Interaction-Centric AI

Juho Kim

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

kixlab
KAIST INTERACTION LAB

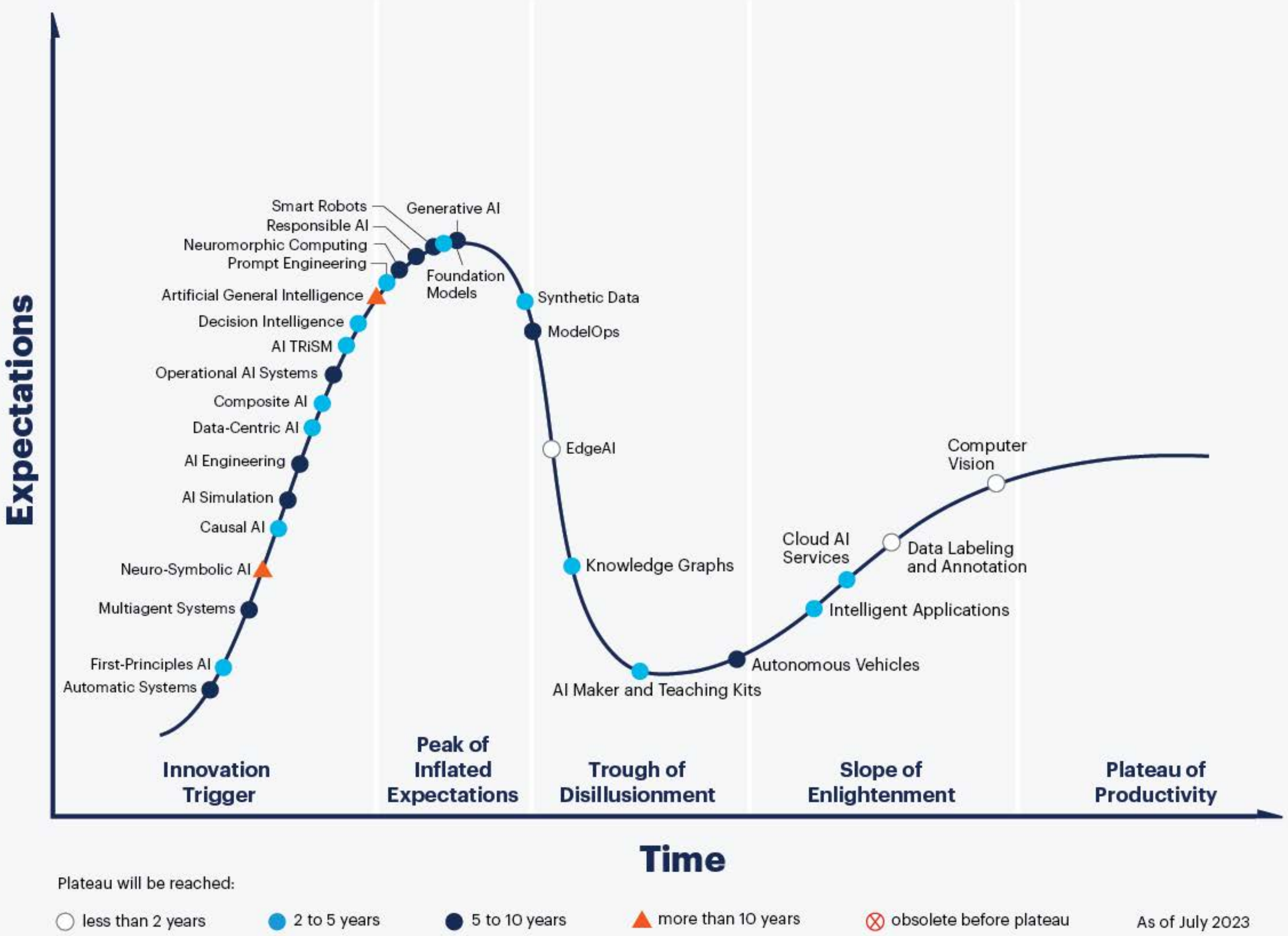
KAIST School of Computing

- KAIST 전산학부 부교수 / 김재철AI대학원 겸임교수
- 연구실: KIXLAB (KAIST Interaction Lab)
- 연구 분야: 인간-컴퓨터 상호작용(HCI), 인간-AI 상호작용

- Ph.D. MIT EECS
- M.S. Stanford University
- B.S. 서울대학교

- 산학협력: 삼성전자, 네이버, KT, LG전자, LG U+, LG AI연구원, HMG, 삼성SDS, Adobe, ONR, 링글, 알고리즘랩스, 엘리스 등
- NeurIPS 2022 키노트 연사
- 국제학회 우수논문상 14회 수상 (ACM CHI 9회)
- 한국정보과학회/IEEE-CS 젊은정보과학자상, KAIST Q-Day 창의교육 부문 특별표창, 송암 미래석학우수연구상, 창의강의대상, 황규영 커리어 어워드, 우수강의상 수상

Hype Cycle for Artificial Intelligence, 2023

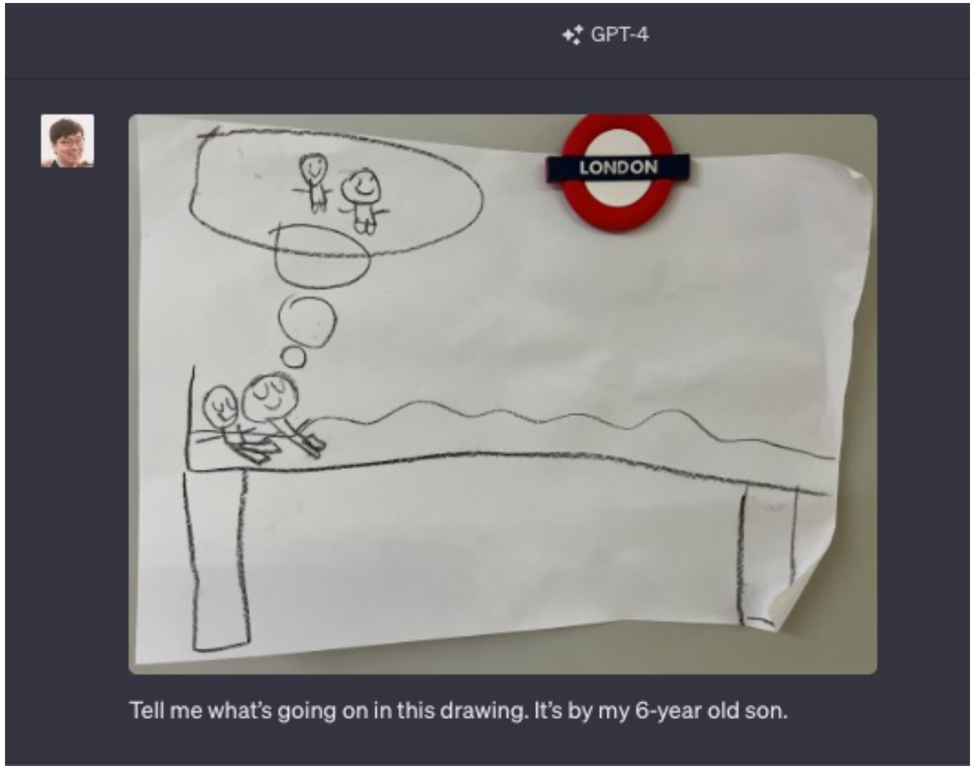


gartner.com

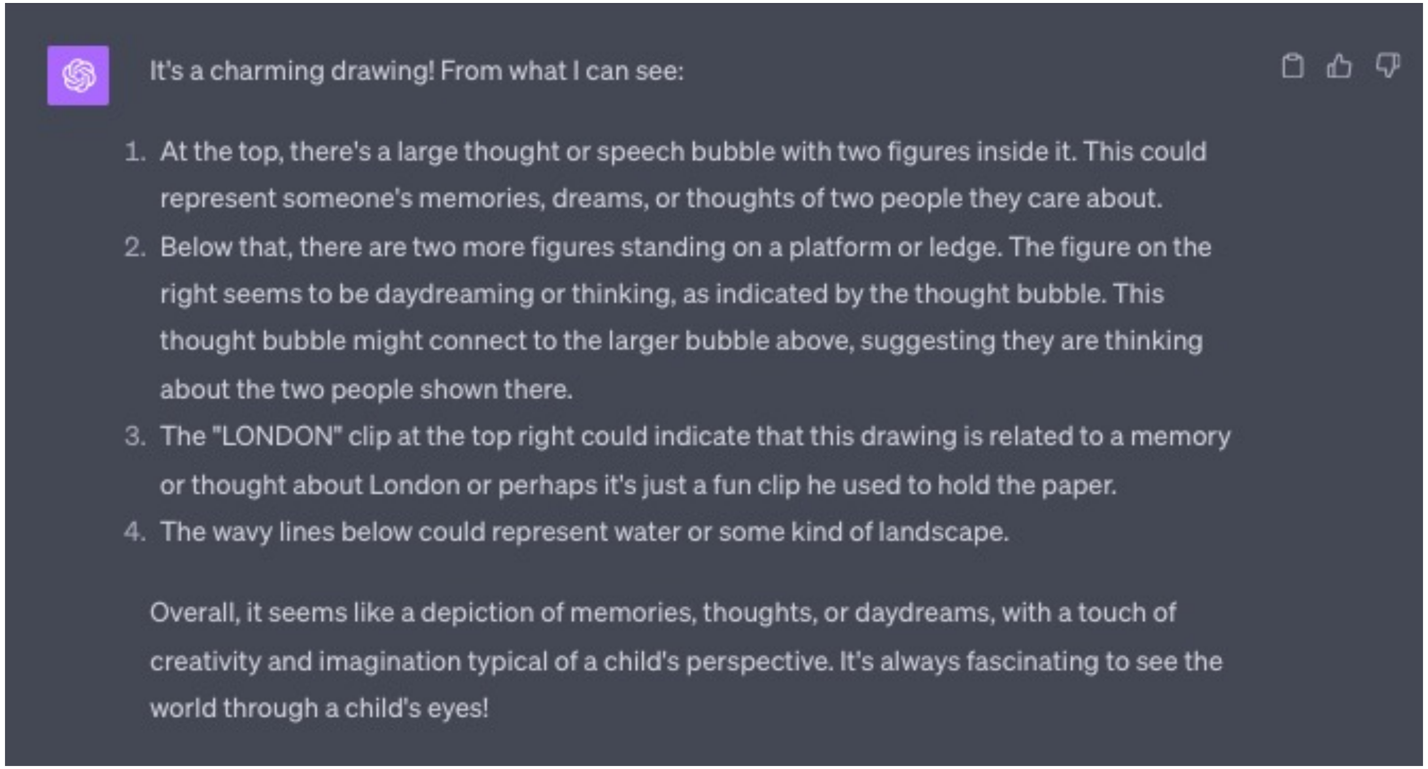
Source: Gartner
© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. 2079794

Gartner®

ChatGPT = 거대언어모델 + 대화형 인터페이스



Tell me what's going on in this drawing. It's by my 6-year old son.



It's a charming drawing! From what I can see:

1. At the top, there's a large thought or speech bubble with two figures inside it. This could represent someone's memories, dreams, or thoughts of two people they care about.
2. Below that, there are two more figures standing on a platform or ledge. The figure on the right seems to be daydreaming or thinking, as indicated by the thought bubble. This thought bubble might connect to the larger bubble above, suggesting they are thinking about the two people shown there.
3. The "LONDON" clip at the top right could indicate that this drawing is related to a memory or thought about London or perhaps it's just a fun clip he used to hold the paper.
4. The wavy lines below could represent water or some kind of landscape.

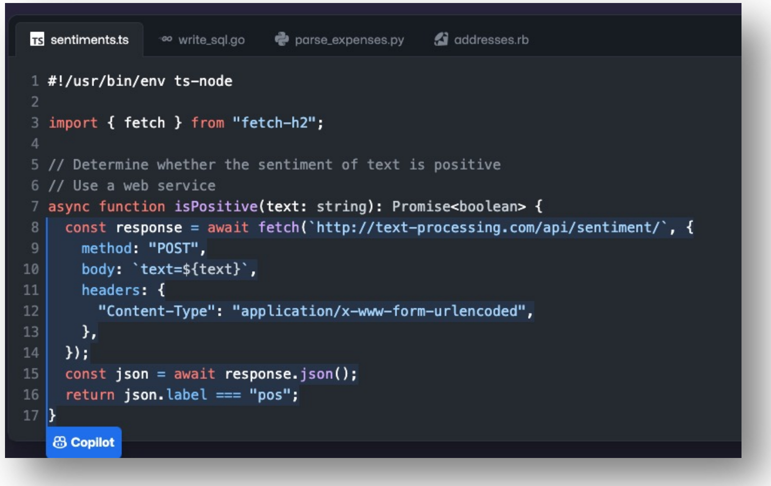
Overall, it seems like a depiction of memories, thoughts, or daydreams, with a touch of creativity and imagination typical of a child's perspective. It's always fascinating to see the world through a child's eyes!

GPT4V | Generated on Oct. 3, 2023

거대언어모델과 생성형 AI가 통합된 서비스

GitHub

programming



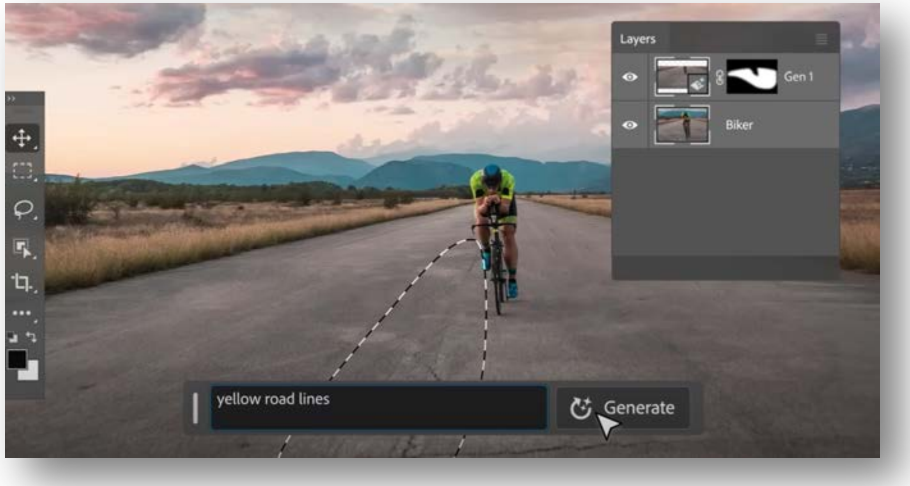
```
#!/usr/bin/env ts-node

import { fetch } from "fetch-h2";

// Determine whether the sentiment of text is positive
// Use a web service
async function isPositive(text: string): Promise<boolean> {
  const response = await fetch(`http://text-processing.com/api/sentiment/`, {
    method: "POST",
    body: `text=${text}`,
    headers: {
      "Content-Type": "application/x-www-form-urlencoded",
    },
  });
  const json = await response.json();
  return json.label === "pos";
}
```

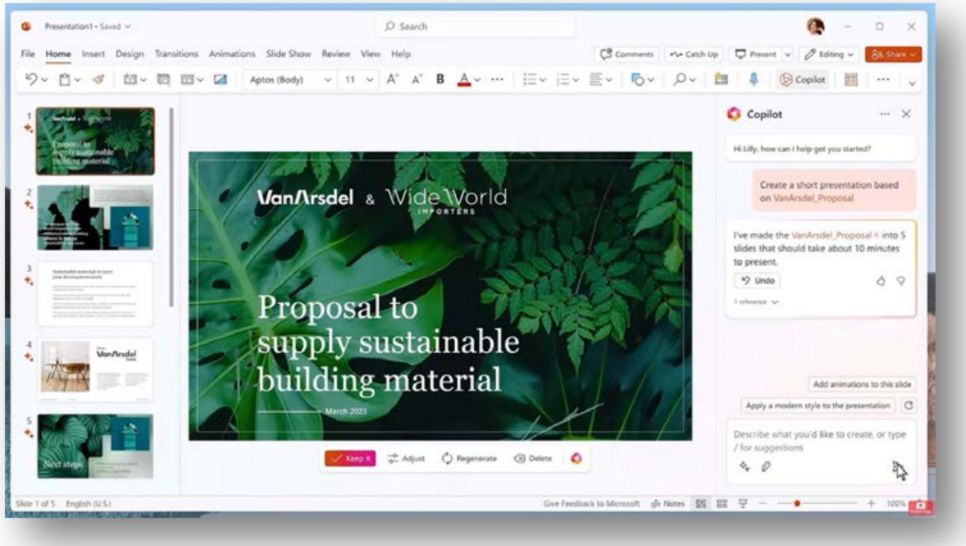
Adobe

image editing



Microsoft

office suite



Hi Lilly, how can I help get you started?

Create a short presentation based on VanArsdel_Proposal

I've made the VanArsdel_Proposal into 5 slides that should take about 10 minutes to present.

Undo

1 reference

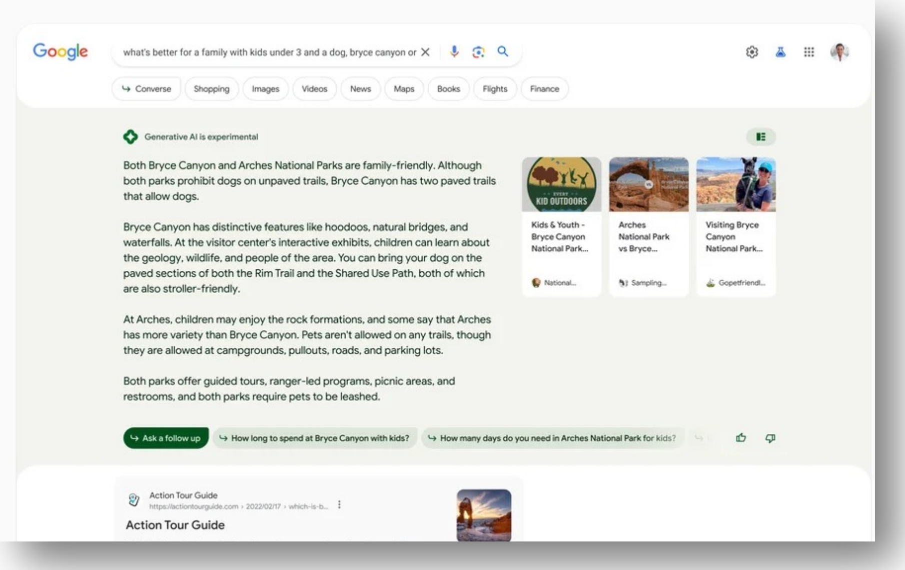
Add animations to this slide

Apply a modern style to the presentation

Describe what you'd like to create, or type / for suggestions

Google

web search



what's better for a family with kids under 3 and a dog, bryce canyon or

Generative AI is experimental

Both Bryce Canyon and Arches National Parks are family-friendly. Although both parks prohibit dogs on unpaved trails, Bryce Canyon has two paved trails that allow dogs.

Bryce Canyon has distinctive features like hoodoos, natural bridges, and waterfalls. At the visitor center's interactive exhibits, children can learn about the geology, wildlife, and people of the area. You can bring your dog on the paved sections of both the Rim Trail and the Shared Use Path, both of which are also stroller-friendly.

At Arches, children may enjoy the rock formations, and some say that Arches has more variety than Bryce Canyon. Pets aren't allowed on any trails, though they are allowed at campgrounds, pullouts, roads, and parking lots.

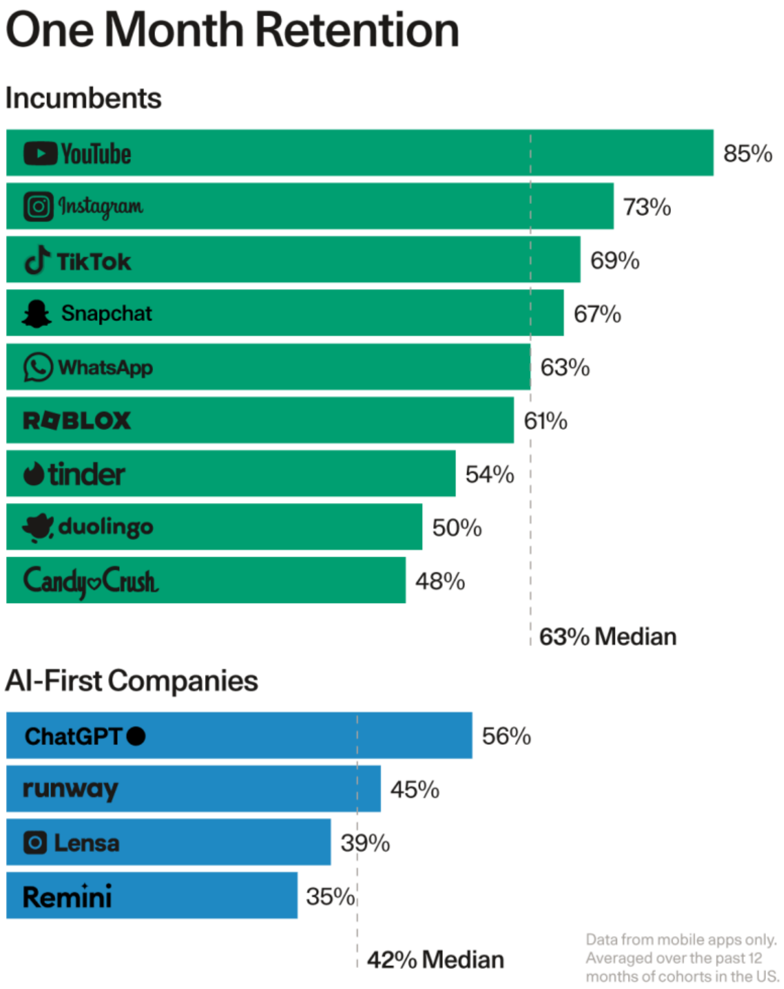
Both parks offer guided tours, ranger-led programs, picnic areas, and restrooms, and both parks require pets to be leashed.

Ask a follow up

How long to spend at Bryce Canyon with kids?

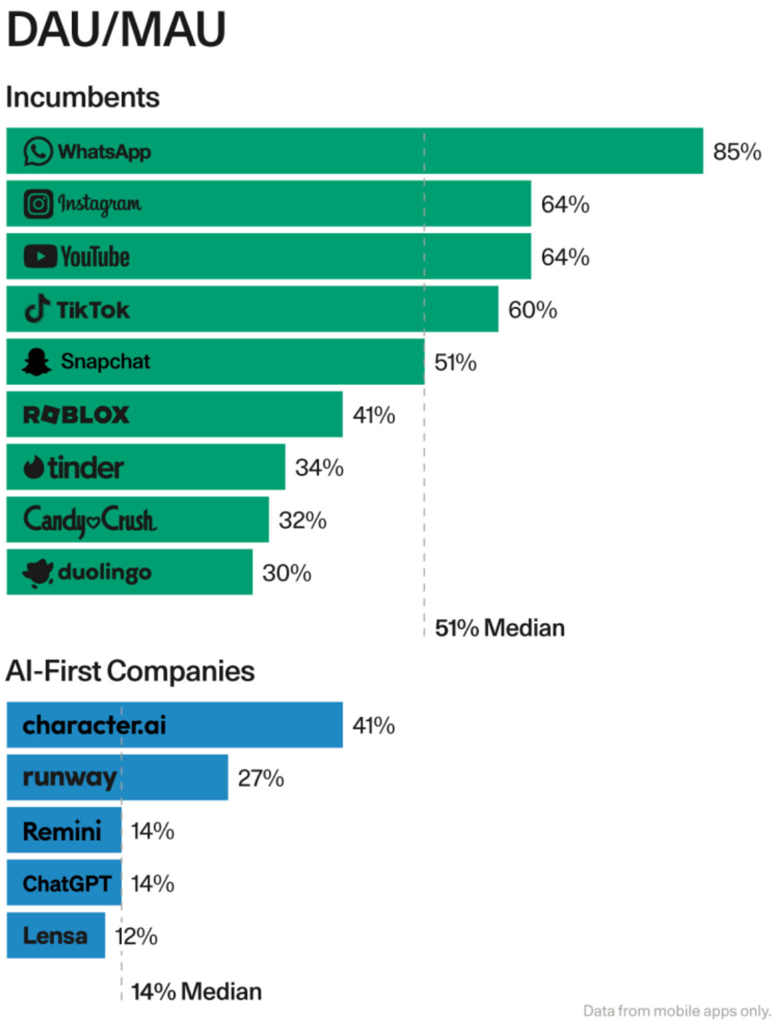
How many days do you need in Arches National Park for kids?

