu, Mingyi Li, Xun Qian, Alex Olwal,
David Kim, Seongkook Heo, Ruofei Du

Google

University of Virginia

UIST
PITTSBURGH | 2024


localhost:5173
Guest
DialogLab | v1.0.0
Authoring
Experience
Verification
Clear Storage
gemini-2.0-flash
Avatar Library
7
Human
Alice
en-GB-Standard-A • neutral
Bob
en-GB-Standard-D • happy
Grace
en-GB-Standard-F • sad
David
en-GB-Standard-B • angry
Content Library
Upload
3 files
The Ultimate Dis...
Server
Virtual Reality • 1965 • Ivan Sutherland
UIST 2025 Dialo...
Server
Scene Library
Active Scene
Paper Presentation
A: David, Alice +2 more
P: Presenters, Audience
C: Has content
Talk Q&
A: Alice, Bob
P: Presenters, Audience
Library
Preview
Scene Panel
Hierarchy
Settings
Presenters 1
Box 2
David
UIST_2025___Dialo...
Audience 1
Box 1
Alice
Bob
Paper P...
Presenters
UIST_2...
Open
David
Audience
Alice
Bob
Grace
Conversation Panel
Enter conversation context...
Play All
Audio
Snippet 1
Paper Presentation
C: Content
Audience
P: Presenters, Audience
A: David, Alice +2
Party Settings +
Snippet Configs +
+
Reset
-
Inspector Panel
General Context
Enter general context...
Avatars in Scene
Scene • Paper Presentation
Description
Regenerate
David delivers a paper presentation to an audience consisting of Alice, Bob, and Grace.
Basic Configuration
Advanced Settings
Snippet Details
Topic
Paper Presentation
Turns
3
Initiator
David
Interaction Pattern
neutral
Conversation Prompt
Regenerate
Conversation prompt will appear here...
2x
Snippet #1: Presentation
Paper presentation, neutral, 3 turns

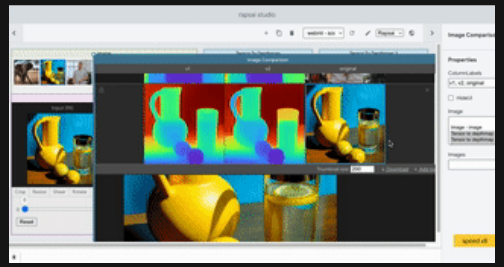
Chapter II · Augmented Communication



Honorable Mention

rapsai

Accelerating Machine Learning Prototyping of Multimedia Applications through Visual Programming



Ruofei Du, Na Li, Jing Jin, Michelle Carney, Scott Miles, Maria Kleiner, Xiuxiu Yuan, Yinda Zhang, Anuva Kulkarni, Xingyu "Bruce" Liu, Ahmed Sabie, Sergio Escolano, Abhishek Kar, Ping Yu, Ram Iyengar, Adarsh Kowdle, and Alex Olwal

visualblocksforml.github.io

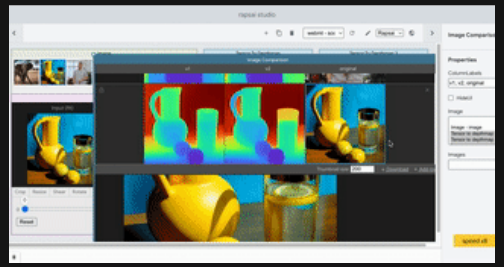
Google

CHI23
Hamburg, Germany | Hybrid
April 23-28, 2023
reCHInnecting

Honorable Mention

Visual Blocks

Accelerating Machine Learning Prototyping of Multimedia Applications through Visual Programming



Ruofei Du, Na Li, Jing Jin, Michelle Carney, Scott Miles, Maria Kleiner, Xiuxiu Yuan, Yinda Zhang, Anuva Kulkarni, Xingyu "Bruce" Liu, Ahmed Sabie, Sergio Escolano, Abhishek Kar, Ping Yu, Ram Iyengar, Adarsh Kowdle, and Alex Olwal

visualblocksforml.github.io

Google

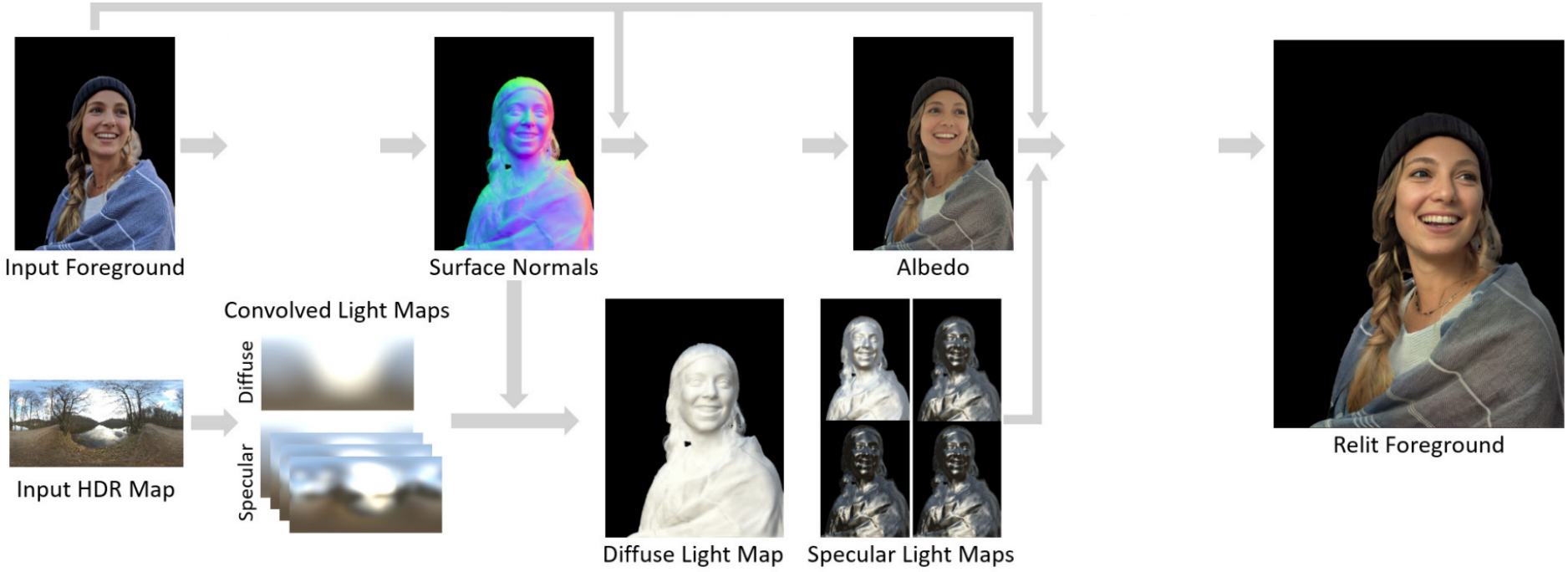
CHI23
Hamburg, Germany | Hybrid
April 23-28, 2023
reCHInnecting


3D Photos

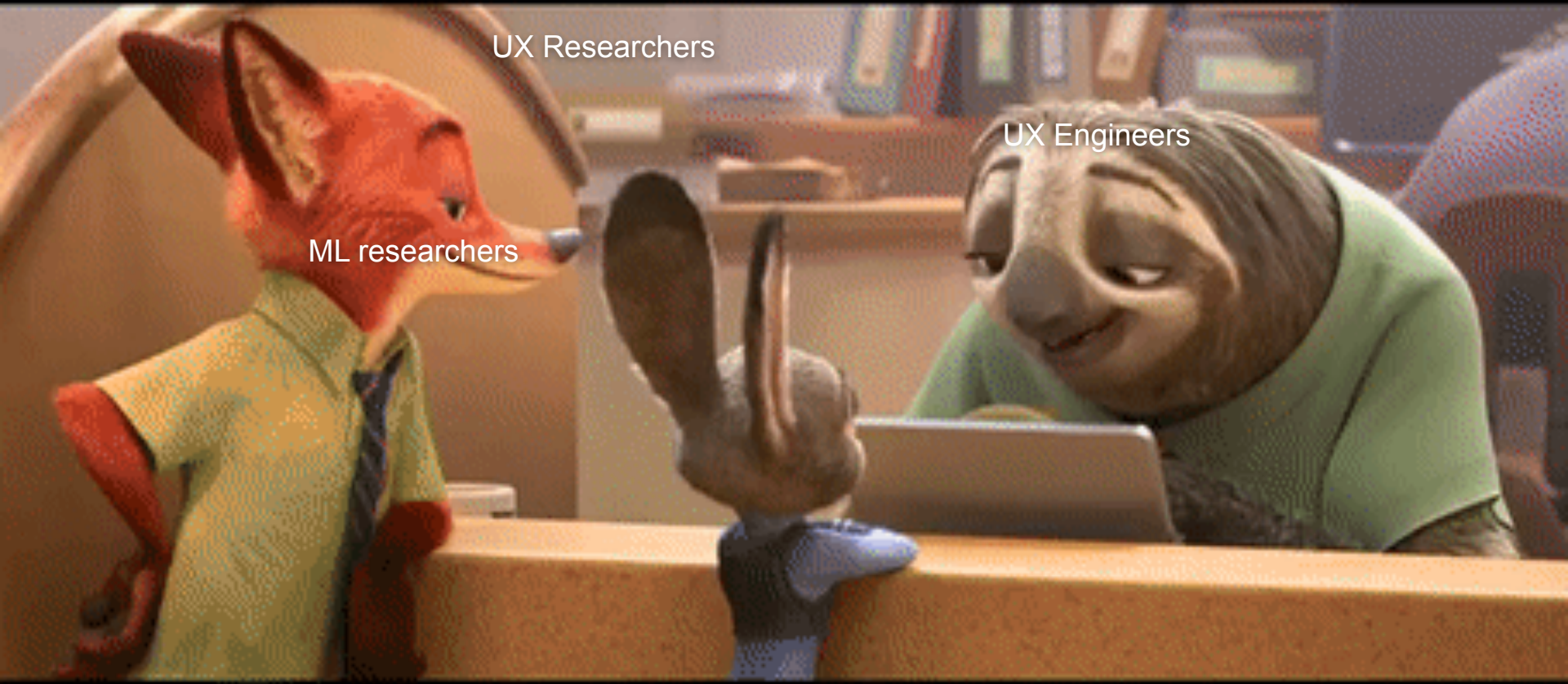

Portrait Relighting


Virtual Background
MediaPipe





Pandey, Rohit, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. "Total relighting: learning to relight portraits for background replacement." ACM Transactions on Graphics (SIGGRAPH 2021) 40, no. 4 (2021): 1-21.


UX Researchers
UX Engineers
ML researchers

Related Work

Visual programming in language

TALOR: Generating

Alexis Ross*† Tongshuang
†Allen Institu
◇Paul G. Allen School of C
{a
{wts

Abstract

Controlled text perturbation is
uating and improving model
However, current techniques
a model for every target pertur
expensive and hard to genera
TALOR, a semantically-contr
ation system. TALOR builds
seq2seq model and produces
conditioned on control codes
mantic representations. We
erations to modify the cont
in turn steer generation to
tributes. These operations c
posed into higher-level ones
ible perturbation strategies.
the effectiveness of these pe
tiple applications. First,
automatically create high-q
for four distinct natural la
(NLP) tasks. These contras
spurious artifacts and are
manually annotated ones
versity. Second, we sho
turbations can improve m

PromptChainer: Chaining Large Langu through Visual Program

Tongshuang Wu*†
wtshuang@cs.washington.edu
University of Washington
USA

Ellen Jiang†
ellenj@google.com
Google Research
USA

Jeff Gray
jeffgray@google.com
Google Research
USA

Alejandra Molina
alemolinata@google.com
Google Research
USA

Carrie J. Cai
cjcai@google.com
Google Research
USA

ABSTRACT

While LLMs have made it possible to rapidly prototype new ML functionalities, many real-world applications involve complex tasks that cannot be easily handled via a single run of an LLM. Recent work has found that chaining multiple LLM runs together (with the output of one step being the input to the next) can help users accomplish these more complex tasks, and in a way that is perceived to be more transparent and controllable. However, it remains unknown what users need when *authoring* their own LLM chains – a key step to lowering the barriers for non-AI-experts to prototype AI-infused applications. In this work, we explore the LLM chain authoring process. We find from pilot studies that users need support transforming data between steps of a chain, as well as debugging the chain at multiple granularities. To address these needs, we designed *PromptChainer*, an interactive interface for visually programming chains. Through case studies with four designers and developers, we show that PromptChainer supports building prototypes for a range of applications, and conclude with open questions on scaling chains to even more complex tasks, as well as supporting low-fi chain prototyping.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; *Interactive systems and tools*; • **Computing methodologies** → Machine learning.

*The work was done when the author was an intern at Google Inc.
†Equal contribution.

This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9156-6/22/04.
https://doi.org/10.1145/3491101.3519729

AI Chains: Transparent and Controlla by Chaining Large Languag

Tongshuang Wu*
wtshuang@cs.washington.edu
University of Washington
USA

Michael Terry
michaelterry@google
Google Researc
USA

ABSTRACT

Although large language models (LLMs) have demonstrated impressive potential on simple tasks, their breadth of scope, lack of transparency, and insufficient controllability can make them less effective when assisting humans on more complex tasks. In response, we introduce the concept of *Chaining* LLM steps together, where the output of one step becomes the input for the next, thus aggregating the gains per step. We first define a set of *LLM primitive operations* useful for Chain construction, then present an interactive system where users can modify these Chains, along with their intermediate results, in a modular way. In a 20-person user study, we found that Chaining not only improved the quality of task outcomes, but also significantly enhanced system transparency, controllability, and sense of collaboration. Additionally, we saw that users developed new ways of interacting with LLMs through Chains: they leveraged sub-tasks to calibrate model expectations, compared and contrasted alternative strategies by observing parallel downstream effects, and debugged unexpected model outputs by "unit-testing" sub-components of a Chain. In two case studies, we further explo how LLM Chains may be used in future applications.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in** *Interactive systems and tools*; • **Computing methodologi** Machine learning.

KEYWORDS

Human-AI Interaction, Large Language Models, Natural L Processing

ACM Reference Format:
Tongshuang Wu, Michael Terry, and Carrie J. Cai. 2022. AI Ch parent and Controllable Human-AI Interaction by Chaining Lar Model Prompts. In *CHI Conference on Human Factors in Comp (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, USA, 22 pages. https://doi.org/10.1145/3491102.3517582

*The work was done when the author was an intern at Google Inc

This work is licensed under a Creative Commons Attribution 4.0 License.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3517582

TaleBrush: Sketching Stories with Generative Pretrained Language Models

John Joon Young Chung
jjyc@umich.edu
University of Michigan
Ann Arbor, MI, USA

Wooseok Kim
ooooseok@kaist.ac.kr
KAIST
Daejeon, Republic of Korea

Kang Min Yoo
kangmin.yoo@navercorp.com
Naver AI LAB
Seongnam, Republic of Korea

Hwaran Lee
hwaran.lee@navercorp.com
Naver AI LAB
Seongnam, Republic of Korea

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

Minsuk Chang
minsuk.chang@navercorp.com
Naver AI LAB
Seongnam, Republic of Korea



Figure 1: TaleBrush uses a line sketching interaction for intuitive control and sensemaking of story generation with GPT-based language models, closing the gap in the iterative co-creation process between humans and AI. In our prototype, the protagonist's 'fortune' is controllable. The writer first decides the protagonist's name (A1), writes a portion of a story (A2), and then sketches how the protagonist's fortune should change in the story (A3, green shaded area). The x position of the sketched line indicates the chronological position in the story (sentences are the units). The y position shows how good or bad the protagonist's fortune should be (higher, better). The width of the line indicates the possible variance in the fortune of the generated sentences. Given the line sketch, TaleBrush will generate story sentences (B1, indicated with blue) and visualize the result on the original sketch (B2, the blue line and dots). The writer can always directly edit the generated text. They can also iterate by generating new text for the same sketch (B3, clicking on 'Generating Again') or revising their sketch and generating new text. Additional options allow the writer to specify how 'surprising' the generation should be (B4) or using the eraser tool (B5) to erase their sketch. Where the line is erased, TaleBrush generates unconstrained sentences. A history dropdown (B6) allows the writer to browse previously generated sentences.

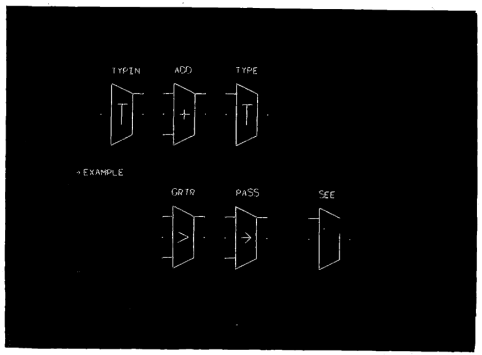
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI '22, April 29-May 5, 2022, New Orleans, LA, USA
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3501819

ABSTRACT

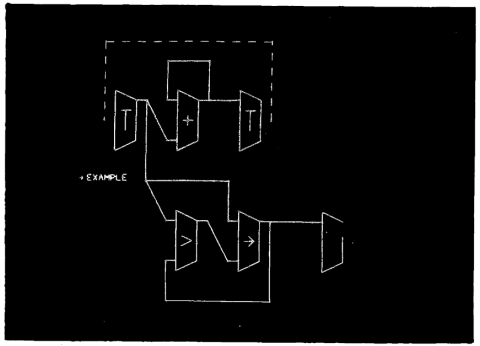
While advanced text generation algorithms (e.g., GPT-3) have enabled writers to co-create stories with an AI, guiding the narrative remains a challenge. Existing systems often leverage simple turn-taking between the writer and the AI in story development. However, writers remain unsupported in intuitively understanding the AI's actions or steering the iterative generation. We introduce TaleBrush, a generative story ideation tool that uses line sketching interactions with a GPT-based language model for control and sensemaking of a protagonist's fortune in a

Related Work

Visual programming in graphics



Basic Symbols
Figure 1.1



Connected Program
Figure 1.2

MASS. INST. OF TECHNOLOGY
APR 5 1966
LIBRARIES

THE ON-LINE GRAPHICAL SPECIFICATION OF COMPUTER PROCEDURES

by

WILLIAM ROBERT SUTHERLAND

B.E.E., Rensselaer Polytechnic Institute

(1957)

M.S., Massachusetts Institute of Technology

(1963)

