



SPEAK2SCENE: Voice-based Storyboarding

Atharva S. Kashyap* , Patrícia Alves-Oliveira 

University of Michigan, Ann Arbor, Michigan, USA

ARTICLE INFO

Keywords:

Inclusive storyboarding
Participatory design
Generative AI
Language to image

ABSTRACT

Storyboarding is a method that supports participatory design by using a series of images to visualize one's experience, allowing them to communicate ideas both visually and through text. Traditionally, storyboarding involves hand-sketching, a modality inaccessible to people with disabilities affecting their upper limbs (e.g., quadriplegia). This restricts their engagement in participatory design to verbal interviews/interactions. We present SPEAK2SCENE, an open-source voice-to-image storyboarding tool powered by Generative Artificial Intelligence that adheres to accessibility standards, to be used as a research tool for inclusive participatory design. Our key insight is that by furthering participatory design research for people with disabilities with the inclusion of visual methods such as storyboarding, they can express their needs, views, and opinions in complementary ways to verbal language, enabling a more nuanced understanding of this population. We tested SPEAK2SCENE with 7 participants as part of a speculative co-design study and concluded that it can empower creative expression, support inclusion, and provide a joyful participatory design experience.

Metadata

Code metadata.

Nr.	Code metadata description	Metadata
C1	Current code version	v1
C2	Permanent link to code/repository used for this code version	https://github.com/studiorobot/Speak2Scene/tree/v2-softwarex
C3	Permanent link to Reproducible Capsule	None
C4	Legal Code License	MIT License
C5	Code versioning system used	git
C6	Software code languages, tools, and services used	JavaScript, CSS, HTML, OpenAI API, Firebase
C7	Compilation requirements, operating environments & dependencies	Node Version 22 or newer, npm Version 10.5 or newer
C8	If available Link to developer documentation/manual	None
C9	Support email for questions	katharva@umich.edu

Software metadata.

Nr.	(Executable) software metadata description	Please fill in this column
S1	Current software version	v1
S2	Permanent link to executables of this version	For example: https://github.com/studiorobot/Speak2Scene/releases/tag/v2-softwarex
S3	Permanent link to Reproducible Capsule	
S4	Legal Software License	MIT License
S5	Computing platforms/Operating Systems	Linux/Unix-based terminal, web: Google Chrome
S6	Installation requirements & dependencies	Node Version 22 or newer, npm Version 10.5 or newer, outlined in Speak2Scene/package-lock.json and Speak2Scene/package.json
S7	Support email for questions	katharva@umich.edu

* Corresponding author.

Email addresses: katharva@umich.edu (A.S. Kashyap), robopati@umich.edu (P. Alves-Oliveira).

<https://doi.org/10.1016/j.softx.2026.102628>

Received 9 December 2025; Received in revised form 3 March 2026; Accepted 24 March 2026