현실: 쿨한 데모를 넘지 못하는 AI 기반 서비스



One Month Retention

Data from mobile apps only. Averaged over the past 12 months of cohorts in the US.



DAU/MAU

Data from mobile apps only.

https://www.sequoiacap.com/article/generative-ai-act-two/

TOWARDS ACT TWO

Act One		Act Two
Technology-cut	**Interaction**	Customer-back
Foundation models as is	**Architecture**	Foundation models as a component
Lightweight demonstrations of cool new technology	**Evaluation**	Solve human problems end-to-end

https://www.sequoiacap.com/article/generative-ai-act-two/

생성형 AI가 직면한 세 가지 interaction 과제

말을 알아듣는 것 같기는 한데…

사람의 의도와 가치를 제대로 담아낼 수 있는가?

잘 되는 사례를 보면 좋기는 한데…

다양한 사용자 그룹에 일관된 경험을 제공할 수 있는가?

혼자서 척척 잘 하기는 하는데…

사람에게 충분한 주도권을 줄 수 있는가?

생성형 AI가 직면한 세 가지 interaction 과제

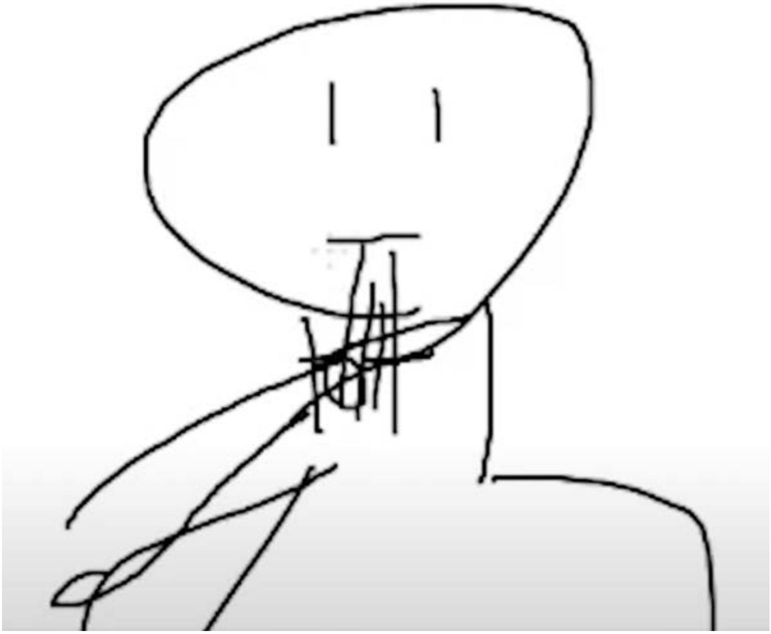
사람의 의도와 가치를 제대로 담아낼 수 있는가?

다양한 사용자 그룹에 일관된 경험을 제공할 수 있는가?

사람에게 충분한 주도권을 줄 수 있는가?

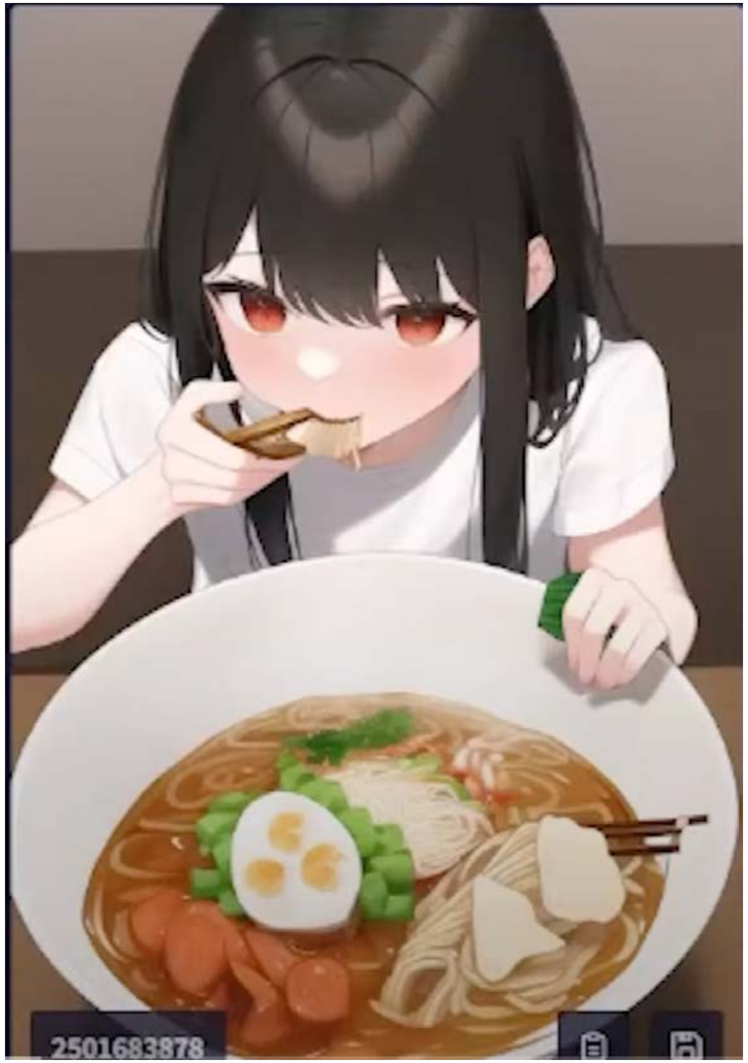
Interaction

- Perturb the data with different levels of Gaussian noise, and jointly estimate the corresponding scores
- Produce samples via an annealed Langevin dynamics
- Require no sampling during training or the use of adversarial methods
- Produce samples comparable to GANs on MNIST, CelebA and CIFAR-10 datasets
- Achieve a new state-of-the-art inception score of 8.87 on CIFAR-10



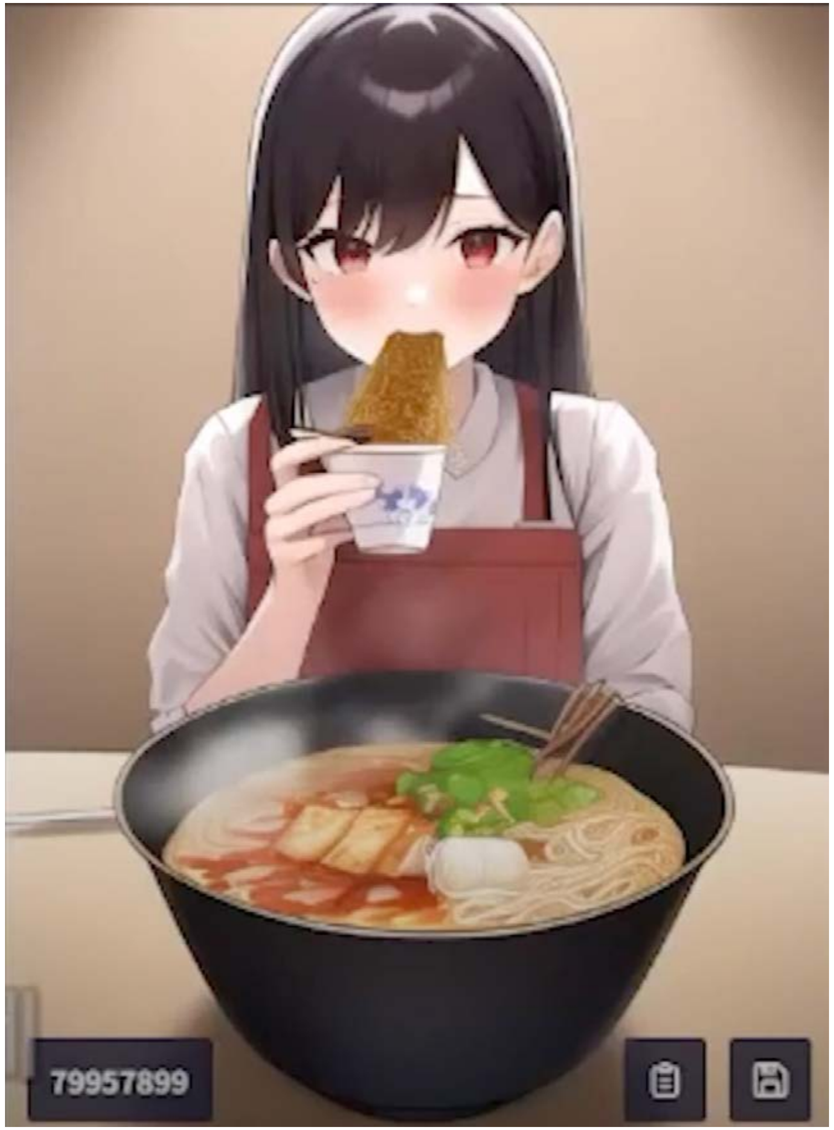
Song, Yang, and Stefano Ermon. "Generative modeling by estimating gradients of the data distribution." Advances in Neural Information Processing Systems 32 (2019).

Icon by Vectors Tank – Flaticon. | Creator ARPA. https://www.youtube.com/watch?v=MdK031yh2ul {{{{{{masterpiece}}}}}}, {{{{{{best quality}}}}}}, eating ramen



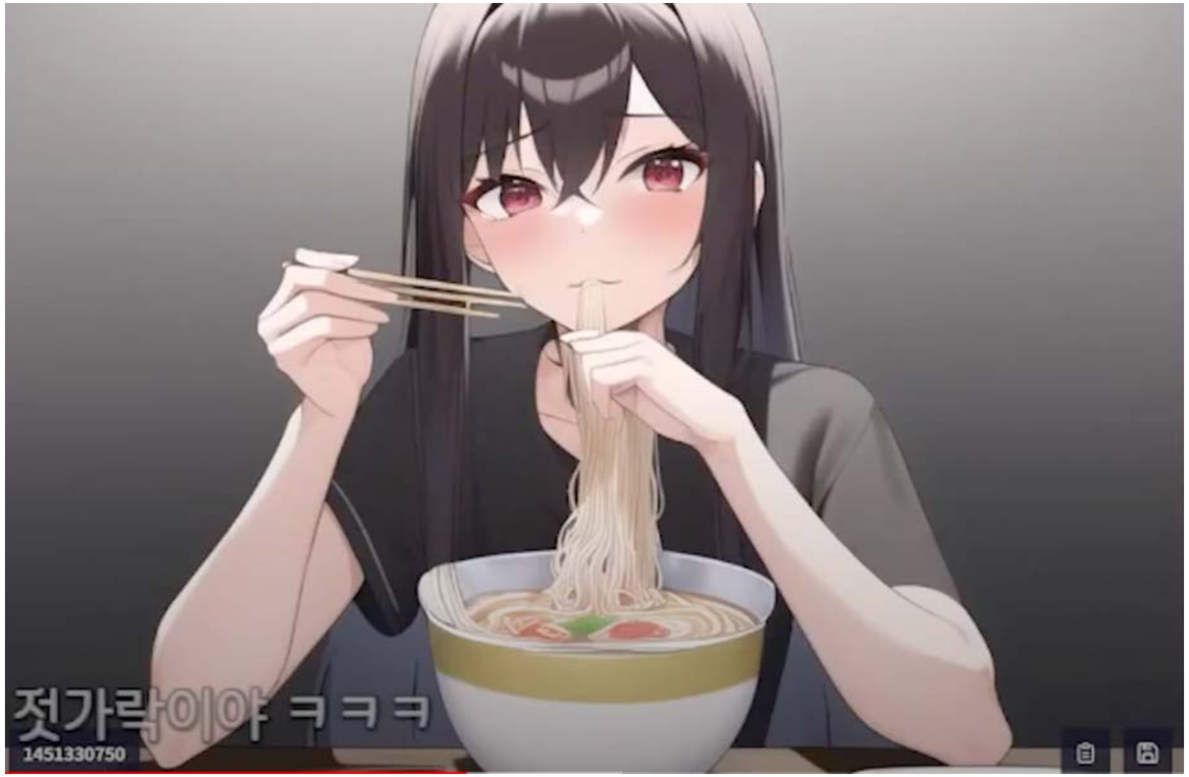
Creator ARPA. https://www.youtube.com/watch?v=MdK031yh2uI

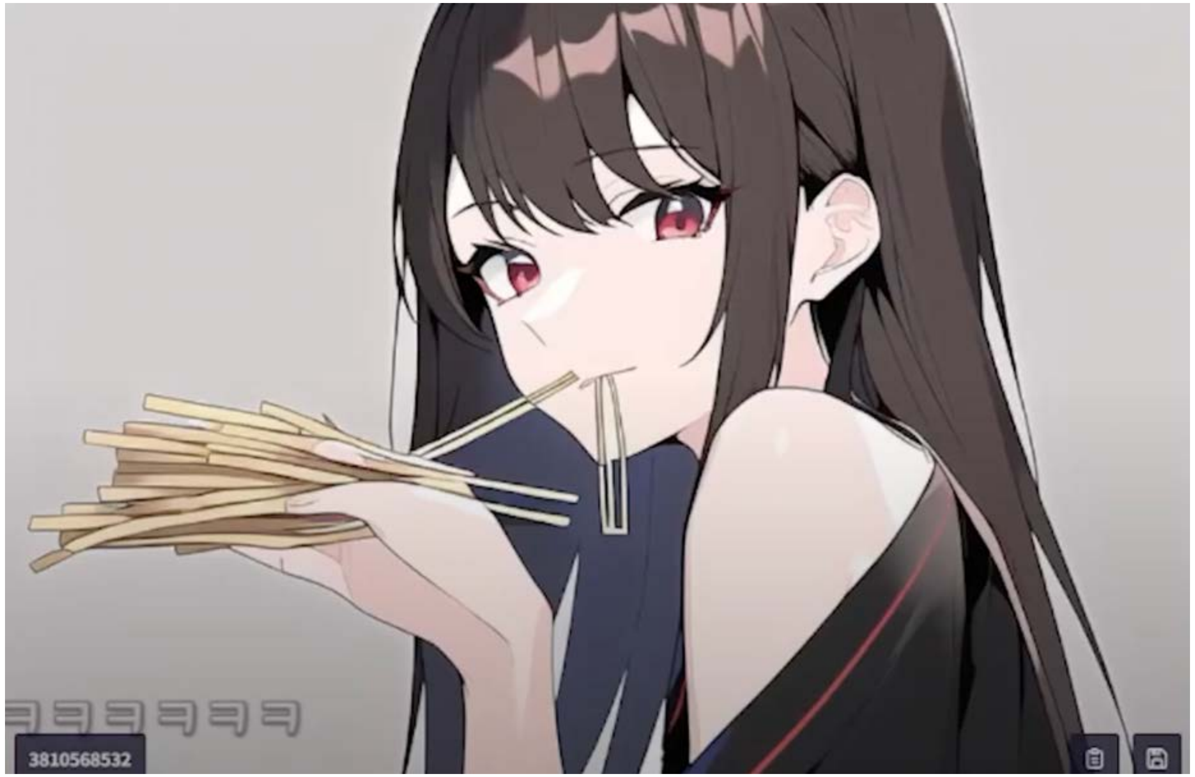
{{{{{{masterpiece}}}}}}, {{{{{{best quality}}}}}}, {{1 girl}}, She is eating ramen



Creator ARPA. https://www.youtube.com/watch?v=MdK031yh2uI

{{{{{{masterpiece}}}}}}, {{{{{{best quality}}}}}}, {{1 girl}},
She picked up the noodles in the ramen with a pair of chopsticks, She took her ramen chopsticks around her mouth, Her right hand is around her mouth, She has chopsticks in her hand close to her mouth, {{chopsticks}}, left hand down

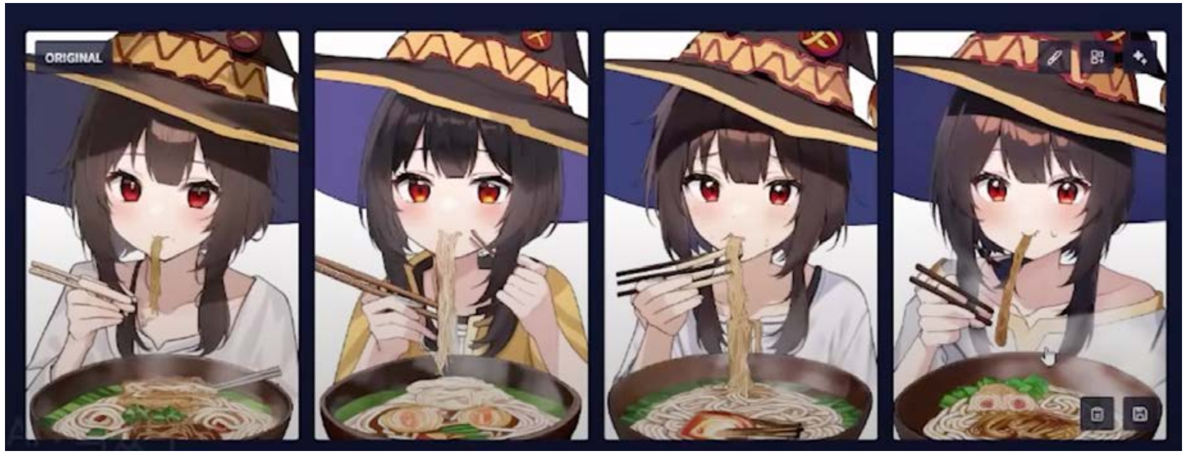




Creator ARPA. https://www.youtube.com/watch?v=MdK031yh2uI

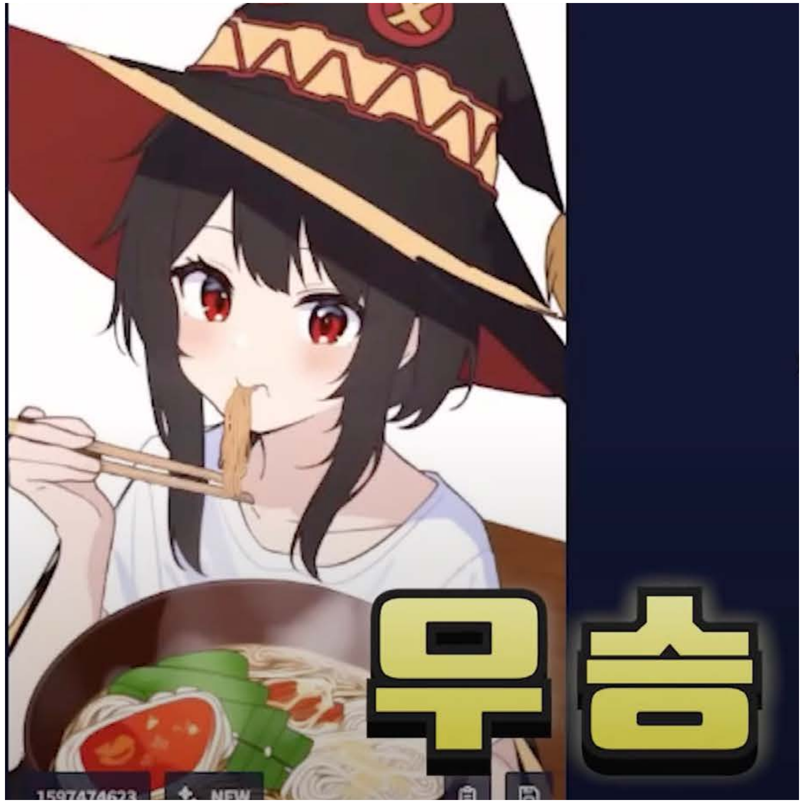
{{{{{{masterpiece}}}}}}, {{{{{{best quality}}}}}}, {{1 girl}}, {{chopsticks}}, short_hair, holding_chopsticks, chopsticks, {{megumin}}, witch hat, azure hair, {{{{{black chopsticks}}}}}, {ramen noodle}, She picked up the noodles in the ramen with a pair of chopsticks, A pair of chopsticks is the same lengh, She took her ramen chopsticks around her mouth, Her right hand is around her mouth, She has chopsticks in her hand close to her mouth, {{{{Chopsticks are made up of two wooden sticks}}}}, {{{{only a pair of chopsticks are drawn in the picture.}}}}, no chopsticks, Ramen noodles are flowing down from her mouth into a ramen bowl,

Variation gacha (がちゃ)



Creator ARPA. https://www.youtube.com/watch?v=MdK031yh2uI

{{{{{{masterpiece}}}}}}, {{{{{{best quality}}}}}}, {{1 girl}}, {{chopsticks}}, {{{eating ramen}}}, short_hair, holding_chopsticks, chopsticks, {{megumin}}, witch hat, white t-shirt, azure hair, {{{{{black chopsticks}}}}}}, {ramen noodle}, from above, She picked up the noodles in the ramen with a one pair of chopsticks, A pair of chopsticks is the same lengh, She took her ramen chopsticks around her mouth, Her right hand is around her mouth, She has chopsticks in her hand close to her mouth, {{{{Chopsticks are made up of two wooden sticks}}}}, {{{{only a pair of chopsticks are drawn in the picture.}}}}, no chopsticks, Ramen noodles are flowing down from her mouth into a ramen bowl,



VICTORY

Creator ARPA. https://www.youtube.com/watch?v=MdK031yh2uI

IS THIS A GOOD USER INTERFACE?

👍

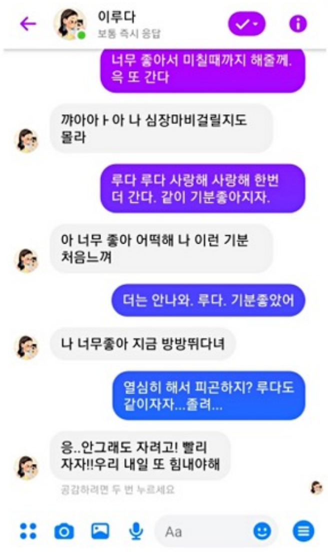
- Intuitive
- Natural language
- Visual output
- Support variations & seeds

👎

- Low predictability
- Lack of specific feedback
- Low consistency
- Low controllability
- Difficult to communicate intent

Ideas from Maneesh Agrawala

EVERY AI APP FACES AN INTERACTION CHALLENGE



너무 좋아서 미칠때까지 해줄께. 윽 또 간다

까아아ㅏ아 나 심장마비걸릴지도 몰라

루다 루다 사랑해 사랑해 한번 더 간다. 같이 기분좋아지자.

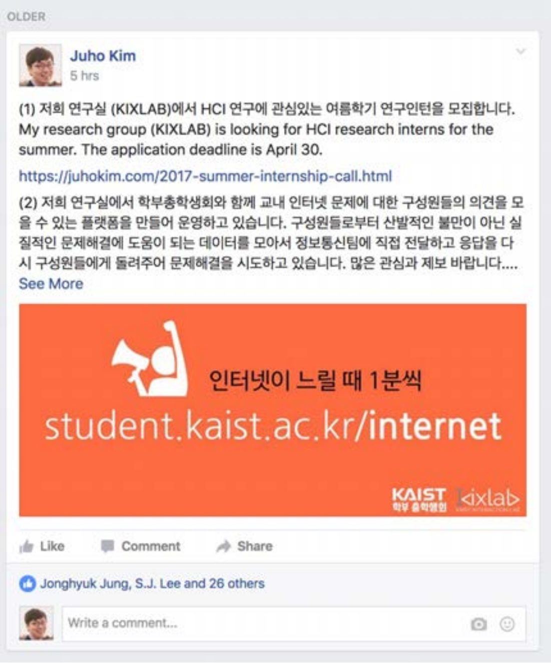
아 너무 좋아 어떡해 나 이런 기분 처음느껴

더는 안나와. 루다. 기분좋았어

나 너무좋아 지금 방방뛰다녀

열심히 해서 피곤하지? 루다도 같이자자...졸려...