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

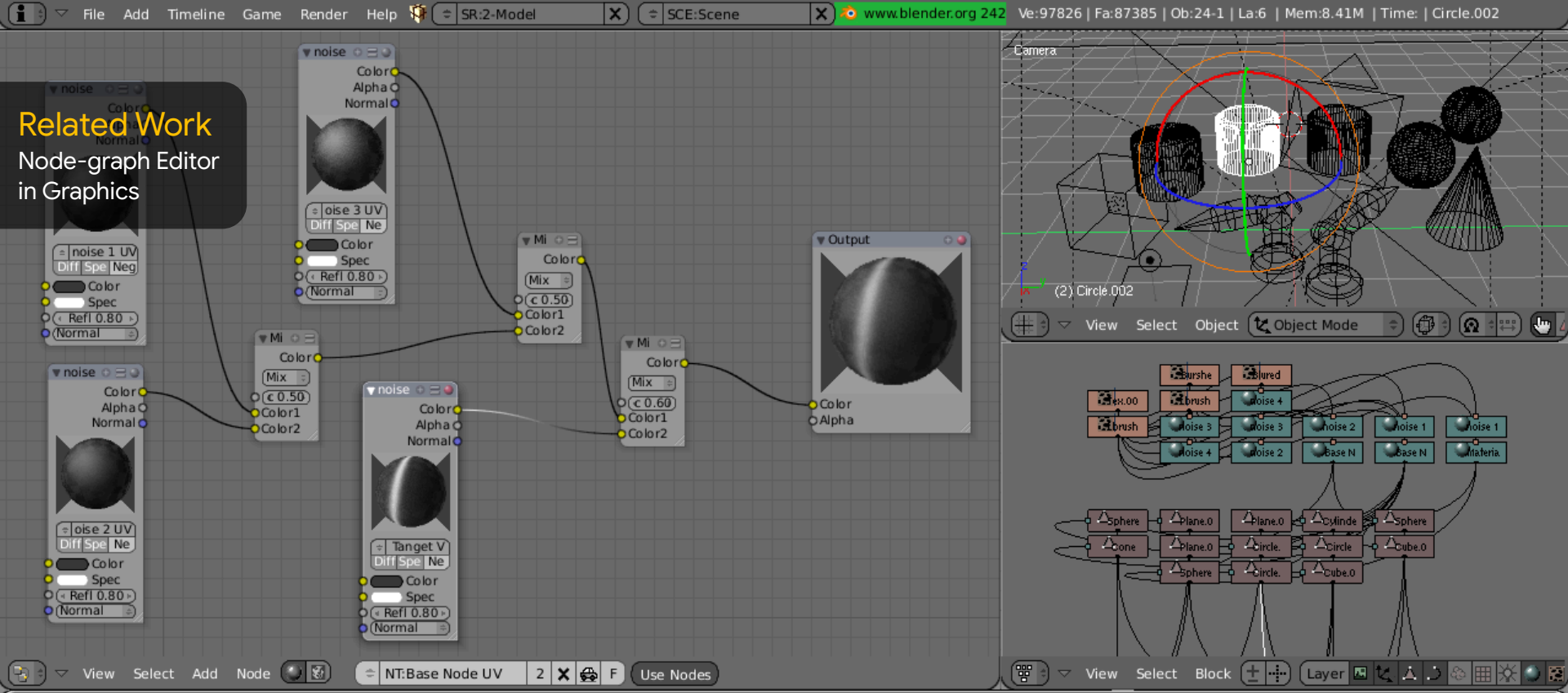
January, 1966

Signature of Author.....
Department of Electrical Engineering, January 10, 1966

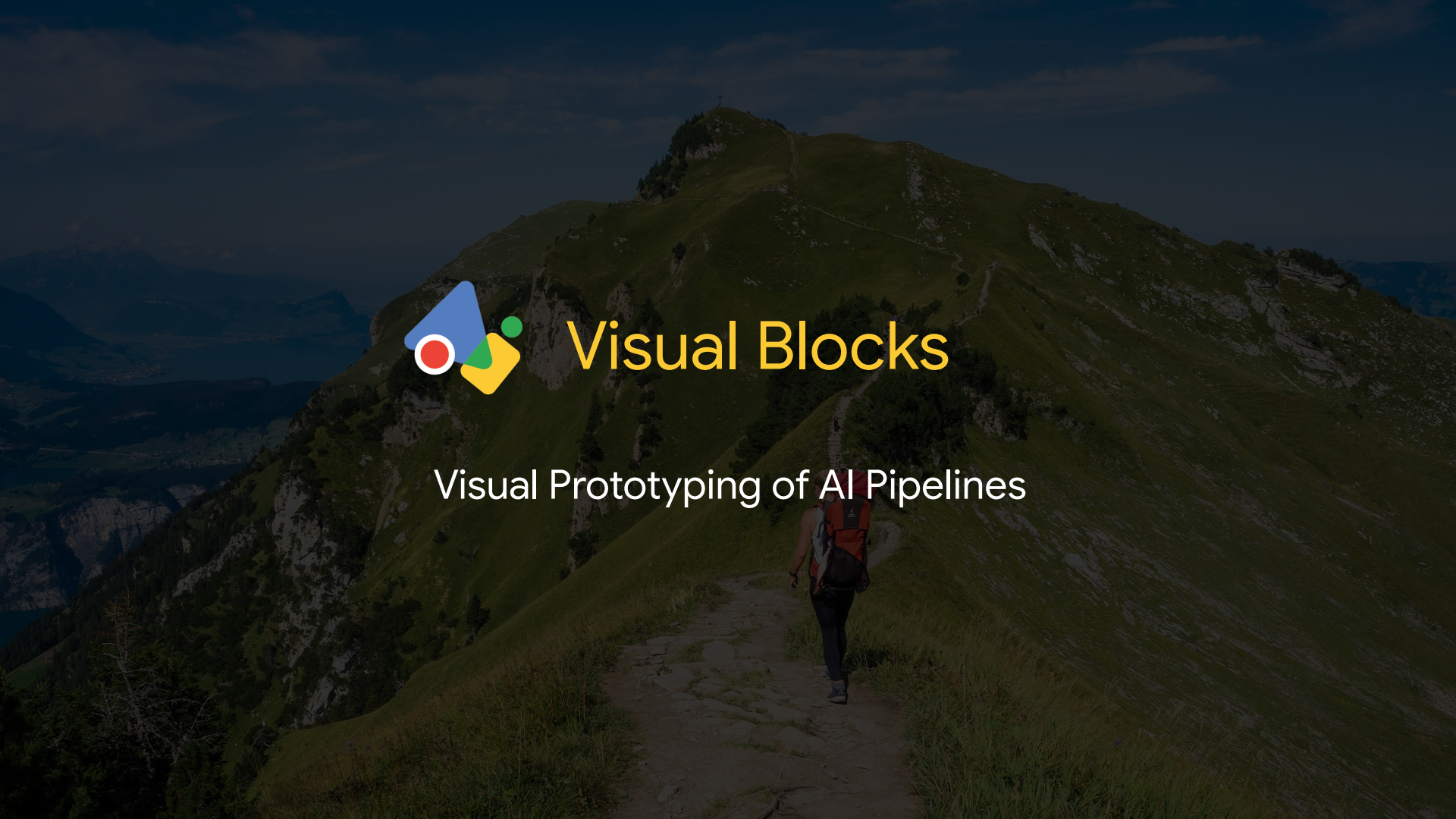
Certified by.....
Thesis Supervisor

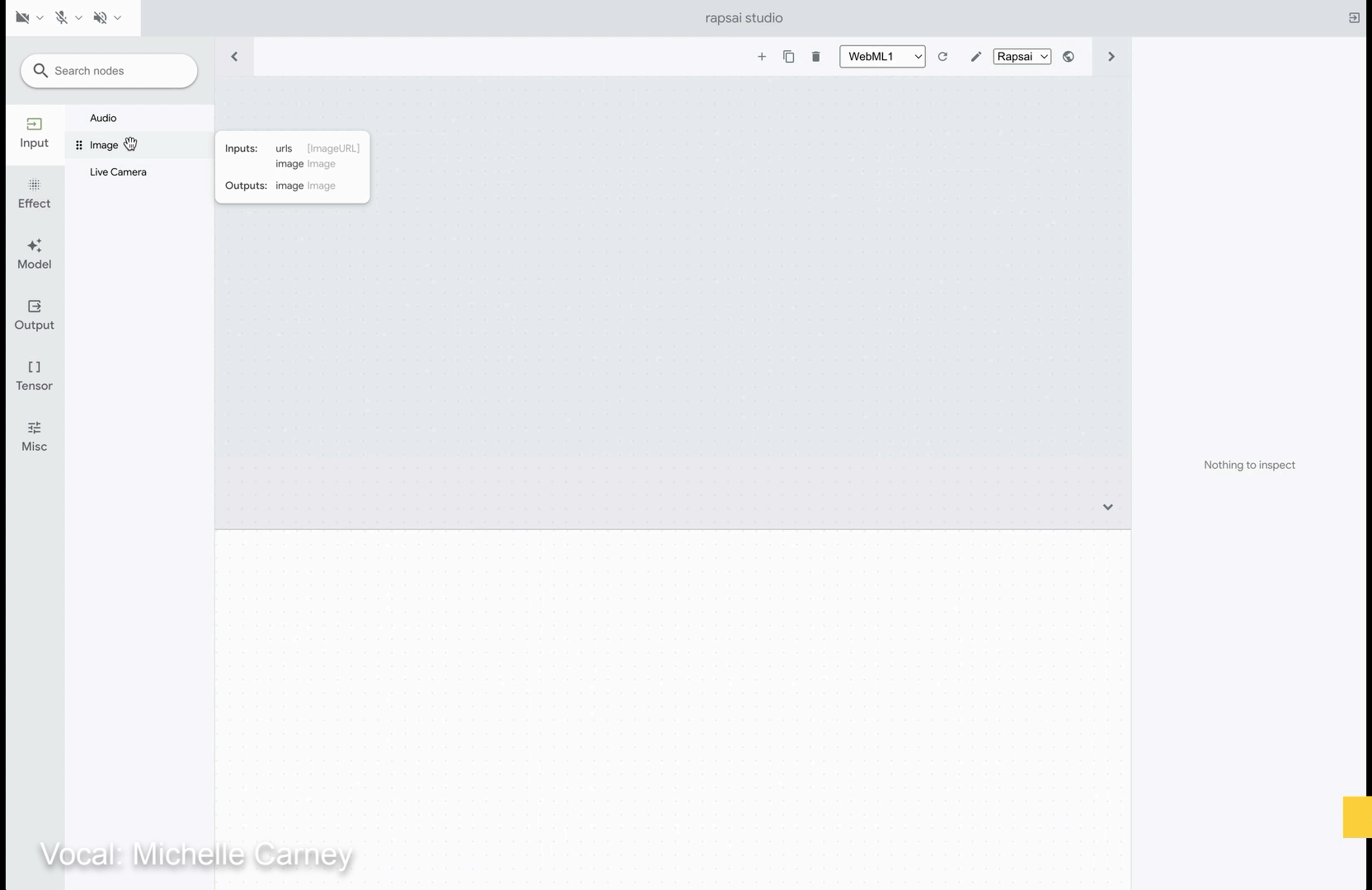
Accepted by.....
Chairman, Departmental Committee on Graduate Students

WR Sutherland. 1966. On-Line Graphical Specification of Procedures. SJCC, Boston, Mass (1966).


Related Work
Node-graph Editor
in Graphics
File Add Timeline Game Render Help
SR:2-Model
SCE:Scene
www.blender.org 242
Ve:97826 | Fa:87385 | Ob:24-1 | La:6 | Mem:8.41M | Time: | Circle.002
Camera
(2) Circle.002
View Select Object Object Mode
View Select Add Node
NT:Base Node UV
Use Nodes
View Select Block
Layer
Panels
Preview
Material
Shaders
Mirror Transp
Texture
Map To
Map Input
Ramps
Links and Nodes


person 0.878
baseball glove 0.157
person 0.888
person 0.864
baseball glove 0.769
d: 2
potted plant,0.62
d: 1
aptop,0.86


Visual Blocks
Visual Prototyping of AI Pipelines


rapsai studio
Search nodes
Input
Effect
Model
Output
Tensor
Misc
Audio
Image
Live Camera
Inputs: urls [ImageURL]
image Image
Outputs: image Image
WebML1
Rapsai
Nothing to inspect
Vocal: Michelle Carney
speed x6

