AI-generated images for speech pathology - an exploratory application to aphasia assessment and intervention materials.

Dr John E. Pierce

Centre of Research Excellence in Aphasia Rehabilitation and Recovery, School of Allied Health Sciences and Sport, La Trobe University, Australia.

j.pierce@latrobe.edu.au

ORCID: 0000-0001-5164-5106

Abstract

Purpose: Images are a core component of aphasia assessment and intervention that require significant resources to produce or source. Text-to-image generation is an Artificial Intelligence (AI) technology that has recently made significant advances and could be a source of low cost, highly customisable images. The aim of this study was to explore the potential of AI image generation for use in aphasia by examining its efficiency and cost during generation of typical images.

Method: Two hundred targets (80 nouns, 80 verbs and 40 sentences) were selected at random from existing aphasia assessments and treatment software. A widely-known image generator, DALL-E 2, was given text prompts for each target. The success rate, number of prompts required, and costs were summarised across target categories (noun/verb/sentence) and compared to frequency and imageability.

Results: Of 200 targets, 189 (94.5%) successfully conveyed the key concept. The process took a mean 2.3 minutes per target at a cost of $0.31 USD each. However, there were aesthetic flaws in many successful images which could impact their utility. Noun images were generated with the highest efficiency and accuracy, followed by verbs, while sentences were more challenging, particularly those with unusual scenes. Patterns of flaws and errors in image generation are discussed.

Conclusions: The ability to rapidly generate low-cost, high-quality images using AI is likely to be a major contribution to aphasia assessment and treatment going forward, particularly as advances in this technology continue.

Introduction

Images are a core component of aphasia assessment and intervention (Brown & Thiessen, 2018). For example, confrontation naming is a key activity in assessment, treatment and self-practice of word retrieval, where images provide a visual stimulus without any inherent linguistic cues. In an analysis of 453 anomia treatment instances, at least 86% used pictures as stimuli (Thomson, 2012). Images are also commonly used to stimulate discourse-level language, such as in picture description and story retelling tasks (e.g., Fromm et al., 2020; Nicholas & Brookshire, 1993). In assessment or treatment of comprehension, one or more images are often presented alongside written words or sentences for matching. Images may also be presented alongside spoken or written messages to augment expression or comprehension (Brown & Thiessen, 2018).

Sourcing suitable images for these purposes can be time consuming, both for the clinician preparing resources for their clients and those developing published treatment materials or assessments. Images need to clearly represent the desired target concept in order to rule out the possibility that errors are due to inaccurate recognition or confusion (Brown & Thiessen, 2018; Heuer, 2016). However, the lower the imageability of the target, the more challenging it can be to find a representative image. In addition, when targets are personalised for an individual, some may be low frequency concepts that can also be challenging to source. Finding clear verb images is also more difficult (Brown & Thiessen, 2018) as the image should encourage a response based on the action rather than the object being used (e.g. ‘turn’ instead of ‘dial’) or may be abstract or low imageability (e.g. ‘know’). Sentences require yet more specific images that show the subject and object (or agent and theme) of the sentence in specified roles. For example, finding an image that depicts, ‘The child throws the paper plane,’ would be more difficult to find than the verb ‘throw,’ a child or a paper plane. In addition, depicting semantically reversible sentences is an important part of grammatical treatments such as Treatment of Underlying Forms (Thompson, 2019), but finding existing images for rarer sentences such as ‘the cat chases the dog’ is unlikely. Overall, images for aphasia management need to be very clear and often depict complex, low frequency concepts, meaning that sourcing suitable images can be challenging.