Available online 6 April 2026

2352-7110/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

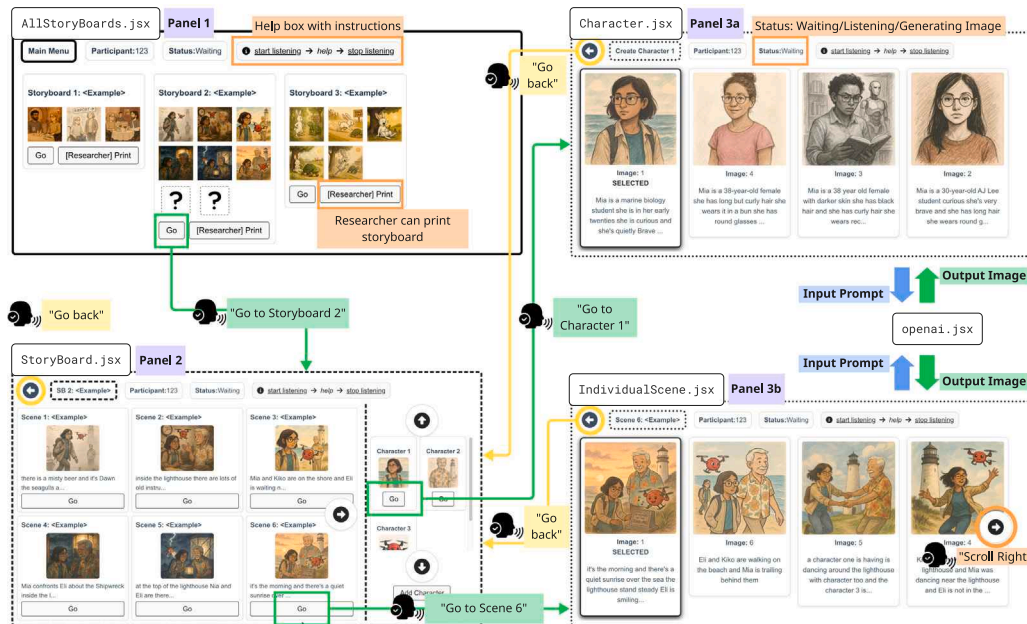


Fig. 1. SPEAK2SCENE Frontend workflow. The landing page of the tool starts with `AllStoryBoards.jsx`, where users can navigate into `Storyboard.jsx`, which will render the particular storyboard they navigated to. From this screen, users may navigate into `Character.jsx`, where they can create their characters, or navigate into `IndividualScene.jsx`, where they can generate images for the particular scene. Within `Character.jsx` or `IndividualScene.jsx`, they may use their voice to interact with `openai.jsx` and generate images.

1. Motivation and significance

Despite the growing emphasis on inclusive design, People with Disabilities (PwD) continue to face barriers to participating as equal contributors in participatory design research, partly due to the inaccessibility of existing design tools [1]. According to Simonsen et al. [2], Participatory Design (PD) methods commonly follow a tell–make–enact framework: participants share stories (tell), create artifacts (make), and perform or simulate scenarios (enact) during a design session. One widely used co-design method that embodies this framework is storyboarding, in which participants create hand-drawn sketches accompanied by short textual descriptions of scenarios [3]. Captions or accompanying semi-structured interviews often function as scripts to articulate experiences (tell), the act of creating the storyboard constitutes a generative process (make), and the storyboard itself serves as a blueprint for enactment (enact).

Storyboarding builds upon storytelling by offering a structured yet expressive medium for participants to visualize experiences and generate design insights across domains [4–7]. However, this method can be inaccessible to PwD with upper-limb mobility impairments (e.g., Parkinson’s disease, quadriplegia), as it traditionally relies on manual sketching. As a result, researchers often default to semi-structured interviews, which may not fully align with the principles of PD [2]. Beyond accessibility concerns, PD and storyboarding should ideally be joyful, playful experiences that invite imagination, co-creation, and a sense of ownership—qualities that interviews alone rarely capture.

In this paper, we introduce SPEAK2SCENE, a storyboarding tool, that leverages Generative Artificial Intelligence (GenAI), particularly `gpt-4o` [8] and `gpt-image-1` [9] models, and ReactJS Speech Recognition library [10] to enable voice-based user interaction to create storyboards. SPEAK2SCENE was adapted and used for a PD study [11]. A broad impact of using this tool is that it fosters accessibility for both PwD and researchers. By lowering participation barriers, SPEAK2SCENE promotes inclusion of diverse populations within PD research, and enables new forms of creative expression that leverage Artificial Intelligence (AI). Further, as an open-source software, it can be easily adapted and extended by researchers to suit their specific needs.

2. SPEAK2SCENE: a tool for inclusive participatory design

In this section, we describe the design process, software architecture, and functionalities of SPEAK2SCENE.

2.1. SPEAK2SCENE software architecture

The overall SPEAK2SCENE software architecture is divided into a frontend developed using ReactJS web framework [12] (summarized in Fig. 1) and a backend for image generation using OpenAI’s `gpt-image-1` [9] and `GPT-4o` [8] models as well as storage using Firebase [13] (summarized in Fig. 2). Detailed documentation about SPEAK2SCENE for use by future researchers is included in the GitHub Repository’s `README.md` file. SPEAK2SCENE relies on OpenAI’s models, which require users to create an account and provision funds for API usage. Associated costs depend on the selected model and the number of prompts generated, and these are the responsibility of the software implementer.