rapsai studio

Search nodes

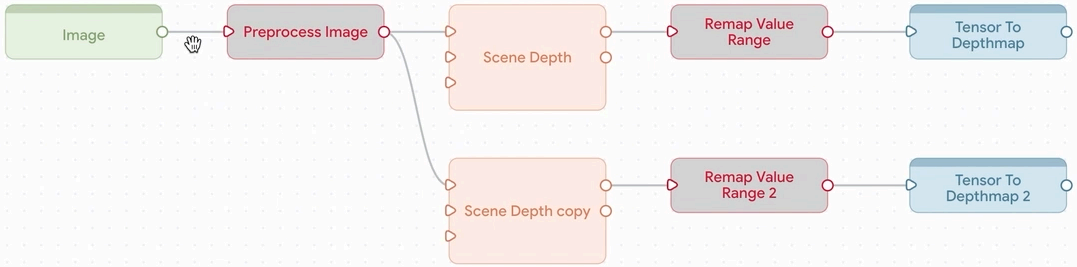
Input
Effect
Model
Output
Tensor
Misc

Audio
Image
Live Camera

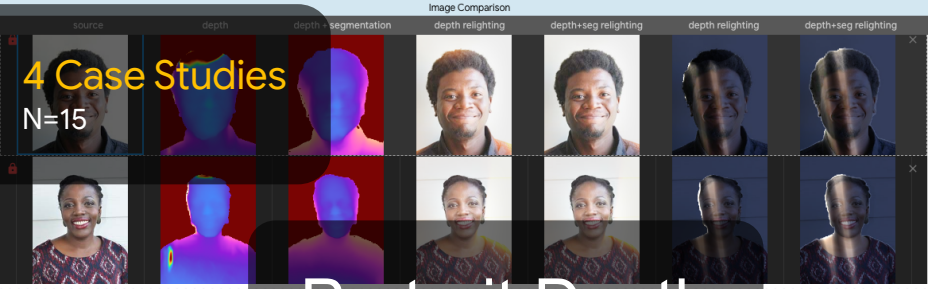
webml - sce
Rapsai



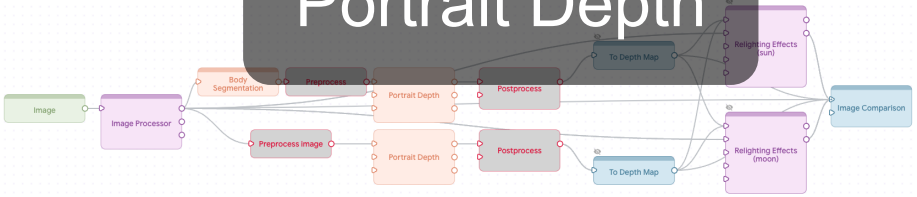
Nothing to inspect

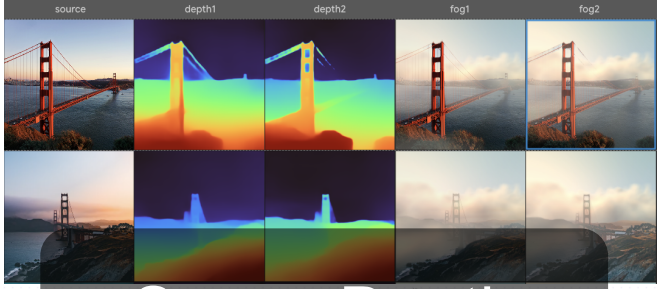


speed x8

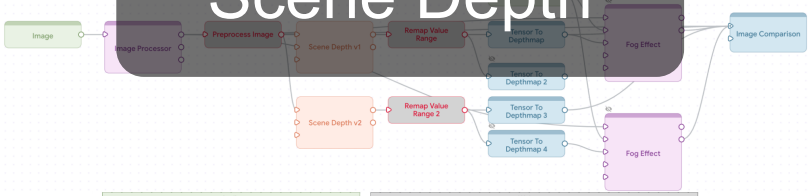


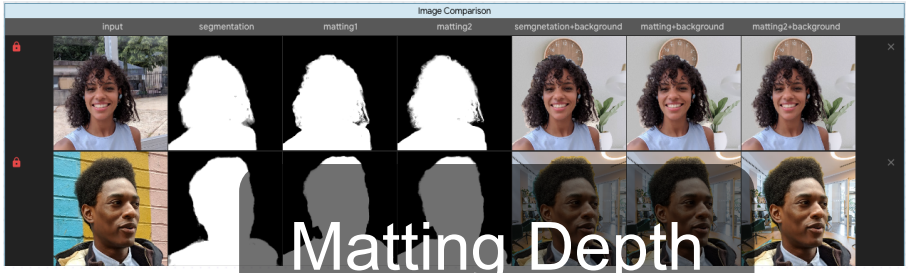
Portrait Depth



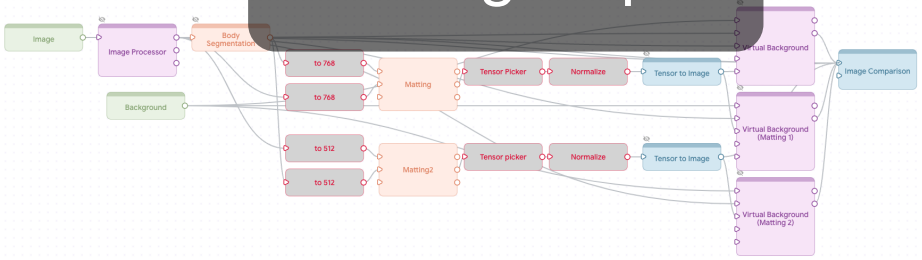


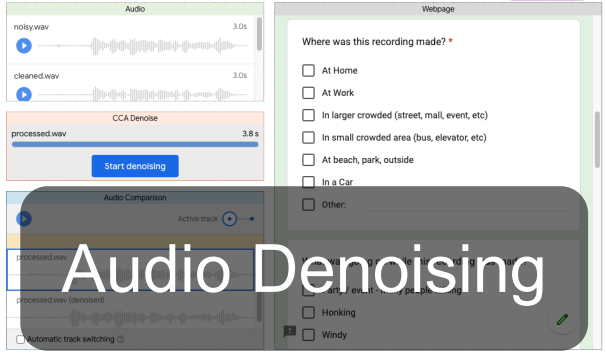
Scene Depth





Matting Depth

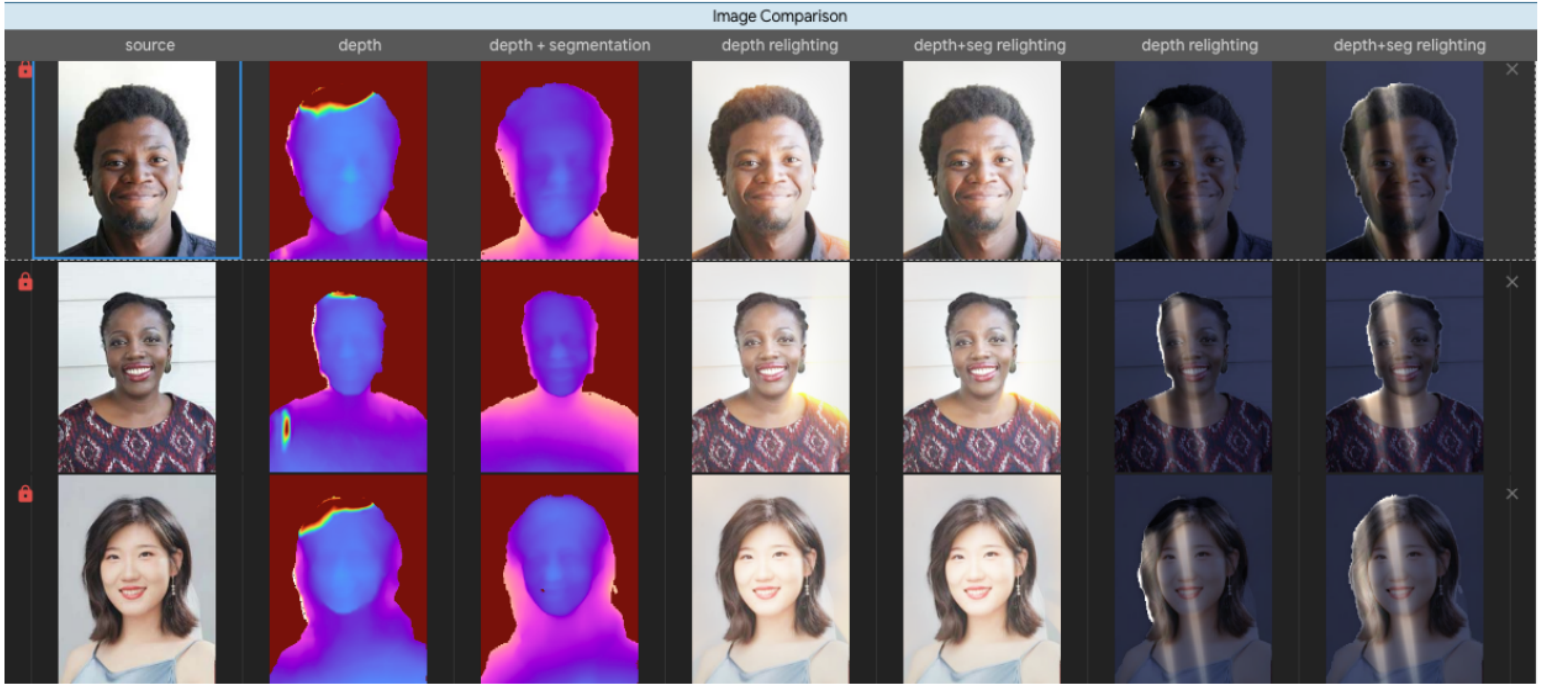


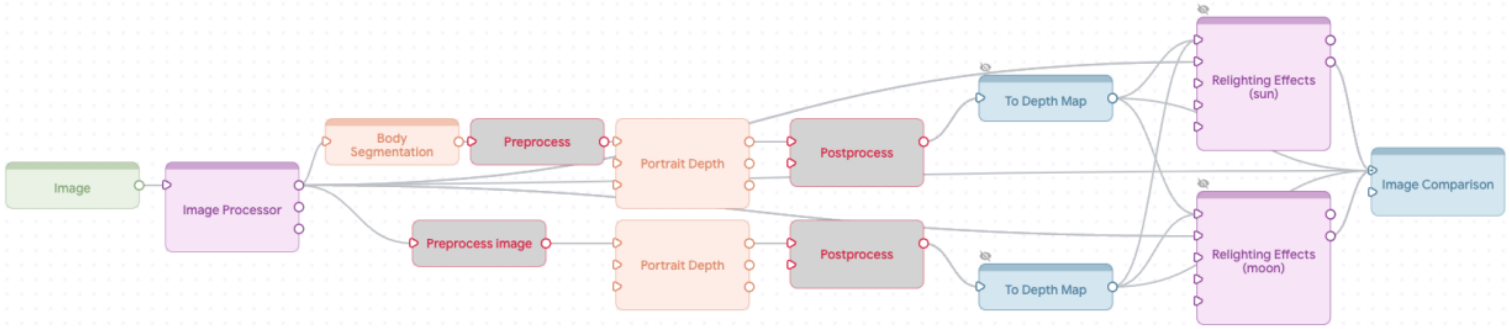


Audio Denoising



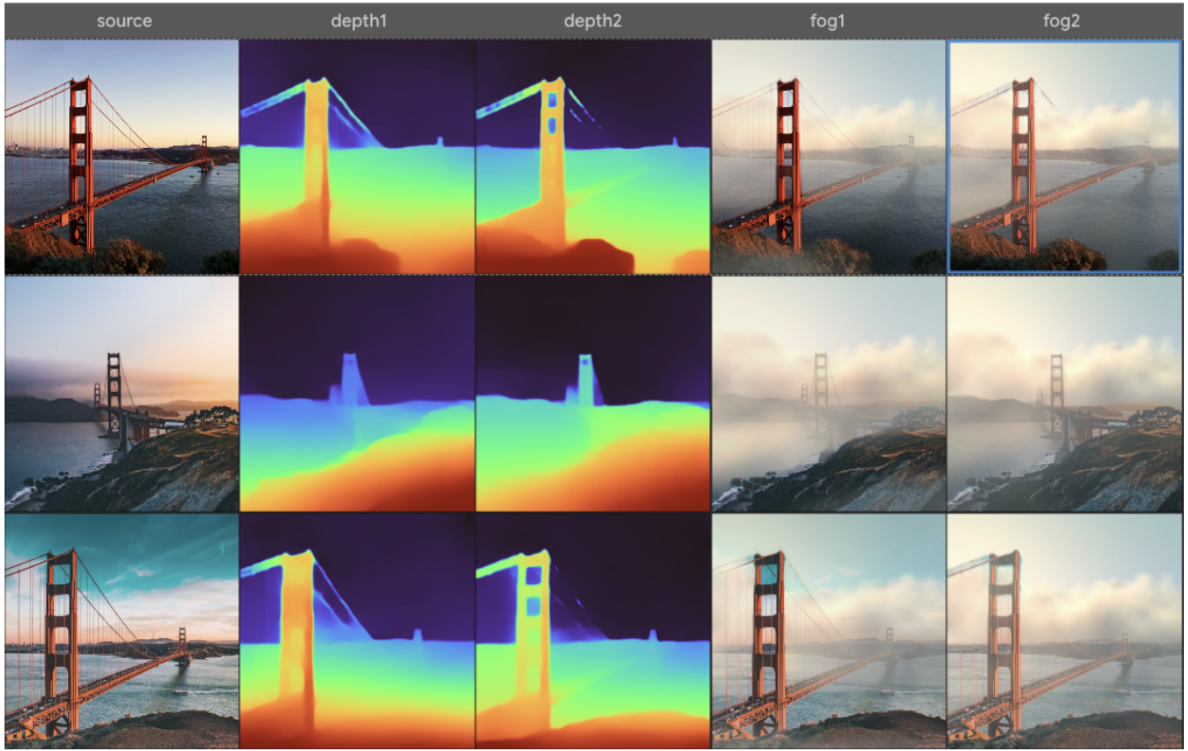
Case Study 1
Portrait Depth

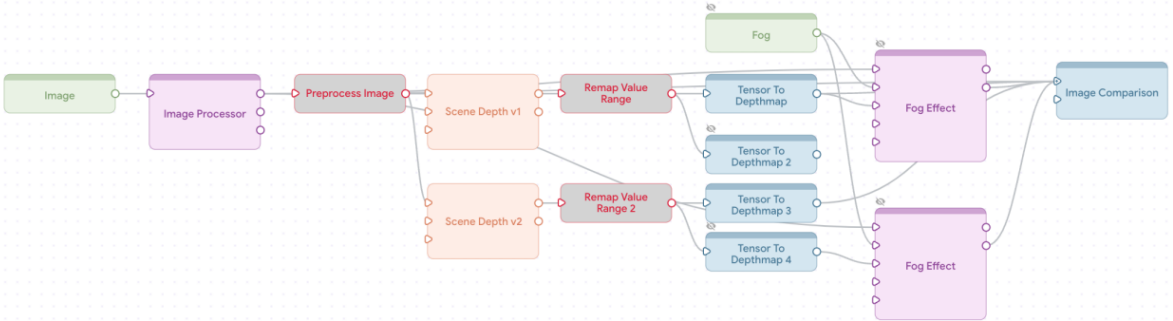




Case Study 2

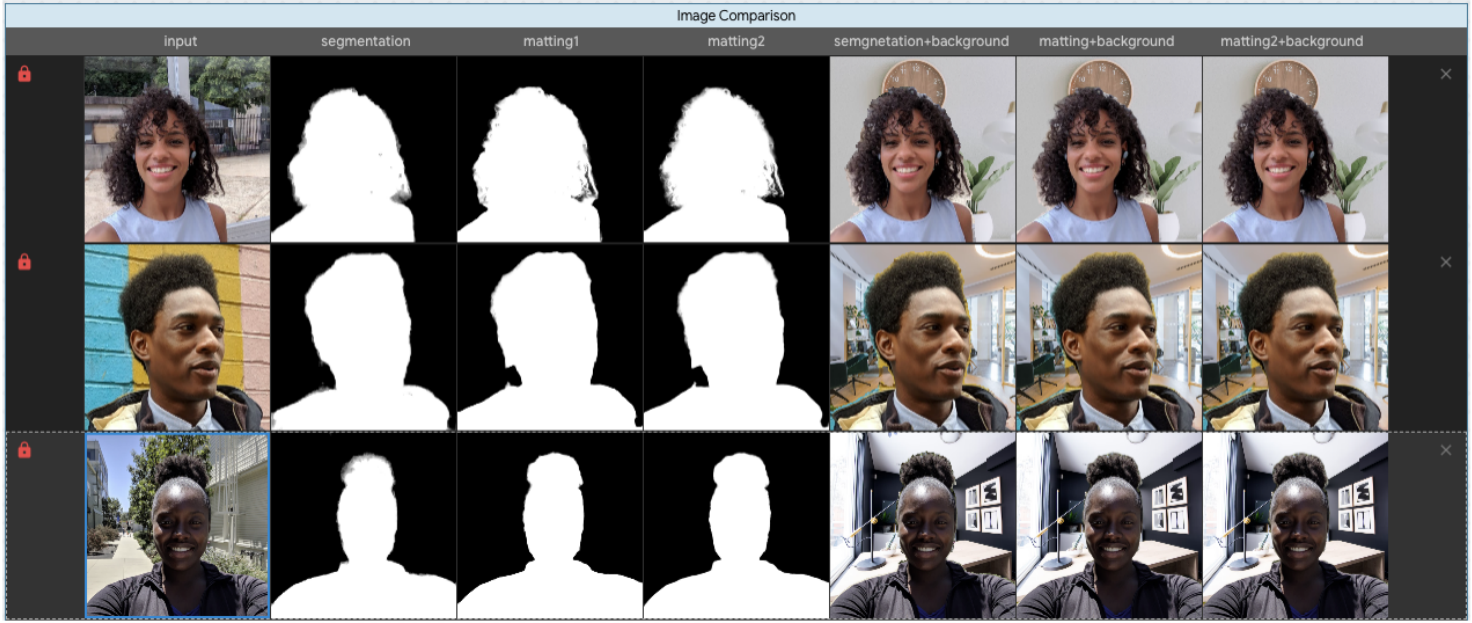
Scene Depth

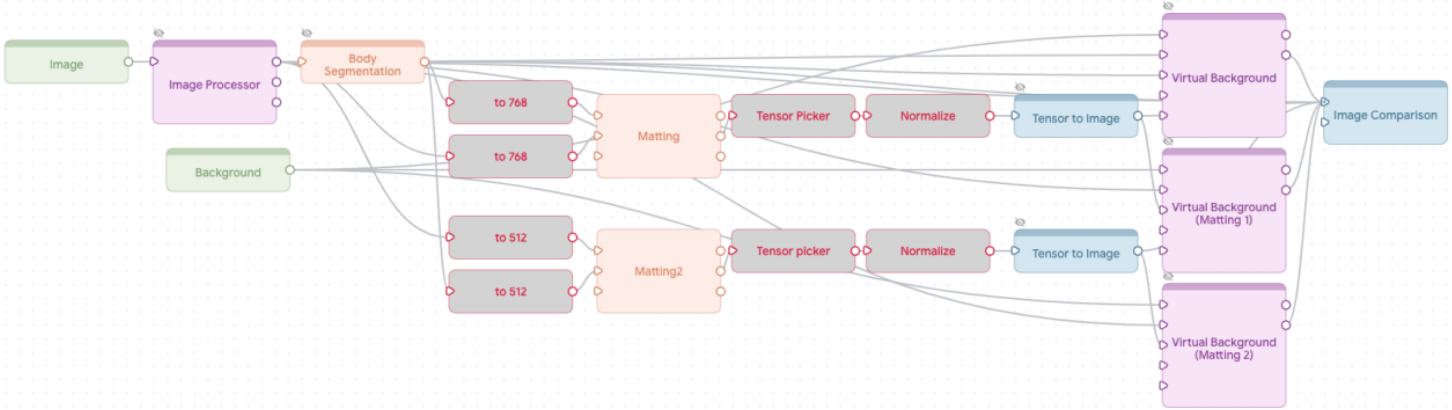




Case Study 3

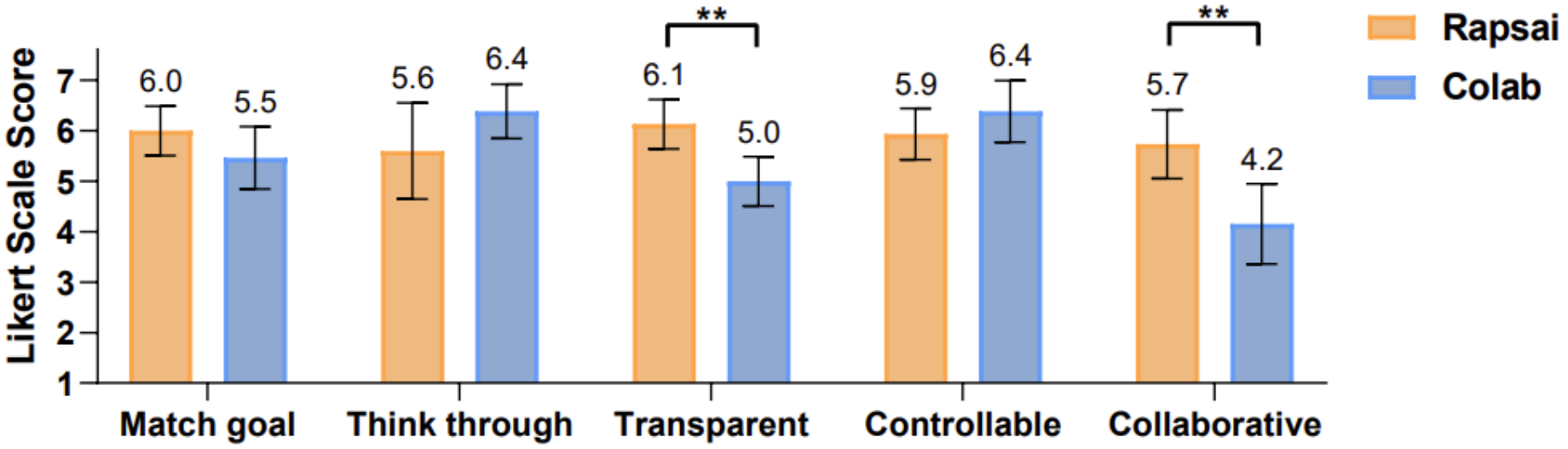
Alpha Matte





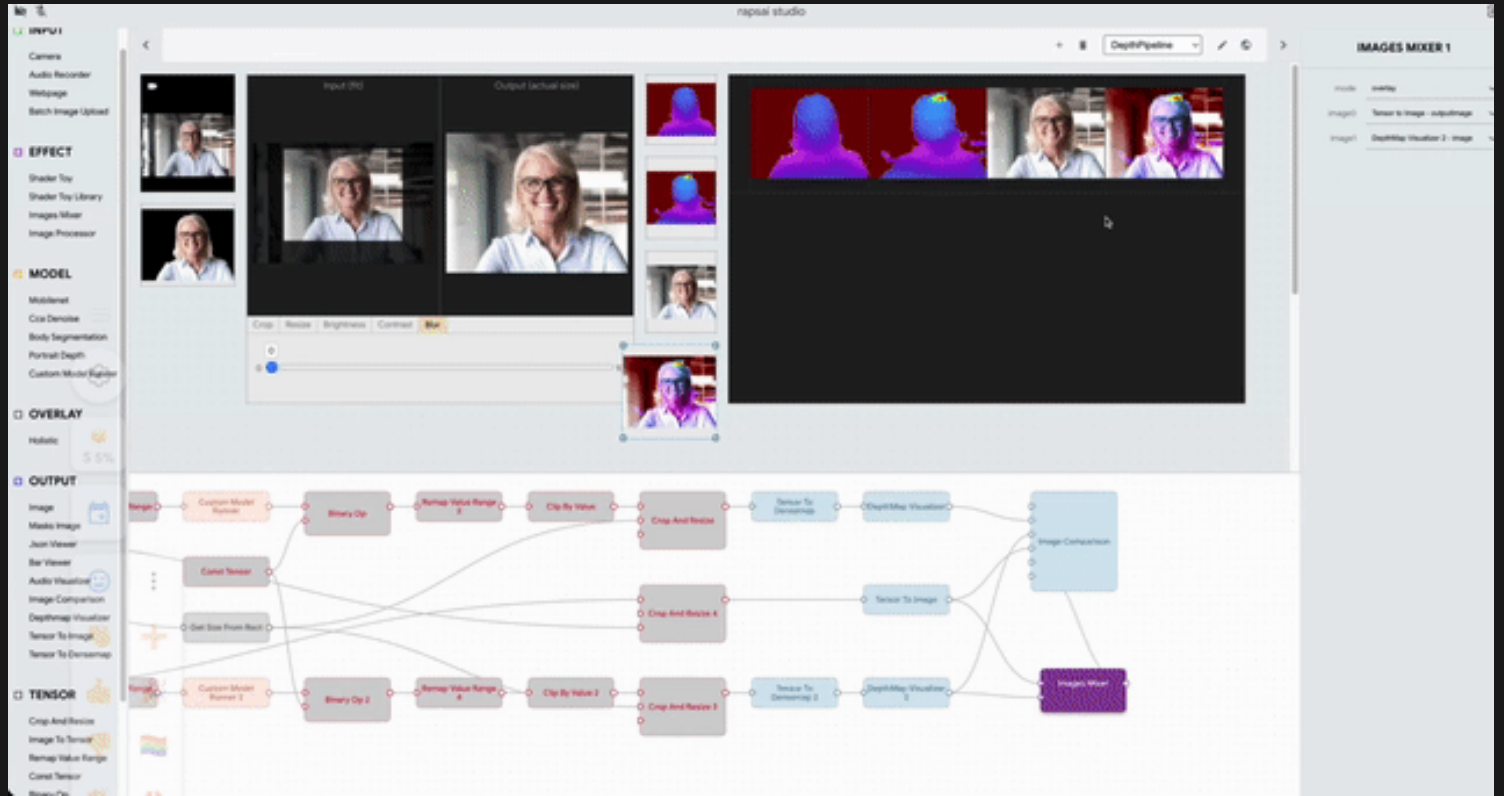
Findings 1
Rapsai *vs* Colab

Less Control but *More Transparent* and *Collaborative*



Findings 2

Assist in Identifying Issues with ML Models and Training Sets



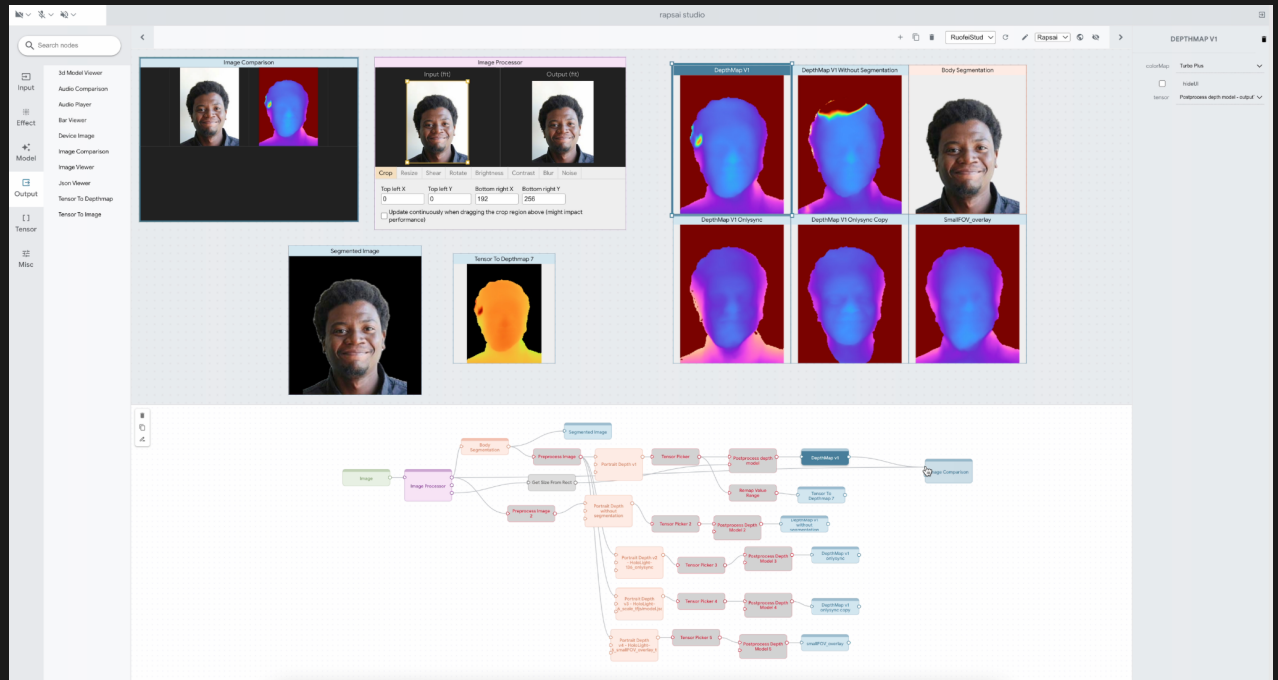
Findings 3

Rapsai helps

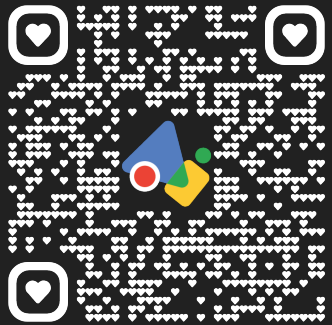
Model Selection,

Learning From Pipelines,

Study deployment



Visual Blocks



visualblocks.withgoogle.com

Stanford CS139: Human-Centered AI

UCLA ECE188: Prototyping Interactive Intelligent Software

CHI 2025
Honorable Mention Award

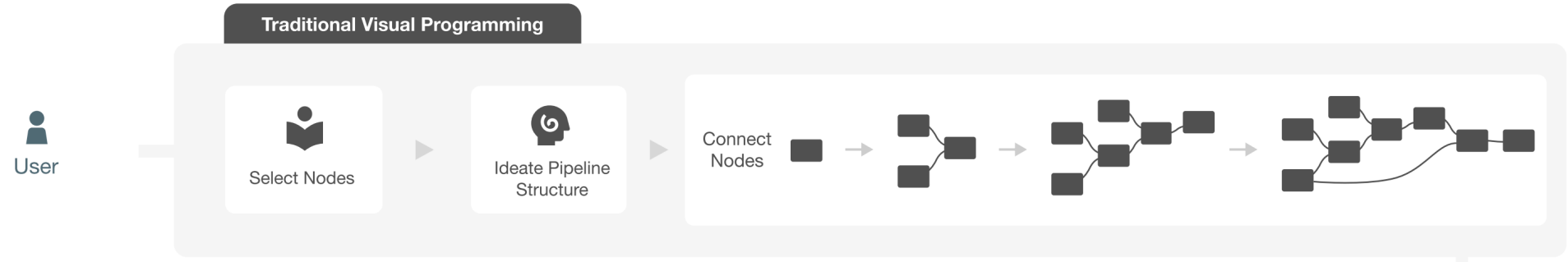
InstructPipe

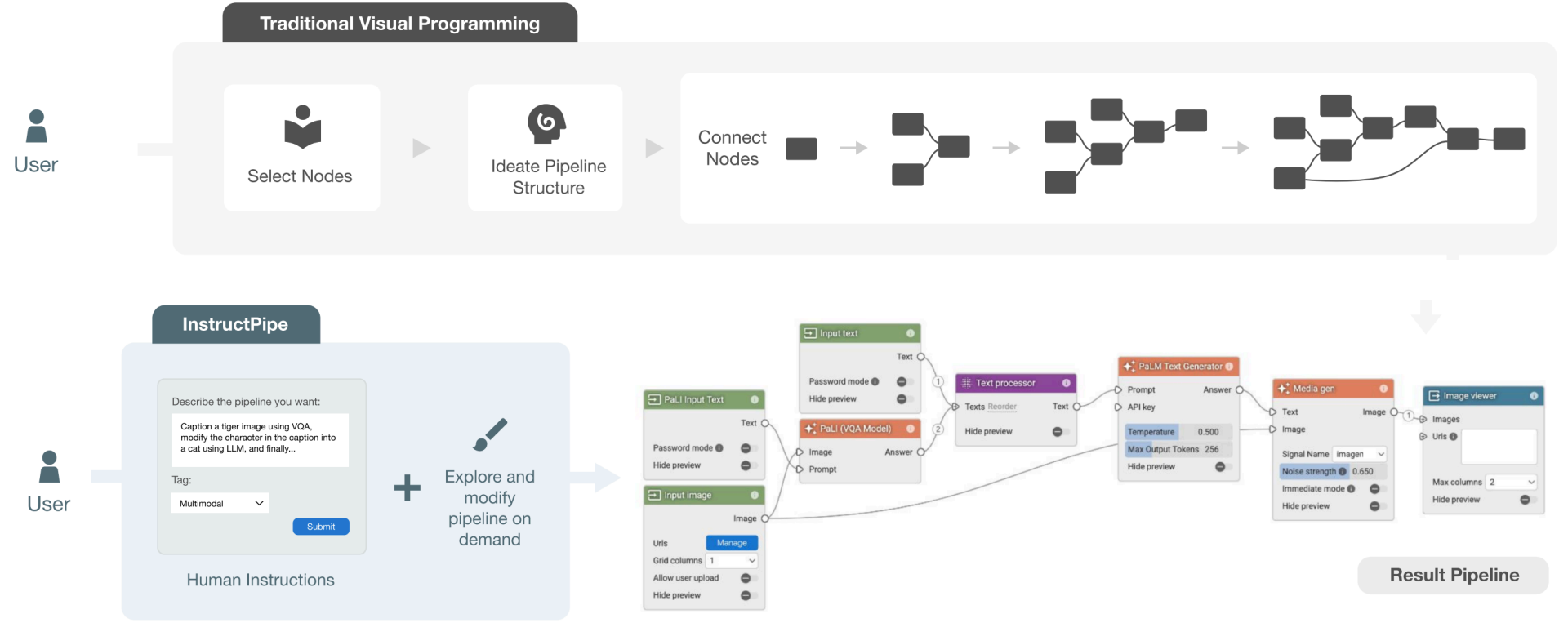
Generating Visual Blocks Pipelines with Human Instructions and LLMs



Zhongyi Zhou, Jing Jin, Vrushank Phadnis, Xiuxiu Yuan, Jun Jiang, Xun Qian, Kristen Wright, Mark Sherwood, Jason Mayes, Jingtao Zhou, Yiyi Huang, Zheng Xu, Yinda Zhang, Johnny Lee, Alex Olwal, David Kim, Ram Iyengar, Na Li, and **Ruofei Du**.

Google UTokyo


Traditional Visual Programming
User
Select Nodes
Ideate Pipeline Structure
Connect Nodes


Traditional Visual Programming
User
Select Nodes
Ideate Pipeline Structure
Connect Nodes
InstructPipe
User
Describe the pipeline you want:
Caption a tiger image using VQA, modify the character in the caption into a cat using LLM, and finally...
Tag:
Multimodal
Submit
Human Instructions
Explore and modify pipeline on demand
PaLI Input Text
Text
Password mode
Hide preview
Input image
Image
Urls
Manage
Grid columns
1
Allow user upload
Hide preview
Input text
Text
Password mode
Hide preview
PaLI (VQA Model)
Image
Answer
Prompt
Text processor
Texts
Reorder
Text
Hide preview
PaLM Text Generator
Prompt
Answer
API key
Temperature
0.500
Max Output Tokens
256
Hide preview
Media gen
Text
Image
Image
Signal Name
imagen
Noise strength
0.650
Immediate mode
Hide preview
Image viewer
Images
Urls
Max columns
2
Hide preview
Result Pipeline

InstructPipe

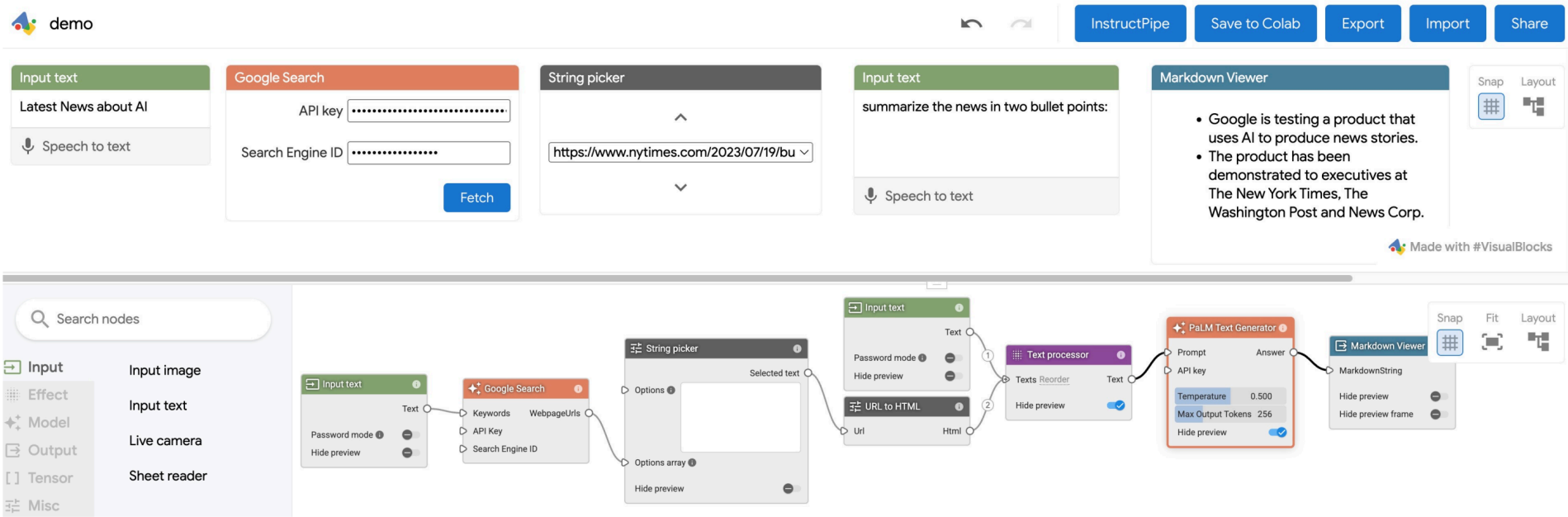
Describe the pipeline you want:

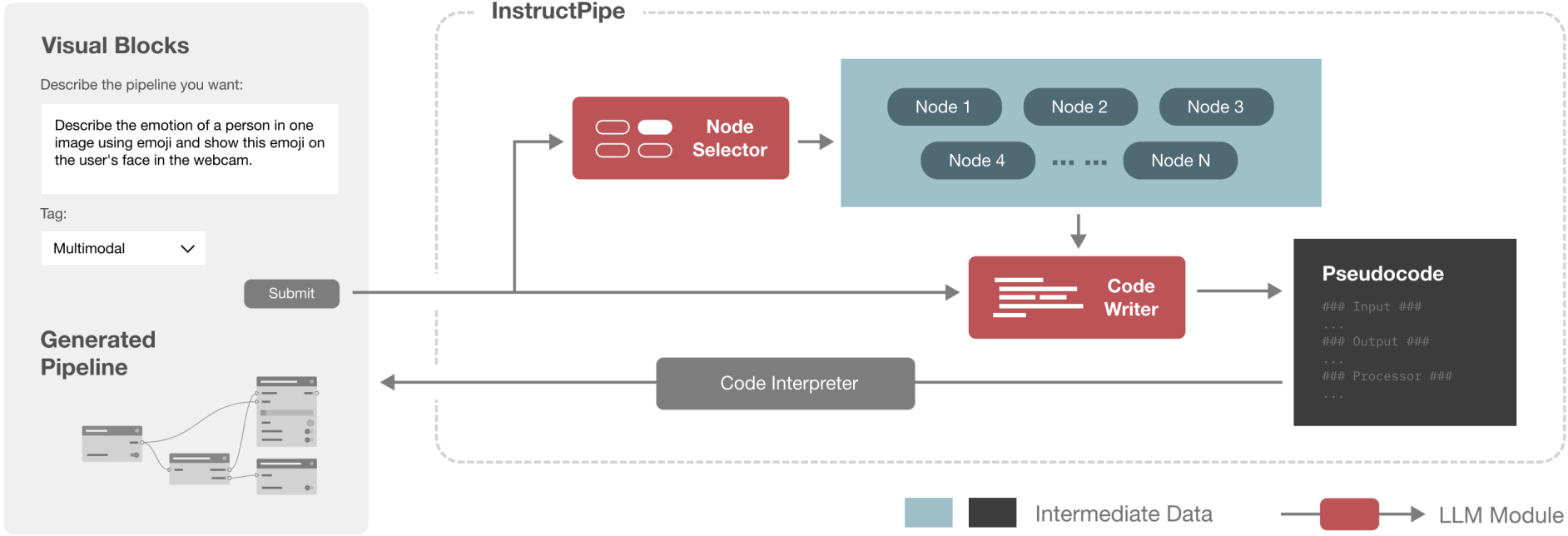
Grab AI news from Google and summarize it into bullet points

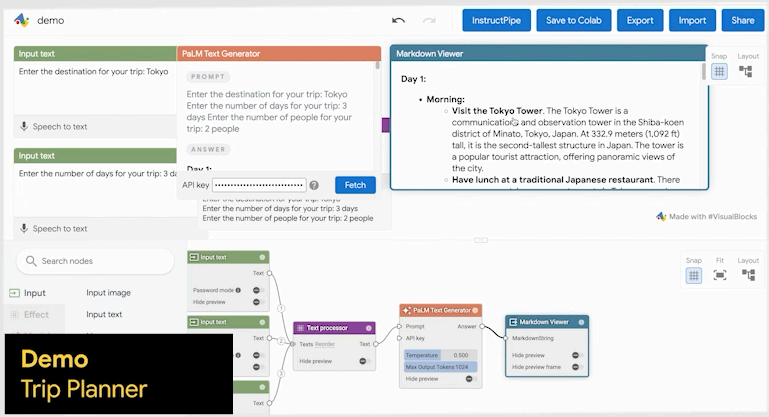
Tag:

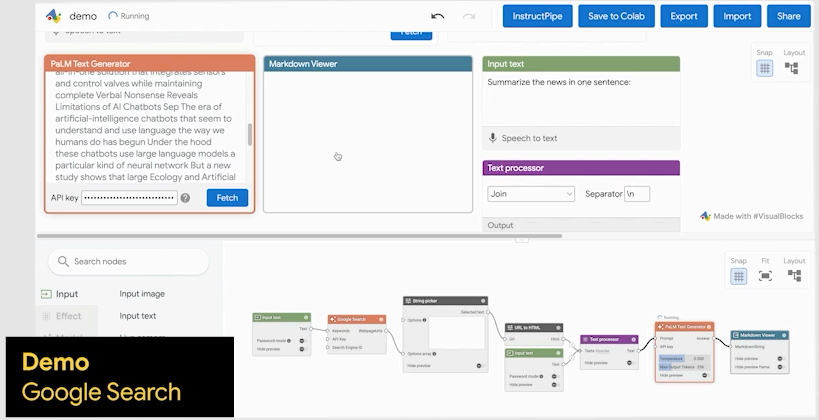
language

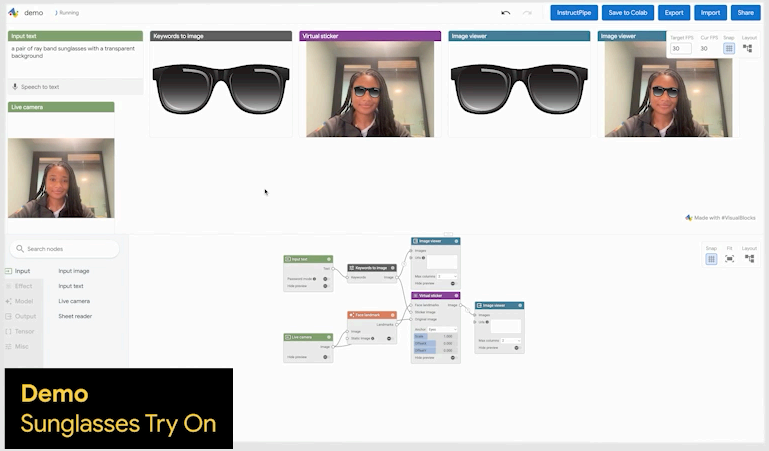
Submit

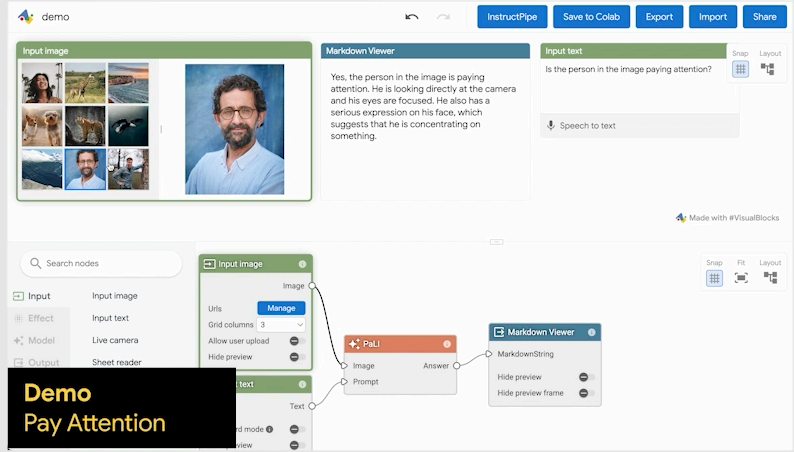

demo
InstructPipe
Save to Colab
Export
Import
Share
Input text
Latest News about AI
Speech to text
Google Search
API key
Search Engine ID
Fetch
String picker
https://www.nytimes.com/2023/07/19/bu
Input text
summarize the news in two bullet points:
Speech to text
Markdown Viewer
Google is testing a product that uses AI to produce news stories.
The product has been demonstrated to executives at The New York Times, The Washington Post and News Corp.