In clinical practice, speech pathologists may repurpose images from published assessments, despite the potential impact on validity, or may conduct image searches online. Although most results in image searches are likely under copyright, clinicians are probably aware that using such images for an individual’s treatment materials is unlikely to result in copyright claims. In contrast, published assessments and intervention materials cannot risk using copyrighted images and so sourcing images is more complicated. In common therapy software for aphasia (e.g. Cuespeak, Aphasia Therapy Online), developers often use stock images, which are typically high-quality but can entail significant cost. An alternative source is public domain or certain Creative Commons licenced images, which have free usage specified by the authors. Such images are available in increasingly large databases but even so, finding suitable images is highly time consuming. Last, many aphasia assessments and interventions use line drawings or other hand illustrations produced by paid artists (e.g. Western Aphasia Battery - Kertesz, 2007; Comprehensive Aphasia Test - Swinburne et al., 2004; Scenario Test - van der Meulen et al., 2010). Hand drawn images have the advantage of being completely customisable, eliminating the problems of obscure scenes or concepts (Reymond et al., 2022), but again come at significant cost and time. The quality of illustrations may also vary. In addition, line drawings have been shown to be less likely to elicit correct responses in people with aphasia compared to photographs or high quality illustrations, as they lack key features that aid recognition such as colour, shading and texture (Brown & Thiessen, 2018; Heuer, 2016; Reymond et al., 2022).

Text-to-image generation is an Artificial Intelligence technology that has made extraordinary advances in efficiency and accuracy in the past few years (Ramesh et al., 2021), and could potentially be applied in speech pathology practice. Given a text prompt, text-to-image models can synthesise matching images, currently with variable, but improving, accuracy. As the models have been scaled up in terms of training data and compute time, results have improved dramatically and models are now able to produce novel combinations and images they were not directly trained on; for example, “an illustration of a baby hedgehog in a Christmas sweater walking a dog” (Ramesh et al., 2021, p. 2). The mechanisms behind AI image generation models are highly complex, but there are two core elements worth describing here. The first is the process of learning connections between language and visual concepts. Training involves machine learning through processing massive datasets of images that have descriptive captions and labels, usually in the order of hundreds of millions of image-caption pairs (Radford et al., 2021). By contrasting different pairs to find commonalities and differences (Contrastive Language-Image Pre-training – CLIP), the AI models can learn the visual meaning of words (Radford et al., 2021). This applies to the objects and characters in images but also the concepts that actions, adjectives and adverbs describe, as well as meta-descriptions such as ‘watercolour art,’ ‘pixelated rendering,’ ‘polaroid photograph’ and even geographical locations and text portrayed within the Images (Radford et al., 2021). The second core element of image generation models is the creation of images in response to the text prompt. Multiple techniques exist, but currently, the most promising and widely used method is diffusion (Dhariwal & Nichol, 2021). Diffusion models are trained to gradually add noise or ‘static’ to a piece of data (image) until it is no longer recognizable. From that point, the model attempts to reconstruct the image to its original form (reverse diffusion). These steps are performed as a training process to learn the generation of an image. The model then learns to remove the noise in data and can eventually generate an image from a starting point of 100% noise (Dhariwal & Nichol, 2021).

If images generated by AI are suitable for aphasia assessment and intervention, this could afford multiple advantages in development of future tools. Importantly, it could save a significant amount of time in searching for images. It could also mean a substantial cost saving. Unlike stock images, no models or photographers are required, nor are artists – once the models are trained, the only cost is computing power. Further, at present, most images generated by AI are not under copyright and may not be legally copyrightable (Murray, 2023), providing a solution to the issue of using image search results[1]. AI generated images have the potential to be very customisable for highly specific or obscure sentences, yet at photorealistic quality, or nearly any style requested. Finally, they may also be customisable to depict gender, race, and culture in a personalised manner for individuals.

The aim of this study was to explore the potential for text-to-image generation to be used in creation of future aphasia assessment and intervention tools by examining its efficiency and cost during generation of typical images. Importantly, AI image generation could have further applications in aphasia, such as development of communication supports, and could extend further still to many domains of speech-language pathology practice. As a non-linguistic medium, images have particular utility in management of communication; for example, enabling communication within high and low tech Augmentative and Alternative Communication systems (Engebretsen et al., 2014; Pak et al., 2023) or assessing and treating pediatric language and phonology (e.g., Carter, 2013; Dunn & Dunn, 2007). The challenges of sourcing images within these fields are likely to be similar. The efficiency and cost findings from this proof-of-concept study of one application, aphasia assessment and therapy, could be similarly applicable to other areas of practice in speech pathology.