2.2. SPEAK2SCENE software functionalities

A primary function of this software is to enable creation of storyboards using voice interaction, such that it can broadly be used by diverse people. Particularly, our target users are people who have disabilities affecting their upper limbs (e.g., Quadriplegia). The interaction is designed in such a way that once the researcher sets up the interface with participant information, the participant can completely use the tool with only the use of their voice. It is also designed to be colorblind-friendly, including the generated images. Participant interaction with interface is summarized in Fig. 1. Examples of some storyboards created using SPEAK2SCENE are highlighted in Fig. 4.

There are primarily four panels. Panel 1, *Main Menu*, features a list of storyboards that the participant will be creating throughout the study. Depending on the needs of the study, the list of storyboards and the scenes within them can be edited by future researchers in the `constants.jsx` file. Navigating into any of the storyboards will take the participant into Panel 2, *SB <num>: <title>*.

Within the storyboard, the participants are able to see the various scenes for their scenario and are able to create characters that are part of



Fig. 2. SPEAK2SCENE backend workflow to generate images using OpenAI APIs. After receiving an input prompt, based on the function call (`generateCharacterImage()` or `generateSceneWithCharacterReference()`), the functions respectively interact with the Firebase database through functions defined in `firebase_helper_functions.jsx` and they interact with OpenAI API through `server.js`.

their story. Participants may further navigate into each individual *character* (Panel 3a) or each individual *scene* (Panel 3b) to create an image for them. Panel 3a and 3b are designed similarly wherein participants are able to create images of their choice and select the one they prefer. In creating images, Panel 3b ensures to use of the context of the list of characters to create images of the scene that have the respective characters in them. Navigation around these items can be done via voice input or through manual control using a computer mouse and the instructions for interacting with the interface are summarized in Fig. 3.

Within Panel 3a and 3b, to generate an image, participants interact with the interface using their own voice (i.e., natural language) to indicate the kind of image they wish to create. As a first step, participants can awaken the interface by saying “start listening”. Upon saying this, the interface will use the browser’s speech recognition capabilities to listen and capture the participant’s words which will be displayed on the screen as “Transcript”. When the interface is listening, the participant may describe their image as if they are simply speaking to another person. The interface was developed to be cognizant of human errors and participants may say “clear transcript” to restart their description. When they are done describing their image, participants can say “stop listening” to complete the interaction and trigger the image creation. During the image generation, the participant’s prompt is sent to OpenAI through an API call in addition to prompt engineering to capture the right input elements and context. While the image is being generated, a progress indicator is displayed to maintain user engagement. In the iteration of SPEAK2SCENE used by Kashyap et al. [11] to design robots for social dining, this generation period was utilized to conduct semi-structured interviews. After approximately 30–45 s, an image based on the prompt is generated and displayed. It is also stored in the database. If the participant is not satisfied with the image, they may repeat the process. At anytime, researchers may print or export any of storyboards as .pdf file at a click of a button.

2.3. SPEAK2SCENE design process

Throughout the design process, we tested the software with five individuals without upper mobility impairments, consulted an accessibility expert, and conducted a pilot test with a PwD. These interactions provided the following key insights:

2.3.1. People can be multiply disabled

A feedback we received was that when working with PwD, individuals may experience multiple disabilities (e.g., someone with Parkinson’s Disease may also have speech stutter). This feedback highlighted the

importance of designing the tool to remain usable across such intersections of disability. While much of the interface assumes that participants do not have significant vision, hearing, or speech impairments, we incorporated this insight by making SPEAK2SCENE colorblind-friendly. Specifically, we ensured that the interface design and prompt engineering processes accounted for color accessibility—not only by using black-and-white elements but also by guiding the model to produce colorblind-friendly visual outputs.

2.3.2. Ethical considerations

AI carries important ethical considerations. For example, Gebru et al. [14] argued that the pursuit of artificial general intelligence risks echoing eugenic ideologies. Biases in language models are well known in research [15–21] and similar issues are also highlighted in text-to-image models [22–24]. On the other end, Azuaje et al. [25] demonstrated that text-to-image models can support expressive writing therapies by helping individuals who have experienced trauma to generate images to positively distract users from their negative emotions. Additionally, Ko et al. [26] performed a study with visual artists on how they might incorporate text-to-image models into their creative work. Further, Liao et al. [27] explored the use of text-to-image models in education and observed an enhancement in student divergent thinking.