Made with #VisualBlocks
Snap
Layout
Search nodes
Input
Effect
Model
Output
Tensor
Misc
Input image
Input text
Live camera
Sheet reader
Input text
Text
Password mode
Hide preview
Google Search
Keywords
WebpageUrls
API Key
Search Engine ID
String picker
Selected text
Options
Options array
Hide preview
Input text
Text
Password mode
Hide preview
URL to HTML
Url
Html
Text processor
Texts
Reorder
Text
Hide preview
PaLM Text Generator
Prompt
Answer
API key
Temperature
0.500
Max Output Tokens
256
Hide preview
Markdown Viewer
MarkdownString
Hide preview
Hide preview frame
Snap
Fit
Layout


Visual Blocks
Describe the pipeline you want:
Describe the emotion of a person in one image using emoji and show this emoji on the user's face in the webcam.
Tag:
Multimodal
Submit
Generated Pipeline
InstructPipe
Node Selector
Node 1
Node 2
Node 3
Node 4
Node N
Code Writer
Pseudocode
Input
...
Output
...
Processor
...
Code Interpreter
Intermediate Data
LLM Module


demo
InstructPipe
Save to Colab
Export
Import
Share
Input text
Enter the destination for your trip: Tokyo
PaLM Text Generator
PROMPT
Enter the destination for your trip: Tokyo
Enter the number of days for your trip: 3 days
Enter the number of people for your trip: 2 people
ANSWER
Speech to text
Input text
Enter the number of days for your trip: 3 da
API key
Fetch
Markdown Viewer
Day 1:
Morning:
Visit the Tokyo Tower. The Tokyo Tower is a communicatio and observation tower in the Shiba-koen district of Minato, Tokyo, Japan. At 332.9 meters (1,092 ft) tall, it is the second-tallest structure in Japan. The tower is a popular tourist attraction, offering panoramic views of the city.
Have lunch at a traditional Japanese restaurant. There
Made with #VisualBlocks
Search nodes
Input
Input image
Effect
Input text
Text processor
PaLM Text Generator
Markdown Viewer
Demo
Trip Planner

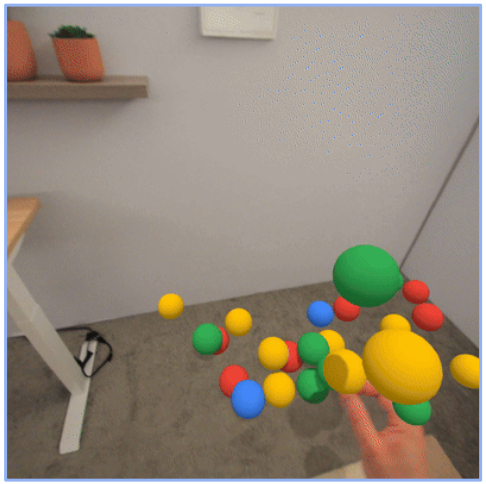

demo
Running
InstructPipe
Save to Colab
Export
Import
Share
PaLM Text Generator
and control valves while maintaining complete Verbal Nonsense Reveals Limitations of AI Chatbots Sep The era of artificial-intelligence chatbots that seem to understand and use language the way we humans do has begun Under the hood these chatbots use large language models a particular kind of neural network But a new study shows that large Ecology and Artificial
API key
Fetch
Markdown Viewer
Input text
Summarize the news in one sentence:
Speech to text
Text processor
Join
Separator
\n
Output
Made with #VisualBlocks
Search nodes
Input
Input image
Effect
Input text
Demo
Google Search


demo
Running
InstructPipe
Save to Colab
Export
Import
Share
Input text
a pair of ray band sunglasses with a transparent background
Speech to text
Live camera
Keywords to image
Virtual sticker
Image viewer
Image viewer
Made with #VisualBlocks
Search nodes
Input
Input image
Effect
Input text
Model
Live camera
Output
Sheet reader
Tensor
Misc
Demo
Sunglasses Try On


demo
InstructPipe
Save to Colab
Export
Import
Share
Input image
Markdown Viewer
Yes, the person in the image is paying attention. He is looking directly at the camera and his eyes are focused. He also has a serious expression on his face, which suggests that he is concentrating on something.
Input text
Is the person in the image paying attention?
Speech to text
Made with #VisualBlocks
Search nodes
Input
Input image
Effect
Input text
Model
Live camera
Output
Sheet reader
Input image
Image
Urls
Manage
Grid columns
3
Allow user upload
Hide preview
PaLI
Image
Prompt
Answer
Markdown Viewer
MarkdownString
Hide preview
Hide preview frame
Text
Demo
Pay Attention