[1] Multiple artificial intelligence companies are currently being sued for copyright infringement, with a group of artists arguing that the models were trained on, and therefore copy, their style without permission (Brittain, 2023)

Method

Image generator

The model DALL-E 2 was chosen as the image generator (OpenAI, 2022a) as it is one of the most well-known examples, though many more are available. DALL-E 2 is also very user friendly, with a simple prompt interface similar to a search engine and no local software required, and therefore potentially the most likely to be used by clinicians and researchers.

Targets

To explore the utility of image generation for typical use, targets were taken from a range of existing aphasia assessments and treatment software, including:

- 40 verbs from Cuespeak (Hunt & Keech, 2017)
- 40 verbs from the Action Naming Test (Obler & Albert, 1979)
- 40 nouns from Aphasia Therapy Online (Pierce, 2013)
- 40 nouns from the Philadelphia Naming Test (Roach et al., 1996)
- 40 sentences from the Comprehensive Aphasia Test–- spoken sentence subtest (Swinburne et al., 2004)

These sources were chosen as commonly used assessments and intervention tools. In addition, Cuespeak and Aphasia Therapy Online were chosen as the author and two contributors to the paper are the developers of these programs and had access to the source data. More single word targets were chosen than sentences, reflecting the greater focus on single words in aphasia management (Hickin et al., 2022).

The 40 items were chosen at random from each source using a randomiser (Haahr, 2010), in order to investigate targets with a range of frequency and imageability. Duplicates (e.g. policeman chases the dancer, dancer is chased by the policeman) were replaced with other randomly chosen items, as were images depicting violence, as these breach content guidelines for DALL-E 2 and cannot be generated. Data from Brysbaert and New (2009) were sourced to check frequency. Both noun and verb categories contained items from low frequency (<1 per million words) to high frequency (>500 per million words). Imageability ratings were taken from the MRC Psycholinguistic Database (1997), which combines multiple imageability data sources. Though ratings were not available for all items, imageability ranged from moderate (322/700) to high (638), which is expected given they were chosen from sources with visual representations.

Procedures

The author generated images using text prompts. A prompt guide specific to DALL-E 2 was used to learn optimal wording for text prompts (Parsons, 2022), with approximately 90 practice runs on concepts not included in the study target list. To limit the scope of the study and to simulate time limits in finding stimuli, a maximum of six prompts were entered per item (each prompt generates four images); thus, each target had a maximum of 24 images generated. If the author judged that a generated image was adequately clear in conveying the concept, no further prompts were entered for that item. Prompts were primarily aimed at generating photographic images that clearly depicted the *concept*, rather than visual copies of source images. The source was only consulted where the meaning of an item was ambiguous from the text alone, e.g. rise, indicate. No advanced tools offered by DALL-E 2 were used, such as inpainting (editing within images using text prompts) or outpainting (expansion of an image beyond its original frame using additional text prompts).

All prompts were recorded, as well as the number of prompts required for each target and whether an acceptable image was generated within six prompts. Results were summarised across category (noun/verb/sentence). Costs were calculated based on the price per prompt–- at the time of data collection, the DALL-E 2 charge was $16.50 USD per 115 prompts ($0.143/prompt). The total time spent on the website was recorded using activity logging software.

Results

Image generation was attempted for 200 targets and was considered unsuccessful for 11 (5.5%), in that a recognisable representation could not be produced within 6 prompts. The prompts and generated images can be viewed online at https://johnepierce.github.io/AI-images-for-aphasia/, and the chosen images for the 'successful' 189 images (94.5%) can be viewed online at https://labs.openai.com/c/AZV0LiHLG3LdUJ7JnqzMmBKY.