Storyboards, moreover, feature characters. Earlier versions of SPEAK2SCENE generated photorealistic images for both characters and scenes. However, feedback indicated that such images could make participants feel misrepresented or uncomfortable. Together, the feedback and ideas on ethics guided our careful consideration of implications of generated images during the development of SPEAK2SCENE. Consequently, we established an avatar-based approach – similar to video games – where participants can create sketch-style representations and use them to compose scenes.

3. Illustrative examples

In Fig. 4(a), re-creation of stories using SPEAK2SCENE is portrayed. An adapted version of SPEAK2SCENE was used to perform a speculative co-design study with seven participants (1 pilot/6 co-design sessions) who have upper mobility impairments to learn about their needs from a robot-assisted feeding system in social dining settings (e.g., restaurants) [11]. While the results of the study were submitted elsewhere, the focus of this paper is SPEAK2SCENE which was used in that study. Some of the storyboards created by them are portrayed in Fig. 4(b). At the end of the study, we interviewed participants to share their perspectives on using SPEAK2SCENE and some are highlighted below:

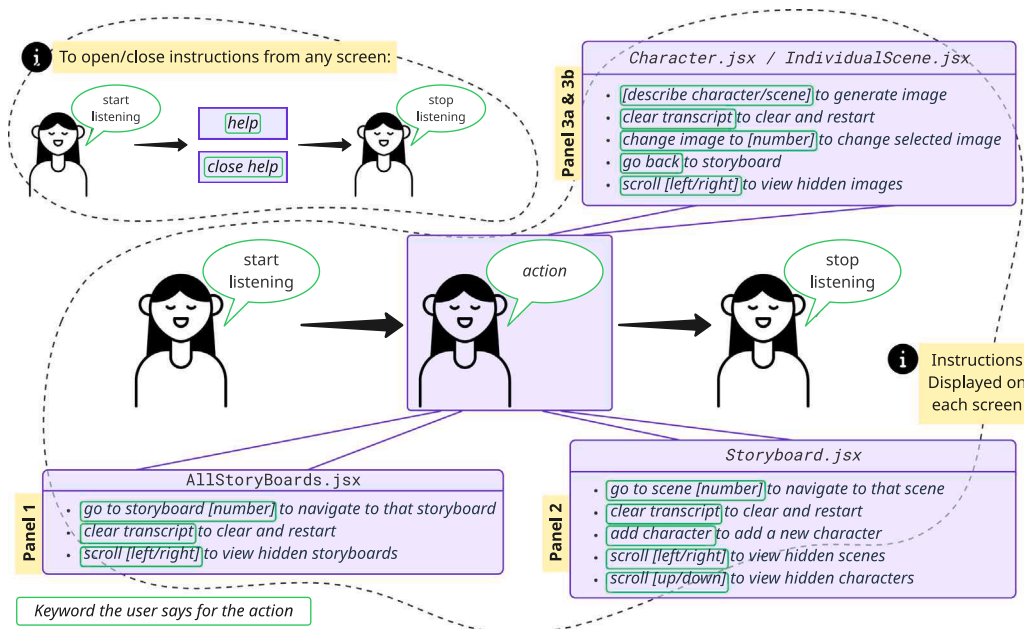


Fig. 3. Instructions for interacting with SPEAK2SCENE. Text outlined in green indicates the keywords that the user needs to say for the action. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1. Functional and inspiring:

Some participants reported that using SPEAK2SCENE positively shaped how they imagined the scenarios. The tool was not only seen as functional but also as inspiring and reflective, helping participants articulate needs they had not previously considered. One participant explained,

"I think the storyboarding tool was really useful. I was really impressed by it, and I thought it really specifically captured the nuances of the [scenario] with the needs of a person with disabilities" (P4, female, age 55–64 y/o, Parkinson's/Partial Paralysis).

This reflection demonstrates how the storyboarding tool extends traditional PD methods by providing an accessible interface that helps participants express complex needs and interactions.