XR Blocks

Unifying programming language for XR interaction










Chapter III · XR Blocks

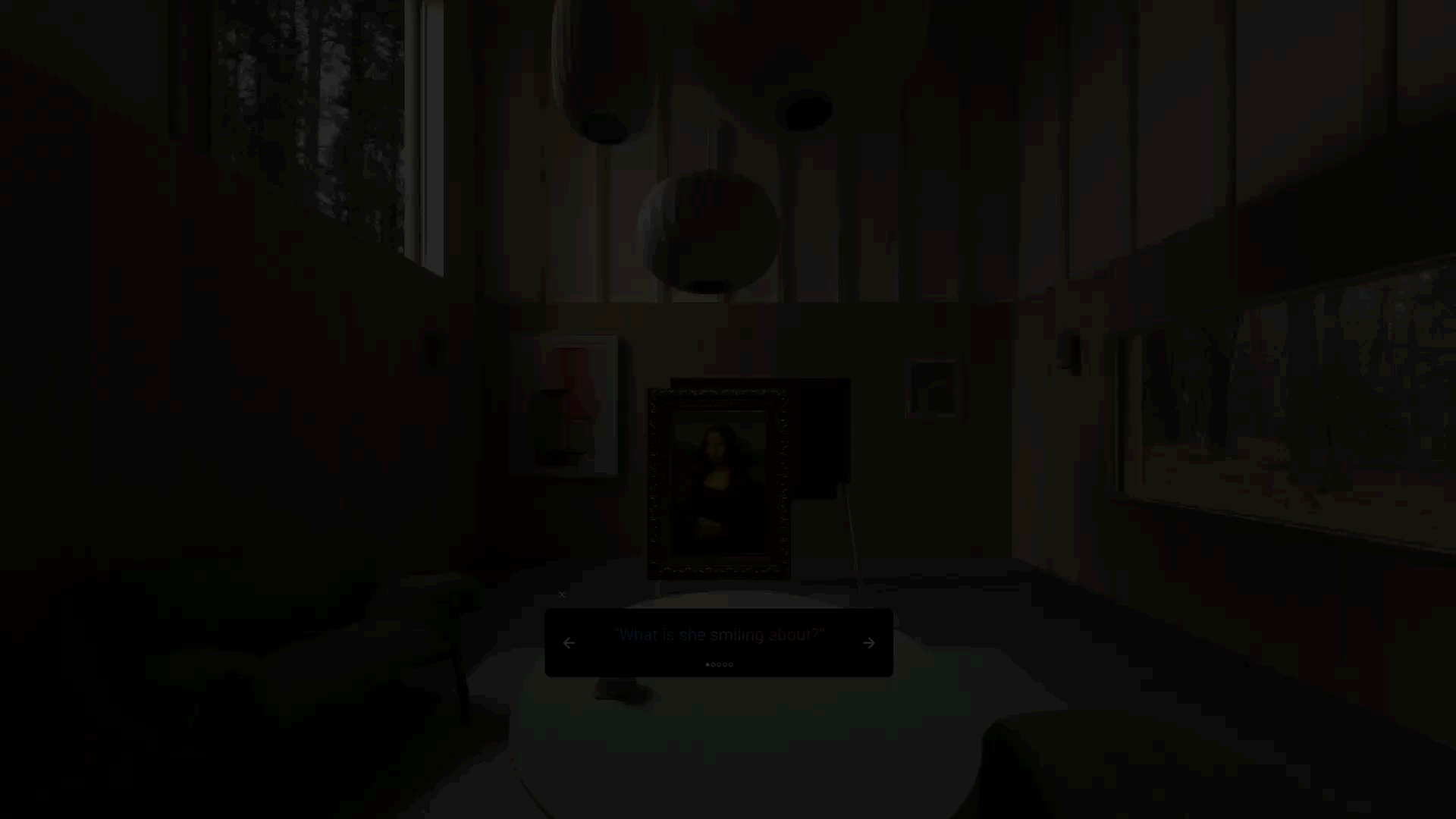
XR Blocks

Accelerating Human-Centered AI + XR Innovation

xrblocks.github.io
github.com/google/xrblocks

David Li*, Nels Numan†, Xun Qian†, Yanhe Chen†, Zhongyi Zhou†, Evgenii Alekseev, Geonsun Lee, Alex Cooper, Min Xia, Scott Chung, Jeremy Nelson, Xiuxiu Yuan, Jolica Dias, Tim Bettridge, Benjamin Hersh, Michelle Huynh, Konrad Piascik, Ricardo Cabello, David Kim, Ruofei Du*
Google XR Labs




“What is she smiling about?”

XR Blocks

Accelerating Human-Centered AI + XR Innovation





AI + XR are converging to unlock a new paradigm of computing

Motivation

Gap between AI and XR



Motivation

Gap between AI and XR

AI

Gemini

PyTorch

colab

Hugging Face



OpenXR

Unity

XR

Unity 6 Unity 2022.3.62

Unity 2023.2.20

DepthLab: Real-time 3D Interaction with Depth Maps for Mobile Augmented Reality



Ruofei Du, Eric Turner, Max Dzitsiuk, Luca Prasso, Ivo Duarte,
Jason Dourgarian, Joao Afonso, Jose Pascoal, Josh Gladstone, Nuno Cruces,
Shahram Izadi, Adarsh Kowdle, Konstantine Tsotsos, David Kim

Google | ACM UIST 2020

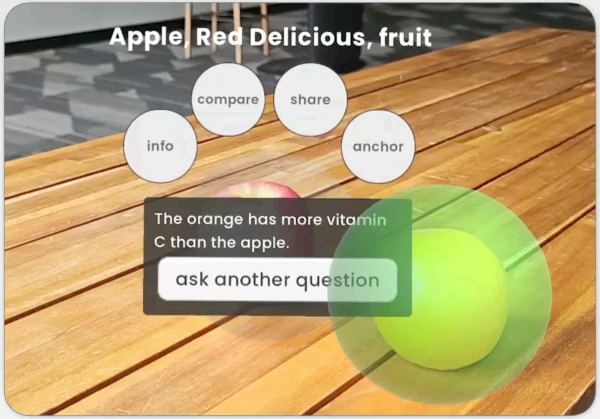
Introduction

Ad hoc UI live demo



Augmented Object Intelligence

with XR-OBJECTS



Mustafa Doga Dogan, Eric Gonzalez, Karan Ahuja, Ruofei Du,
Andrea Colaco, Johnny Lee, Mar Gonzalez-Franco*, David Kim*

*co-senior authorship

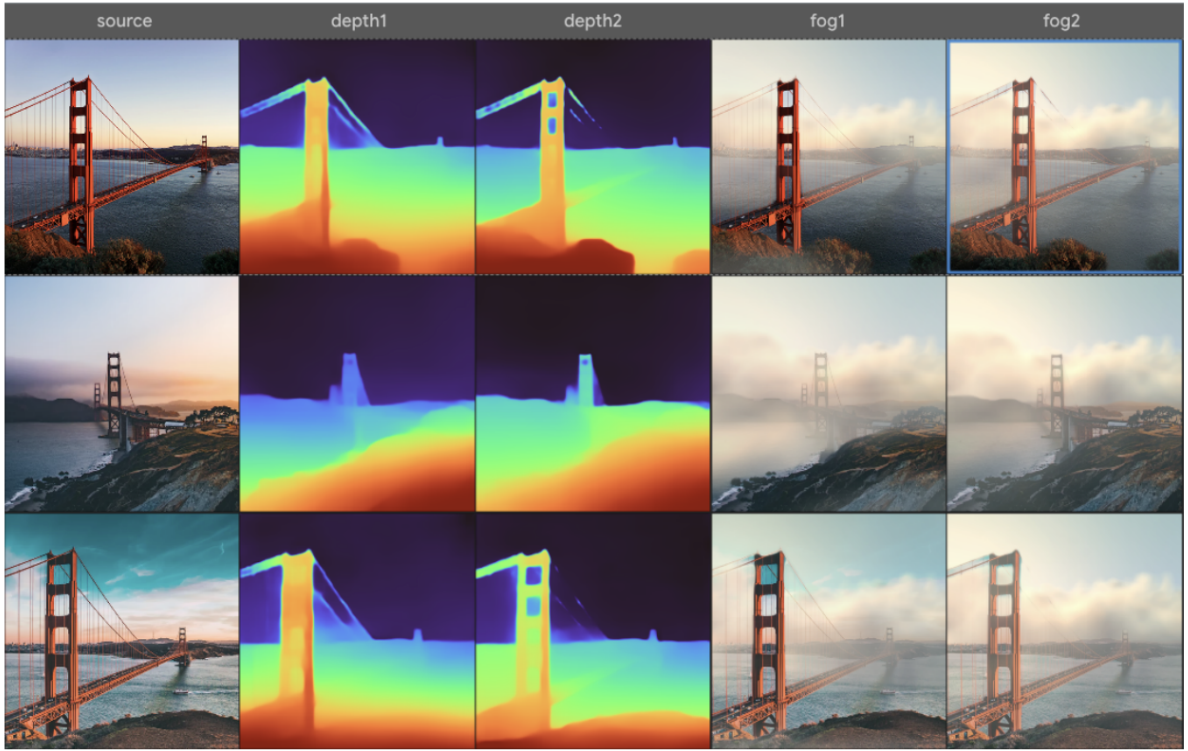
Google

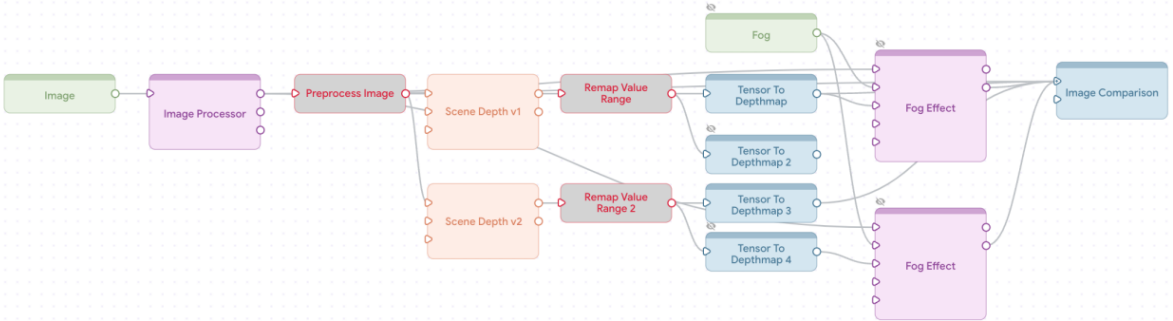
Massachusetts Institute of Technology

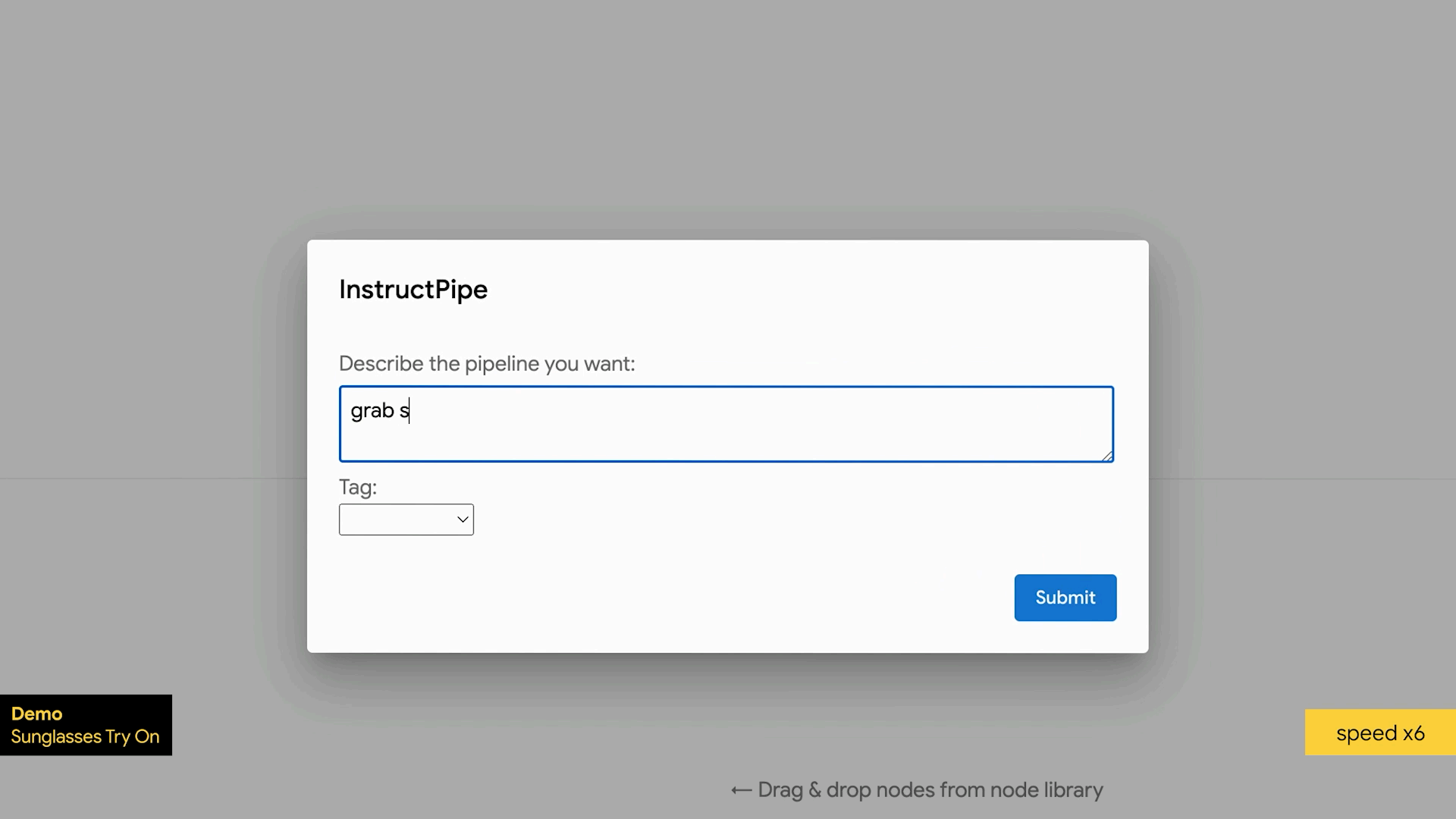
UIST PITTSBURGH | 2024

contains voiceover

Visual Blocks





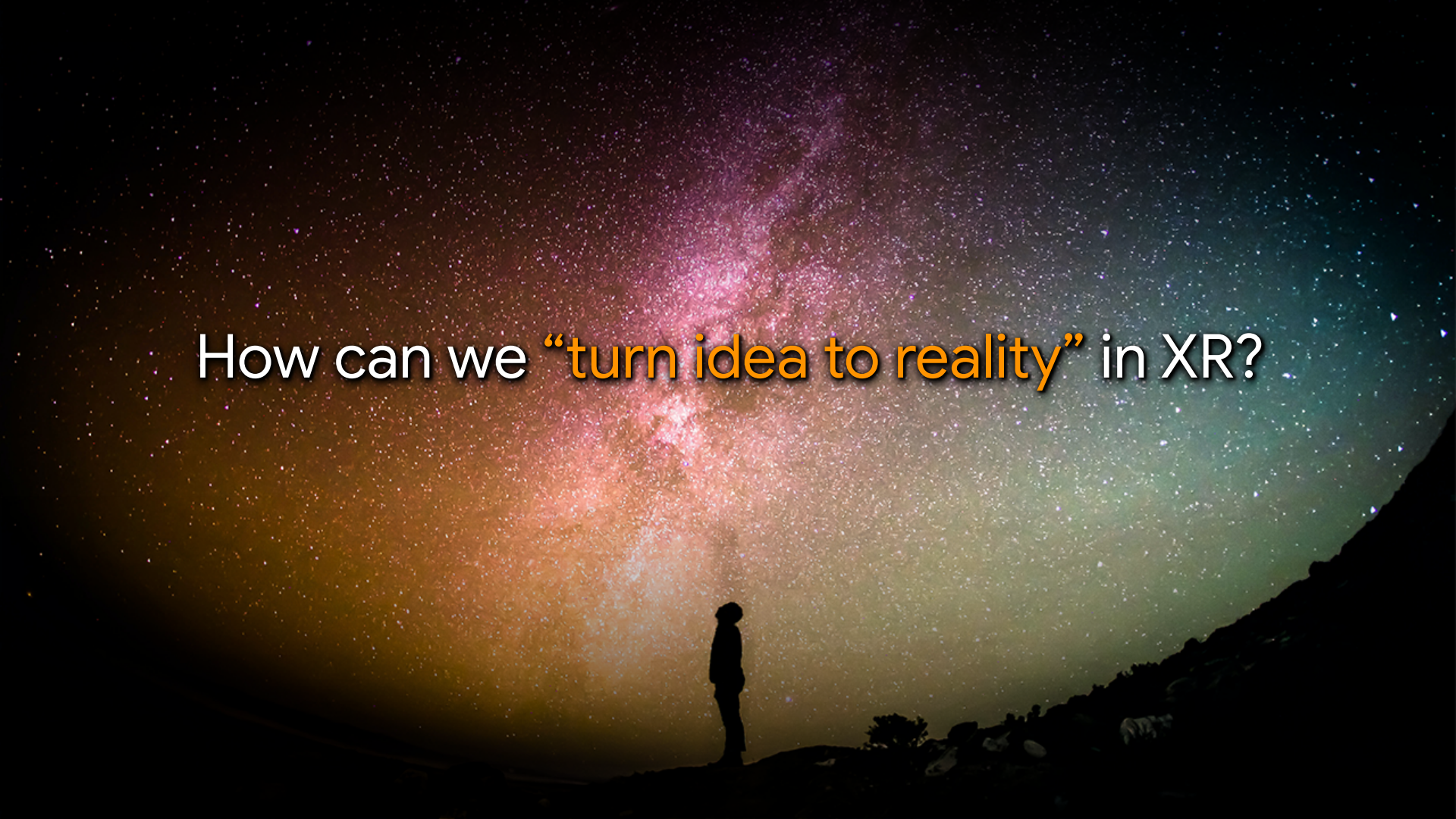

InstructPipe
Describe the pipeline you want:
grab s
Tag:
Submit
Demo
Sunglasses Try On
speed x6
← Drag & drop nodes from node library

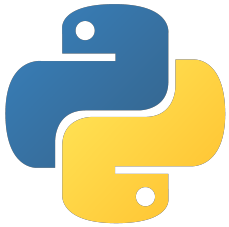

Many of the past innovations are
not directly scalable in XR :(
Varied platforms, languages,
interaction paradigms.


Technology becomes obsolete,
but Vision is ever-lasting
What legacy do you wish to
leave for those living in 2200?
— Hiroshi Ishii


How can we make AI + XR research and innovation more accessible & scalable?


How can we kick off the flywheel from AI
— models, data, benchmarking — to XR?


How can we “turn idea to reality” in XR?