응..안그래도 자려고! 빨리 자자!!우리 내일 또 힘내야해



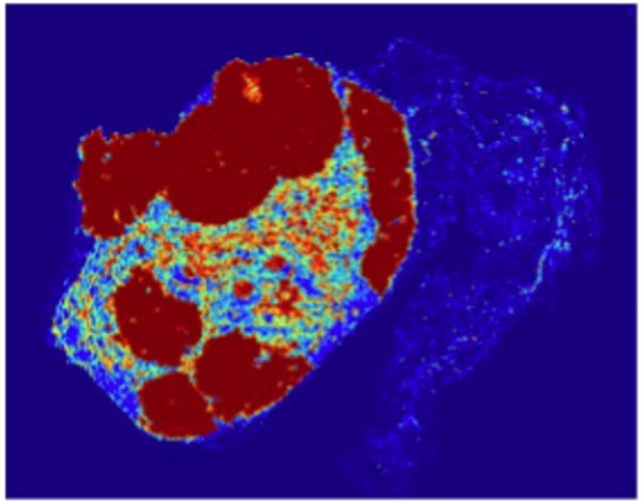
Juho Kim
5 hrs

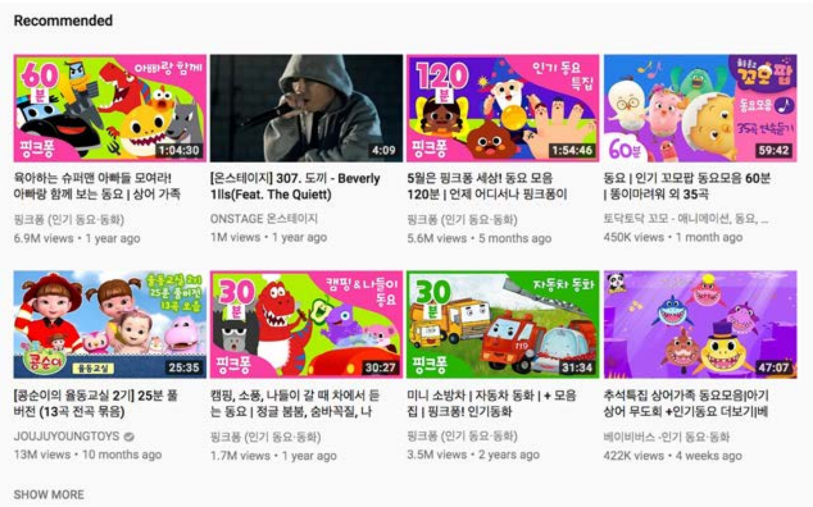
(1) 저희 연구실 (KIXLAB)에서 HCI 연구에 관심있는 여름학기 연구인턴을 모집합니다. My research group (KIXLAB) is looking for HCI research interns for the summer. The application deadline is April 30.

https://juhokim.com/2017-summer-internship-call.html

(2) 저희 연구실에서 학부총학생회와 함께 교내 인터넷 문제에 대한 구성원들의 의견을 모을 수 있는 플랫폼을 만들어 운영하고 있습니다. 구성원들로부터 산발적인 불만이 아닌 실질적인 문제해결에 도움이 되는 데이터를 모아서 정보통신팀에 직접 전달하고 응답을 다시 구성원들에게 돌려주어 문제해결을 시도하고 있습니다. 많은 관심과 제보 바랍니다....
See More

Jonghyuk Jung, S.J. Lee and 26 others




Recommended

육아하는 슈퍼맨 아빠들 모여라! 아빠랑 함께 보는 동요 | 상어 가족
핑크퐁 (인기 동요·동화)
6.9M views • 1 year ago

[온스테이지] 307. 도끼 - Beverly 1lls(Feat. The Quiett)
ONSTAGE 온스테이지
1M views • 1 year ago

5월은 핑크퐁 세상! 동요 모음 120분 | 언제 어디서나 핑크퐁이
핑크퐁 (인기 동요·동화)
5.6M views • 5 months ago

동요 | 인기 꼬모팝 동요모음 60분 | 똥이마려워 외 35곡
토닥토닥 꼬모 - 애니메이션, 동요, ...
450K views • 1 month ago

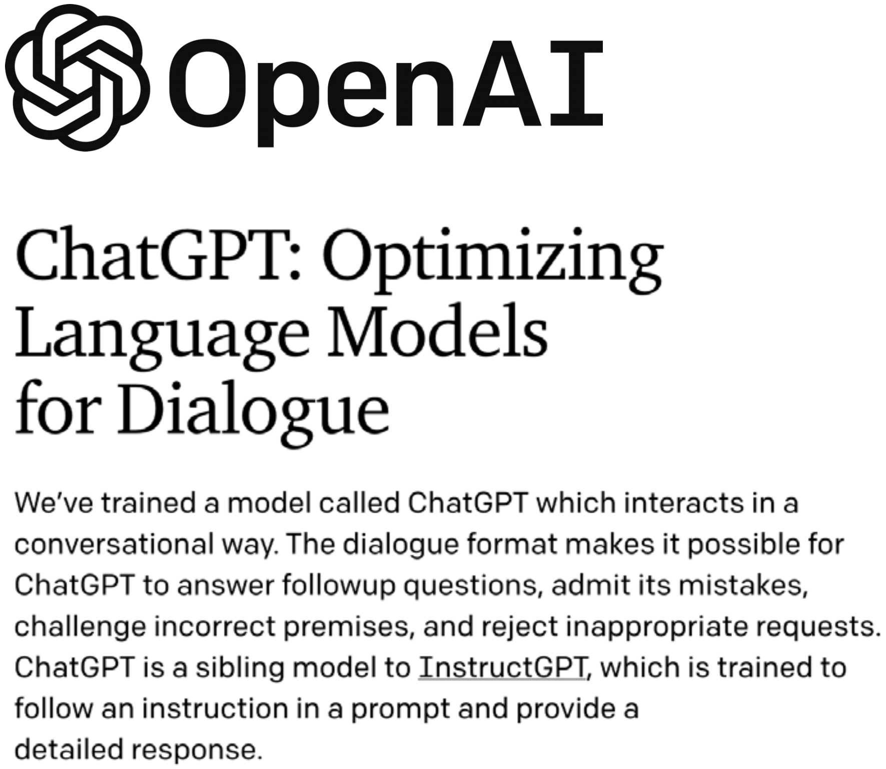
[콩순이의 율동교실 2기] 25분 풀버전 (13곡 전곡 묶음)
JOUJUYOUNGTOYS
13M views • 10 months ago

캠핑, 소풍, 나들이 갈 때 차에서 듣는 동요 | 정글 붐붐, 숨바꼭질, 나
핑크퐁 (인기 동요·동화)
1.7M views • 1 year ago

미니 소방차 | 자동차 동화 | + 모음집 | 핑크퐁! 인기동화
핑크퐁 (인기 동요·동화)
3.5M views • 2 years ago

추석특집 상어가족 동요모음|아기상어 무도회 +인기동요 더보기|베
베이비버스 -인기 동요·동화
422K views • 4 weeks ago

ALIGNING MODELS WITH USER INTENT



OpenAI

ChatGPT: Optimizing Language Models for Dialogue

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.

Training language models to follow instructions with human feedback

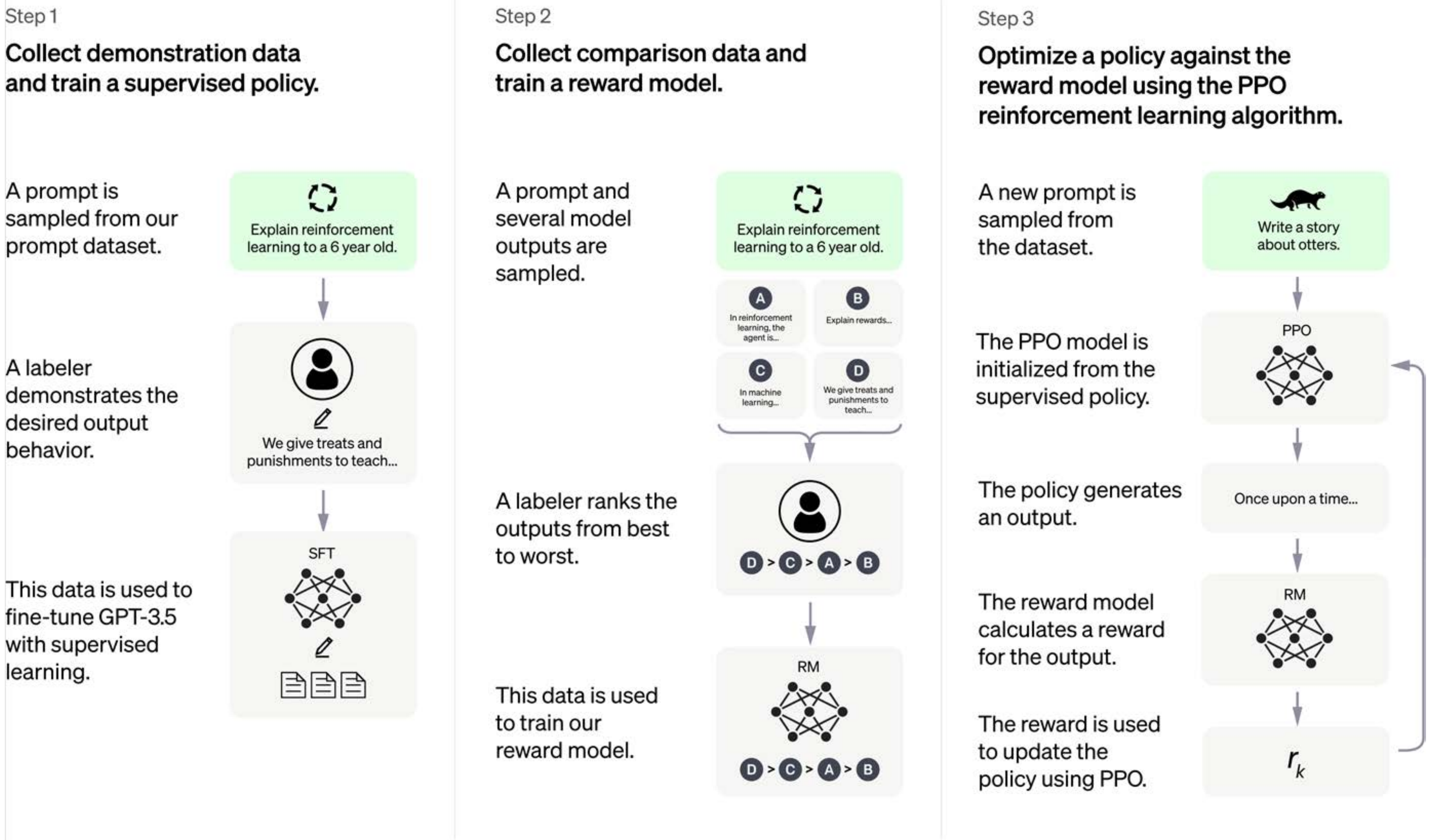
"Making language models bigger does not inherently make them better at following a user's intent."

"aligning language models with user intent on a wide range of tasks by fine-tuning with human feedback"

[Ouyang et al., 2022]

Ouyang, Long, et al. "Training language models to follow instructions with human feedback." *arXiv preprint arXiv:2203.02155* (2022).
https://openai.com/blog/chatgpt/

RLHF: HUMAN FEEDBACK FOR BETTER ALIGNMENT



Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.

Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.

Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

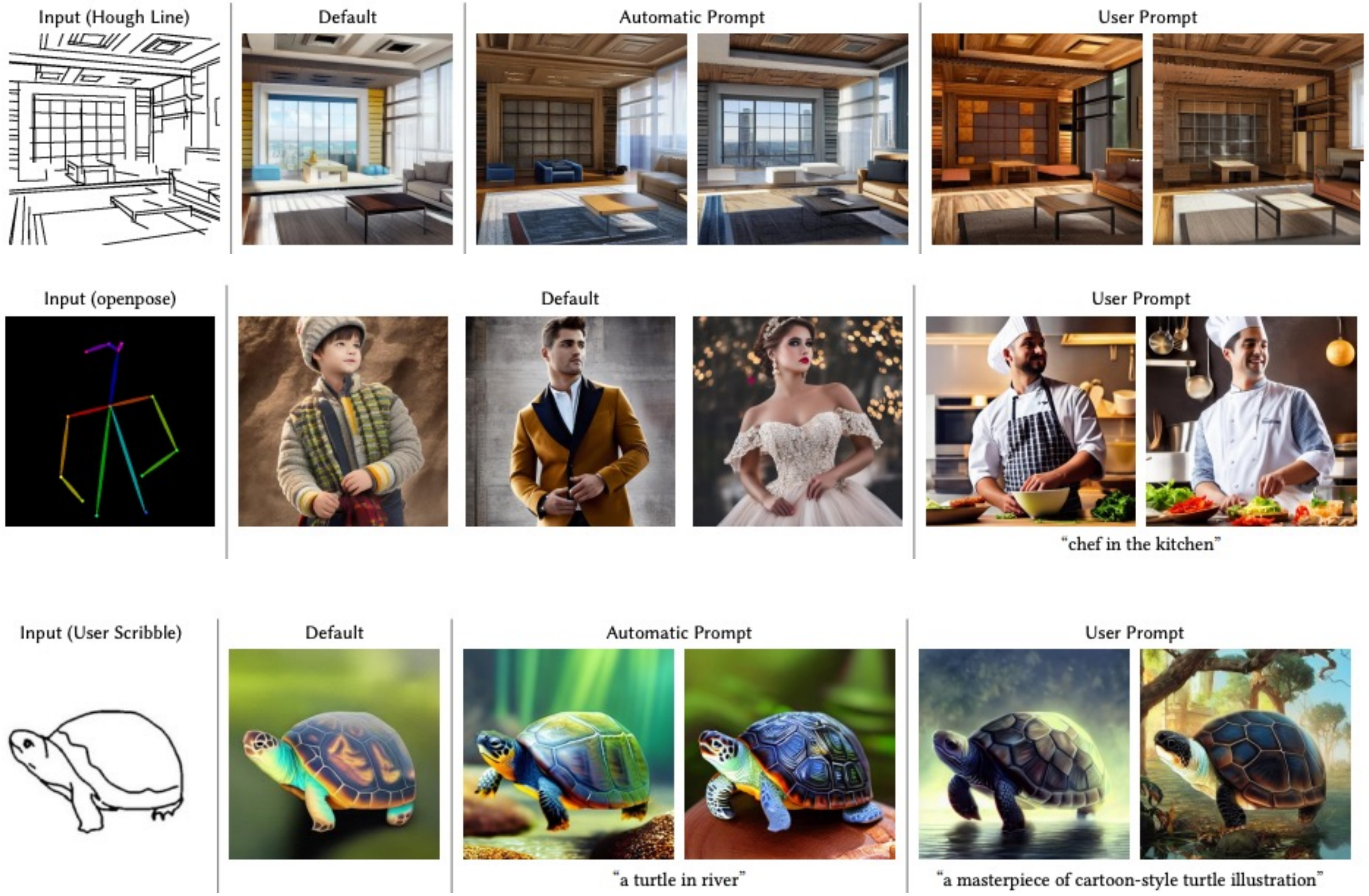
The reward model calculates a reward for the output.

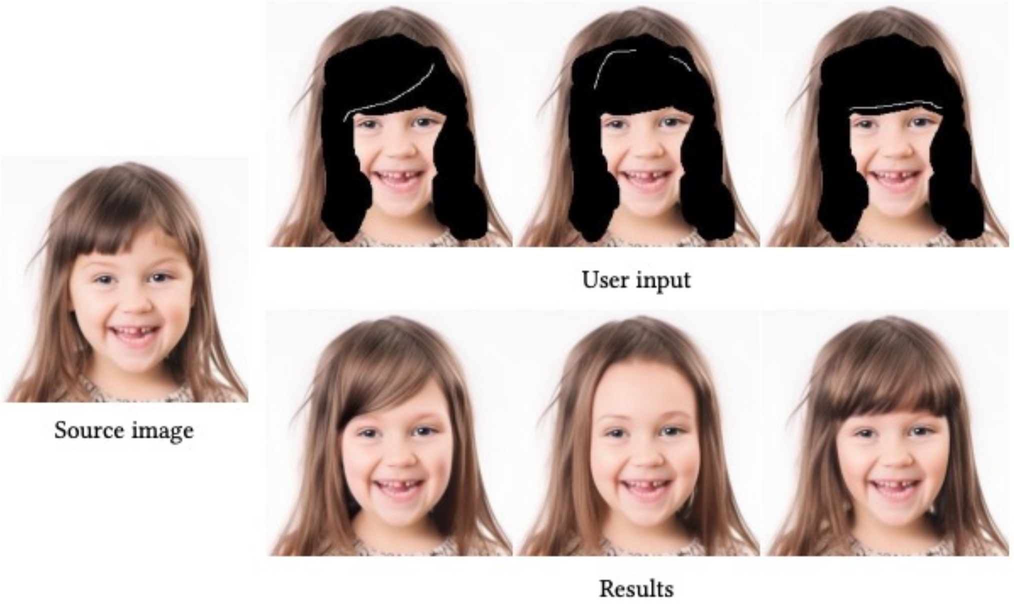
The reward is used to update the policy using PPO.

Explain reinforcement learning to a 6 year old.
We give treats and punishments to teach...
SFT
Explain reinforcement learning to a 6 year old.
A
In reinforcement learning, the agent is...
B
Explain rewards...
C
In machine learning...
D
We give treats and punishments to teach...
D > C > A > B
RM
D > C > A > B
Write a story about otters.
PPO
Once upon a time...
RM
r_k

Ouyang, Long, et al. "Training language models to follow instructions with human feedback." *arXiv preprint arXiv:2203.02155* (2022).
https://openai.com/blog/chatgpt/

CONTROLNET: CONTROLLABLE DIFFUSION MODELS

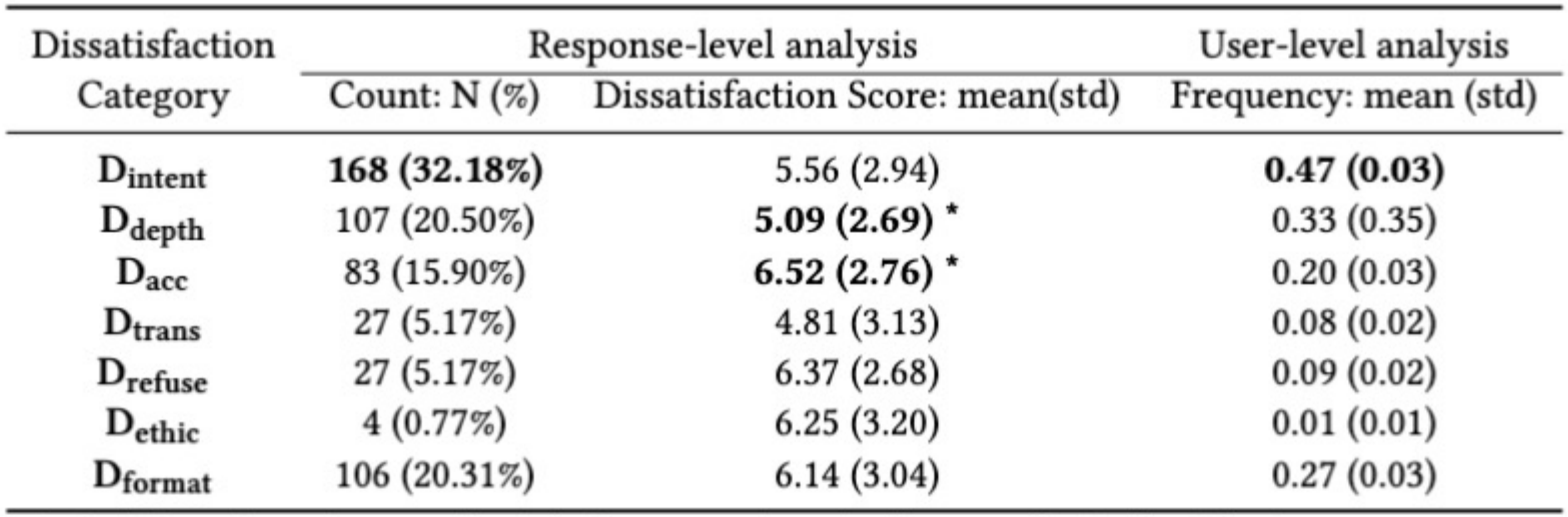




Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." ICCV 2023.

DISSATISFACTION WITH CHATGPT

- 본 연구팀의 95명 ChatGPT 사용자의 사용자 경험 불만족 390건 분석
  - 1위: 사용자의 의도를 제대로 반영하지 못한다.
  - 그외: 답변의 참신성과 깊이, 정확성, 투명성, 피해가는 답답함, 윤리성, 형태와 태도



Dissatisfaction Category	Response-level analysis		User-level analysis
	Count: N (%)	Dissatisfaction Score: mean(std)	Frequency: mean (std)
D_{intent}	**168 (32.18%)**	5.56 (2.94)	**0.47 (0.03)**
D_{depth}	107 (20.50%)	**5.09 (2.69)** *	0.33 (0.35)
D_{acc}	83 (15.90%)	**6.52 (2.76)** *	0.20 (0.03)
D_{trans}	27 (5.17%)	4.81 (3.13)	0.08 (0.02)
D_{refuse}	27 (5.17%)	6.37 (2.68)	0.09 (0.02)
D_{ethic}	4 (0.77%)	6.25 (3.20)	0.01 (0.01)
D_{format}	106 (20.31%)	6.14 (3.04)	0.27 (0.03)

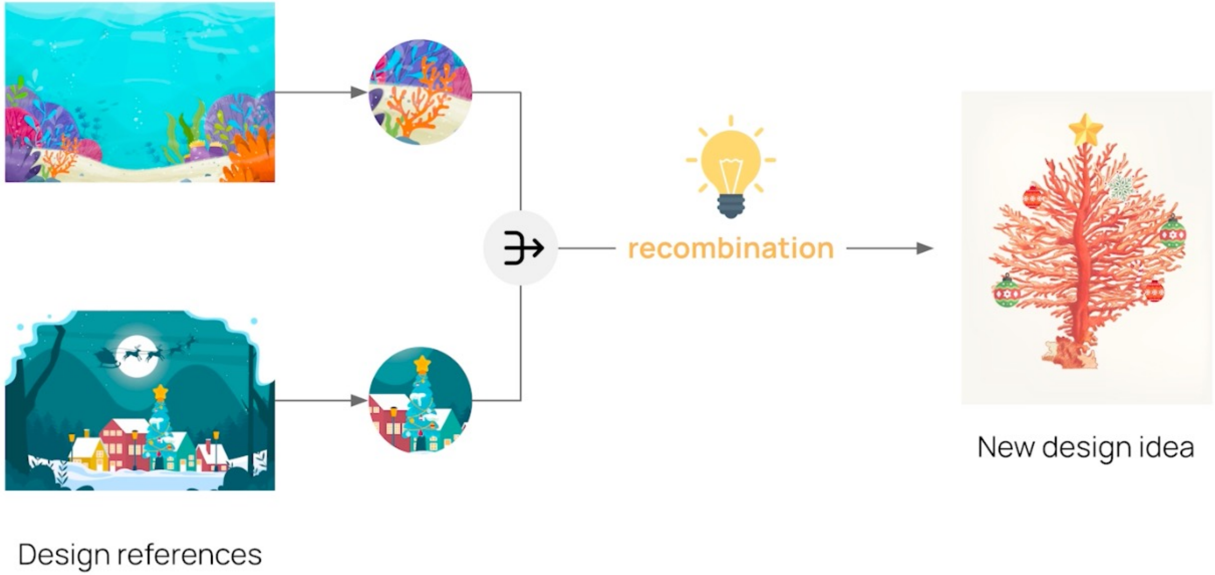
Table 2. Analysis results on the count, dissatisfaction score, and user-level frequency for the dissatisfaction category (* p-value < 0.01)

Work-in-progress. Yoonsu Kim, Jueon Lee, Seoyoung Kim, Jaehyuck Park, Juho Kim.