In total, image generation took 7h 40m (average of 2.3 minutes per target). A mean of 2.2 prompts were required to produce a successful image for each target. The total cost was $59.66, or $0.31 per successful target.

Table 1 shows the results across target classes. Nouns used the lowest number of prompts to produce a suitable image, perhaps reflecting higher frequency and imageability. Sentences required the most prompts; more than double that of nouns. The success rates were comparable between nouns and verbs but considerably lower for sentences.

Table 1

Type	n	Median frequency, range	Mean imageability (sd), range	Mean prompts	Mean cost (USD)	Success rate
Noun	80	18.4, 0.04-514	591.3 (39.4), 451-637	1.50	$0.22	98.75%
Verb	80	10.2, 0.04-4583	514.9 (73.4), 322-638	2.33	$0.33	97.50%
Sentence	40	n/a	n/a	3.25	$0.47	80.00%
All	**200**	**15.47, 0.04-4583**	**555.2 (69.2), 322-638**	**2.18**	**$0.31**	**94.50%**

Images were judged as 'successful' by the author when they clearly depicted the target. Many successful images were produced in excellent quality (Figure 1); however, there were aesthetic flaws in a majority (Figure 2). The flaws varied considerably in number and degree from image to image, which is likely attributable to the random starting noise used for each image; however, some trends are described here. Most notably, human hands were very often anatomically incorrect (e.g. too many hands or fingers, odd angles of digits) and faces frequently looked bizarre (see online supplementary material for a selection of particularly unusual unsuccessful images produced during data collection). Fine or repeating details were also poorly rendered, including knitting needles, piano keys, or buttons on a telephone or ATM, as well as written words. Targets with fewer features were generated at very high quality (e.g. bell, orange juice) in comparison to those with many required elements (e.g. restaurant), where the quality of each element suffered. Uncommon details were also less likely to appear without careful prompt engineering – for example, in sentences requiring a blue flower, the colour was sometimes applied to other objects in the scene, necessitating a colour label for each. For a prompt requiring a green flower beneath a cup, in some images, plants and grass were rendered instead of a flower or the cup was green. Similarly, sentences describing typical scenes, such as a man painting a picture, were more accurately generated than sentences with unusual roles, such as a singer photographing a soldier.

Figure 1 – Selection of high quality generated images

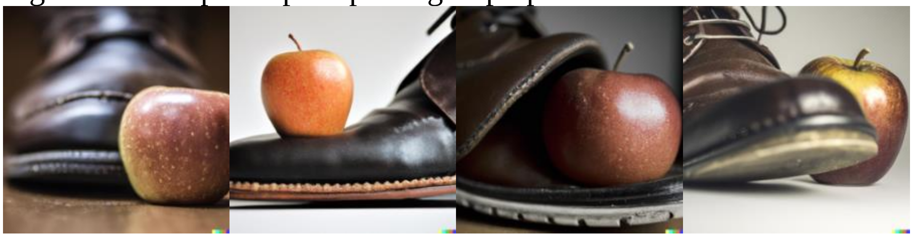
No obvious imperfections, likely immediately recognizable as representing the target

Figure 2 – Selection of successful images with imperfections

Likely recognizable as the target but strange features may draw attention away from the task

Finally, important syntactic properties of prompts were not well parsed, again more often where the target was unusual. Prepositions were often ignored (Figure 3) – as one example, a pen under paper could not be successfully generated, instead the pen would always appear on top of the paper – and frequently, the patient and theme of a sentence were seemingly randomly assigned to the nouns in the prompt. For instance, prompts requesting a dancer chasing a policeman, and vice versa, resulted in images where both figures ran, but sometimes in opposite directions, sometimes toward one another, and sometimes in the same direction, with no consistent pattern (Figure 4).

Figure 3. Example of poor parsing of prepositions



Results of prompt, "An apple underneath a leather shoe, stock photography"

Figure 4. Example of inadequate understanding of directionality.