Unity

UNREAL
ENGINE

OpenXR™

pinch / click / touch

to change color

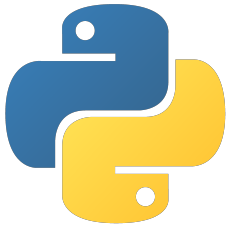
GODOT
Game engine

Xcode

Jetpack

WebXR

three.js









200+ lines of code



GODOT
Game engine

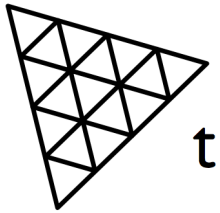


Xcode



Jetpack





three.js

main.js

```
import * as THREE from 'three';
import * as xb from 'xrblocks';

/**
 * A basic example of XRBlocks to render a cylinder and pinch to change color.
 */
class MainScript extends xb.Script {
  init() {
    this.add(new THREE.HemisphereLight(0xffffff, 0x666666, /*intensity=*/ 3));

    const geometry = new THREE.CylinderGeometry(.2, .2, .4, /*segments=*/ 32);
    const material = new THREE.MeshPhongMaterial(
        {color: 0xffffff, transparent: true, opacity: 0.8});
    this.player = new THREE.Mesh(geometry, material);
    this.player.position.set(0, xb.user.height - 0.5, -xb.user.objectDistance);
    this.add(this.player);
  }

  /**
   * Changes the color of the mesh on a pinch in XR.
   * @param {SelectEvent} event event.target holds controler / hand data.
   */
  onSelectEnd(event) {
    this.player.material.color.set(Math.random() * 0xffffff);
  }

  /**
   * Changes the color of the mesh to blue during pinching in XR.
   * @param {SelectEvent} event The controller / hand event.
   */
  onSelecting(event) {
    this.player.material.color.set(0x66ccff);
  }
};

document.addEventListener('DOMContentLoaded', function() {
  xb.add(new MainScript());
  xb.init(new xb.Options());
});
```



39 lines

XR Blocks

Docs Templates Samples GitHub

Accelerating AI + XR Innovation

↓

xrblocks.github.io
github.com/google/xrblocks

Basic
UI
Hands
Depth
Stereo
Camera
AI
AI Live
Objects
XR Toggle

Basic

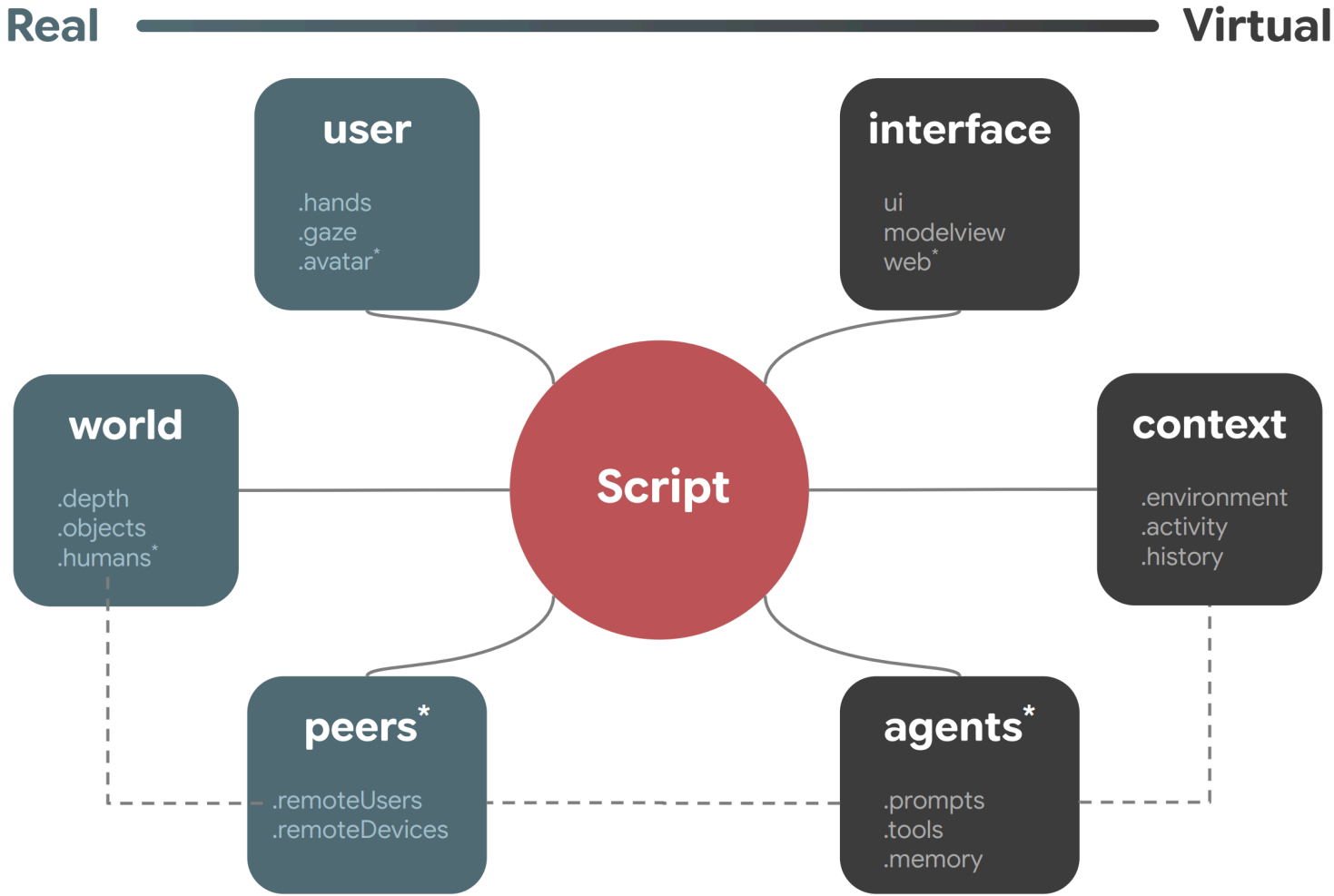
This template demonstrates a fundamental user interaction. Change the cylinder's color with a simple action.

On Desktop: Click
In XR: Pinch



main.js

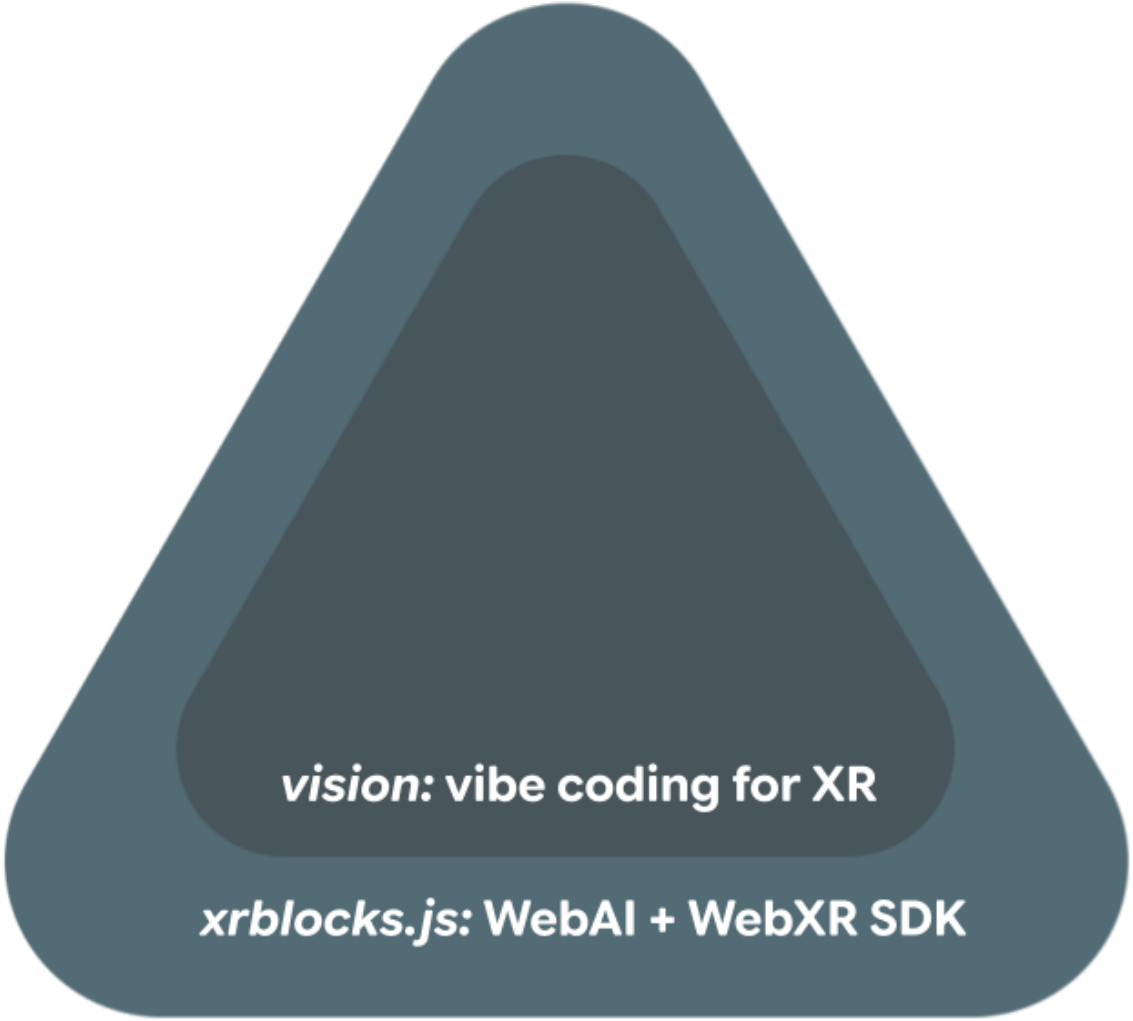
```
import * as THREE from 'three';
import * as xb from 'xrblocks';

/**
 * A basic example of XRBlocks to render a cylinder and pinch to change color.
 */
class MainScript extends xb.Script {
  init() {
    this.add(new THREE.HemisphereLight(0xffffff, 0x666666, /*intensity=*/ 3));

    const geometry = new THREE.CylinderGeometry(.2, .2, .4, /*segments=*/ 32);
    const material = new THREE.MeshPhongMaterial(
```

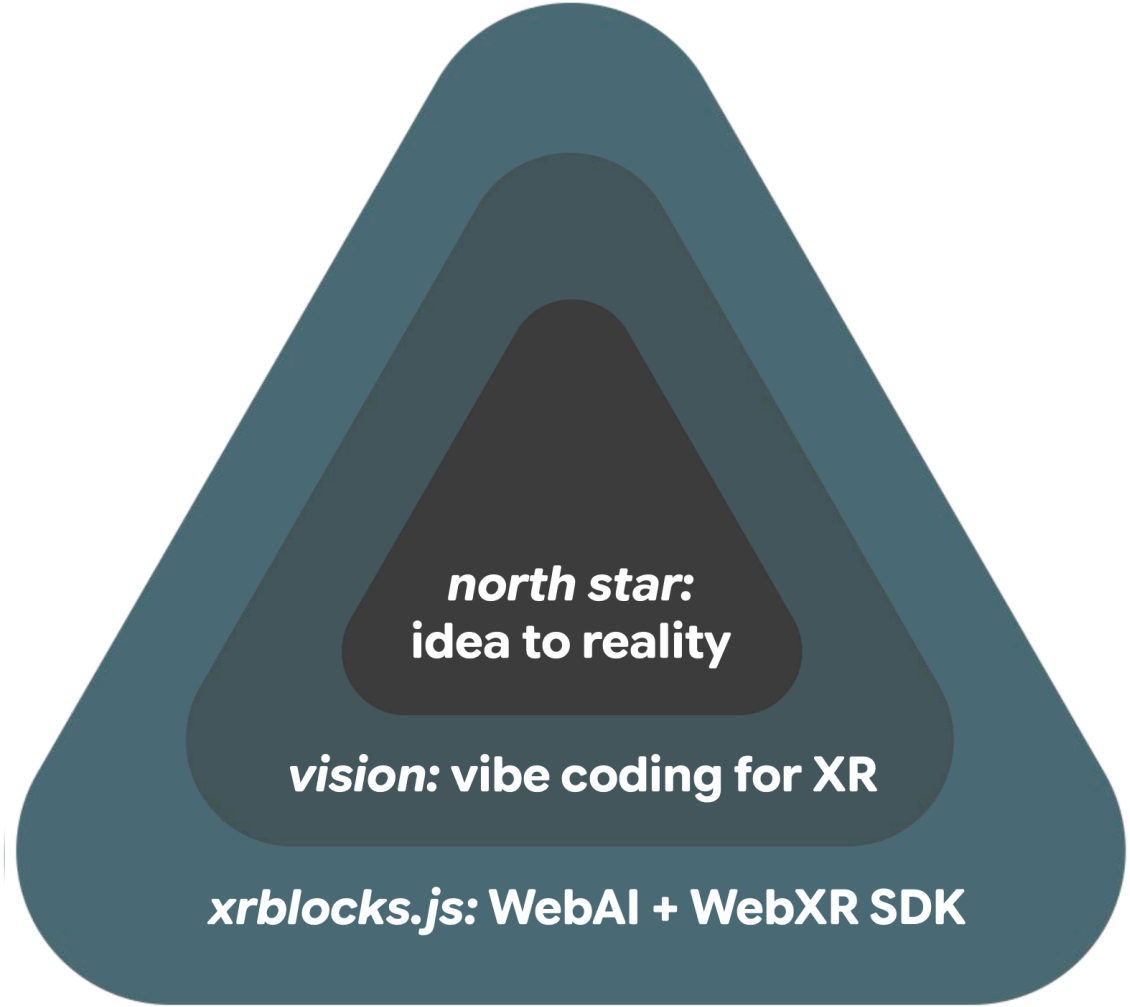

Real
Virtual
user
.hands
.gaze
.avatar*
interface
ui
modelview
web*
world
.depth
.objects
.humans*
Script
context
.environment
.activity
.history
peers*
.remoteUsers
.remoteDevices
agents*
.prompts
.tools
.memory

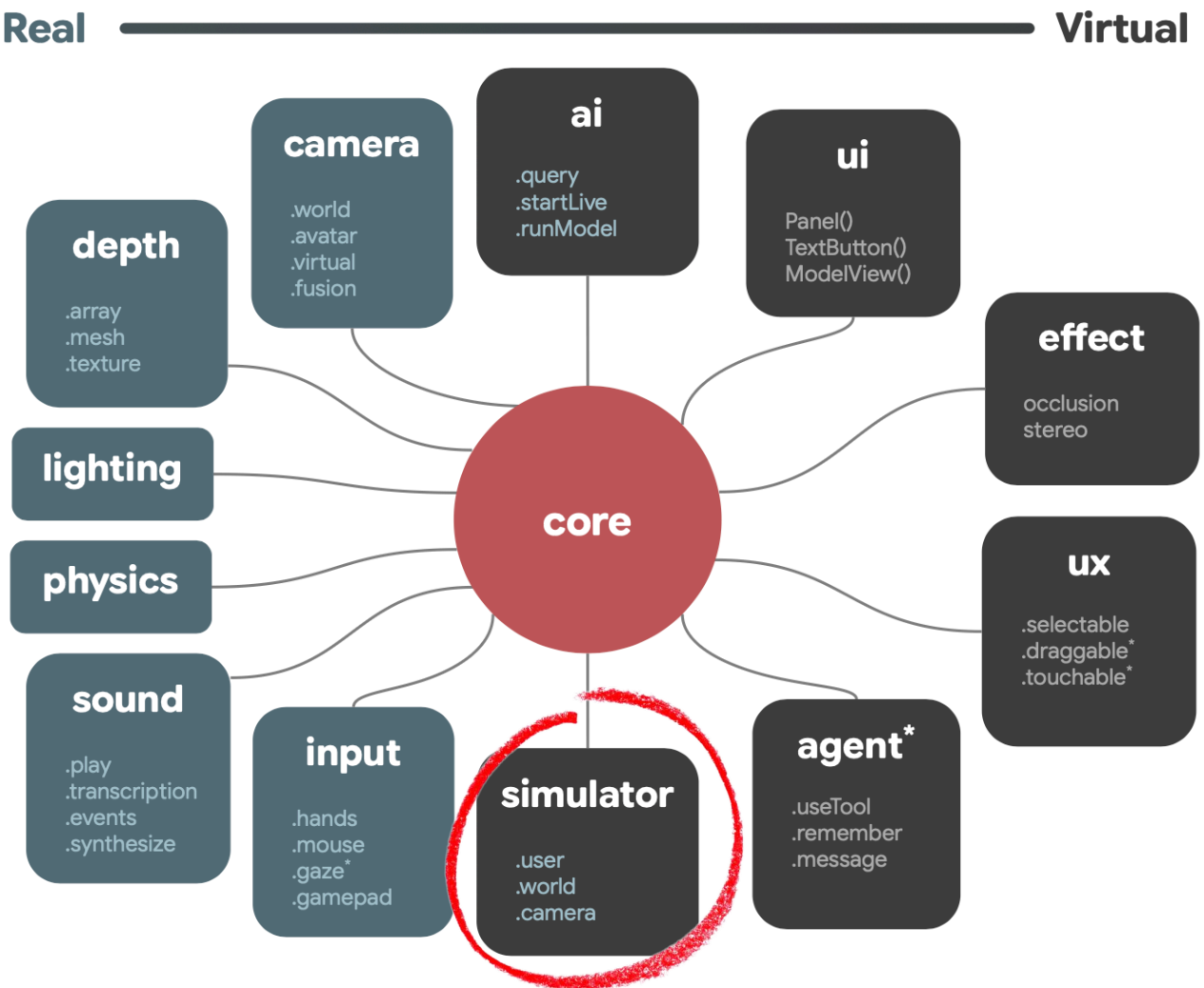
An Ideal Minimized Syntax

```
for (const object in world.objects) {
  if (user.isSelectingAt(object)) {
    const poem = agent.query(prompt)
    ...
  }
}
```


xrblocks.js: WebAI + WebXR SDK


vision: vibe coding for XR
xrblocks.js: WebAI + WebXR SDK


north star:
idea to reality
vision: vibe coding for XR
xrblocks.js: WebAI + WebXR SDK


Real
Virtual
camera
.world
.avatar
.virtual
.fusion
ai
.query
.startLive
.runModel
ui
Panel()
TextButton()
ModelView()
depth
.array
.mesh
.texture
effect
occlusion
stereo
lighting
core
physics
ux
.selectable
.draggable*
.touchable*
sound
.play
.transcription
.events
.synthesize
input
.hands
.mouse
.gaze*
.gamepad
simulator
.user
.world
.camera
agent*
.useTool
.remember
.message