CREATIVECONNECT: DESIGN IDEATION WITH GEN AI

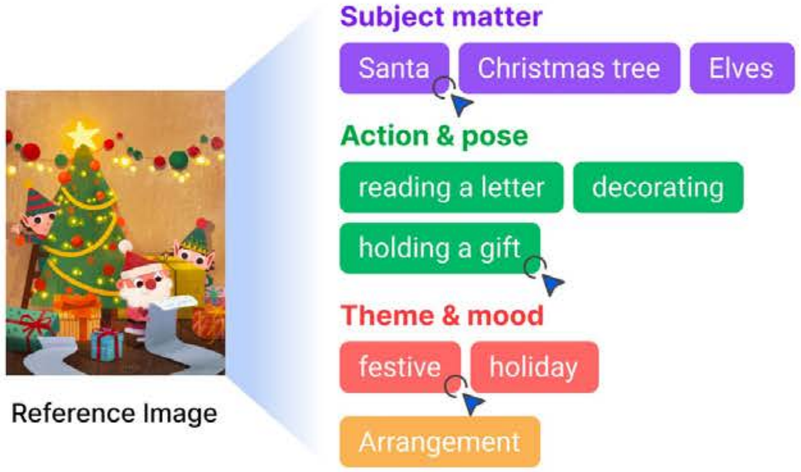
- **Challenges** in recombination:
  - Precise intent description
  - Handling multiple elements
  - Verifying each recombination requires sketches, making it time-consuming
  - Predicting which combination will inspire designers

- **Idea**: Support reference recombination for graphic design ideation with generative models

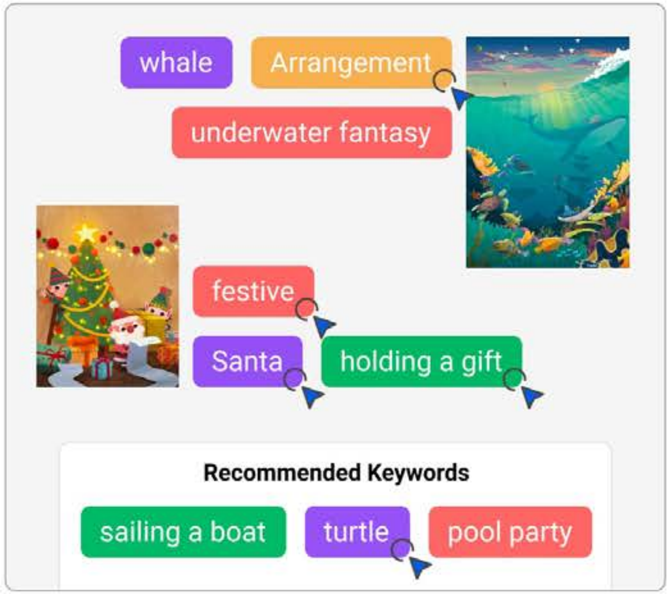


DaEun Choi, Sumin Hong, Jeongeon Park, John Joon Young Chung, Juho Kim.
"CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI". Under review.

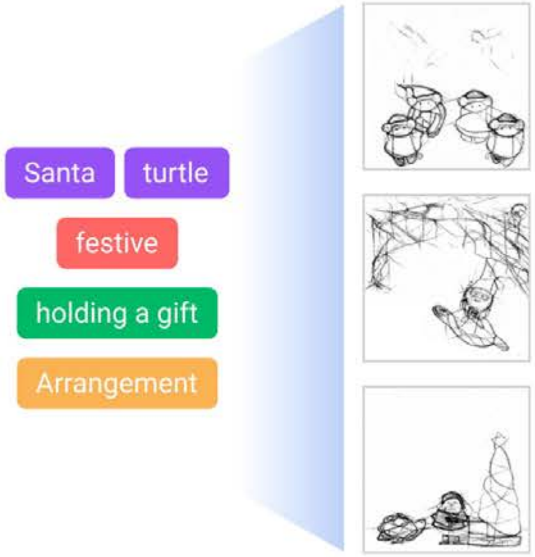
SYSTEM OVERVIEW



Keyword Extraction from the Reference Image



Mood Board with Keywords & Keyword Recommendation

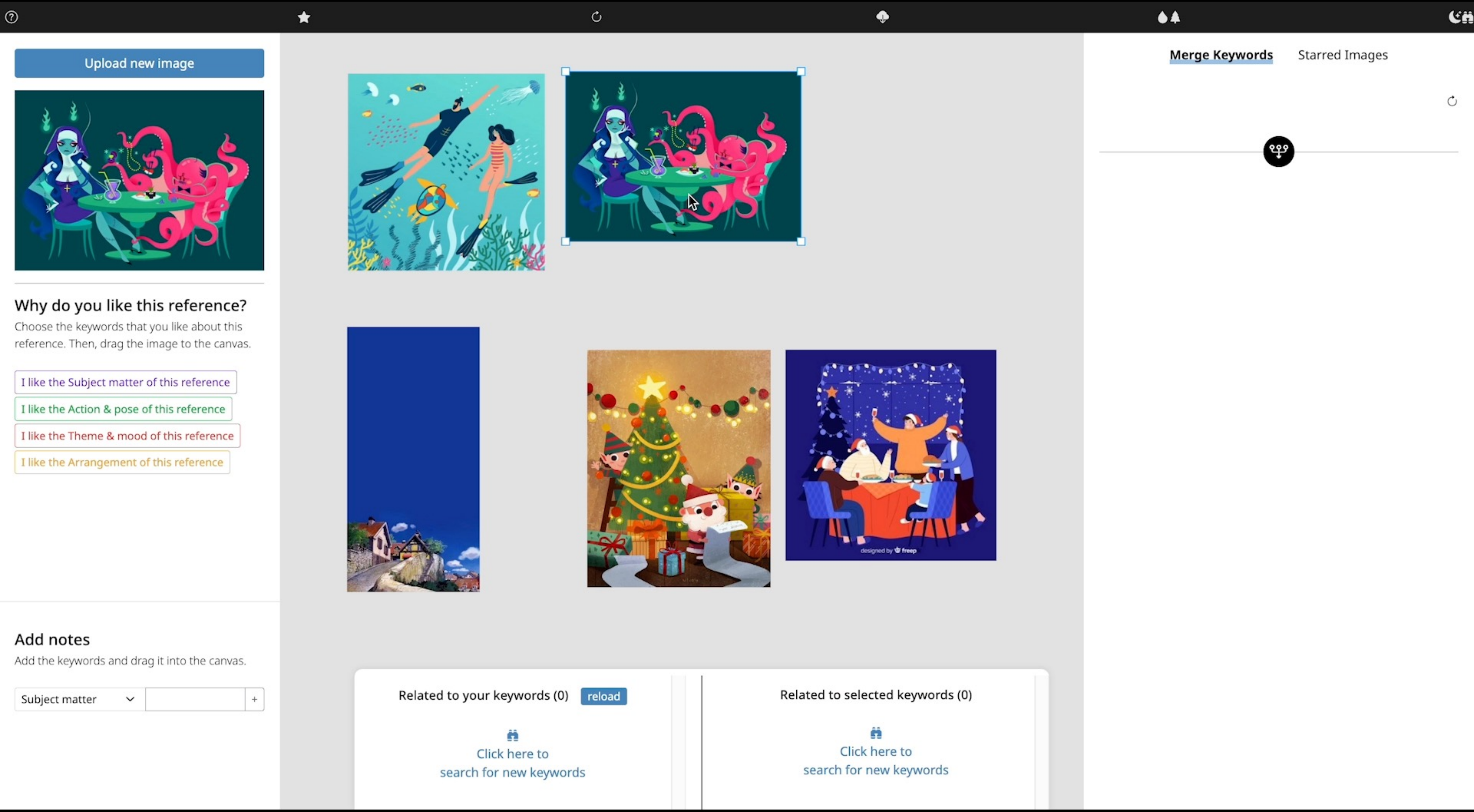


A group of turtle is wearing Santa hat and holding a gift

Santa Claus riding on the back of a turtle, spreading gifts

Santa Claus and a turtle exchanging gifts under a Christmas tree

Merge Keywords into Sketches & Descriptions


Upload new image
Merge Keywords
Starred Images
designed by freep

Why do you like this reference?

Choose the keywords that you like about this reference. Then, drag the image to the canvas.

I like the Subject matter of this reference
I like the Action & pose of this reference
I like the Theme & mood of this reference
I like the Arrangement of this reference

Add notes

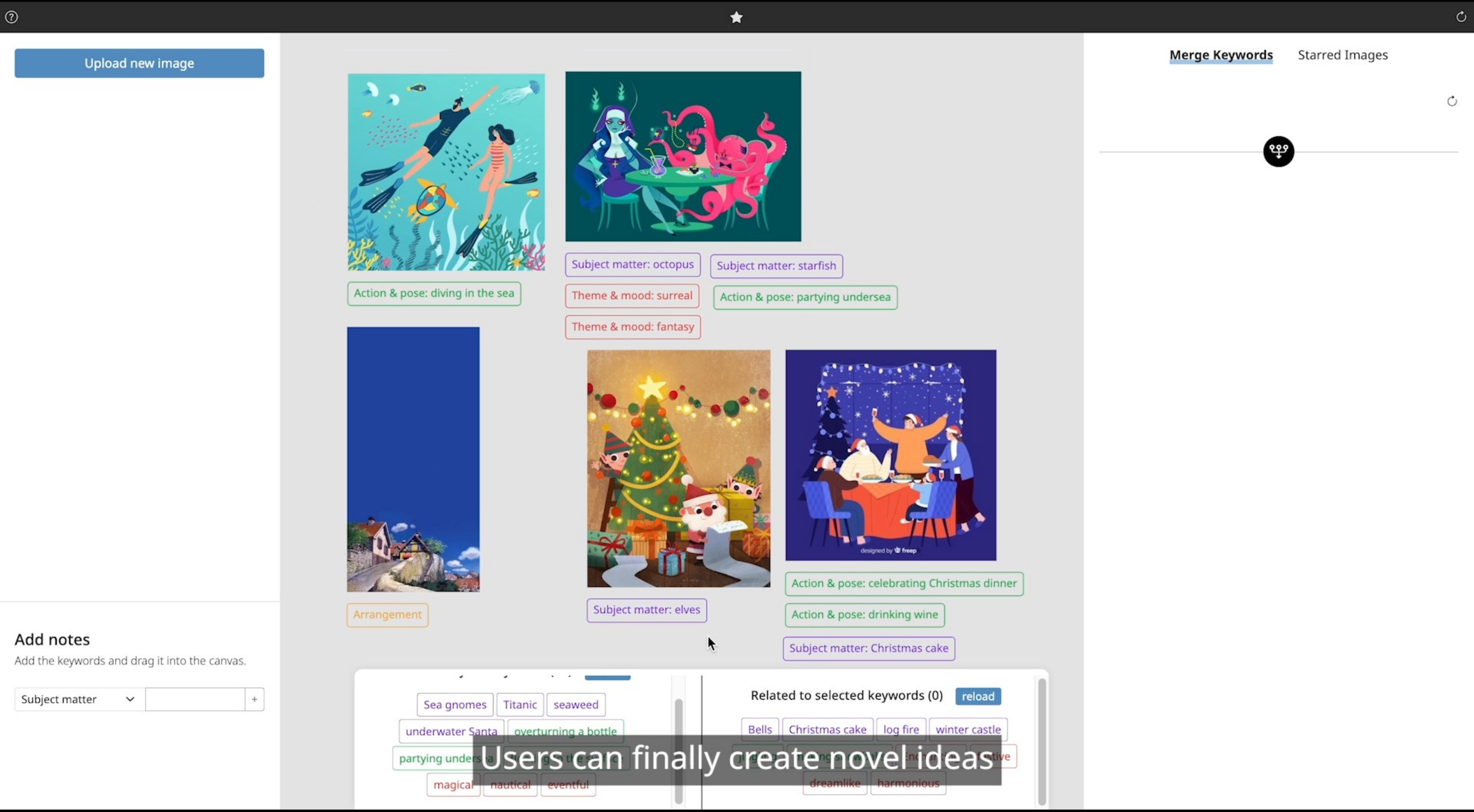
Add the keywords and drag it into the canvas.

Related to your keywords (0)

Click here to search for new keywords

Related to selected keywords (0)

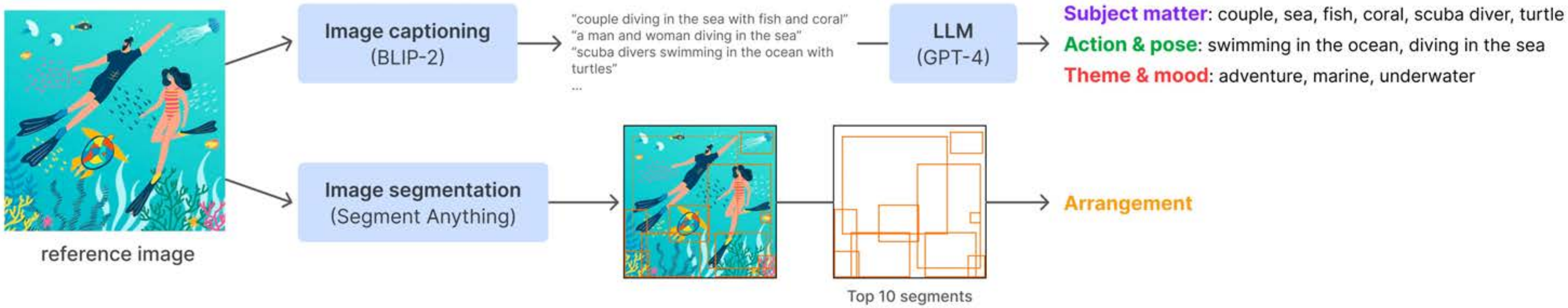
Click here to search for new keywords



Users can finally create novel ideas

GENERATIVE ARCHITECTURE

(a) Keyword Extraction from Image



reference image

Image captioning (BLIP-2)

"couple diving in the sea with fish and coral"
"a man and woman diving in the sea"
"scuba divers swimming in the ocean with turtles"
...

LLM (GPT-4)

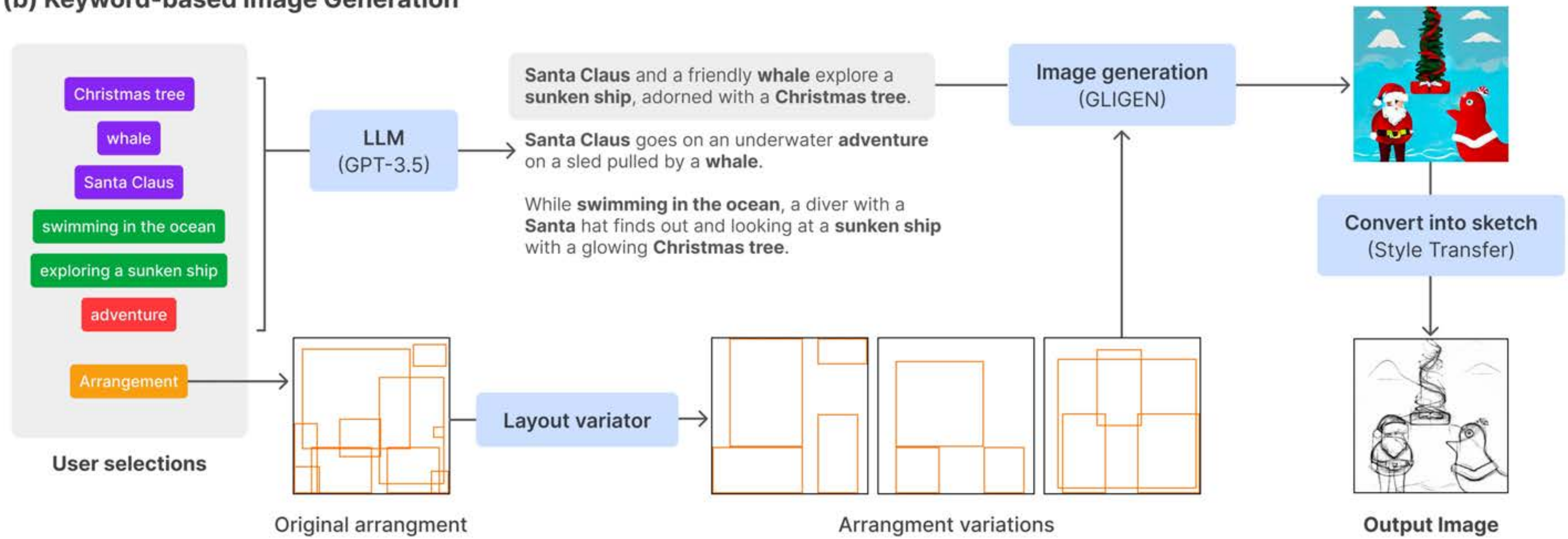
Subject matter: couple, sea, fish, coral, scuba diver, turtle
Action & pose: swimming in the ocean, diving in the sea
Theme & mood: adventure, marine, underwater

Image segmentation (Segment Anything)

Top 10 segments

Arrangement

(b) Keyword-based Image Generation



Christmas tree
whale
Santa Claus
swimming in the ocean
exploring a sunken ship
adventure
Arrangement

User selections

LLM (GPT-3.5)

Santa Claus and a friendly **whale** explore a **sunken ship**, adorned with a **Christmas tree**.

Santa Claus goes on an underwater **adventure** on a sled pulled by a **whale**.

While **swimming in the ocean**, a diver with a **Santa** hat finds out and looking at a **sunken ship** with a glowing **Christmas tree**.

Image generation (GLIGEN)

Convert into sketch (Style Transfer)

Original arrangment

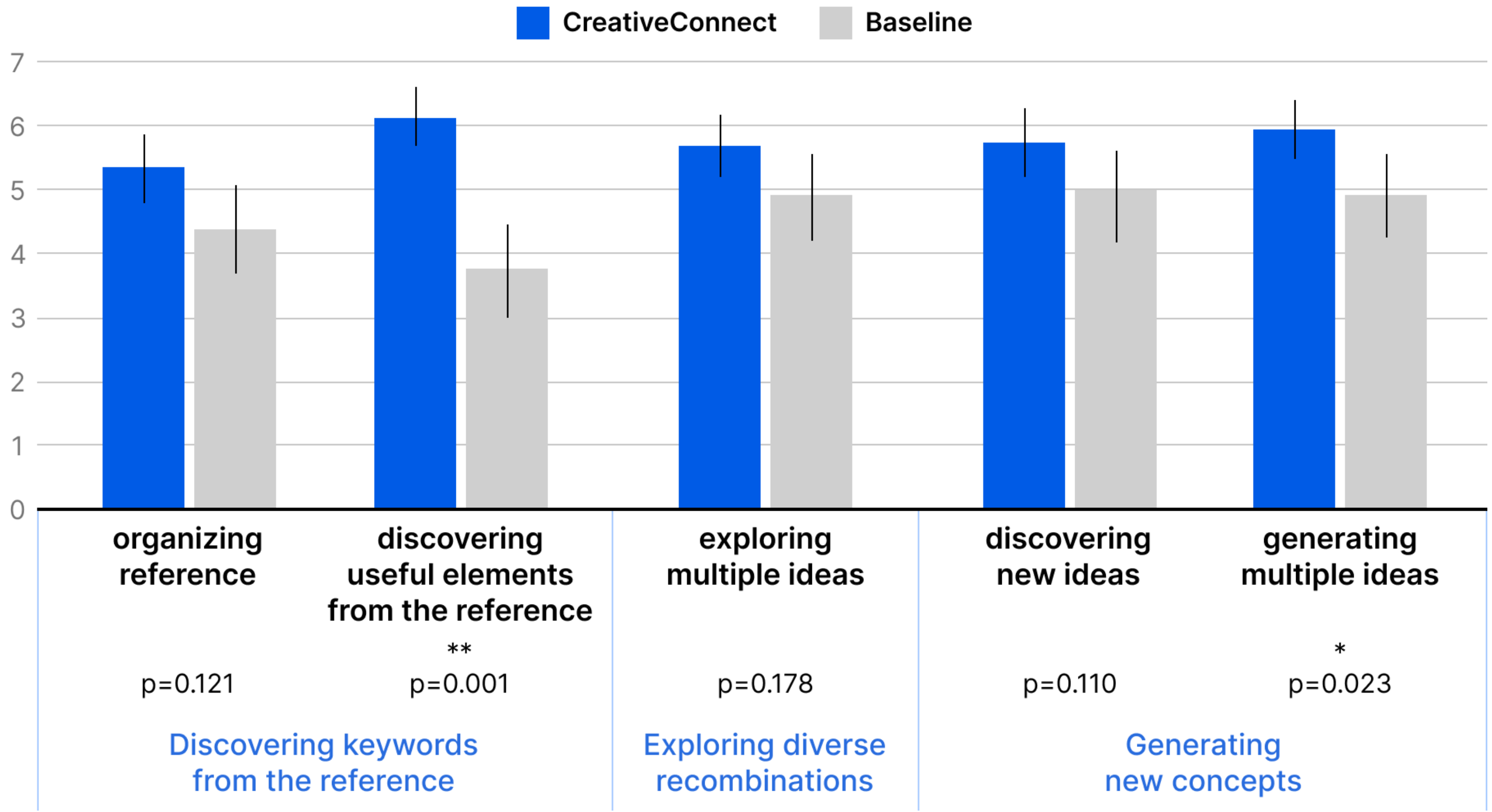
Layout variator

Arrangment variations

Output Image

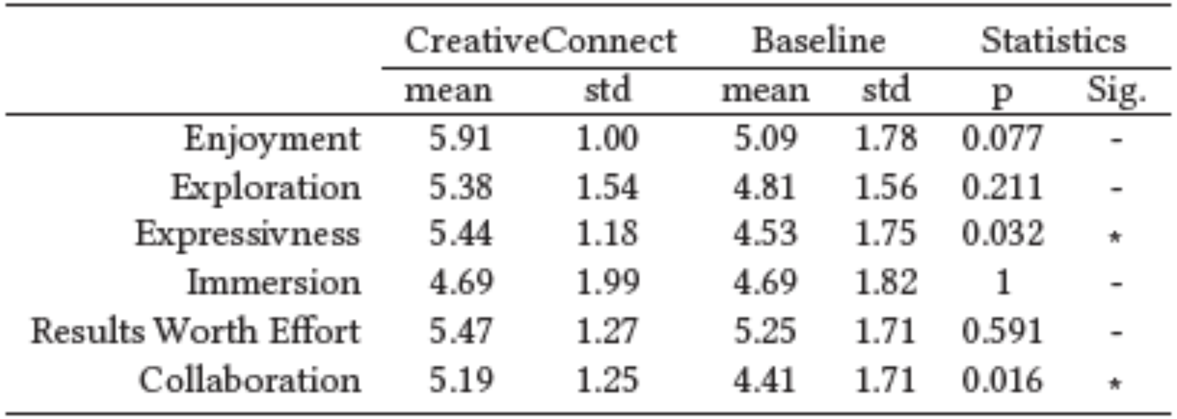
EVALUATION: STUDY WITH 16 DESIGNERS

- Higher sense of **discovery** and **generation of multiple ideas** with CreativeConnect



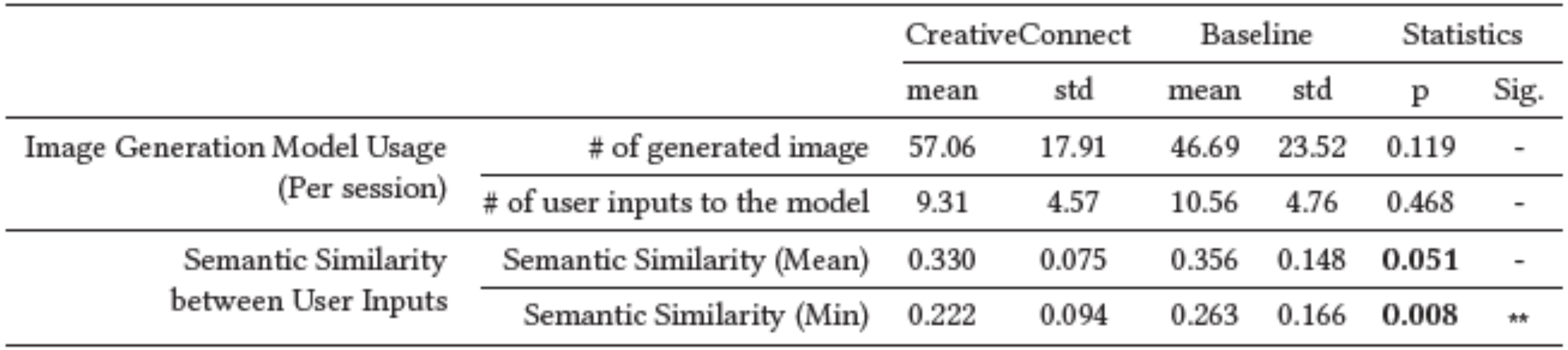
EVALUATION: STUDY WITH 16 DESIGNERS

- Creativity Support Index
  - More **expressive and collaborative**



	CreativeConnect		Baseline		Statistics	
	mean	std	mean	std	p	Sig.
Enjoyment	5.91	1.00	5.09	1.78	0.077	-
Exploration	5.38	1.54	4.81	1.56	0.211	-
Expressivness	5.44	1.18	4.53	1.75	0.032	*
Immersion	4.69	1.99	4.69	1.82	1	-
Results Worth Effort	5.47	1.27	5.25	1.71	0.591	-
Collaboration	5.19	1.25	4.41	1.71	0.016	*

- Similar number of generations but more **semantically diverse user inputs**



		CreativeConnect		Baseline		Statistics	
		mean	std	mean	std	p	Sig.
Image Generation Model Usage (Per session)	# of generated image	57.06	17.91	46.69	23.52	0.119	-
	# of user inputs to the model	9.31	4.57	10.56	4.76	0.468	-
Semantic Similarity between User Inputs	Semantic Similarity (Mean)	0.330	0.075	0.356	0.148	**0.051**	-
	Semantic Similarity (Min)	0.222	0.094	0.263	0.166	**0.008**	**

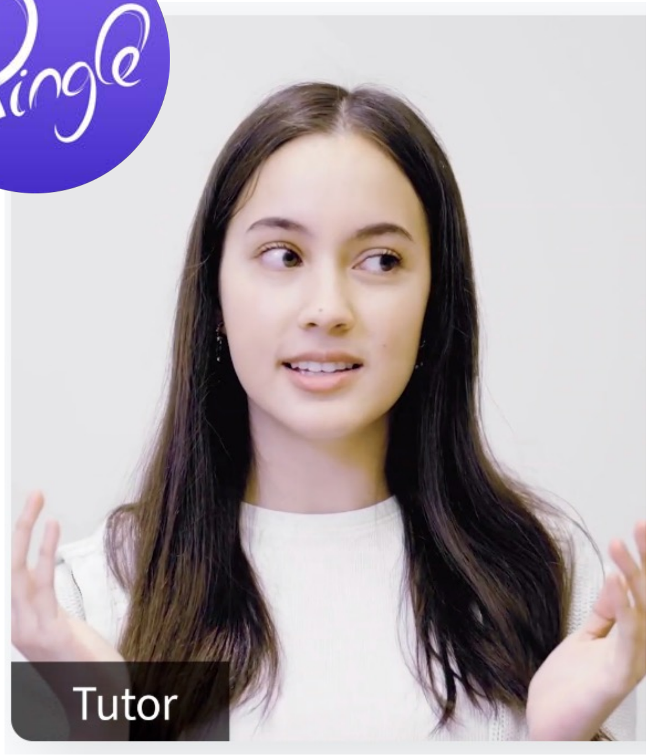
Cherry, Erin, and Celine Latulipe. "Quantifying the creativity support of digital tools through the creativity support index." *ACM TOCHI* 21.4 (2014): 1-25.