Results of prompt, "A ballerina runs behind a policeman, high speed shutter"

Discussion

This study explored one potential application of the novel and rapidly developing technology of text-to-image generation. Overall, DALL-E 2 shows promise as a helpful tool in aphasia. Some images were rendered to a very high, photorealistic standard and could be suitable for immediate use. More often, images looked acceptable at first glance but when enlarged, imperfections became obvious. While such flaws were not predicted to affect interpretation of the target, they could be distracting to the viewer, or even provoke a sense of unease (Wang et al., 2015).

Nouns were most easily represented, followed by verbs. Sentence targets required higher specificity and these were substantially less accurate and efficient to generate. This may be due to the CLIP embedding being less successful at learning concepts as abstractness increases.

While DALL-E 2 is known for successfully producing novel combinations of concepts, it nonetheless struggled with less common requests such as green flowers. Through careful adjustments to the text prompts, these targets sometimes could be generated, suggesting that the skill and experience of the prompt generator is crucial. Some tips for speech pathologists wishing to trial DALL-E 2 are outlined in Box 1.

Box 1

Tips for enhancing results in DALL-E 2

- Expect randomness – the same prompt will produce better and worse results across multiple attempts.
- Picture the type of result you want before creating the prompt. This encourages a more specific prompt.
- Prompt as if you are captioning an existing image in a newspaper. Read stock photograph descriptions to get a feel for wording and style as DALL-E 2 was trained on image-caption pairs. Present tense seems to work best.
- Multiple clauses can be used to specify additional requirements:
    - Medium
      *A ballpoint pen lying on a desk, **stock photograph***
      *Portrait of a king wearing a golden crown, head and shoulders, **renaissance painting***
    - Source
      *An astronaut spacesuit in a museum, **tourist's photograph***
      *Photograph of a family listening to the record player, wide shot, **life magazine 1970***
    - Lighting
      *A whole green cucumber and slices, **studio lighting***
      *A croquet game on a green lawn, **warm outdoor photograph**, calm*
    - Camera attributes
      *Closeup of a wooden lattice, garden visible in background, **shallow depth of field***
      *A man bowling at a ten pin bowling alley, **action shot***
- Specify camera zoom and angle as DALL-E 2 often defaults to closeups
  ***Wide shot** of a restaurant, diners and wait staff visible*
  ***Full shot** of a man in fireman's uniform and hat, studio lighting, stock photograph*
  *Chicken schnitzel **closeup***
- Adjectives can be very effective but are not consistently applied to the correct noun.
- Duplication has been reported to be effective at focusing on a particular description and improving its quality, e.g. A smiling girl is tickled, laughing, bright lighting, happy

It was clear from results that DALL-E 2 was not able to consistently parse the syntax of the text prompts. While there was evidence that sentences were understood beyond individual words, prompts with specific prepositions resulted in a range of arrangements; e.g. “An apple underneath a leather shoe” resulted in an apple on, beside or inside a shoe at random (Figure 3). Poor accuracy of text-to-image generation has been demonstrated for a range of syntactical features, including negation, numbers, passive sentences, comparative sentences, and ambiguous sentences that require contextual reasoning (Leivada et al., 2022). Spatial relations are poorly understood, adjectives may be applied to both entities in an image (Rassin et al., 2022), and often the generators will render only the first of two objects in the sentence (Gokhale et al., 2022).

A curious strength of DALL-E 2 was its limited ability to form written words, producing visually similar imitations of text instead. This might prove useful where the impression of a sign or label is needed without providing an orthographic prompt to the person with aphasia. For example, food items or some buildings would look unusual without labels or signs.