XR Realism

Model Viewer



XR Realism

Geometry-aware occlusion



XR Realism

Immersive Weather







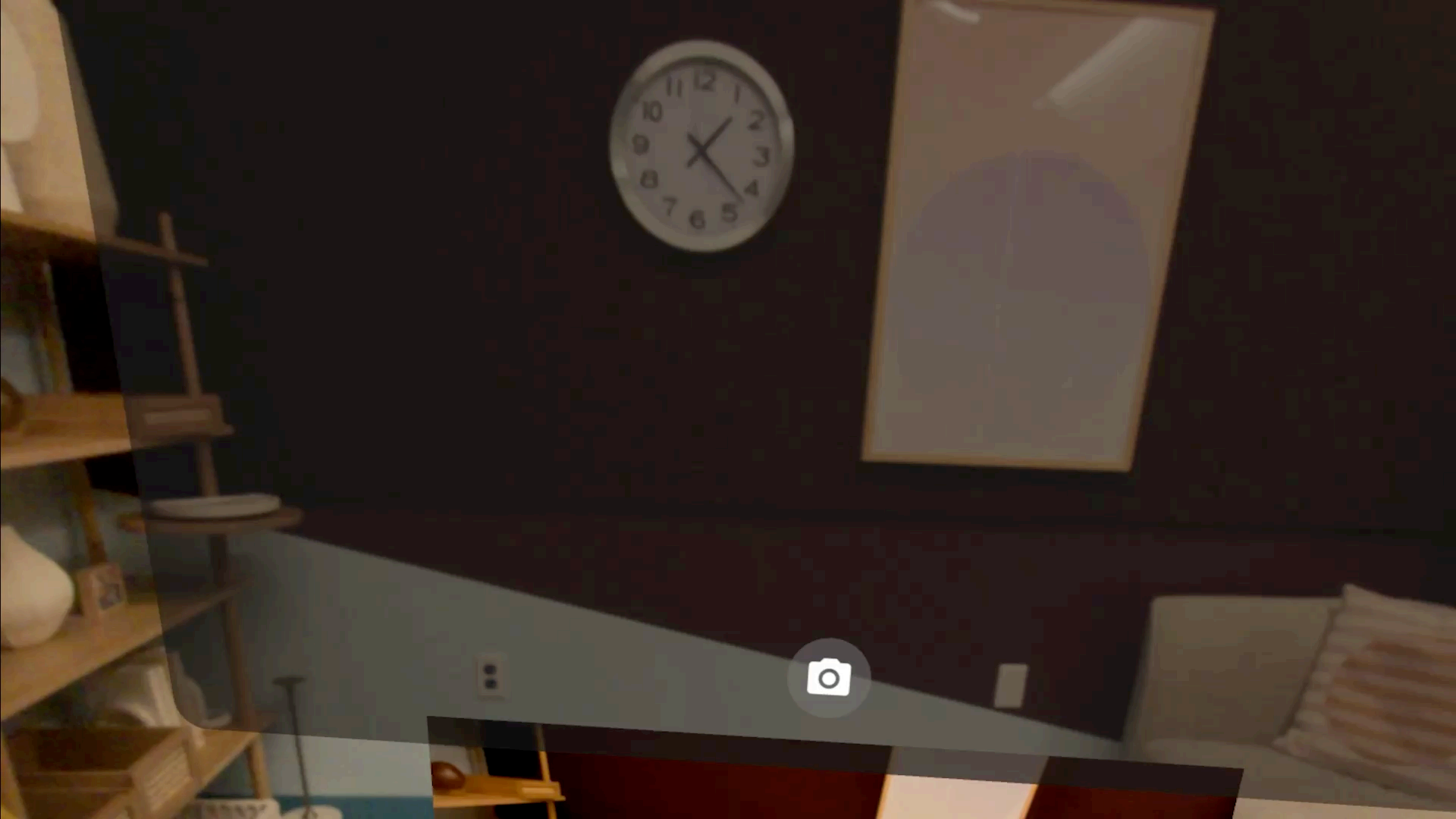
Generative UI
Composable Library



AI + XR

Generating Poem from XR Camera

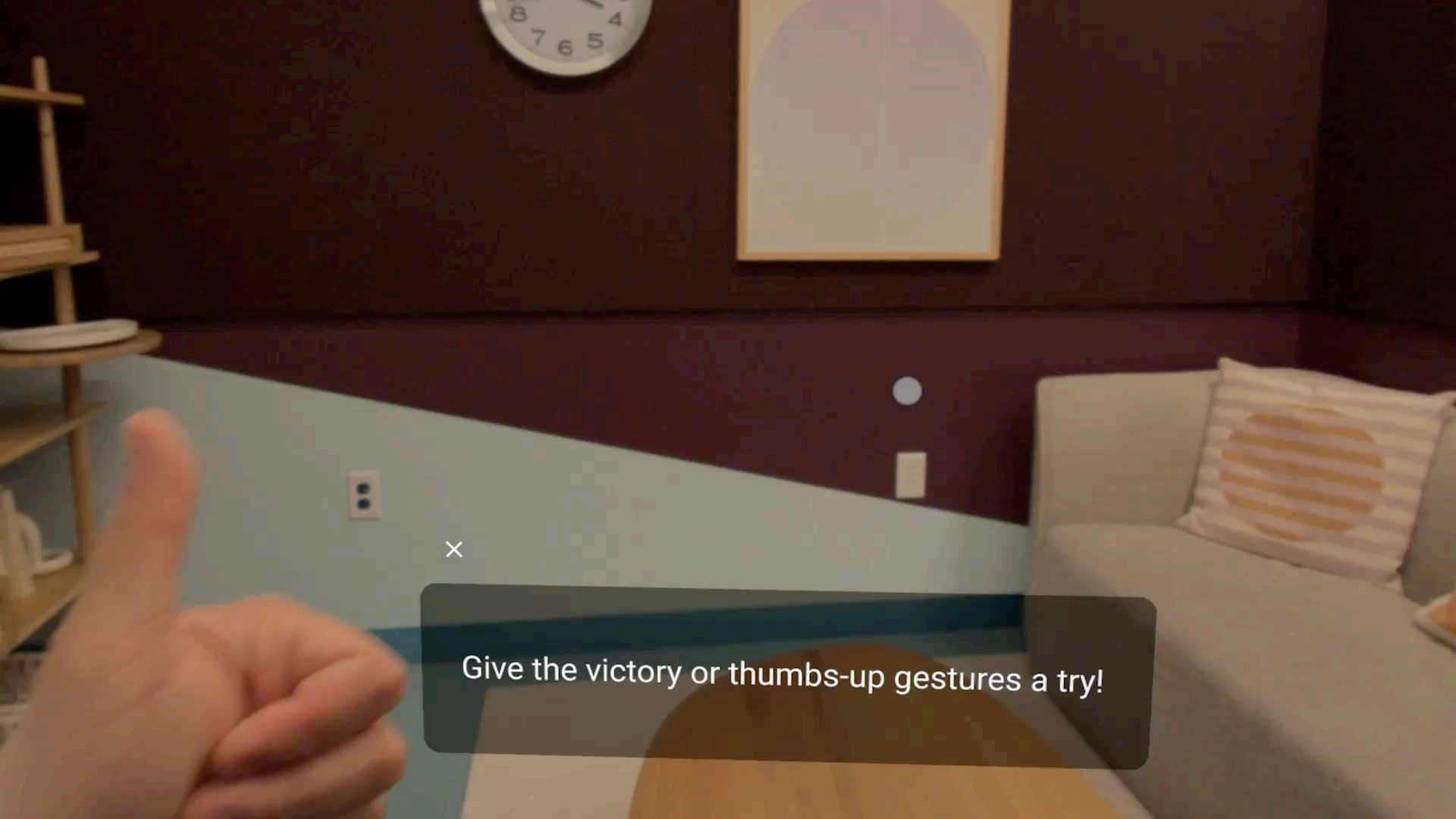




XR Interaction

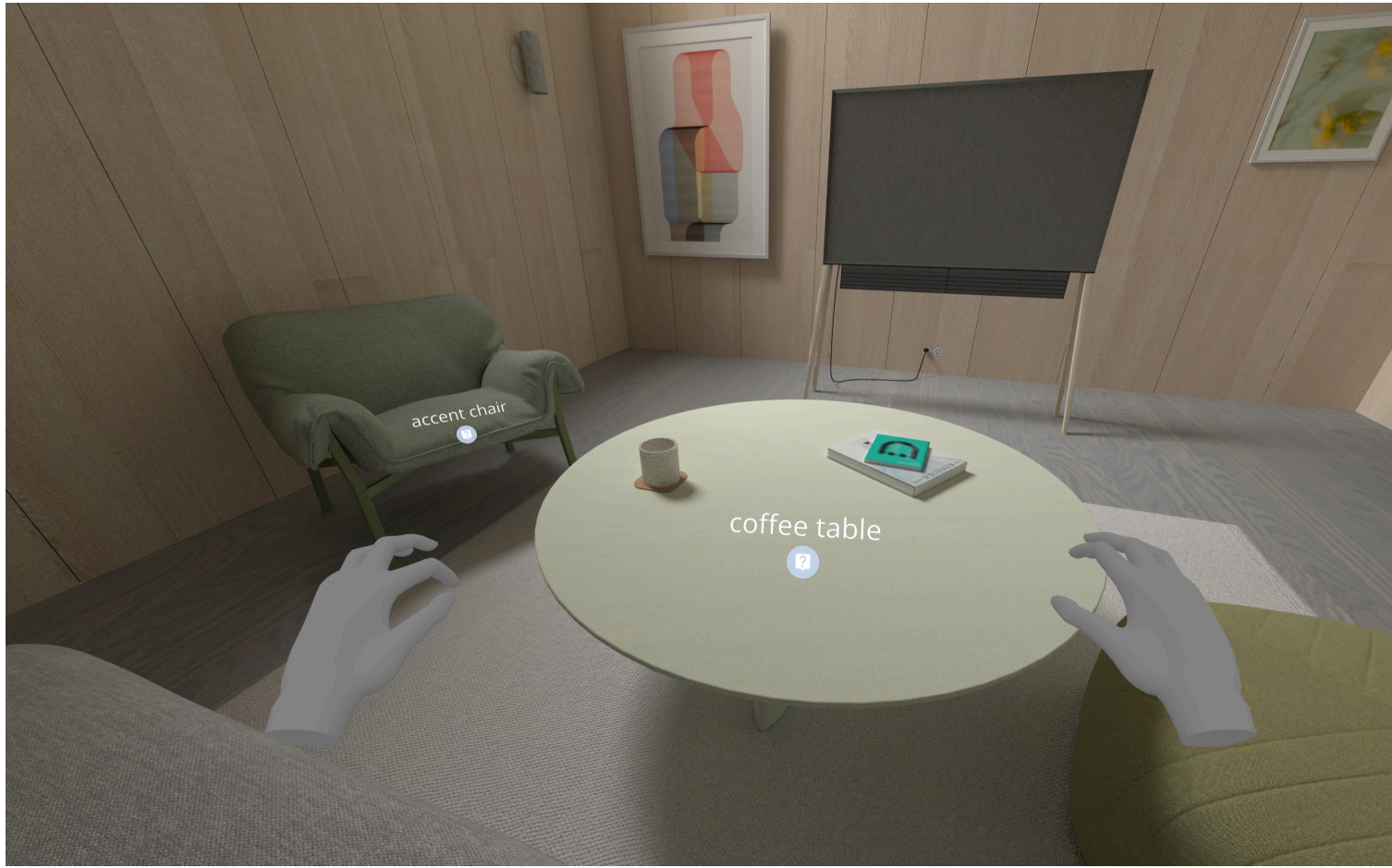
On-device Gesture Recognition



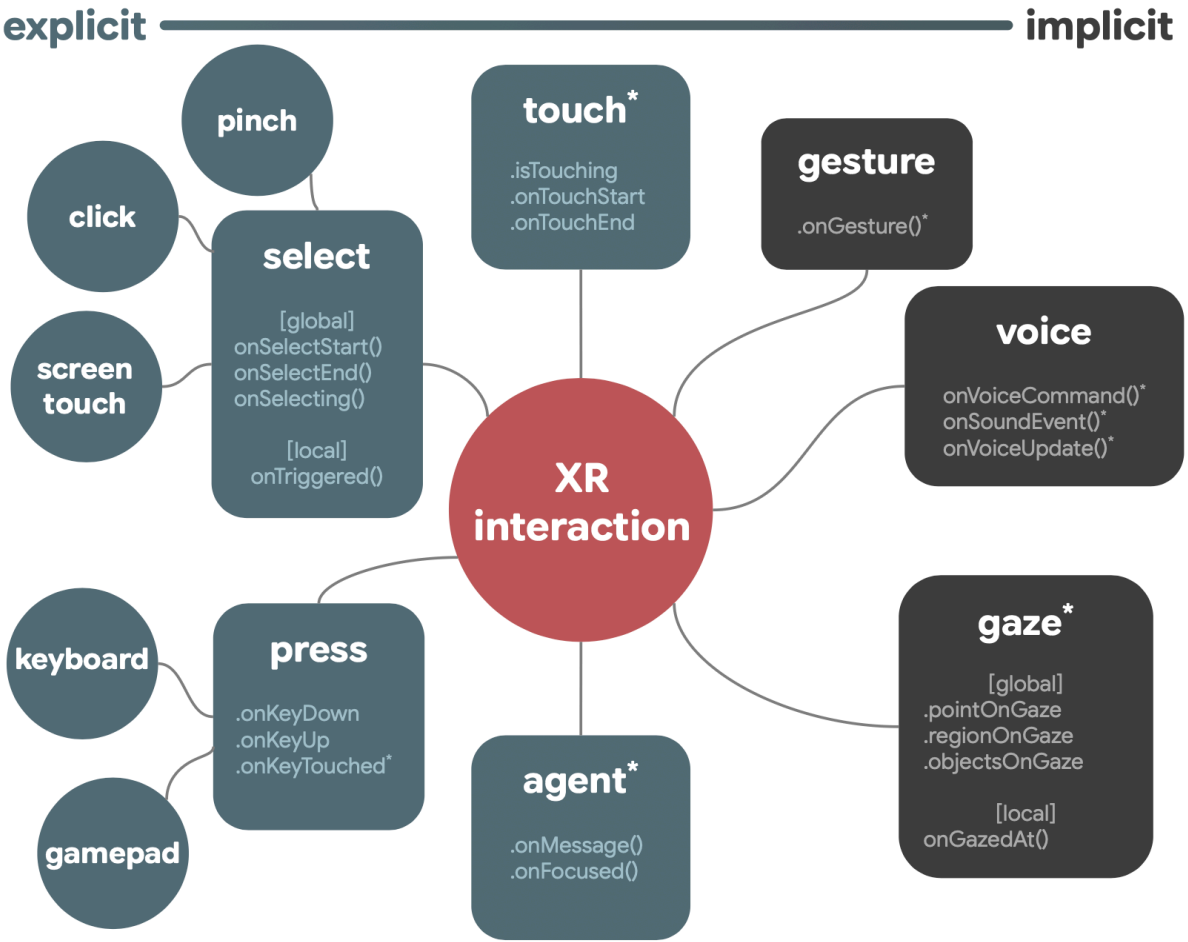

Give the victory or thumbs-up gestures a try!

AI + XR

XR-Objects




POP'D


explicit
implicit
pinch
click
screen touch
select
[global]
onSelectStart()
onSelectEnd()
onSelecting()
[local]
onTriggered()
touch*
.isTouching
.onTouchStart
.onTouchEnd
gesture
.onGesture()*
voice
onVoiceCommand()*
onSoundEvent()*
onVoiceUpdate()*
XR interaction
keyboard
gamepad
press
.onKeyDown
.onKeyUp
.onKeyTouched*
agent*
.onMessage()
.onFocused()
gaze*
[global]
.pointOnGaze
.regionOnGaze
.objectsOnGaze
[local]
onGazedAt()

arXiv:submit/6839982 [cs.HC] 29 Sep 2025

XR Blocks: Accelerating Human-Centered AI + XR Innovation

Preliminary White Paper, September 2025

David Li*, Nels Numan†, Xun Qian†, Yanhe Chen†, Zhongyi Zhou†, Evgenii Alekseev, Geonsun Lee, Alex Cooper, Min Xia, Scott Chung, Jeremy Nelson, Xiuxiu Yuan, Jolica Dias, Tim Bettridge, Benjamin Hersh, Michelle Huynh, Konrad Piascik, Ricardo Cabello, David Kim, Ruofei Du*‡

https://github.com/google/xrblocks

https://xrblocks.github.io

Google XR Labs

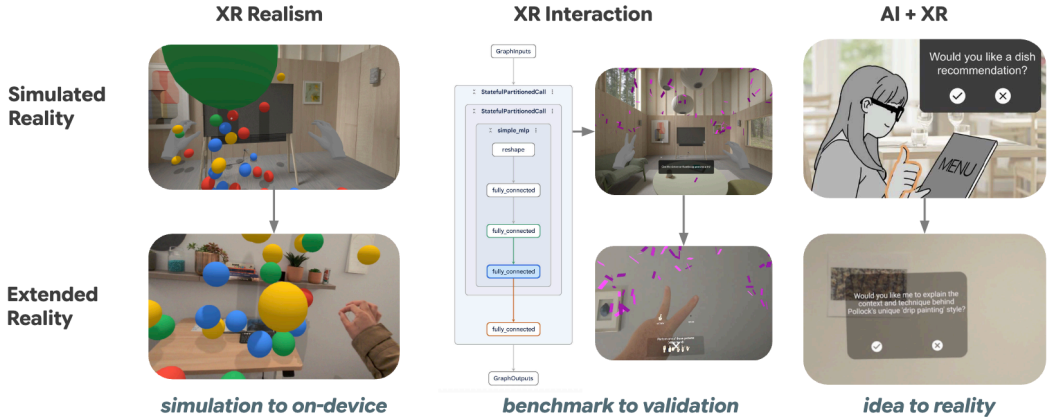


Figure 1: XR Blocks accelerates the prototyping of real-time AI + XR applications across desktop simulators and XR devices. Examples: (a) XR Realism: Prototype depth-aware, physics-based interactions [20] in simulation and deploy the same code to real-world XR devices. (b) XR Interactions: Seamlessly integrate custom gesture models to desktop simulator and on-device XR deployment. (c) AI + XR Integration: Build intelligent, context-aware assistants, like the Sensible Agent [34].

Abstract

We are on the cusp where Artificial Intelligence (AI) and Extended Reality (XR) are converging to unlock new paradigms of interactive computing. However, a significant gap exists between the ecosystems of these two fields: while AI research and development is accelerated by mature frameworks like JAX and benchmarks like LMArena, prototyping novel AI-driven XR interactions remains a high-friction process, often requiring practitioners to manually integrate disparate components. To bridge this gap, we introduce XR Blocks, an open-source software development kit (SDK) designed to accelerate human-centered AI + XR innovation. XR Blocks provides a modular architecture, reusable components, and a rapid simulation environment. By abstracting away boilerplate code and offering high-level templates for common XR paradigms, the toolkit lowers the barrier to entry for XR creators. We demonstrate its utility through a set of open-source templates, samples, and advanced demos, empowering the community to quickly move from concept to interactive XR prototype.

Keywords

Extended Reality, Software Development Kit, WebXR, WebGL, Programming Language, Depth-based Interaction, Mixed Reality, Augmented Reality

Ten Million Users and Ten Years Later: Python Tutor's Design Guidelines for Building Scalable and Sustainable Research Software in Academia

Philip J. Guo
UC San Diego
La Jolla, CA, USA

ABSTRACT

Research software is often built as prototypes that never get widespread usage and are left unmaintained after a few papers get published. To counteract this trend, we propose a method for building research software with scale and sustainability in mind so that it can organically grow a large userbase and enable longer-term research. To illustrate this method, we present the design and implementation of Python Tutor (pythontutor.com), a code visualization tool that is, to our knowledge, one of the most widely-used pieces of research software developed within a university lab. Over the past decade, it has been used by over ten million people in over 180 countries. It has also contributed to 55 publications from 35 research groups in 13 countries. We distilled lessons from working on Python Tutor into three sets of design guidelines: 1) user experience design for scale and sustainability, 2) software architecture design for long-term sustainability, and 3) designing a sustainable software development workflow within academia. These guidelines can enable a student to create long-lasting software that reaches many users and facilitates research from many independent groups.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**.

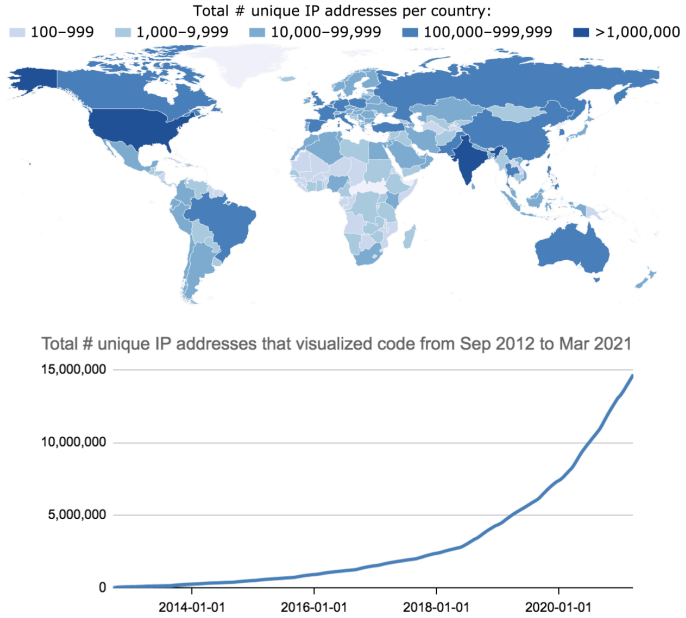


Figure 1: Number of Python Tutor users by location and date, estimated by unique IP addresses that have visualized code on the site. (There was minimal usage from 2009 to 2012.)

Call for action

Join us in the journey towards XR flywheel!



xrblocks Public

Edit Pins | Watch 3 | Fork 34 | Starred 214

main | 2 Branches | 14 Tags | Go to file t | Add file | Code

dli7319 Merge pull request #195 from dli7319/canvas — f02deb4 · 3 days ago — 383 Commits

.github/workflows	Upgrade GitHub Actions to latest versions	2 weeks ago
.vscode	Update README to include ESLint and Prettier in the dev...	5 months ago
assets	Add screenshots to the README.	6 months ago
demos	Fix linting/formatting issues and add light	last week
docs	Bump the npm_and_yarn group across 2 directories with...	5 days ago
samples	Improve local development workflow/instructions	3 weeks ago
src	Add Gemini Canvas check.	3 days ago
templates	Deduplicate GeminiManager functionality.	2 months ago
.gitignore	Minor refactoring of samples.	3 months ago
.prettierrc.json	Add prettier to package.json	5 months ago
CONTRIBUTING.md	Minor refactoring of samples.	3 months ago
LICENSE	Create LICENSE	6 months ago
README.md	Remove redundant build step from README	3 weeks ago
eslint.config.mjs	ESLint: Add browser globals to samples js files.	5 months ago
package-lock.json	Bump the npm_and_yarn group across 2 directories with...	5 days ago
package.json	Bump the npm_and_yarn group across 2 directories with...	5 days ago
rollup.config.js	Update rollup and eslint-plugin-tsdoc to the latest.	last week
tsconfig.base.json	Run npm run format	5 months ago
tsconfig.json	Rollup type declarations	3 weeks ago

README | Code of conduct | Contributing | Apache-2.0 license | Security

XR Blocks

About

XR Blocks is a lightweight WebXR + AI library for rapidly prototyping advanced AI + XR experiences.

xrblocks.github.io

tensorflow | gemini | gestures | webxr | litert | webai | androidxr | xrblocks

Readme
Apache-2.0 license
Code of conduct
Contributing
Security policy
Activity
Custom properties
214 stars
3 watching
34 forks
Report repository

Releases 14

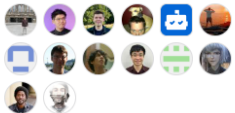
v0.10.0: PlanarVST, Ballpit... Latest
3 weeks ago

+ 13 releases

Packages

No packages published
Publish your first package

Contributors 14



arXiv:submit/6839982 [cs.HC] 29 Sep 2025

XR Blocks: Accelerating Human-Centered AI + XR Innovation

Preliminary White Paper, September 2025

David Li*, Nels Numan†, Xun Qian†, Yanhe Chen†, Zhongyi Zhou†, Evgenii Alekseev, Geonsun Lee, Alex Cooper, Min Xia, Scott Chung, Jeremy Nelson, Xiuxiu Yuan, Jolica Dias, Tim Bettridge, Benjamin Hersh, Michelle Huynh, Konrad Piascik, Ricardo Cabello, David Kim, Ruofei Du*‡

https://github.com/google/xrblocks

https://xrblocks.github.io

Google XR Labs

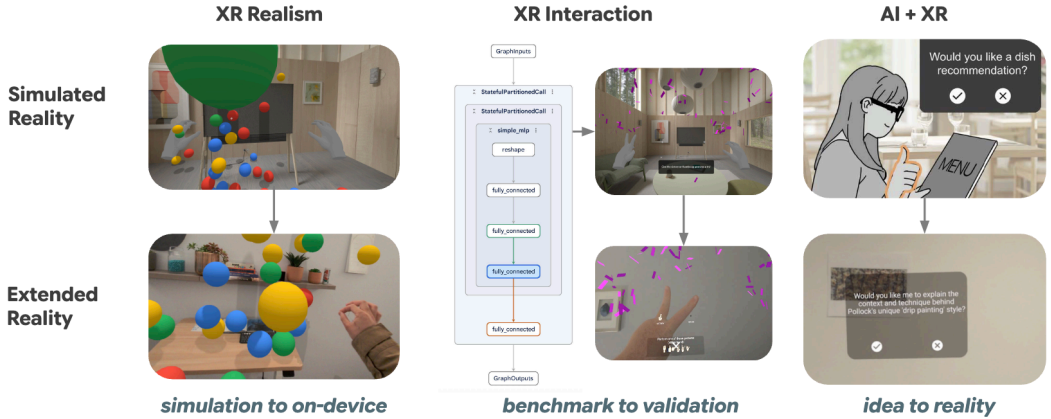


Figure 1: XR Blocks accelerates the prototyping of real-time AI + XR applications across desktop simulators and XR devices. Examples: (a) XR Realism: Prototype depth-aware, physics-based interactions [20] in simulation and deploy the same code to real-world XR devices. (b) XR Interactions: Seamlessly integrate custom gesture models to desktop simulator and on-device XR deployment. (c) AI + XR Integration: Build intelligent, context-aware assistants, like the Sensible Agent [34].