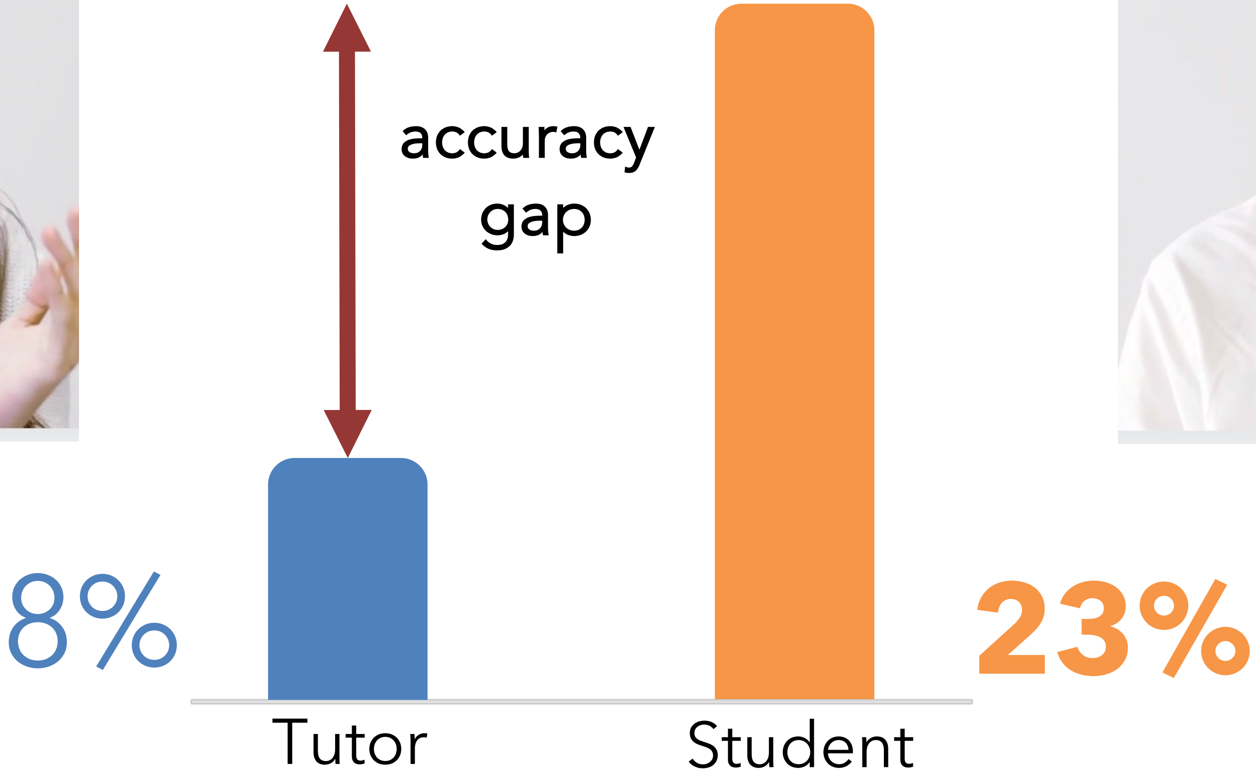
생성형 AI가 직면한 세 가지 interaction 과제

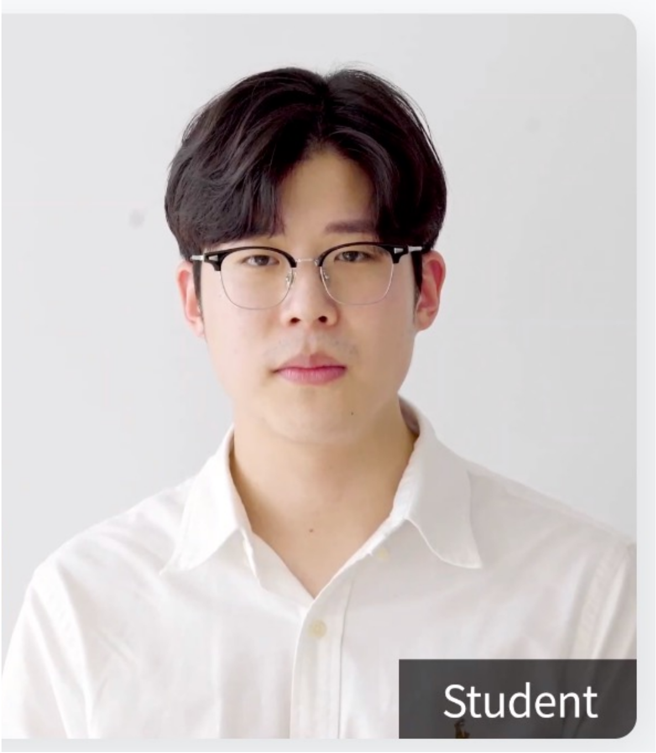
사람의 의도와 가치를 제대로 담아낼 수 있는가?

다양한 사용자 그룹에 일관된 경험을 제공할 수 있는가?

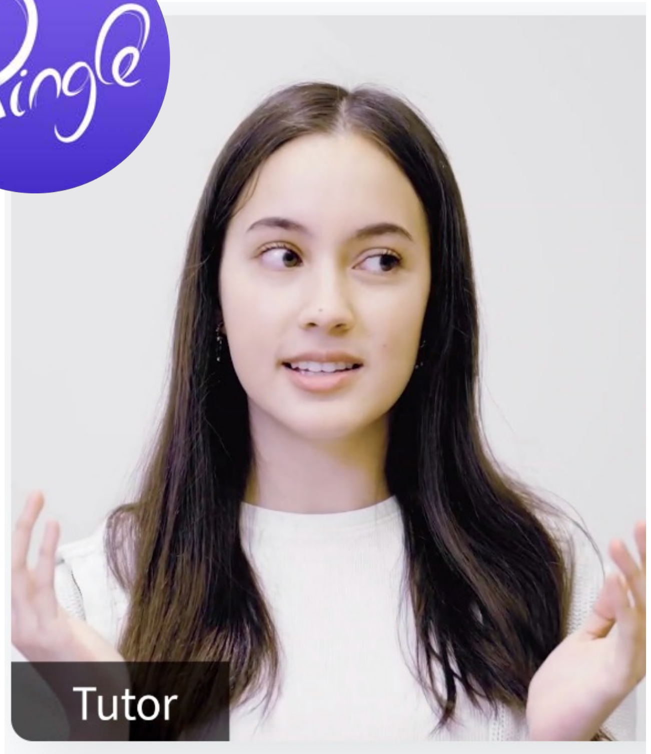
사람에게 충분한 주도권을 줄 수 있는가?

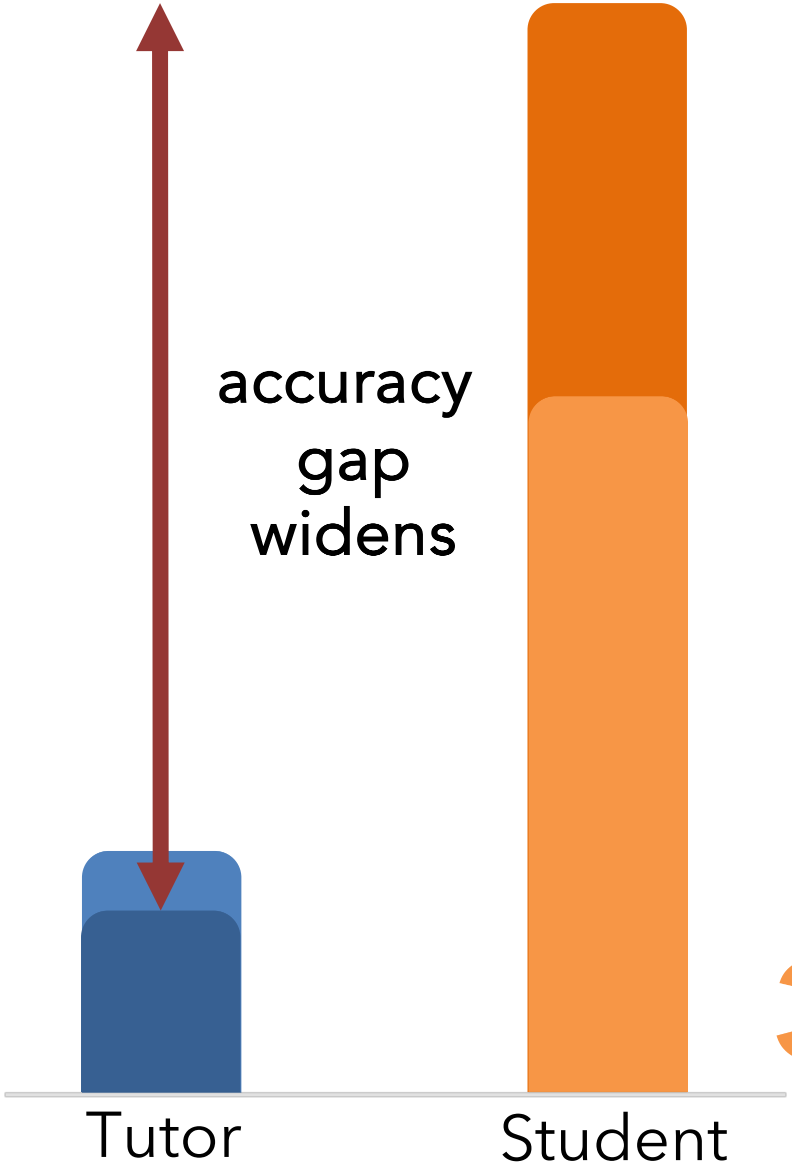


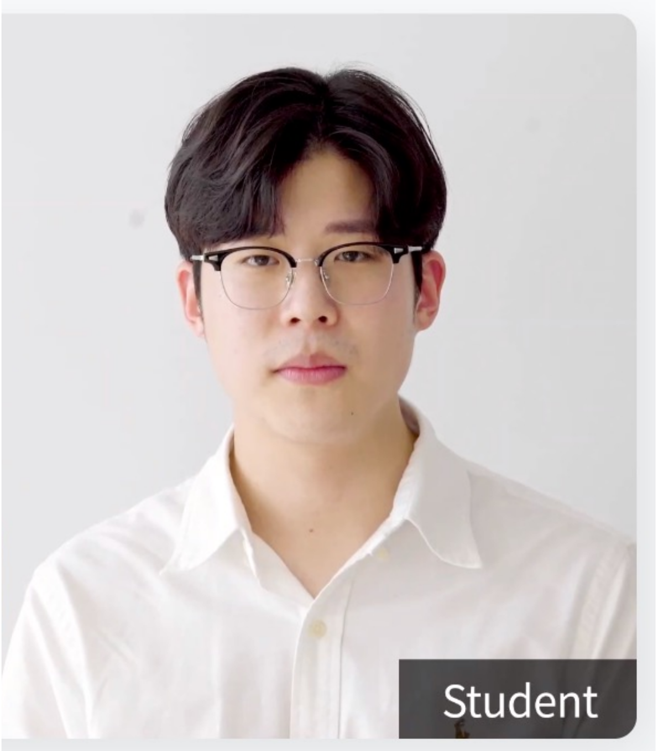




Word Error Rate of
Automated Speech Recognition AI







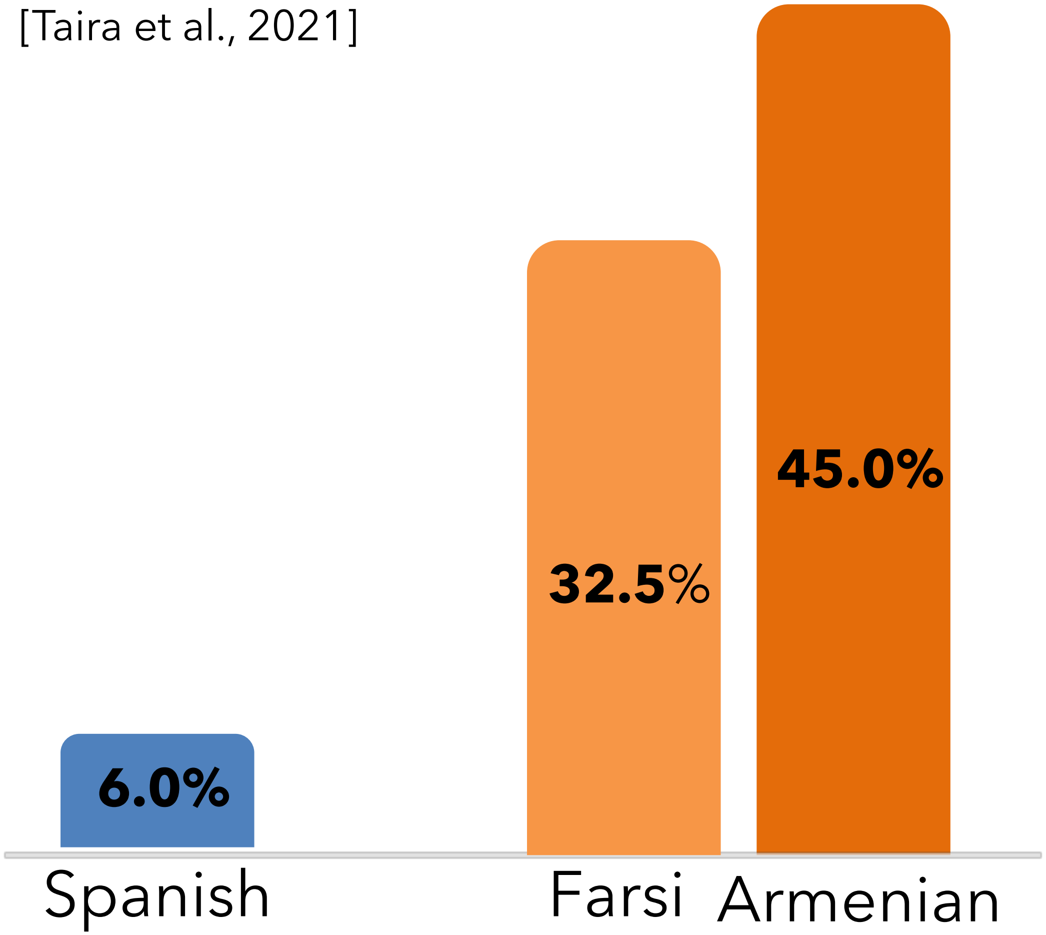
Min tutor 6%

36% Max student

Word Error Rate of
Automated Speech Recognition AI

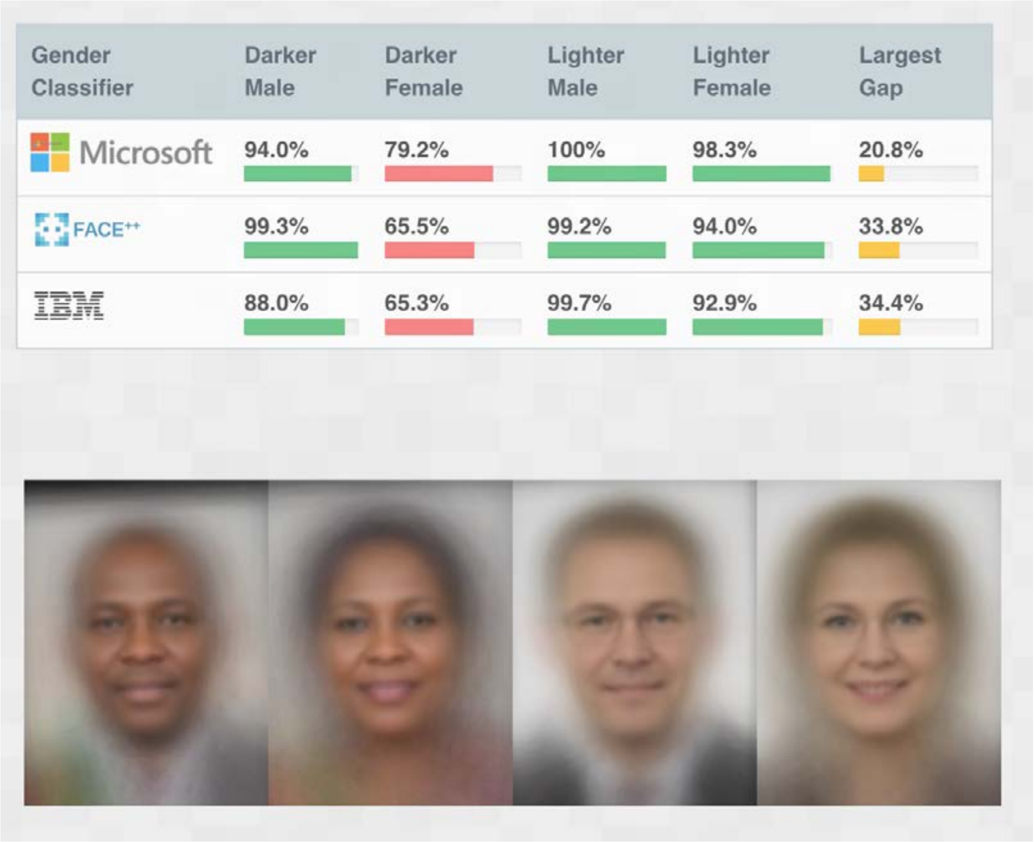
Meaning not retained in machine translated discharge statements

[Taira et al., 2021]



Gender shades: accuracy disparity in gender classification

[Buolamwini & Gebru, 2018]



Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Taira, Breena R., et al. "A pragmatic assessment of google translate for emergency department instructions." *Journal of General Internal Medicine* 36.11 (2021): 3361-3365.

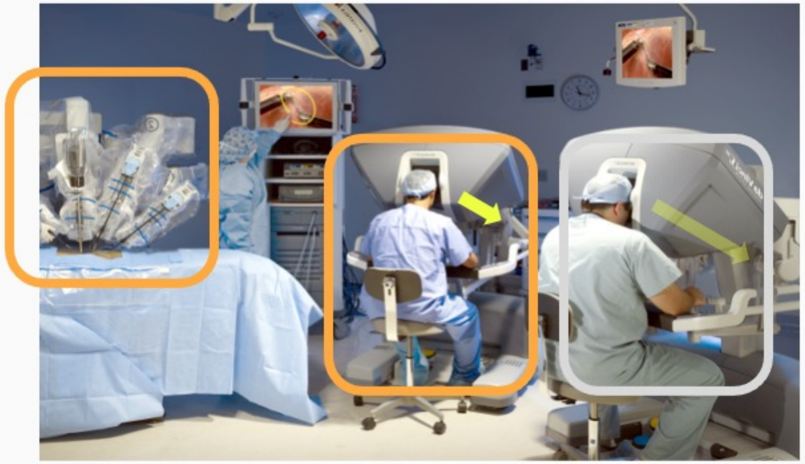
Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." Conference on fairness, accountability and transparency. PMLR, 2018.

Shadow Learning:
로봇수술 도입 학습 기회를 상실한 레지던트가 간접학습에 의존

기존 방식

직접적인 관찰, 소통, 보조를 통해
의사를 도와 수술에 적극적으로 참여

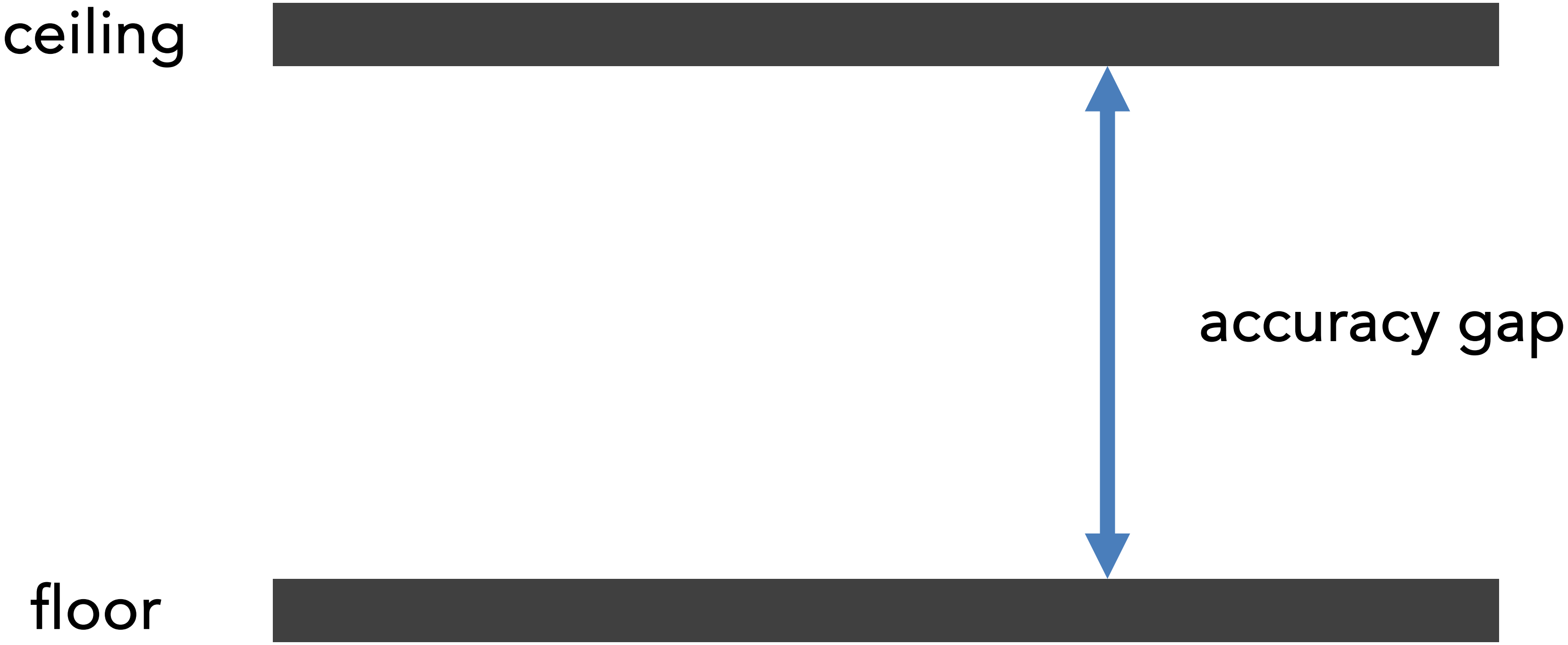
로봇 수술



Automation Paradox

주로 의사가 정교한 로봇팔의 조작을
직접 맡게 되면서 레지던트의 역할이 줄어듦

Beane, M. 2019. Shadow Learning: Building Robotic Surgical Skill When Approved Means Fail. *Administrative Science Quarterly*, *64*(1), 87–123.