It is also worth noting that the technology, despite not being provided with race-related prompts, generated images of humans representing diverse races. The developer, OpenAI, intentionally programmed Dall-E 2 to increase racial diversity, in response to earlier models heavily biasing white humans (Offert & Phan, 2022; OpenAI, 2022b). Nonetheless, the data used to train Dall-E 2 was predominantly English-language, Western culture images and captions (Bianchi et al., 2023), meaning generating images for other cultures may be substantially less accurate and efficient. Indeed, there is evidence that image generation technology may specifically default to American norms of visual appearance (Bianchi et al., 2023). The ability to generate images for diverse cultures for speech pathology purposes is worth investigation as it is these images that may be most difficult to source elsewhere. In addition, gender stereotypes were present; for example, the nurse was depicted as female and the butchers and soldiers were male, without exception. Speech pathologists should consider the impact of this technology ‘defaulting’ to white, Western culture and to gender stereotypes, if it is to be implemented in clinical practice (Luccioni et al., 2023; Offert & Phan, 2022). Importantly, Artificial Intelligence is not inherently biased; rather, it reflects and often exaggerates existing bias within the human-generated data used to train it (Ali et al., 2023; Bianchi et al., 2023; Luccioni et al., 2023).

Limitations of this study

This research note is an exploratory report providing an initial estimate of the utility of a new technology. As such, not all methods were operationalised. Just one person created the prompts and evaluated the suitability of images, without the use of objective criteria for ‘successful’. The extent to which the images represent the targets is therefore speculative. Formal measurement of the acceptability of images to people with aphasia, clinicians or therapy/assessment authors is needed. Future work could investigate the comparative acceptability of AI-generated and existing images within these groups as well as any impact on accuracy in assessments or treatment materials.

The limit of six prompts per target was introduced to test the efficiency of the process but this likely means a trade off on quality – more prompts would have produced a higher success rate. Equally, the author’s subjective decision to stop attempting further text prompts after deciding a particular image was ‘successful’ meant fewer chances to generate an ideal image. Improved results might also have been possible with the additional tools provided within DALL-E 2 – *variations, outpainting* and *inpainting* – which would have allowed regeneration of images, or selected parts of an image, until satisfactory. These tools were not explored, as it was hoped the technology might generate accurate images with only text prompts for maximum efficiency but should be examined in future research.

While DALL-E 2 performs well in open-ended uses such as graphic design, illustration, and art, the ability to render highly specific and atypical images is most valuable within aphasia. However, in this study, DALL-E 2 could only consistently produce flawless images depicting high frequency, concrete nouns, with more variability for verbs and sentences. This limits its current utility, as these easily generated items are also readily available elsewhere. Nonetheless, the high-quality images were produced rapidly and at vastly lower cost. Improvements could be possible very soon, given the extremely rapid advancements in text-to-image technology (Ramesh et al., 2021) and other engines may already be able to produce more accurate results. Even during this study, new models and technology have progressed. Some can be run on local hardware free of charge, which would decrease costs further. The method employed in this paper could be used to compare different generators and processes in future.

AI has many potential uses in aphasia assessment, treatment, and management (Adikari et al., 2023), but given the core role of images in aphasia and many other the ability to rapidly generate low cost, high quality images, is likely to be a major contribution going forwards. Assuming continued advancements in this technology, this process could be a preview of the future of image sourcing in aphasia and other areas of speech pathology practice.

Acknowledgments

The author would like to acknowledge the valuable contributions of Jon Hunt and Robin Keech from Cuespeak. Their support, feedback, and suggestions during the data collection and write-up stages were greatly appreciated and instrumental in shaping the final outcome.

Data Availability Statement

The data that support the findings of this study are openly available at https://johnepierce.github.io/AI-images-for-aphasia/ and in the supplementary materials.