Abstract

We are on the cusp where Artificial Intelligence (AI) and Extended Reality (XR) are converging to unlock new paradigms of interactive computing. However, a significant gap exists between the ecosystems of these two fields: while AI research and development is accelerated by mature frameworks like JAX and benchmarks like LMArena, prototyping novel AI-driven XR interactions remains a high-friction process, often requiring practitioners to manually integrate disparate components. To bridge this gap, we introduce XR Blocks, an open-source software development kit (SDK) designed to accelerate human-centered AI + XR innovation. XR Blocks provides a modular architecture, reusable components, and a rapid simulation environment. By abstracting away boilerplate code and offering high-level templates for common XR paradigms, the toolkit lowers the barrier to entry for XR creators. We demonstrate its utility through a set of open-source templates, samples, and advanced demos, empowering the community to quickly move from concept to interactive XR prototype.

Keywords

Extended Reality, Software Development Kit, WebXR, WebGL, Programming Language, Depth-based Interaction, Mixed Reality, Augmented Reality

XR Blocks

Accelerating Human-Centered AI + XR Innovation

xrblocks.github.io
github.com/google/xrblocks

David Li*, Nels Numan†, Xun Qian†, Yanhe Chen†, Zhongyi Zhou†, Evgenii Alekseev, Geonsun Lee, Alex Cooper, Min Xia, Scott Chung, Jeremy Nelson, Xiuxiu Yuan, Jolica Dias, Tim Bettridge, Benjamin Hersh, Michelle Huynh, Konrad Piascik, Ricardo Cabello, David Kim, Ruofei Du*
Google XR Labs



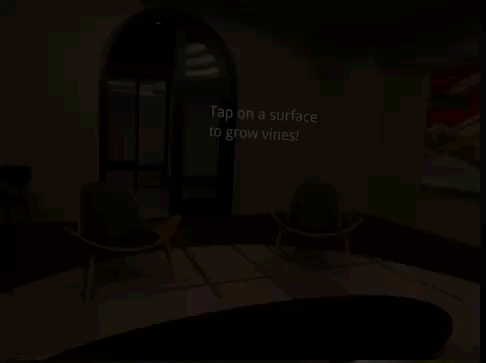
XR Blocks

Accelerating Human-Centered AI + XR Innovation

xrblocks.github.io
github.com/google/xrblocks

David Li*, Nels Numan†, Xun Qian†, Yanhe Chen†, Zhongyi Zhou†, Evgenii Alekseev, Geonsun Lee, Alex Cooper, Min Xia, Scott Chung, Jeremy Nelson, Xiuxiu Yuan, Jolica Dias, Tim Bettridge, Benjamin Hersh, Michelle Huynh, Konrad Piascik, Ricardo Cabello, David Kim, Ruofei Du*
Google XR Labs



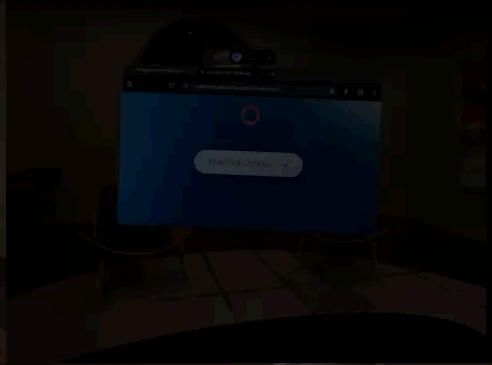

Tap on a surface
to grow vines!


Site can use your camera


Room
Chair
Table
Mode



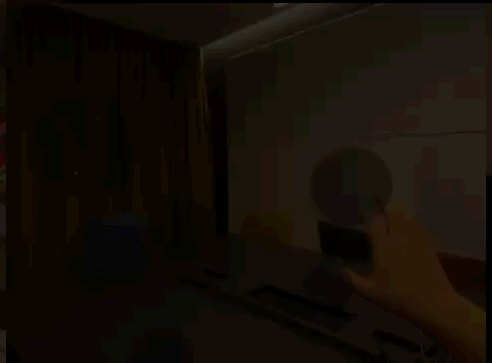




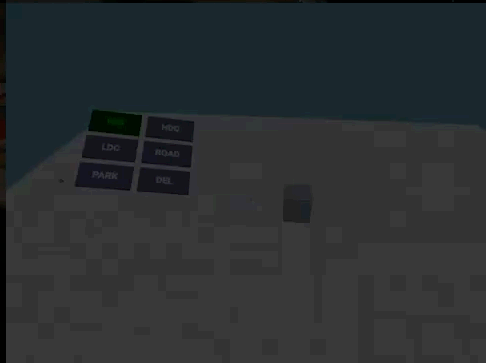


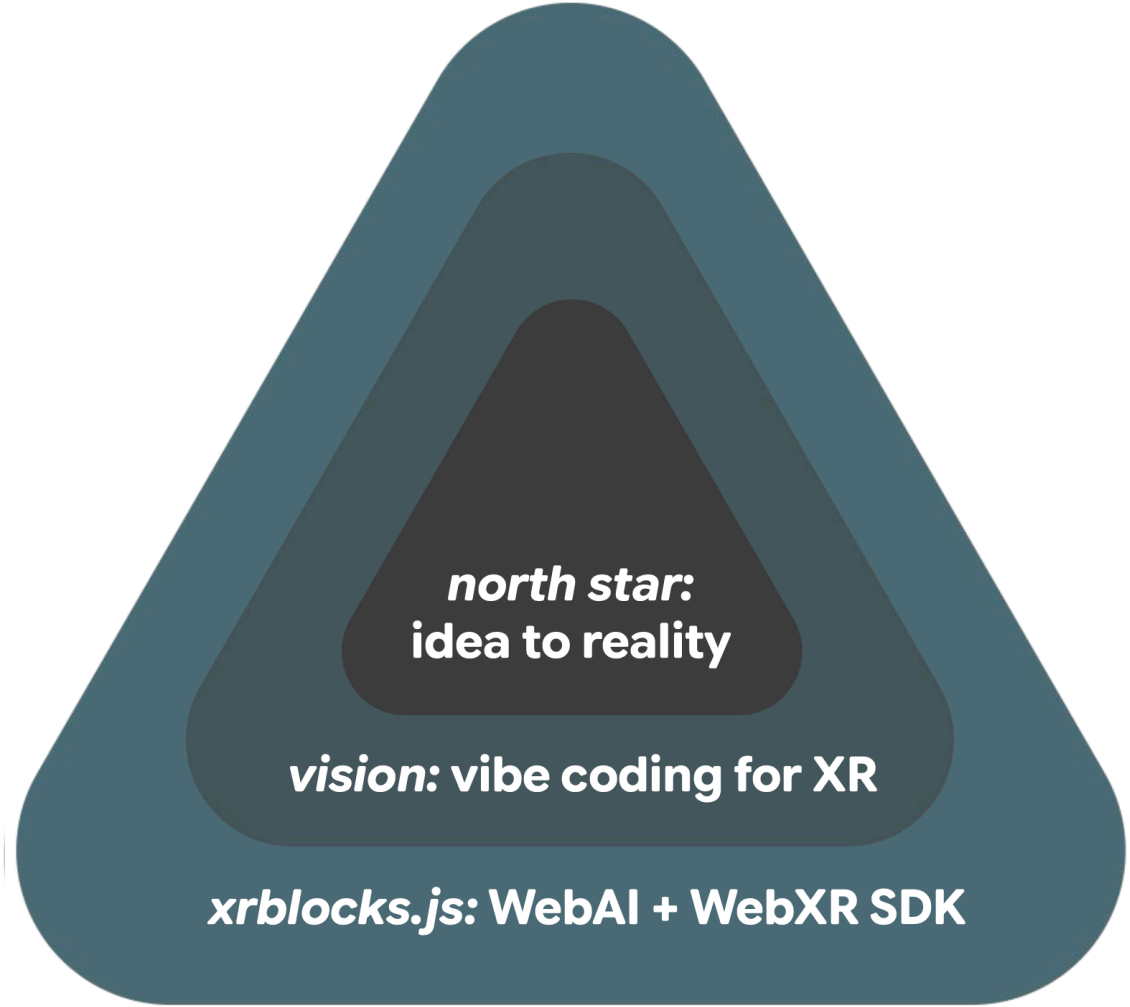


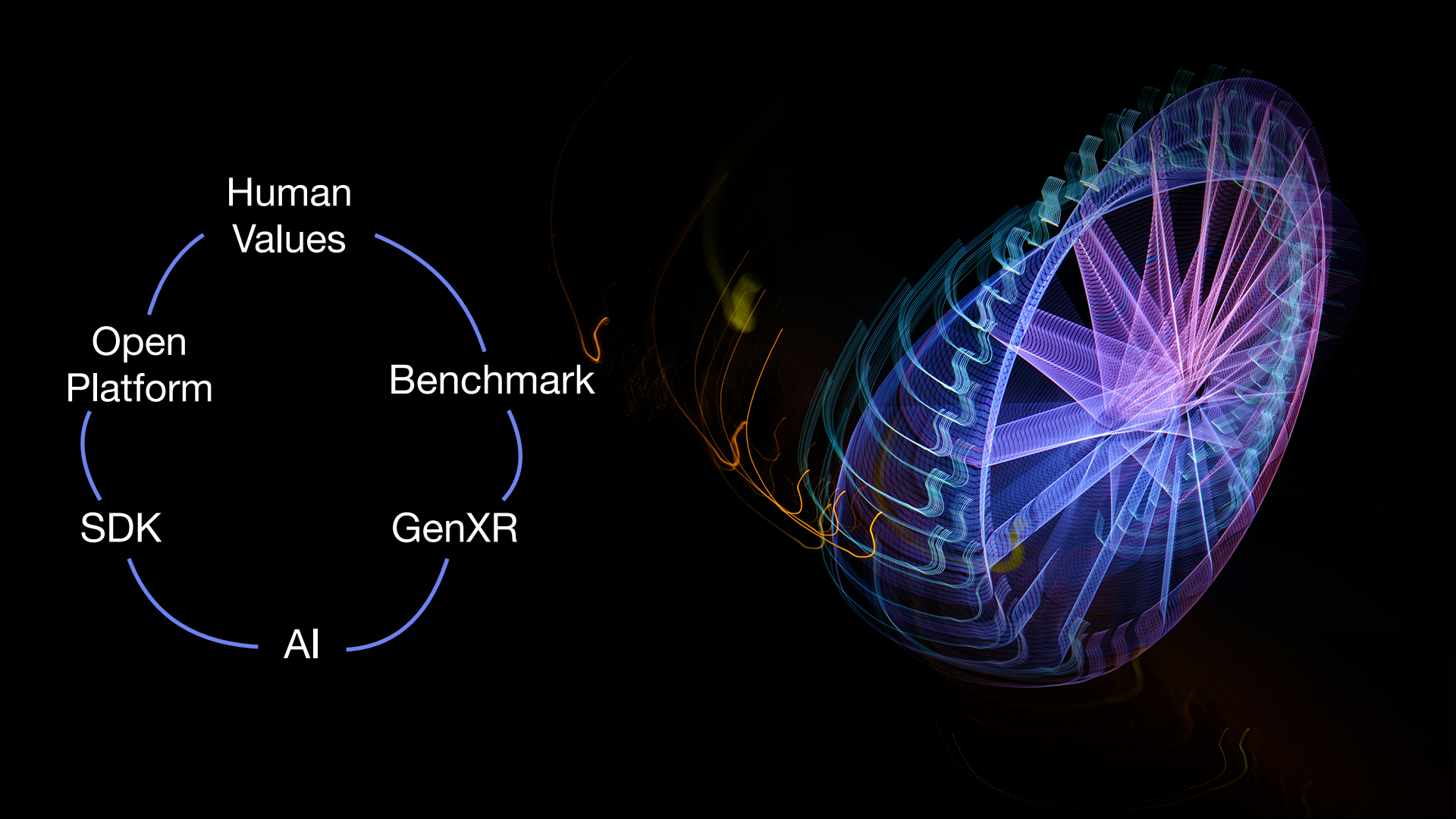

SCORE: 909
DODGE!

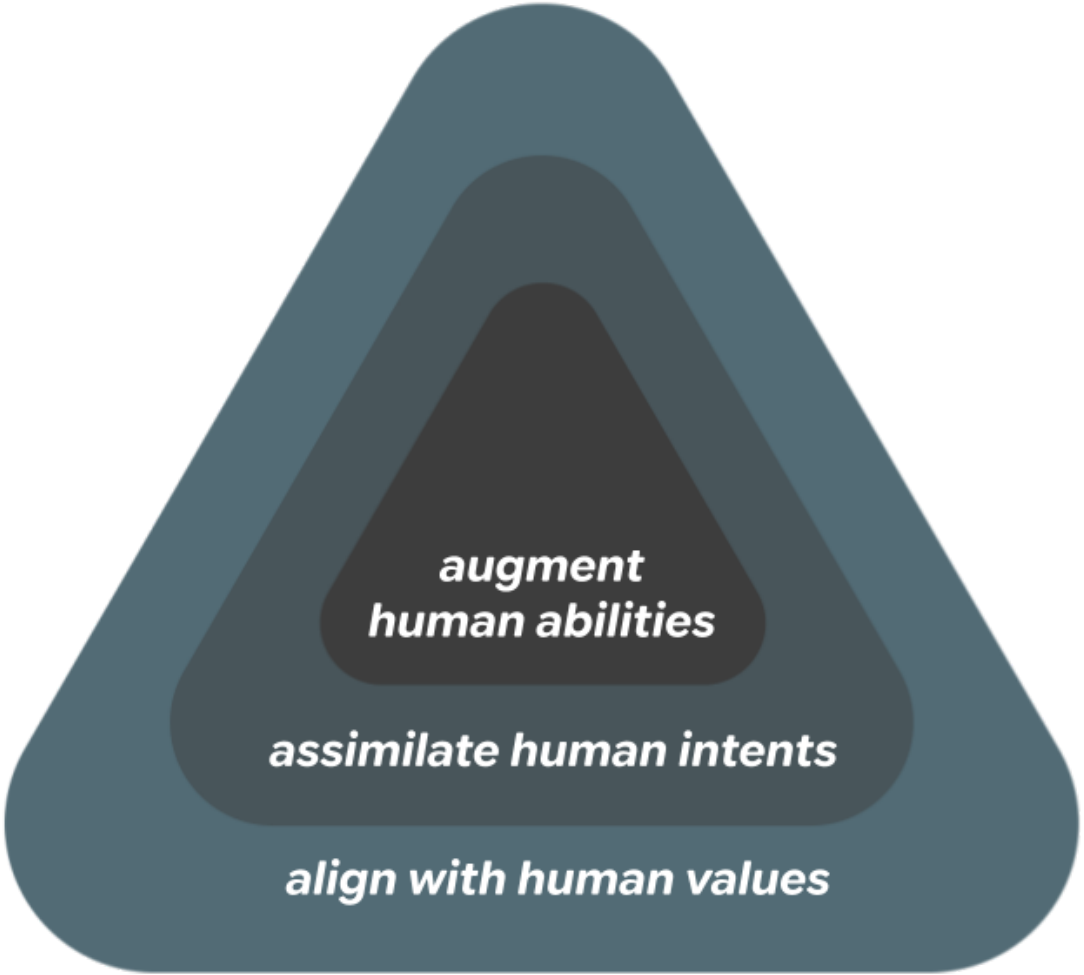



Site can use your camera


HDC
LDC
ROAD
PARK
DEL


north star:
idea to reality
vision: vibe coding for XR
xrblocks.js: WebAI + WebXR SDK


Human
Values
Benchmark
GenXR
AI
SDK
Open
Platform


augment
human abilities
assimilate human intents
align with human values


So, how can we predict the future of XR?


TED

Co-presence

Future Directions

The Ultimate XR Platform



Future Directions

Fuses Past Events



Immersive Learning



Immersive Tourism


Future Directions
And look into the future
Predict
the future


Future Directions
Change the way we communicate in 3D and consume the information
No gap in space / language

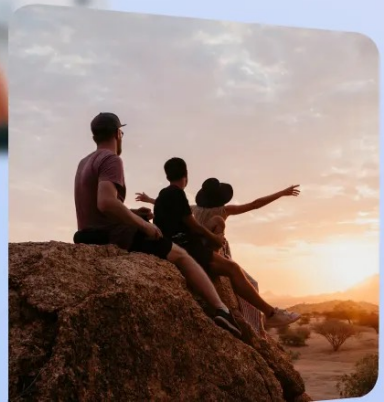
Future Directions

Make a better world through concrete inventions





Android XR





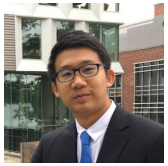
Acknowledgement

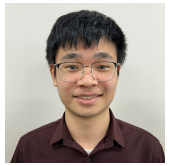
Advisors & Collaborators & Students & more…


David Kim


Shahram Izadi


Alex Olwal


Xun Qian

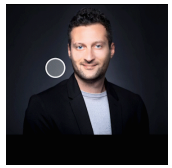

Zhongyi Zhou


Mahdi Tayarani


Max Dzitsiuk


David Li


Yinda Zhang


Sean Fanello

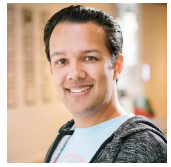

Adarsh Kowdle


Na Li

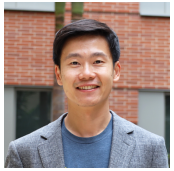

Xiuxiu Yuan


Vladimir Kirilyuk


Jing Jin


Jason Mayes


Amitabh Varshney


Yang Zhang


Xiang 'Anthony' Chen


Seongkook Heo

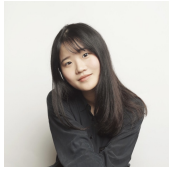

Ken Perlin


Karthik Ramani


Jens Emil Sloth Grønbæk


Jon Froehlich


Geonsun Lee


Erzhen Hu

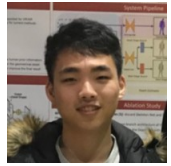

Safa C. Medin


Ziyi Liu


Hugues Hoppe

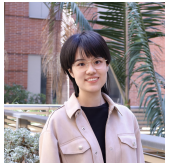

Zhenyi He


Xiaoxu Meng


Feitong Tan


Faraz Faruqi


Xingyu Bruce Liu


Siyou Pei


Dhruv Jain


Ananta Balaji


Doga Dogan


Fengyuan Zhu


Sujal Bista

Kicking Off the Flywheel of AI + XR Innovations with XR Blocks



Ruofei Du
XR Labs Lead / Mgr @ Google
March 10, 2026 @ MIT HCI seminar
www.ruofeidu.com
X: DuRuofei@
me@duruofei.com