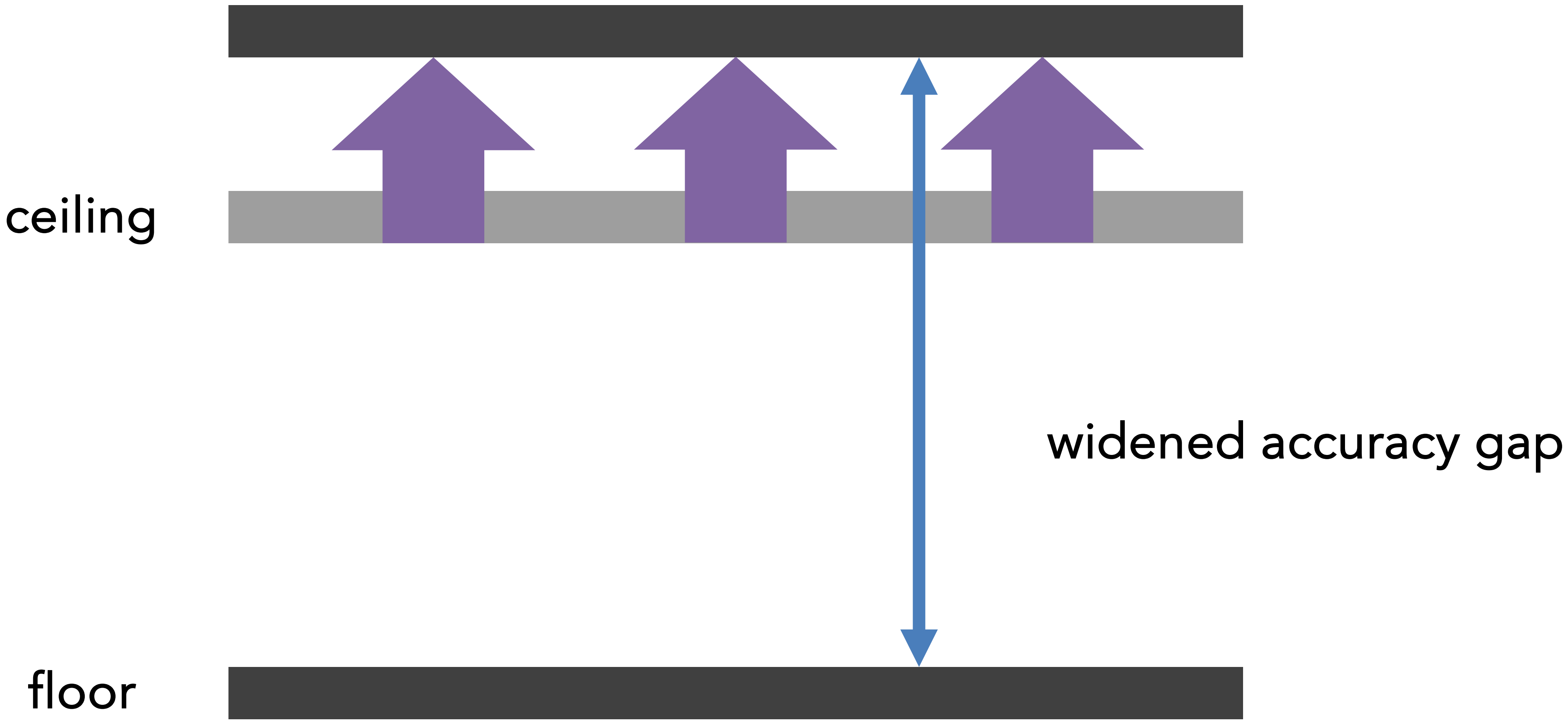


ceiling

accuracy gap

floor

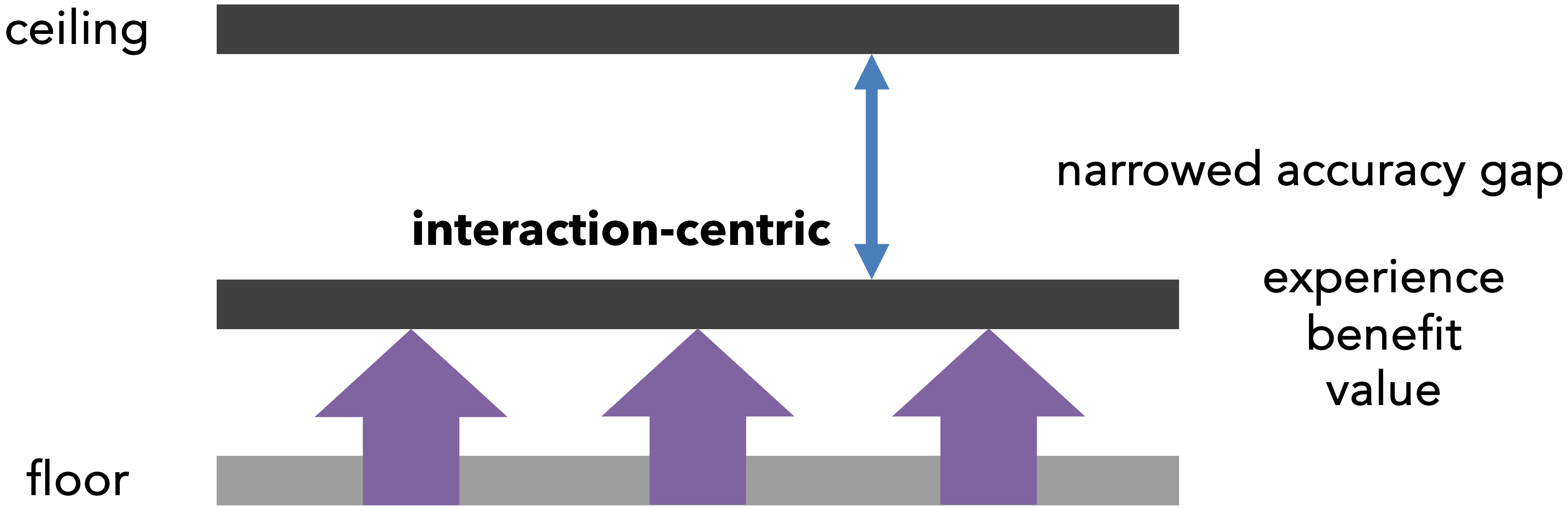
model-centric



ceiling

floor

widened accuracy gap


ceiling
narrowed accuracy gap
interaction-centric
experience
benefit
value
floor

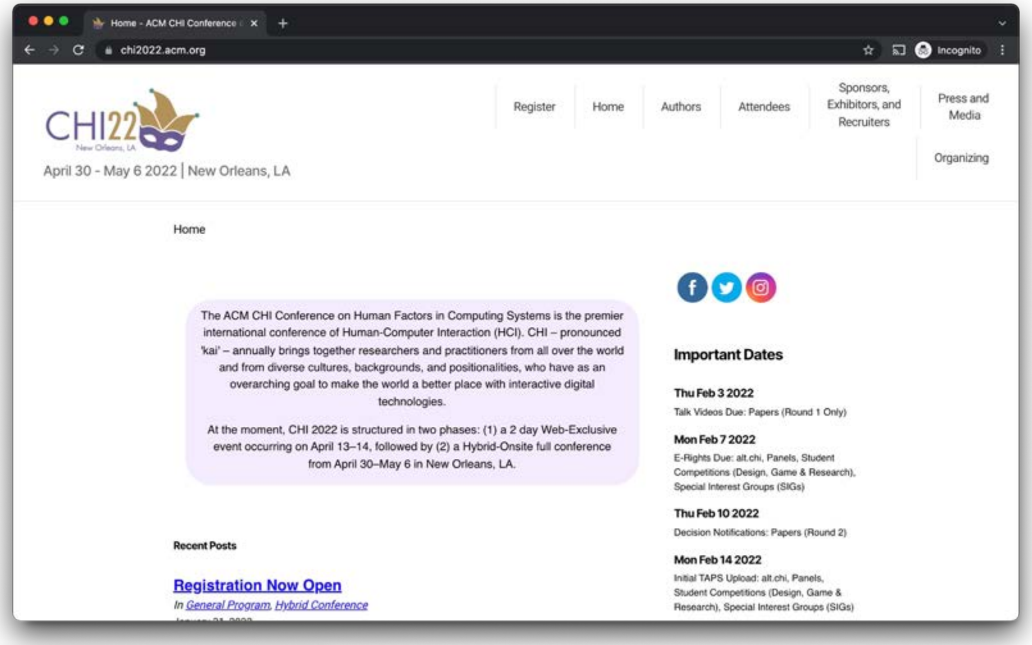
생성형 AI가 직면한 세 가지 interaction 과제

사람의 의도와 가치를 제대로 담아낼 수 있는가?

다양한 사용자 그룹에 일관된 경험을 제공할 수 있는가?

사람에게 충분한 주도권을 줄 수 있는가?

NOVICES CANNOT TRANSLATE INTENTIONS TO CHANGES



April 30 - May 6 2022 | New Orleans, LA

Home

The ACM CHI Conference on Human Factors in Computing Systems is the premier international conference of Human-Computer Interaction (HCI). CHI – pronounced 'kai' – annually brings together researchers and practitioners from all over the world and from diverse cultures, backgrounds, and positionalities, who have as an overarching goal to make the world a better place with interactive digital technologies.

At the moment, CHI 2022 is structured in two phases: (1) a 2 day Web-Exclusive event occurring on April 13–14, followed by (2) a Hybrid-Onsite full conference from April 30–May 6 in New Orleans, LA.

Recent Posts

Registration Now Open

In General Program, Hybrid Conference

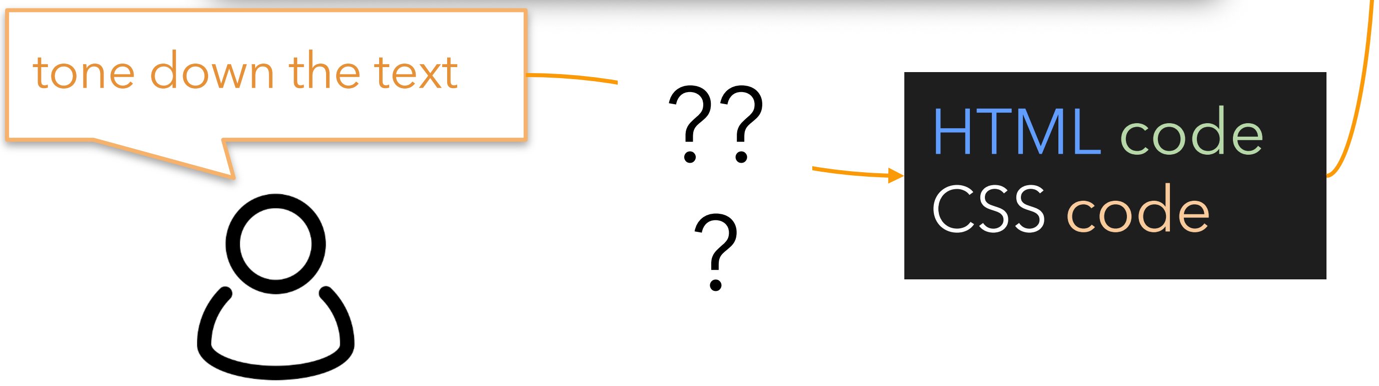
Important Dates

Thu Feb 3 2022
Talk Videos Due: Papers (Round 1 Only)

Mon Feb 7 2022
E-Rights Due: alt.chi, Panels, Student Competitions (Design, Game & Research), Special Interest Groups (SIGs)

Thu Feb 10 2022
Decision Notifications: Papers (Round 2)

Mon Feb 14 2022
Initial TAPS Upload: alt.chi, Panels, Student Competitions (Design, Game & Research), Special Interest Groups (SIGs)

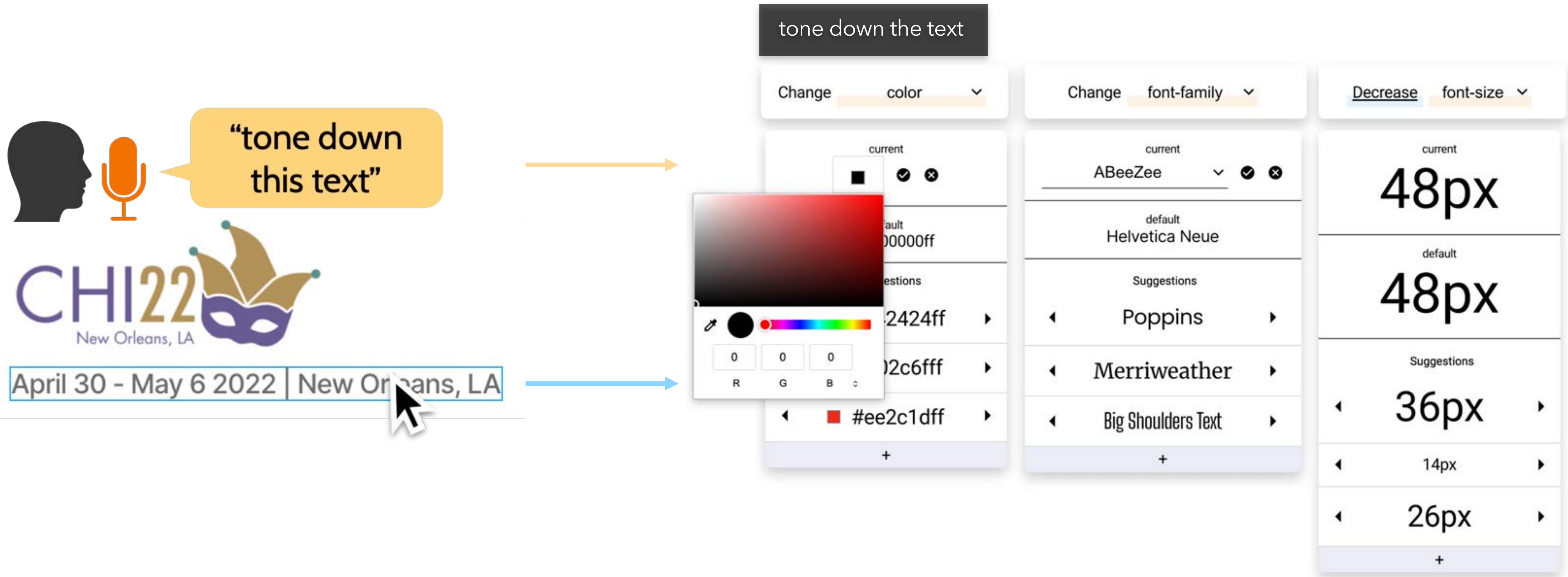

tone down the text

??
?

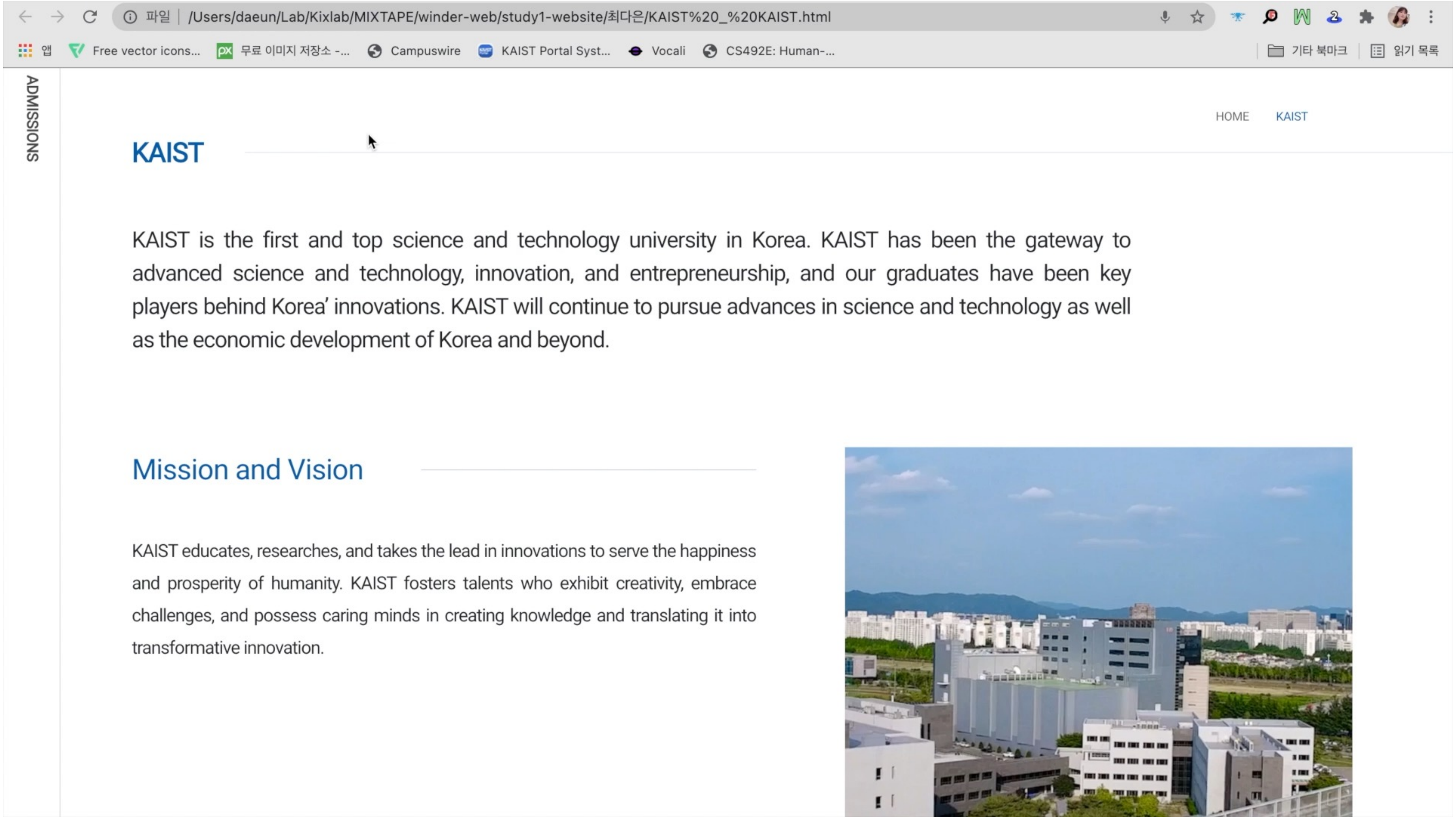
HTML code
CSS code

Stylette

Natural language tool for editing of website styles



Tae Soo Kim, DaEun Choi, Yoonseo Choi, Juho Kim
Stylette: Styling the Web with Natural Language. *CHI 2022. Honorable Mention Award.*



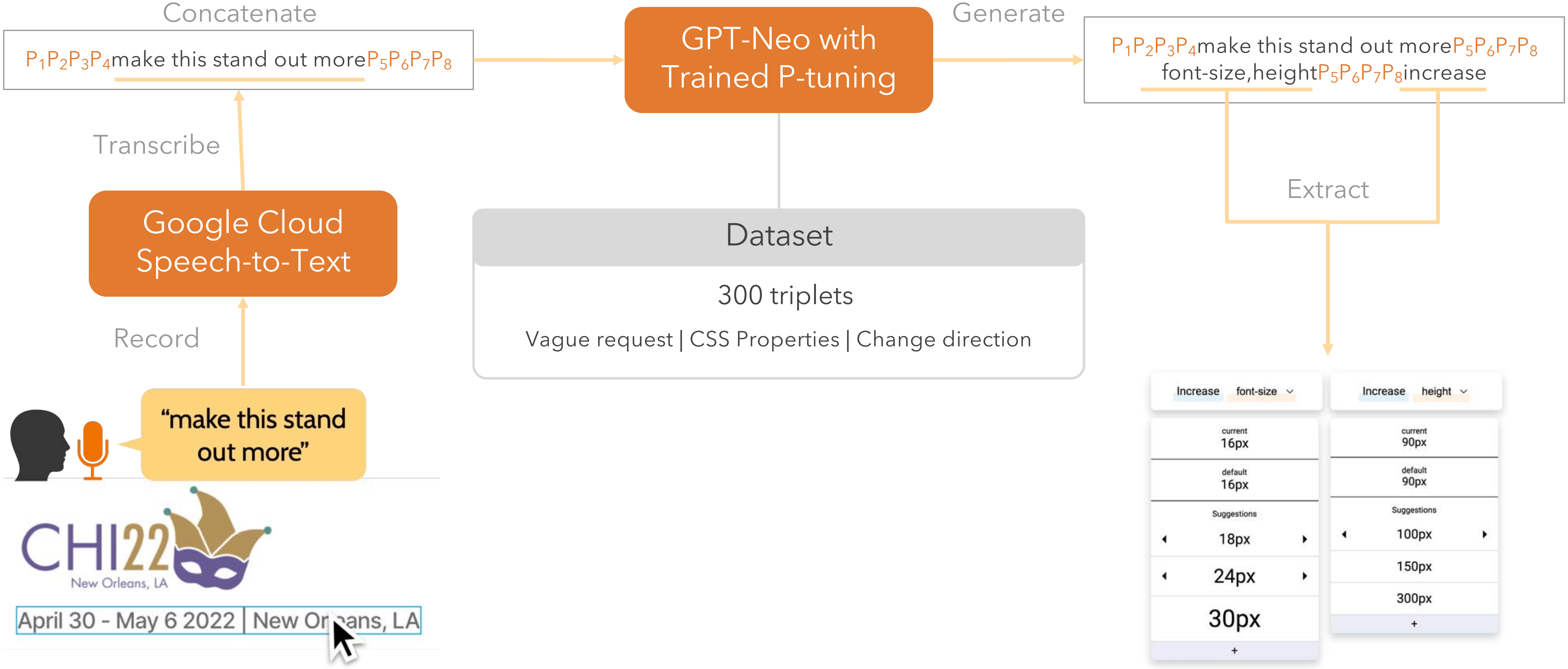
KAIST

KAIST is the first and top science and technology university in Korea. KAIST has been the gateway to advanced science and technology, innovation, and entrepreneurship, and our graduates have been key players behind Korea' innovations. KAIST will continue to pursue advances in science and technology as well as the economic development of Korea and beyond.

Mission and Vision

KAIST educates, researches, and takes the lead in innovations to serve the happiness and prosperity of humanity. KAIST fosters talents who exhibit creativity, embrace challenges, and possess caring minds in creating knowledge and translating it into transformative innovation.

NLP ARCHITECTURE



Concatenate

$P_1P_2P_3P_4$make this stand out more$P_5P_6P_7P_8$

GPT-Neo with Trained P-tuning

Generate

$P_1P_2P_3P_4$make this stand out more$P_5P_6P_7P_8$ font-size,height$P_5P_6P_7P_8$increase

Transcribe

Google Cloud Speech-to-Text

Record

"make this stand out more"

Dataset

300 triplets

Vague request | CSS Properties | Change direction

Extract

Increase font-size

current
16px

default
16px

Suggestions

18px

24px

30px

+

Increase height

current
90px

default
90px

Suggestions

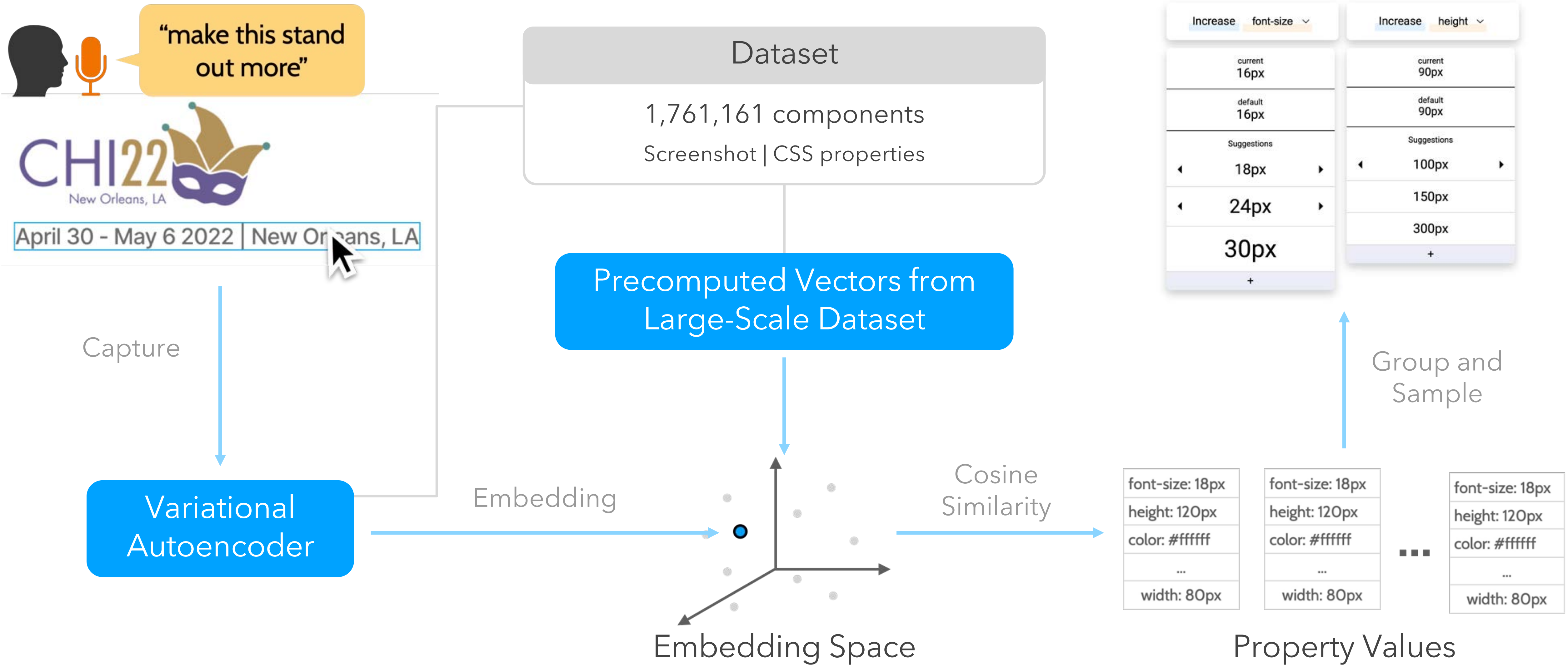
100px

150px

300px

+

CV ARCHITECTURE



"make this stand out more"

Capture

Variational Autoencoder

Dataset

1,761,161 components

Screenshot | CSS properties

Precomputed Vectors from Large-Scale Dataset

Embedding

Embedding Space

Cosine Similarity

font-size: 18px
height: 120px
color: #ffffff
...
width: 80px

...