3.2. Joyful storyboarding experience:

Beyond its practical utility, the tool also carried an affective impact. As one participant shared,

"It made me feel hopeful that there's much more help than I would have ever imagined. And that there's people that care about that. That, and they research it and study it. Because I'm not just thinking of myself, but I'm thinking of people that... have MS [Multiple Sclerosis] and they deal with a lot" (P2, female, age 65–74 y/o, Osteoarthritis/Shoulder Injury)

This statement highlights how speculative PD through a tool like SPEAK2SCENE can foster a sense of possibility and collective relevance, extending their imagination of what is possible. One participant who had an enjoyable time in the study expressed that she had a

"wonderful experience to use [SPEAK2SCENE] as a tool and it really did help [her] increase [her] imagination" (P4, female, age 55–64 y/o, Parkinson's/Partial Paralysis).

Using an adapted version of SPEAK2SCENE made the study sessions enjoyable for participants. PD sessions are most effective when participants feel comfortable expressing their experiences and ideas. In the speculative co-design study, using SPEAK2SCENE helped bridge the gap between participants and researchers, fostering a sense of shared

creativity. At times, participants even proposed humorous ideas for their robot designs, reflecting genuine engagement during the sessions.

3.3. Empower out-of-the-box thinking:

The goal of storyboarding in PD is to enable participants to imagine situations or realities beyond the ordinary. Several participants echoed this intent, with one participant noting that using SPEAK2SCENE helped him *"think differently"* (P6, male, age 25–34 y/o, Multiple Sclerosis) and generate new ideas for designing within the presented scenarios.

As another participant shared her experience using the tool helped her gain more design ideas on how she could control her robot,

"As I went along, I got more ideas about how to... program [the robot]" (P1, female, age 35–44 y/o, Quadriplegia)

Broadly, participants found SPEAK2SCENE both functional and inspiring, as it helped them articulate their needs, imagine new possibilities, and engage joyfully in the PD process. The tool fostered creativity, comfort, and collaboration, and empowered participants to think beyond traditional boundaries to envision solutions.

4. Impact

The impact of this software is manifold and we highlight some here:

4.1. Increased inclusion

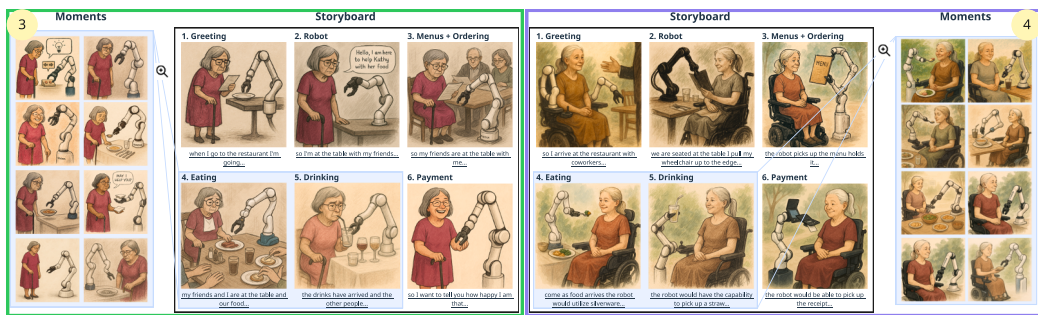
Inclusion in PD can take many forms and the impact of SPEAK2SCENE lies in its ability to foster inclusion both through accessibility and experience. The tool lowers barriers to participation by providing a voice-based interface that supports users with upper-body mobility impairments or limitations. Further, when participants find PD activities enjoyable, they are more likely to feel included and motivated to contribute. SPEAK2SCENE not only enables participation from diverse populations but also cultivates a sense of joy and belonging throughout the design process.

4.2. Creative expression

SPEAK2SCENE allows individuals who might not consider themselves "creative" to actively engage in storyboarding and design ideation.



(a) Example 1 (Left) is a re-creation of the popular *Hare and Turtle* story from Aesop’s Fables [28]. Example 2 (Right) is re-creation of a different short story (generated using *gpt-4o* and included in Appendix A).



(b) Example 3 (Left) is created by P2 (female, age 65-74 y/o, Osteoarthritis/Shoulder Injury). Example 4 (Right) is created by P3 (female, age 55-64 y/o, Spinal Muscular Atrophy).