Supplemental information

1. Full supplementary dataset – all generated images

   https://johnepierce.github.io/AI-images-for-aphasia/

2. Full set of 'successful' images

   https://labs.openai.com/sc/AZV0LiHLG3LdUJ7JnqzMmBKY

3. Unusual images

   https://labs.openai.com/sc/X3vPwJmq9G3CrK416OJRlmFm

References

Adikari, A., Hernandez, N., Rose, M. L., Alahakoon, D., & Pierce, J. E. (2023). From Concept to Practice: A scoping review of the application of AI to Aphasia diagnosis and management. *Disability and Rehabilitation*. https://doi.org/10.1080/09638288.2023.2199463

Ali, R., Tang, O. Y., Connolly, I. D., Abdulrazeq, H. A., Mirza, F. N., Lim, R. K., Johnston, B. R., Groff, M. W., Williamson, T., Svokos, K., Libby, T. J., Shin, J. H., Gokaslan, Z. L., Doberstein, C. E., Zou, J., & Asaad, W. F. (2023). *The Face of a Surgeon: An Analysis of Demographic Representation in Three Leading Artificial Intelligence Text-to-Image Generators* (p. 2023.05.24.23290463). medRxiv. https://doi.org/10.1101/2023.05.24.23290463

Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., & Caliskan, A. (2023). Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1493–1504. https://doi.org/10.1145/3593013.3594095

Brittain, B. (2023, January 17). Lawsuits accuse AI content creators of misusing copyrighted work. *Reuters*. https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/

Brown, J., & Thiessen, A. (2018). Using Images With Individuals With Aphasia: Current Research and Clinical Trends. *American Journal of Speech-Language Pathology, 27*(1S), 504–515. https://doi.org/10.1044/2017_AJSLP-16-0190

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Carter, M. (2013). Reconsidering Overlap-Based Measures for Quantitative Synthesis of Single-Subject Data: What They Tell Us and What They Don't. *Behavior Modification*, *37*(3), 378–390. https://doi.org/10.1177/0145445513476609

Dhariwal, P., & Nichol, A. (2021). *Diffusion Models Beat GANs on Image Synthesis* (arXiv:2105.05233). arXiv. http://arxiv.org/abs/2105.05233

Dunn, L. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test—Third Edition*. Pearson.

Engebretsen, K., Hartman, R., Beukelman, D., & Hux, K. (2014). The Role of Photographs in Face-to-Face Interactions involving Younger and Older Neurotypical Adults. *Perspectives on Augmentative and Alternative Communication*, *23*(1), 55–59. https://doi.org/10.1044/aac23.1.55

Fromm, D., Forbes, M., Holland, A., & MacWhinney, B. (2020). Using AphasiaBank for Discourse Assessment. *Seminars in Speech and Language*, *41*(01), 010–019. https://doi.org/10.1055/s-0039-3399499

Gokhale, T., Palangi, H., Nushi, B., Vineet, V., Horvitz, E., Kamar, E., Baral, C., & Yang, Y. (2022). *Benchmarking Spatial Relationships in Text-to-Image Generation* (arXiv:2212.10015). arXiv. http://arxiv.org/abs/2212.10015

Haahr, M. (2010). *Random.org—True Random Number Service*. http://www.random.org

Heuer, S. (2016). The influence of image characteristics on image recognition: A comparison of photographs and line drawings. *Aphasiology*, *30*(8), 943–961. https://doi.org/10.1080/02687038.2015.1081138

Hickin, J., Cruice, M., & Dipper, L. (2022). A feasibility study of a novel computer-based treatment for sentence production deficits in aphasia, delivered by a combination of clinician-led and self-managed treatment sessions. *Aphasiology*, *0*(0), 1–23. https://doi.org/10.1080/02687038.2022.2116928

Hunt, J., & Keech, R. (2017). *Cuespeak* [IOS app].

Kertesz, A. (2007). *Western Aphasia Battery (Revised)*. PsychCorp.

Leivada, E., Murphy, E., & Marcus, G. (2022). *DALL-E 2 Fails to Reliably Capture Common Syntactic Processes* (arXiv:2210.12889). arXiv. https://doi.org/10.48550/arXiv.2210.12889

Luccioni, A. S., Akiki, C., Mitchell, M., & Jernite, Y. (2023). *Stable Bias: Analyzing Societal Representations in Diffusion Models* (arXiv:2303.11408). arXiv. http://arxiv.org/abs/2303.11408

Murray, M. (2023). Generative and AI Authored Artworks and Copyright Law. *Hastings Communications and Entertainment Law Journal*, *45*(1), 27.