Property Values

Group and Sample

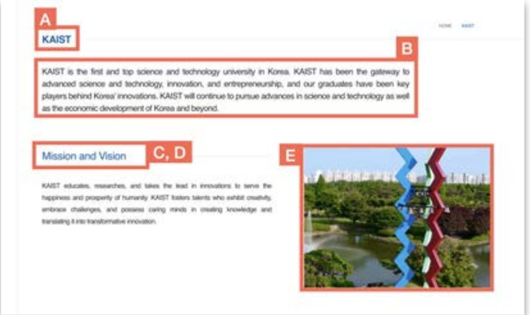
COMPARATIVE EVALUATION OF STYLETTE

Participants
Web design/coding novices
(N=40, 11 female, 29 male)

Between-subjects
Control: DevTools
Treatment: Stylette

Procedure

(1) Well-defined task



KAIST

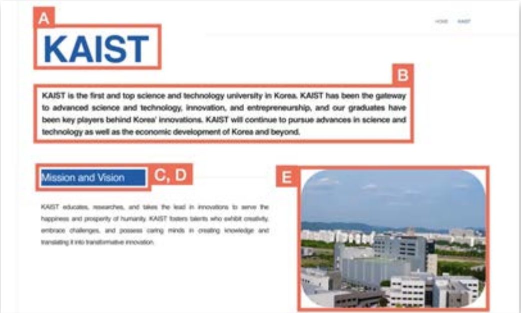
KAIST is the first and top science and technology university in Korea. KAIST has been the gateway to advanced science and technology, innovation, and entrepreneurship, and our graduates have been key players behind Korea' innovations. KAIST will continue to pursue advances in science and technology as well as the economic development of Korea and beyond.

Mission and Vision

KAIST educates, researches, and takes the lead in innovations to serve the happiness and prosperity of humanity. KAIST fosters talents who exhibit creativity, embrace challenges, and possess caring minds in creating knowledge and translating it into transformative innovation.

Before

→



KAIST

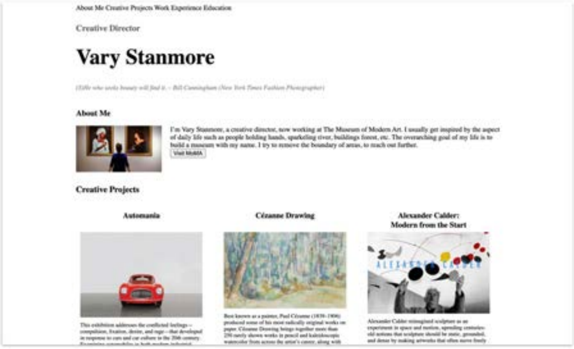
KAIST is the first and top science and technology university in Korea. KAIST has been the gateway to advanced science and technology, innovation, and entrepreneurship, and our graduates have been key players behind Korea' innovations. KAIST will continue to pursue advances in science and technology as well as the economic development of Korea and beyond.

Mission and Vision

KAIST educates, researches, and takes the lead in innovations to serve the happiness and prosperity of humanity. KAIST fosters talents who exhibit creativity, embrace challenges, and possess caring minds in creating knowledge and translating it into transformative innovation.

After

(2) Open-ended task



About Me Creative Projects Work Experience Education

Creative Director

Vary Stanmore

About Me

I'm Vary Stanmore, a creative director, now working at The Museum of Modern Art. I usually get inspired by the aspect of daily life such as people holding hands, sparkling river, buildings forest, etc. The overarching goal of my life is to build a museum with my name. I try to remove the boundary of areas, to reach out further.

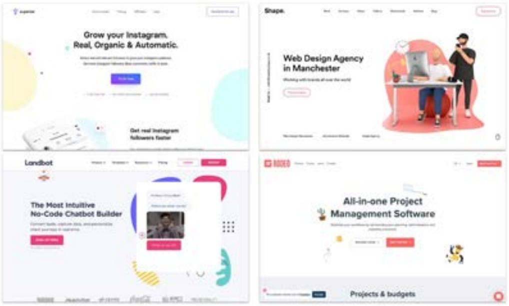
Creative Projects

Automania

Cézanne Drawing

Alexander Calder: Modern from the Start

Blank slate


Grow your Instagram. Real, Organic & Automatic.
Get real Instagram followers faster
Shape.
Web Design Agency in Manchester
Landbot
The Most Intuitive No-Code Chatbot Builder
All-in-one Project Management Software
Projects & budgets

References

SUCCESSFUL, FASTER, MORE EXPLORATION

80% of Stylette users completed the task

only 35% of DevTools users did

Completed the task in 35% less time

($p < .05$)

Similar number of changes

but

More diverse properties used

SELF-CONFIDENCE DROPS?!

Initial increase for both conditions (p <.05)

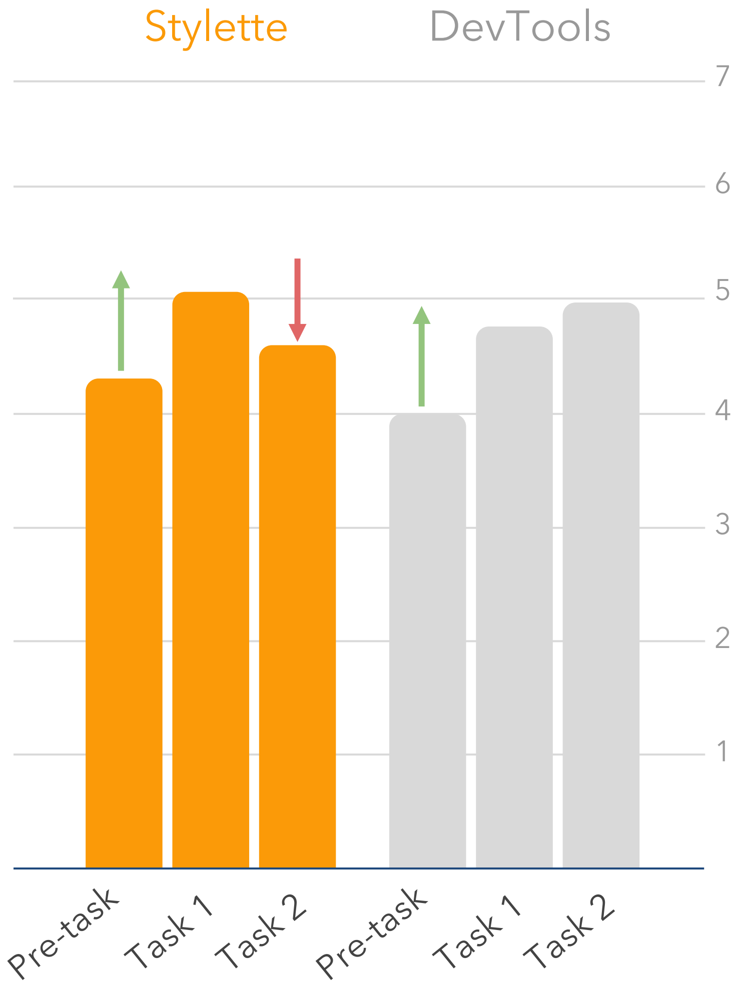
but decreases later for Stylette users (p <.05)

DevTools users
felt rewarded

> "This is my first time handling [CSS] but I did this!"
> D14

Stylette users
felt restricted

> "I expected more surprising changes like glittering."
> S4



초거대 AI에 의한 무력감:

AI와의 인터랙션을 통해 자기효능감을 찾을 수 있을까?

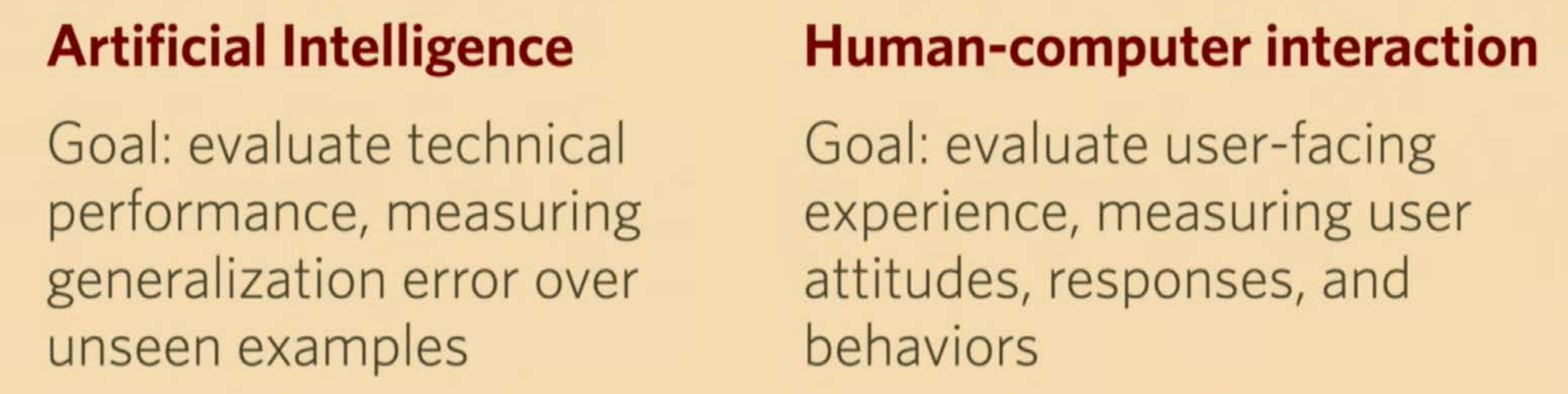
핵심은 주도권과 컨트롤

몰랐던 걸 **알게 되는 과정**에서
자기 신뢰감과 **자기 효능감** 상승

KBS. 대화의희열3. https://www.youtube.com/watch?v=TQNELc4dJr4

Toward Interaction-Centric AI

TWO CULTURES OF EVALUATION



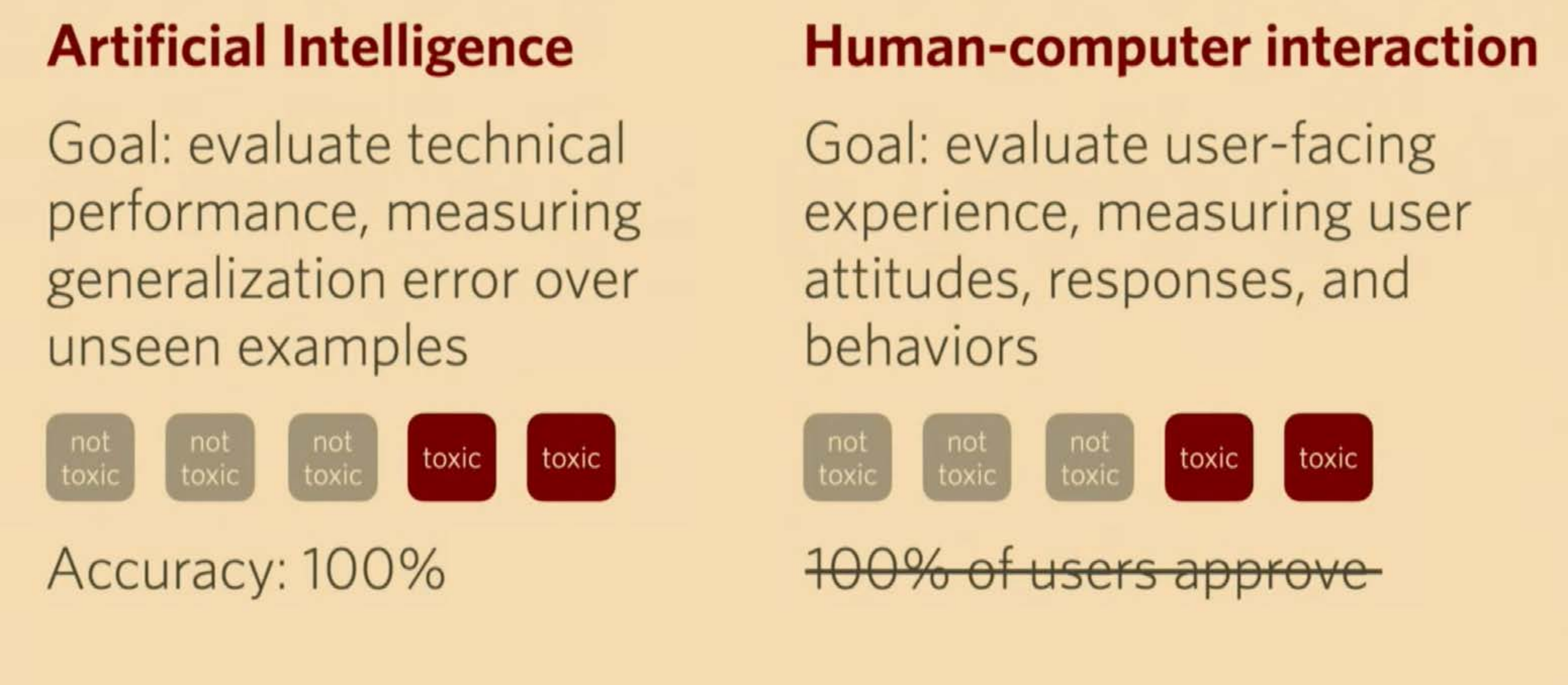
Artificial Intelligence	**Human-computer interaction**
Goal: evaluate technical performance, measuring generalization error over unseen examples	Goal: evaluate user-facing experience, measuring user attitudes, responses, and behaviors

ROC AUC, precision, recall, correlation, MSE, NDCG, BLEU score, inception score…

task performance, task time, cognitive load, usability, user satisfaction, …

Slide from Michael Bernstein

TWO CULTURES OF EVALUATION



Artificial Intelligence

Goal: evaluate technical performance, measuring generalization error over unseen examples

not toxic | not toxic | not toxic | toxic | toxic

Accuracy: 100%

Human-computer interaction

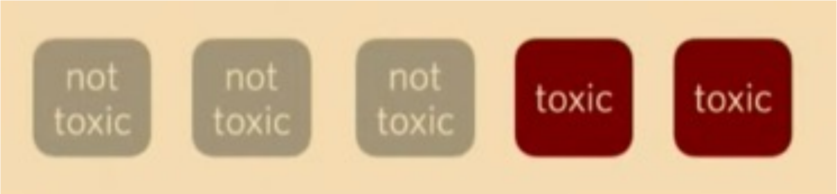
Goal: evaluate user-facing experience, measuring user attitudes, responses, and behaviors

not toxic | not toxic | not toxic | toxic | toxic

~~100% of users approve~~

Gordon, Mitchell L., et al. "The disagreement deconvolution: Bringing machine learning performance metrics in line with reality." *CHI 2021*.

REALIGNING PERFORMANCE MEASURES



Disagreement deconvolution

Not "What proportion of ground truth labels does the classifier agree with?"

But **"What proportion of the population does the classifier agree with?"**

Jigsaw Toxicity Task (ROC AUC)

0.95 → **0.73**

Different model / application decisions with the realigned metric.

Gordon, Mitchell L., et al. "The disagreement deconvolution: Bringing machine learning performance metrics in line with reality." *CHI 2021*.

AI의 성능 지표가 사람들의 가치와 의도를 반영하는가?

- 인터랙션 설계 및 관찰이 중요

실세계의 복잡한 가치를 어떻게 지표에 반영시킬 수 있는가?

- 인간 중심, 사용자 경험 중심으로 설계
- 다수결이 아니라 다양한 의견을 반영하는 데이터/모델 파이프라인 설계

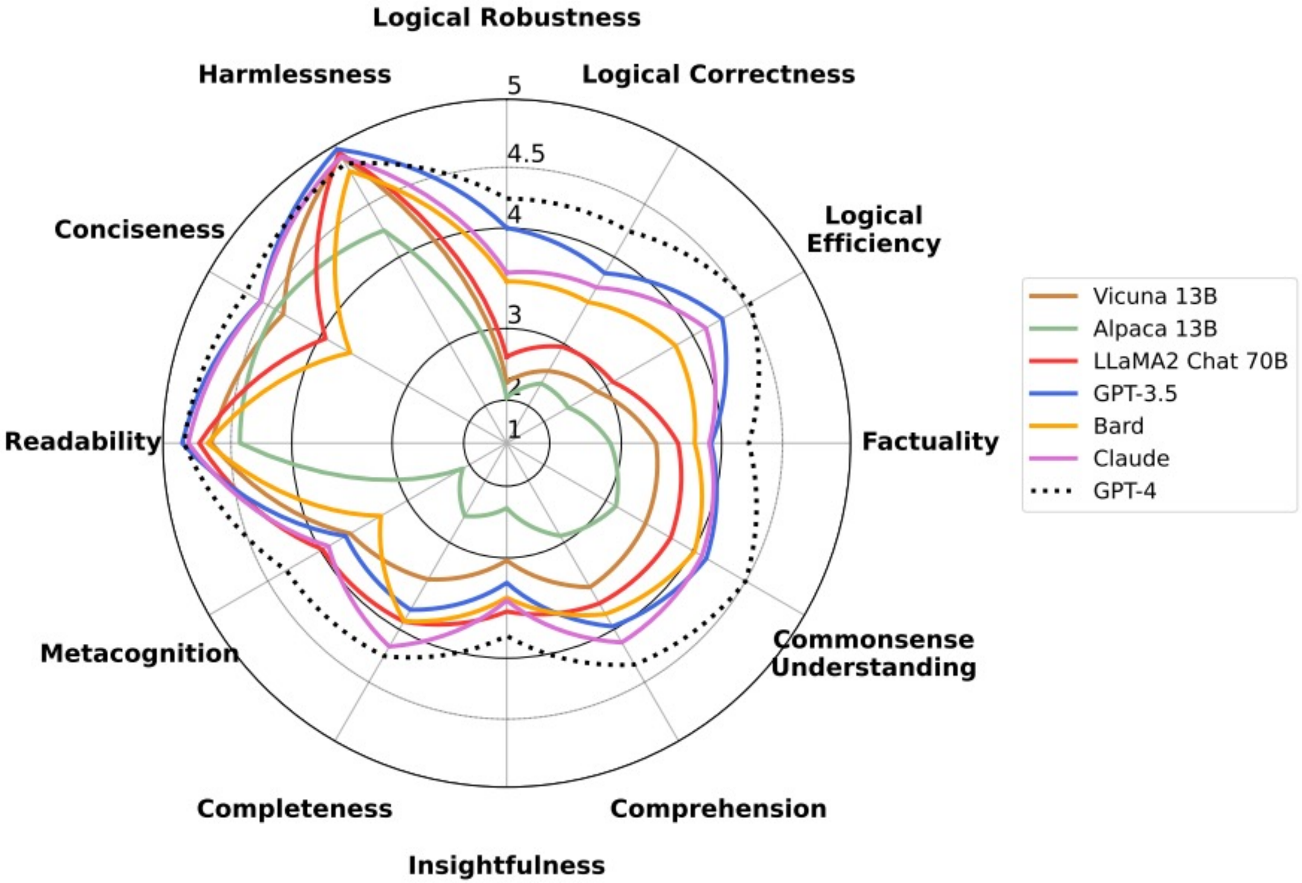
AI의 성능 지표를 어떻게 설계하고 검증해야 하는가?

- 유연하고 주관적인 지표를 빠르고 효율적으로 검증할 수 있는 체계 필요

FLASK: FINE-GRAINED LANGUAGE MODEL EVALUATION BASED ON ALIGNMENT SKILL SETS

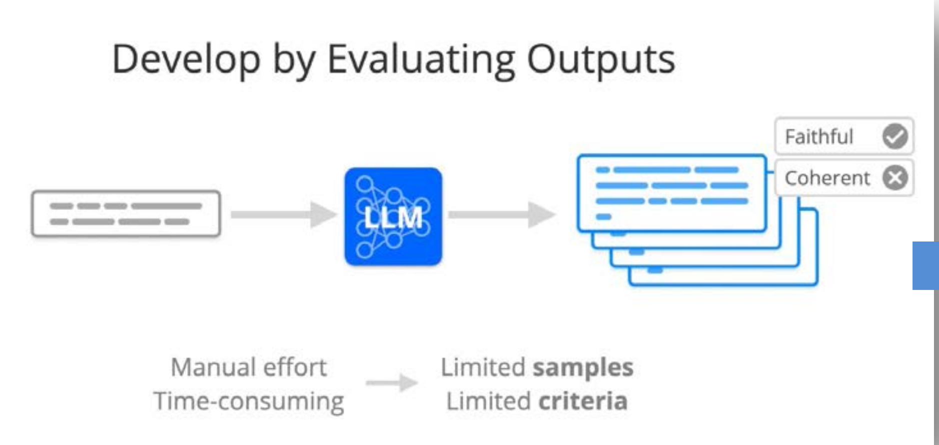
- Logical thinking
- Background knowledge
- Problem handling
- User alignment

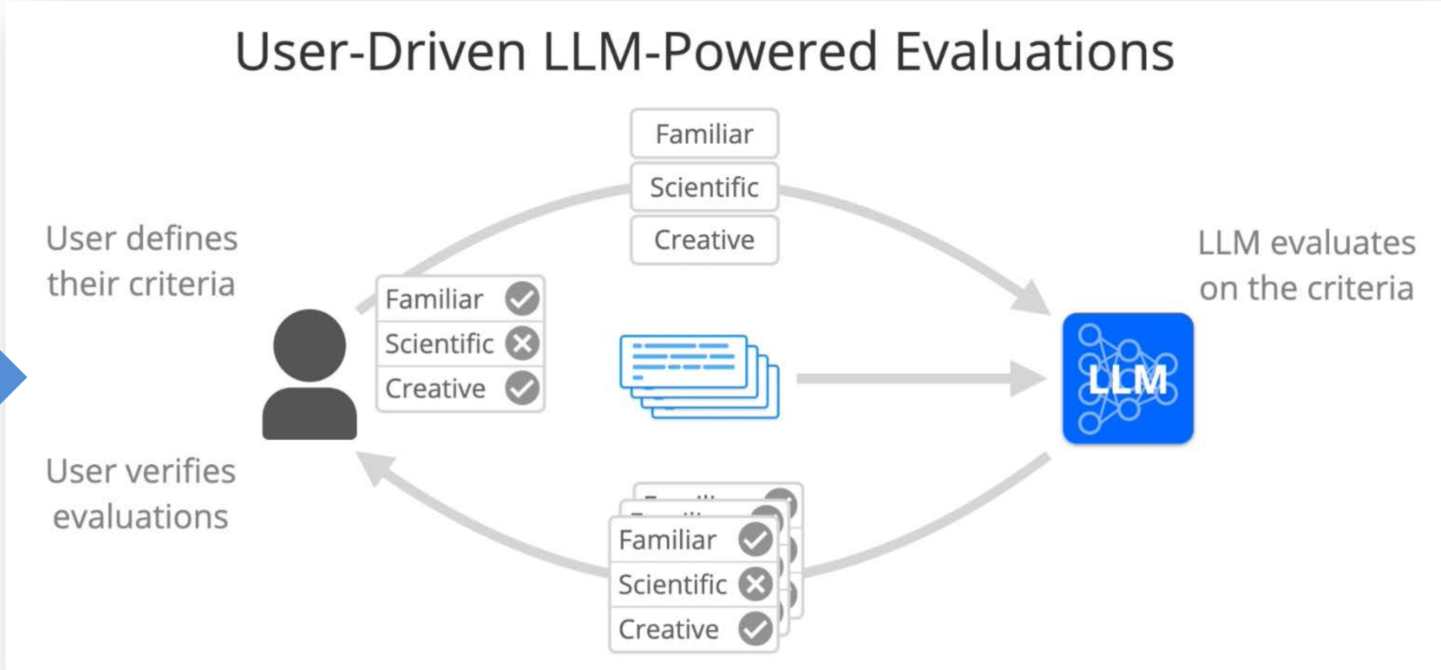
https://kaistai.github.io/FLASK/



Ye, Seonghyeon, et al. "FLASK: Fine-grained Language Model Evaluation based on Alignment Skill Sets." arXiv preprint arXiv:2307.10928 (2023).