Fig. 4. (a) Stories re-created using SPEAK2SCENE [28]. (b) Storyboards and key Moments created using an adapted version of SPEAK2SCENE by participants of a co-design study.

Furthermore, it promotes expressive inclusion by allowing users to concentrate on articulating their ideas rather than on the artistic or technical skills typically required in traditional storyboarding. Moreover, it allows people to use their voice, which can sometimes be preferred for expressing their ideas.

4.3. Accessible software

SPEAK2SCENE is an open-source software. Researchers with minimal software development experience can use the tool with minimal changes to the software. They will need to follow steps to create an OpenAI API key and setup a project on Firebase. Using those details, they can setup the application. Detailed instructions for setup are outlined in the README.md. Developers can also extend the tool to integrate features from popular design and image-editing platforms such as Miro [29], Adobe Express [30], or Figma [31]. Such integrations would allow storyboards and generated images to be directly incorporated into broader user experience (UX) design and research workflows. Moreover, by bridging SPEAK2SCENE with these established platforms, the accessibility and usability of professional design tools could be further democratized for PwD.

5. Conclusions & limitations

In this paper, we introduce SPEAK2SCENE, an open-source storyboarding tool that can be beneficial for researchers and participants engaging in PD. SPEAK2SCENE can be adapted to meet the needs of diverse use cases, expanding accessibility in participatory design. SPEAK2SCENE leverages the use of GenAI to create scenes and storyboards. This increases inclusion of diverse participants within PD, has the potential to boost creativity, and can create an enjoyable PD experience for participants.

Using GenAI has limitations, such as capturing biases (outlined in Section 2.3) or hallucinations [32–34]. Some of these limitations get carried over to SPEAK2SCENE. In some of the images generated, for instance, there were times when participants’ entire descriptions were not completely captured or there were details not mentioned being added.

SPEAK2SCENE is developed with accessibility at the forefront, which is the reason we chose to use voice-based interface. However, voice interaction can have limitations. For example, it is not always adept at handling unique accents or speech impairments (e.g., stuttering). While the interface is developed to handle errors in prompts, there are keywords (e.g., “start listening”) that need to be said in a way such that they are captured by the interface. Additionally, because SPEAK2SCENE relies on voice-based input, users can say the keyword “clear transcript” to restart the interaction; however, the system does not currently support partial corrections. Notably, since the system leverages GenAI, minor errors (e.g., grammar or spelling) are often automatically interpreted correctly by the model. Future work could explore adding multi-modal interaction and avenues to support partial corrections. Overall, our goal with the tool was to introduce a novel method of conducting PD, especially for a population that can be excluded from existing methods.

CRedit authorship contribution statement

Atharva S. Kashyap: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis. **Patrícia Alves-Oliveira:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

During the preparation of this work, the author(s) used GPT-4o to edit grammar. All ideas are original to the authors. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article. Icons from the Noun Project (<https://thenounproject.com/>) were used in Figs. 1 and 3. Thank you to Ugne, Connor, Francesca, Grace, and Laura Murphy for testing and providing feedback on versions of SPEAK2SCENE.

Appendix A. Verbose story: “The last seed”

“In a quiet village surrounded by dry, cracked earth, a young girl named Lena lived with her grandfather, Mason, a retired gardener. The rains had not come for months, and the once-vibrant fields had turned to dust. Every morning, Lena followed her grandfather to the abandoned garden, where only brittle stems remained.

One evening, while exploring the shed, Lena found a small metal box tucked beneath old tools. Inside it lay a single seed—round, smooth, and glimmering faintly under the lantern light. She showed it to her grandfather, who gazed at it silently before guiding her to their barren garden. There, together, they planted the seed in the cracked soil.

Days passed. The sun burned relentlessly, and their bucket of water grew lighter each day as they shared it between themselves and the tiny patch of ground where the seed slept. Just when it seemed hopeless, dark clouds gathered. Rain began to fall—first a few drops, then a downpour.

By dawn, the seed had sprouted into a small green shoot, trembling but alive. Lena and her grandfather stood in silence as the garden slowly began to breathe again. The single seed had become the heart of rebirth—proof that patience and care could outlast even drought.”