Nicholas, L. E., & Brookshire, R. H. (1993). A System for Quantifying the Informativeness and Efficiency of the Connected Speech of Adults With Aphasia. *Journal of Speech, Language, and Hearing Research*, *36*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

Obler, L., & Albert, M. (1979). *The Action Naming Test*. Boston: VA Medical Center.

Offert, F., & Phan, T. (2022). *A Sign That Spells: DALL-E 2, Invisual Images and The Racial Politics of Feature Space* (arXiv:2211.06323). arXiv. http://arxiv.org/abs/2211.06323

OpenAI. (2022a). *DALL·E 2* [Computer software]. https://openai.com/product/dall-e-2

OpenAI. (2022b). *Reducing bias and improving safety in DALL·E 2*. https://openai.com/blog/reducing-bias-and-improving-safety-in-dall-e-2

Pak, N. S., Bailey, K. M., Ledford, J. R., & Kaiser, A. P. (2023). Comparing Interventions With Speech-Generating Devices and Other Augmentative and Alternative Communication Modes: A Meta-Analysis. *American Journal of Speech-Language Pathology*, *32*(2), 786–802. https://doi.org/10.1044/2022_AJSLP-22-00220

Parsons, G. (2022). *The DALL·E 2 Prompt Book*. https://dallery.gallery/the-dalle-2-prompt-book/

Pierce, J. E. (2013). *Aphasia Therapy Online*. http://www.aphasiatherapyonline.com/

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision* (arXiv:2103.00020). arXiv. https://doi.org/10.48550/arXiv.2103.00020

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). *Zero-Shot Text-to-Image Generation* (arXiv:2102.12092). arXiv. http://arxiv.org/abs/2102.12092

Rassin, R., Ravfogel, S., & Goldberg, Y. (2022). *DALLE-2 is Seeing Double: Flaws in Word-to-Concept Mapping in Text2Image Models* (arXiv:2210.10606). arXiv. https://doi.org/10.48550/arXiv.2210.10606

Reymond, C., Widmer Beierlein, S., Müller, C., Reutimann, R., Kuntner, K. P., Falcon Garcia, N., Grumbinaite, I., Hemm Ode, S., Degen, M., Parrillo, F., Karlin, S., Park, S., Renner, M., & Blechschmidt, A. (2022). Naming Images in Aphasia: Effects of Graphic Representations and Photographs on Naming Performance in Persons With and Without Aphasia. *Aphasiology*, 1–23. https://doi.org/10.1080/02687038.2022.2064421

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology*, *24*, 121–133.

Swinburne, K., Porter, G., & Howard, D. (2004). *Comprehensive Aphasia Test*. Psychology Press.

Thompson, C. K. (2019). Neurocognitive Recovery of Sentence Processing in Aphasia. *Journal of Speech, Language, and Hearing Research*, *62*(11), 3947–3972. https://doi.org/10.1044/2019_JSLHR-L-RSNP-19-0219

Thomson, J. (2012). *Assessing the benefits of multimodal rehabilitation therapy for aphasia [Masters Dissertation]* (pp. 1–143). https://www.escholar.manchester.ac.uk/uk-ac-man-scw:188802

University of Western Australia School of Psychology. (1997). *MRC Psycholinguistic Database*. https://websites.psychology.uwa.edu.au/school/mrcdatabase/uwa_mrc.htm

van der Meulen, I., van de Sandt-Koenderman, W. M. E., Duivenvoorden, H. J., & Ribbers, G. M. (2010). Measuring verbal and non-verbal communication in aphasia: Reliability, validity, and sensitivity to change of the Scenario Test. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *45*(4), 424–435. https://doi.org/10.3109/13682820903111952

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, *19*, 393–407. https://doi.org/10.1037/gpr0000056