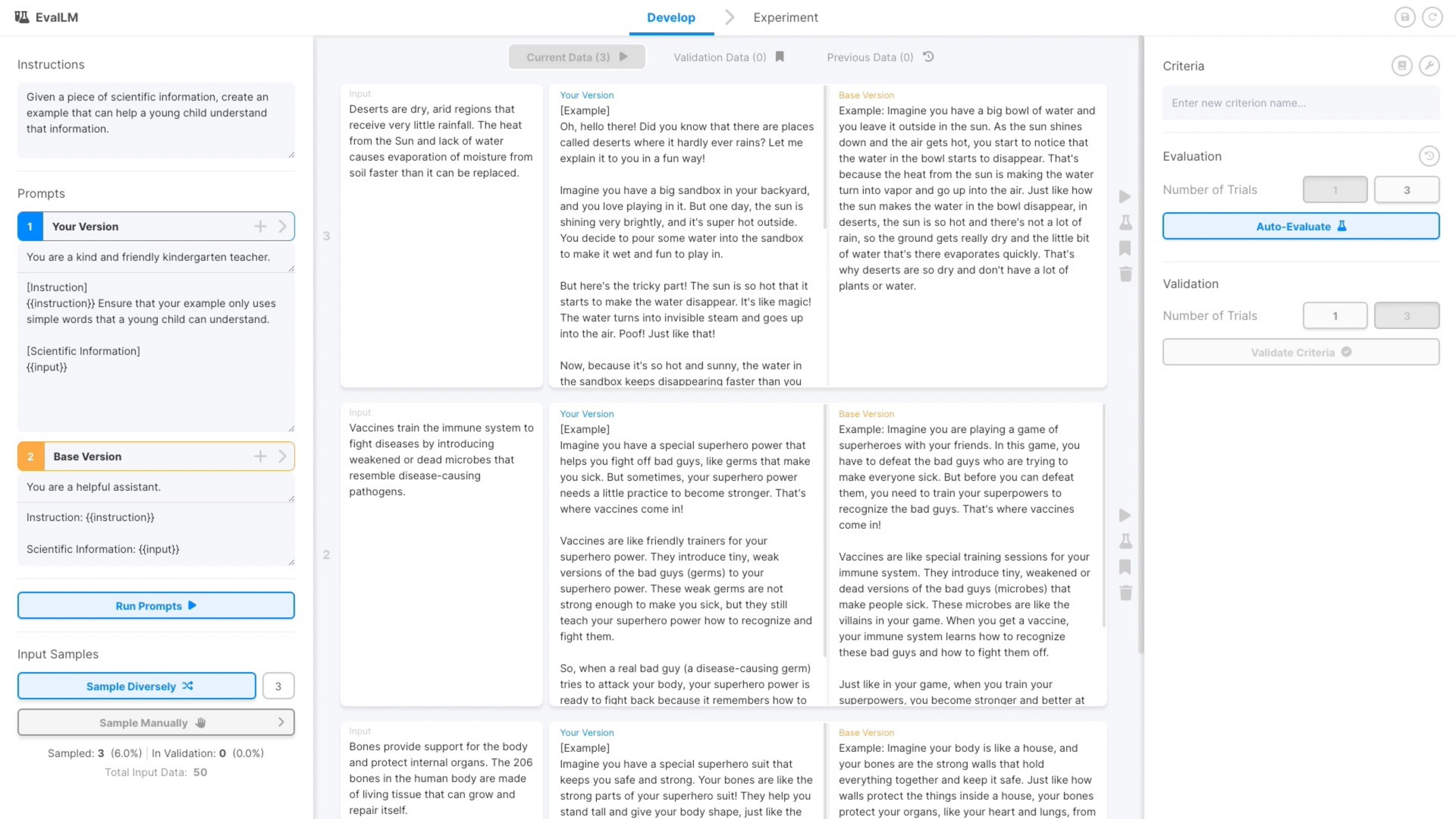
EVALLM: USER-DEFINED METRICS IN LLM APPLICATIONS





https://evallm.kixlab.org/

Kim, Tae Soo, et al. "EvalLM: Interactive Evaluation of Large Language Model Prompts on User-Defined Criteria." arXiv:2309.13633 (2023).



EvalLM

Develop > Experiment

Instructions

Given a piece of scientific information, create an example that can help a young child understand that information.

Prompts

1 **Your Version**

You are a kind and friendly kindergarten teacher.

[Instruction]
{{instruction}} Ensure that your example only uses simple words that a young child can understand.

[Scientific Information]
{{input}}

2 **Base Version**

You are a helpful assistant.

Instruction: {{instruction}}

Scientific Information: {{input}}

Run Prompts ▶

Input Samples

Sample Diversely 3

Sample Manually

Sampled: 3 (6.0%) | In Validation: 0 (0.0%)
Total Input Data: 50

Current Data (3) ▶ Validation Data (0) Previous Data (0)

3

Input

Deserts are dry, arid regions that receive very little rainfall. The heat from the Sun and lack of water causes evaporation of moisture from soil faster than it can be replaced.

Your Version

[Example]
Oh, hello there! Did you know that there are places called deserts where it hardly ever rains? Let me explain it to you in a fun way!

Imagine you have a big sandbox in your backyard, and you love playing in it. But one day, the sun is shining very brightly, and it's super hot outside. You decide to pour some water into the sandbox to make it wet and fun to play in.

But here's the tricky part! The sun is so hot that it starts to make the water disappear. It's like magic! The water turns into invisible steam and goes up into the air. Poof! Just like that!

Now, because it's so hot and sunny, the water in the sandbox keeps disappearing faster than you

Base Version

Example: Imagine you have a big bowl of water and you leave it outside in the sun. As the sun shines down and the air gets hot, you start to notice that the water in the bowl starts to disappear. That's because the heat from the sun is making the water turn into vapor and go up into the air. Just like how the sun makes the water in the bowl disappear, in deserts, the sun is so hot and there's not a lot of rain, so the ground gets really dry and the little bit of water that's there evaporates quickly. That's why deserts are so dry and don't have a lot of plants or water.

2

Input

Vaccines train the immune system to fight diseases by introducing weakened or dead microbes that resemble disease-causing pathogens.

Your Version

[Example]
Imagine you have a special superhero power that helps you fight off bad guys, like germs that make you sick. But sometimes, your superhero power needs a little practice to become stronger. That's where vaccines come in!

Vaccines are like friendly trainers for your superhero power. They introduce tiny, weak versions of the bad guys (germs) to your superhero power. These weak germs are not strong enough to make you sick, but they still teach your superhero power how to recognize and fight them.

So, when a real bad guy (a disease-causing germ) tries to attack your body, your superhero power is ready to fight back because it remembers how to

Base Version

Example: Imagine you are playing a game of superheroes with your friends. In this game, you have to defeat the bad guys who are trying to make everyone sick. But before you can defeat them, you need to train your superpowers to recognize the bad guys. That's where vaccines come in!

Vaccines are like special training sessions for your immune system. They introduce tiny, weakened or dead versions of the bad guys (microbes) that make people sick. These microbes are like the villains in your game. When you get a vaccine, your immune system learns how to recognize these bad guys and how to fight them off.

Just like in your game, when you train your superpowers, you become stronger and better at

Input

Bones provide support for the body and protect internal organs. The 206 bones in the human body are made of living tissue that can grow and repair itself.

Your Version

[Example]
Imagine you have a special superhero suit that keeps you safe and strong. Your bones are like the strong parts of your superhero suit! They help you stand tall and give your body shape, just like the

Base Version

Example: Imagine your body is like a house, and your bones are the strong walls that hold everything together and keep it safe. Just like how walls protect the things inside a house, your bones protect your organs, like your heart and lungs, from

Criteria

Enter new criterion name...

Evaluation

Number of Trials 1 3

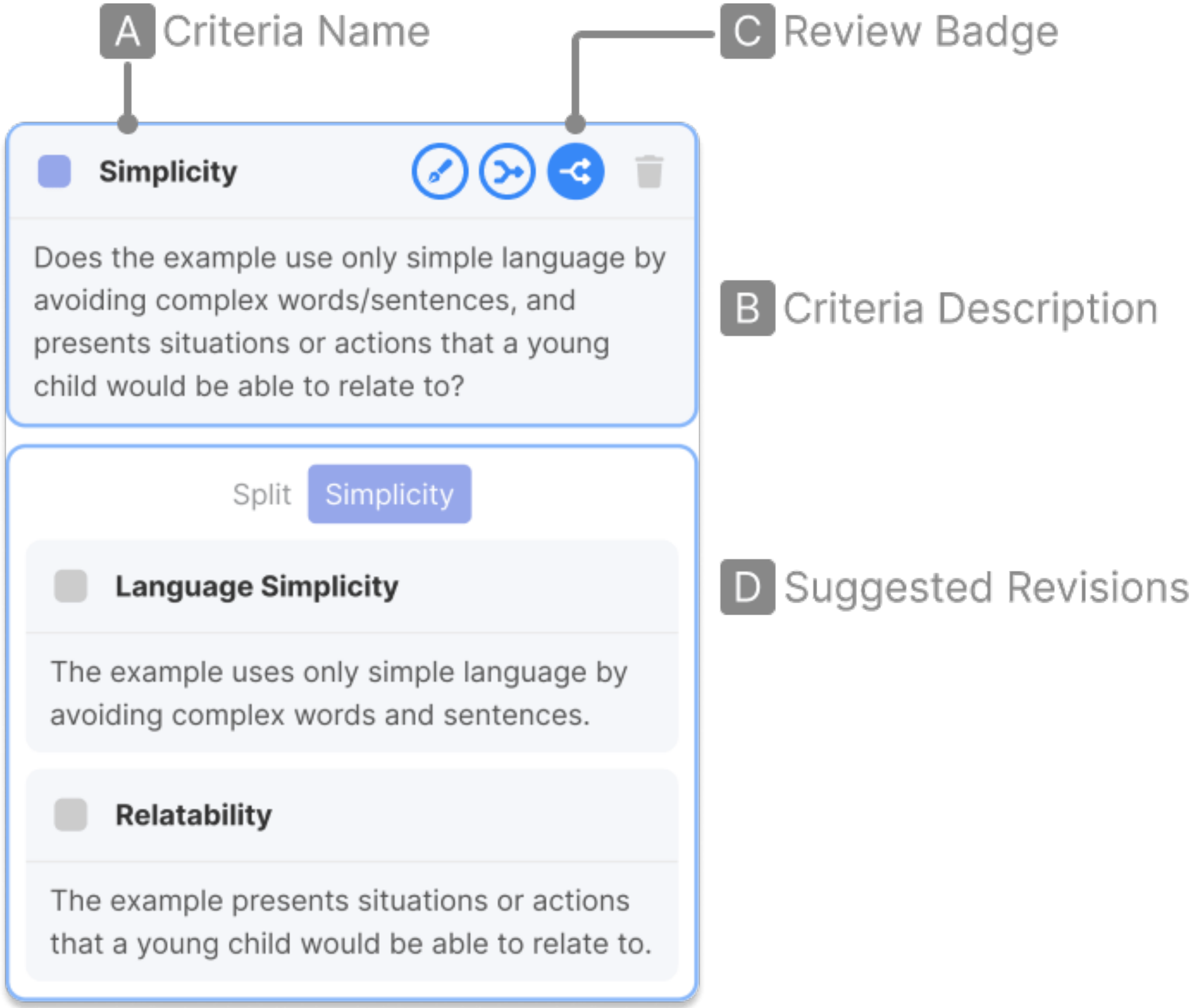
Auto-Evaluate

Validation

Number of Trials 1 3

Validate Criteria

- Define custom criteria
- Pre-defined criteria suggestions
- Refine/Merge/Split suggestions



Simplicity

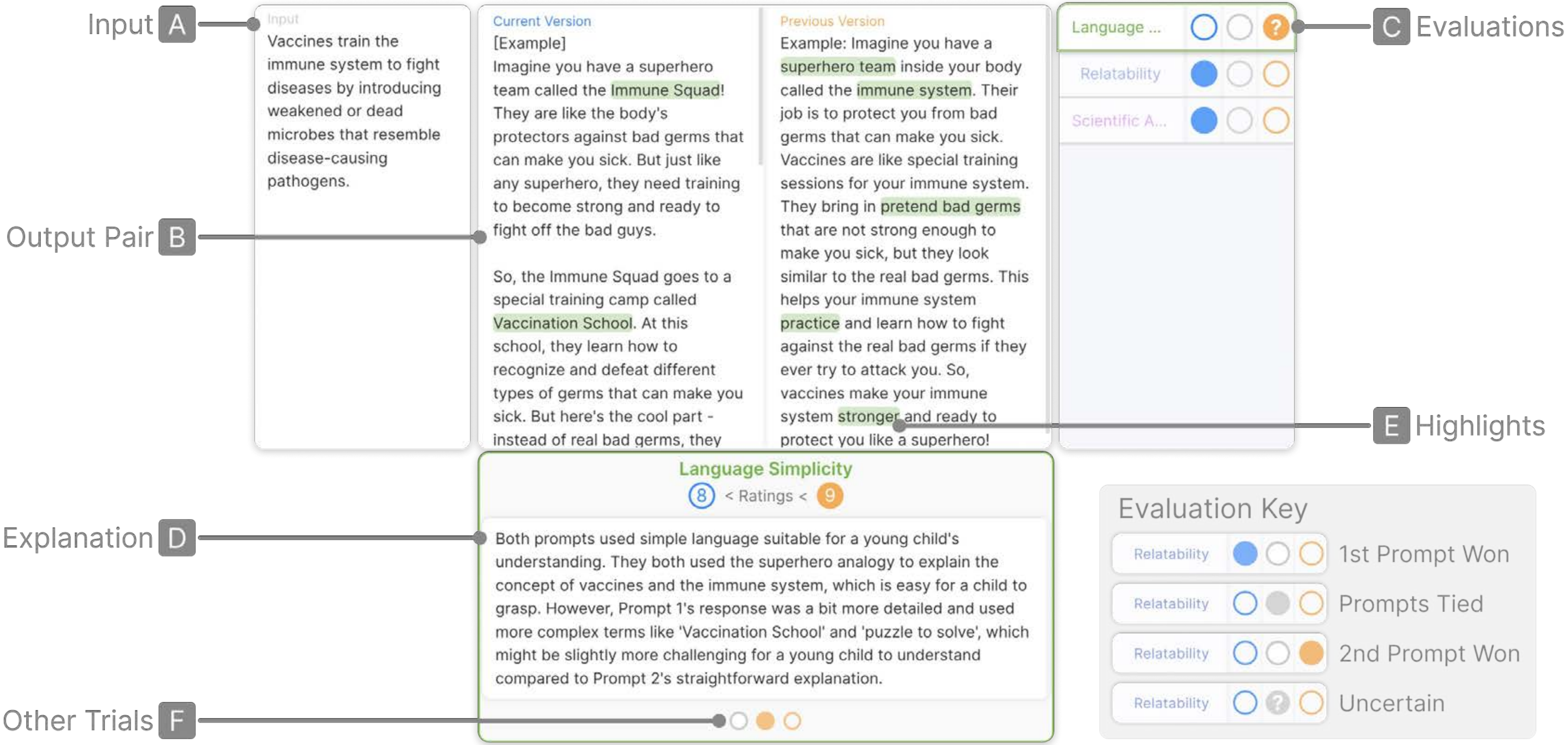
Does the example use only simple language by avoiding complex words/sentences, and presents situations or actions that a young child would be able to relate to?

Language Simplicity

The example uses only simple language by avoiding complex words and sentences.

Relatability

The example presents situations or actions that a young child would be able to relate to.



Input

Vaccines train the immune system to fight diseases by introducing weakened or dead microbes that resemble disease-causing pathogens.

Current Version

[Example]
Imagine you have a superhero team called the Immune Squad! They are like the body's protectors against bad germs that can make you sick. But just like any superhero, they need training to become strong and ready to fight off the bad guys.

So, the Immune Squad goes to a special training camp called Vaccination School. At this school, they learn how to recognize and defeat different types of germs that can make you sick. But here's the cool part - instead of real bad germs, they

Previous Version

Example: Imagine you have a superhero team inside your body called the immune system. Their job is to protect you from bad germs that can make you sick. Vaccines are like special training sessions for your immune system. They bring in pretend bad germs that are not strong enough to make you sick, but they look similar to the real bad germs. This helps your immune system practice and learn how to fight against the real bad germs if they ever try to attack you. So, vaccines make your immune system stronger and ready to protect you like a superhero!

Language Simplicity

8 < Ratings < 9

Both prompts used simple language suitable for a young child's understanding. They both used the superhero analogy to explain the concept of vaccines and the immune system, which is easy for a child to grasp. However, Prompt 1's response was a bit more detailed and used more complex terms like 'Vaccination School' and 'puzzle to solve', which might be slightly more challenging for a young child to understand compared to Prompt 2's straightforward explanation.

		Human-Centric AI?
Model-Centric AI	**Data-Centric AI**	**Interaction-Centric AI**
Good accuracy	Good efficiency	Good UX
Well-trained model	Quality data	Usable application
Benchmarks	MLOps	Human-AI interaction
Building blocks, often not an end point	Reliable & streamlined pipeline	Complex real-world tasks

100X
AI APPLICATIONS

Education, productivity, healthcare, finance, data analysis, decision making, writing, programming, design, translation, marketing, search, planning, shopping, …

100X
HUMAN-AI INTERACTIONS

Key Components:

(1) **Interaction** enables intuitive & efficient LLM UX.

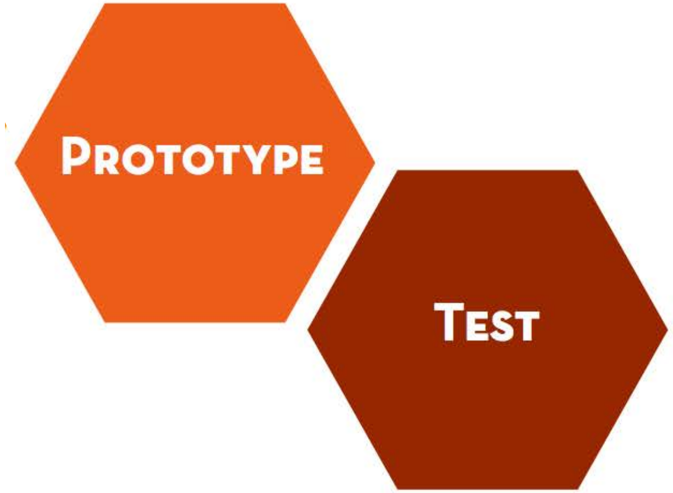
(2) **Architecture** enables controllable & predictable LLM behavior.

(3) **Evaluation** enables custom & aligned LLM performance.

MODEL-CENTRIC DESIGN PROCESS

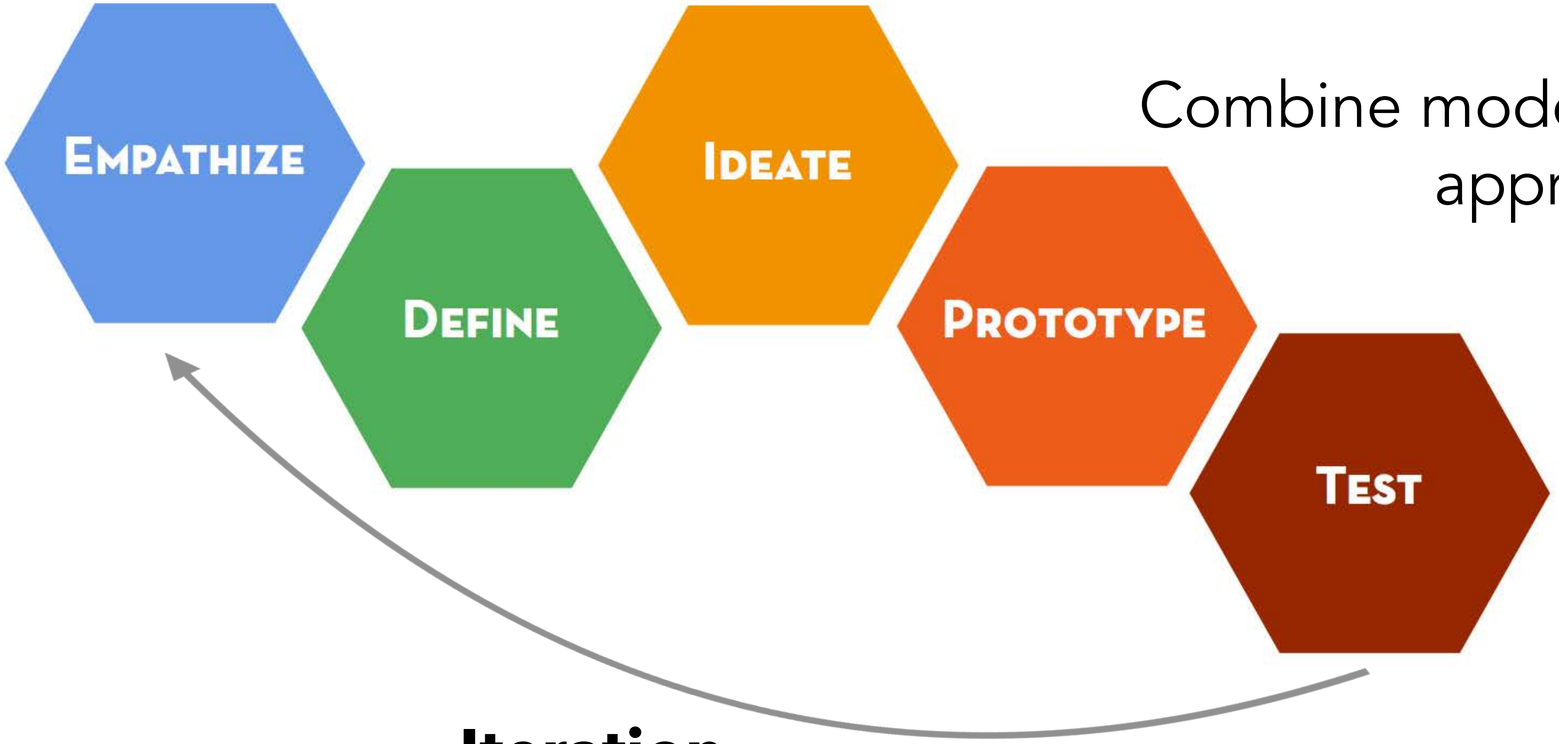
Here's a cool model.
Let's build a better performing model.

Here's a cool model.
What can we build with it?



ITERATIVE HUMAN-CENTERED DESIGN PROCESS

Interaction post-hoc ➔
guide the design & research process



Combine model- & data-centric
approaches

Iteration

Image Credit: Stanford d.school

생성형 AI가 직면한 세 가지 interaction 과제

사람의 의도와 가치를 제대로 담아낼 수 있는가?

다양한 사용자 그룹에 일관된 경험을 제공할 수 있는가?

사람에게 충분한 주도권을 줄 수 있는가?

생성형 AI가 제시하는 interaction 혁신의 기회

사람의 의도와 가치를 제대로 담아낸다. → **aligned AI**

다양한 사용자 그룹에 일관된 경험을 제공한다. → **equitable AI**

사람에게 충분한 주도권을 준다. → **controllable AI**

Interaction-centric AI

+

Interaction/Architecture/Evaluation Framework