Appendix B. Supplementary material

Supplementary material for this article can be found online at doi:10.1016/j.softx.2026.102628.

References

- Rios D, Magasi S, Novak C, Harniss M. Conducting accessible research: including people with disabilities in public health, epidemiological, and outcomes studies. *Am J Public Health* 2016;106(12):2137–44, PMID: 27736212. <https://doi.org/10.2105/AJPH.2016.303448>
- Simonsen J, Robertson T. *Routledge International Handbook of Participatory Design*, vol. 711. Routledge New York; 2013.
- Young R, Sage K, Broom D, Hext A, Smith C. Effective use of storyboarding as a co-design method to enhance power assisted exercise equipment for people with stroke. *Des Health* 2022;6(2):244–75. <https://doi.org/10.1080/24735132.2022.2101257>
- Cross R, Warwick-Booth L. Using storyboards in participatory research. *Nurse Res* 2016;23(3):8. <https://doi.org/10.7748/NR.23.3.8.S3>
- Yu X, Ramli SH, Hamid HA, Norowi NM, Surip SS, Li Z, You D, Yu G, Chu Q. Visualizing user demands through storyboarding in the mHealth app development for pregnant women: a conceptual framework. *mHealth* 2025;11. <https://mhealth.amegroups.org/article/view/135588>.
- Rodda J, Ranscombe C, Kuys B. A method to explore strategies to communicate user experience through storyboards: an automotive design case study. *Artif Intell Eng Des Anal Manuf* 2022;36:e16. <https://doi.org/10.1017/S0890060421000287>
- Young R, Sage K, Broom D, Hext A, Smith C. Effective use of storyboarding as a co-design method to enhance power assisted exercise equipment for people with stroke. *Des Health* 2022. this is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 6. <https://doi.org/10.1080/24735132.2022.2101257>. <http://creativecommons.org/licenses/by-nc-nd/4.0/>, which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.
- OpenAI. GPT-4o model, Multimodal large language model capable of text, image, and audio understanding and generation, 2025. <https://openai.com/gpt-4o> [accessed: 25 July 2025].
- OpenAI. Image generated by GPT-image-1, Prompt: Describe the setting for a fantasy adventure, a bustling city built into a giant tree, with unique shops and characters, 2025. <https://openai.com/gpt-image-1> [accessed: 25 July 2025].
- React-Speech-Recognition Maintainers. *React-speech-recognition*. 2025. <https://www.npmjs.com/package/react-speech-recognition>, version 4.0.1, published 4 months ago.
- Kashyap AS, Morkute UA, Alves-Oliveira P. Robot-assisted social dining as a white glove service. In: *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26, New York, NY, USA: Association for Computing Machinery; 2026. <https://doi.org/10.1145/3772318.3790481>
- Facebook Open Source. *React – a JavaScript library for building user interfaces*; 2025. <https://react.dev/> [accessed: 25 July 2025].
- Google. *Firestore*. 2025. <https://firebase.google.com/>.
- Gebru T, Torres EP. The Tescree bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday* Apr 2024;29(4). <https://doi.org/10.5210/fm.v29i4.13636>. <https://firstmonday.org/ojs/index.php/fm/article/view/13636>.
- Gadiraju V, Kane S, Dev S, Taylor A, Wang D, Denton R, Brewer R. I wouldn't say offensive but...: disability-centered perspectives on large language models. In: *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*, FAccT '23, New York, NY, USA: Association for Computing Machinery; 2023. p. 205–16. <https://doi.org/10.1145/3593013.3593989>
- Navigli R, Conia S, Ross B. Biases in large language models: origins, inventory, and discussion. *J Data Inf Qual* Jun 2023;15(2). <https://doi.org/10.1145/3597307>
- Gallegos IO, Rossi RA, Barrow J, Tanjim MM, Kim S, Derroncourt F, Yu T, Zhang R, Ahmed NK. Bias and fairness in large language models: a survey. *Comput Linguist* 2024;50(3):1097–179.
- Kotek H, Dockum R, Sun D. Gender bias and stereotypes in large language models. In: *Proceedings of the ACM Collective Intelligence Conference*, CI '23, New York, NY, USA: Association for Computing Machinery; 2023. p. 12–24. <https://doi.org/10.1145/3582269.3615599>
- Bang Y, Chen D, Lee N, Fung P. Measuring political bias in large language models: what is said and how it is said. In: Ku L-W, Martins A, Srikumar V, editors. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (volume 1: long papers). Bangkok, Thailand: Association for Computational Linguistics; 2024. p. 11142–59. <https://doi.org/10.18653/v1/2024.acl-long.600>. <https://aclanthology.org/2024.acl-long.600/>.
- Gupta V, Narayanan Venkit P, Wilson S, Passonneau R. Sociodemographic bias in language models: a survey and forward path. In: Faleńska A, Basta C, Costa-Jussà M, Goldfarb-Tarrant S, Nozza D, editors. *Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing* (GeBNLP). Bangkok, Thailand: Association for Computational Linguistics; 2024. p. 295–322. <https://doi.org/10.18653/v1/2024.gebnlp-1.19>. <https://aclanthology.org/2024.gebnlp-1.19/>.
- Bird C, Ungless E, Kasirzadeh A. Typology of risks of generative text-to-image models. In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*; 2023. p. 396–410.
- Bianchi F, Kalluri P, Durmus E, Ladhak F, Cheng M, Nozza D, Hashimoto T, Jurafsky D, Zou J, Caliskan A. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, New York, NY, USA: Association for Computing Machinery; 2023. p. 1493–504. <https://doi.org/10.1145/3593013.3594095>
- Elsharif W, Alzubaidi M, Agus M. Cultural bias in text-to-image models: a systematic review of bias identification, evaluation, and mitigation strategies. *IEEE Access* 2025;13:122636–59. <https://doi.org/10.1109/ACCESS.2025.3585745>
- Wan Y, Subramonian A, Ovalle A, Lin Z, Suvarna A, Chance C, Bansal H, Pattichis R, Chang K-W. Survey of bias in text-to-image generation: Definition, evaluation, and mitigation. [arXiv preprint] [arXiv:2404.01030](https://arxiv.org/abs/2404.01030). 2024.
- Azuaje J, Liew K, Buening R, She WJ, Siriaraya P, Wakamiya S, Aramaki E. Exploring the use of AI text-to-image generation to downregulate negative emotions in an expressive writing application. *Royal Soc Open Sci* 2023;10(1):220238.
- Ko H-K, Park G, Jeon H, Jo J, Kim J, Seo J. Large-scale text-to-image generation models for visual artists' creative works. In: *Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23, New York, NY, USA: Association for Computing Machinery; 2023. p. 919–33. <https://doi.org/10.1145/3581641.3584078>
- Liao C-W, Chen H-W, Chen B-S, Wang I-C, Ho W-S, Huang W-L. Exploring the application of text-to-image generation technology in art education at vocational senior high schools in Taiwan. *Information* 2025;16(5). <https://doi.org/10.3390/info16050341>. <https://www.mdpi.com/2078-2489/16/5/341>.
- Aesop. *The hare & the tortoise*. <https://read.gov.aesop/025.html> [accessed 22 October 2025]. (n.d.)
- Miro. 2025. <https://miro.com/>, online collaboration whiteboard platform.
- Adobe Express. 2025. <https://www.adobe.com/express/>, design and content creation platform.
- Figma. 2025. <https://www.figma.com/>, collaborative interface design tool.
- Gunjal A, Yin J, Bas E. Detecting and preventing hallucinations in large vision language models. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38. 2024. p. 18135–43.
- Li Y, Du Y, Zhou K, Wang J, Zhao X, Wen J-R. Evaluating object hallucination in large vision-language models. In: Bouamor H, Pino J, Bali K, editors. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Singapore: Association for Computational Linguistics; 2023. p. 292–305. <https://doi.org/10.18653/v1/2023.emnlp-main.20>. <https://aclanthology.org/2023.emnlp-main.20/>.
- Reddy GP, Pavan Kumar YV, Prakash KP. Hallucinations in large language models (llms). In: *2024 IEEE Open Conference of Electrical, Electronic and Information Sciences* (eStream); 2024. p. 1–6. <https://doi.org/10.1109/eStream61684.2024.10542617>