

Conversations in Content and Context: Can We Teach AI to Chat About Drawings?

David IRELAND^{a,1}, Nicole ANDREWS^b, and Dana BRADFORD^a

^a Australian E-Health Research Centre, CSIRO, Australia

^b RECOVER Injury Research Centre, University of Queensland, Australia

ORCID ID: David Ireland <https://orcid.org/0000-0003-2189-4624>, Nicole Andrews

<https://orcid.org/0000-0002-7739-1738>, Dana Bradford

<https://orcid.org/0000-0002-2939-3018>

Abstract. An innovative chatbot incorporates a drawing tool allowing users to draw pictures that symbolise the nature of their chronic pain. Rather than simply ask ‘What did you draw?’ what if the chatbot could engage the user to provide insight into symbols and images related to pain and suffering? The system was tested on images that included features of clinical interest in a first step toward a chatbot able to ‘see’ and discuss visual input. Such a chatbot has the potential for retrospective and prospective applications. *Contains childhood sexual abuse references*

Keywords. generative artificial intelligence, art therapy, child abuse, chronic pain, wellbeing

1. Introduction

The old adage, a picture is worth a thousand words, is particularly relevant in the health domain. Asking a person to draw their feelings, sensations and state of mind is clinically useful when an individual cannot put an experience into words due to age, reluctance, culture or pathology, and it takes the pressure off the individual to verbalise. This mechanism also allows the brain to access implicit (or nonverbal) memories, which are focused on emotional salience and physical sensations of an experience rather than on facts and sequences of events.

Drawing interpretation is probably best known in the field of art therapy for children. Nationally and internationally, advocates for child protection and child abuse education aim to use drawings to identify children who have been abused [1,2]. Libraries of image features indicative of abuse are being developed to assist with identifying drawings worthy of exploring or reporting. Drawings tend to be symbolic with specific colour choices. The meaning of drawings, or elements within them, taken at face value, are subject to the beliefs and biases of the interpreter (often parents, teachers, or clinicians). Recently, there has been emerging interest in using Artificial Intelligence (AI) to analyse drawings. When provided with self-figure drawings to identify those drawn by abused children, AI models did achieve high accuracy but were still outperformed by human experts [3]. Potentially, interpretations could be richer and more accurate if drawings are seen as a starting point for a conversation that allows a suitably

¹ Corresponding Author: David Ireland, d.ireland@csiro.au.

qualified individual to probe a specific line of inquiry. Identification of images for follow-up may be a more appropriate application of AI.

An interest emerged in whether generative AI that collected, recognised, and began conversations about drawings could elicit confirmation of content and potential context from which a health professional could make more confident inferences and interpretations. Dolores [4], a chatbot for people living with chronic pain, has shown potential for a digital drawing tool to gain insights into consistent symbols of pain. When the drawing tool was used, Dolores only asked a single generic question: “What did you draw?”. A missed opportunity was realised from these dialogues for the chatbot to interact further to understand the drawings better. This prompted current work to see how well a large, open-source generative AI model could recognise specific features and engage in conversations that provided content and context for features of clinical interest.

2. Methods

The system uses generative AI to commence a conversation around a drawing, beginning with a description of the drawing using natural language text and entering into a visual question-answering (VQA) interface for follow-up questions. It was tested on publicly available images drawn by children who have experienced abuse to see if the software could detect and describe features, including those known to be associated with abuse. The potential of the software to recognise features in drawings obtained from Dolores was then explored [4].

2.1. *Generative AI*

The LAVIS framework [5], a deep-learning library for Language and VISion intelligence applications, was used. This library provides a one-stop solution to rapidly develop models for multi-modal applications. The LAVIS comes with various pretrained models, and the Bootstrapping Language-Image Pre-training (BLIP-2) model off-the-shelf without any refinement or retraining was chosen. This model provides a natural language text description when given an image and a VQA component. VQA combines visual perception and language understanding, allowing a question-answering dialogue about the image to take place. The interrogator can be a human or another digital agent, like a chatbot.

The system was tested on a typical desktop machine running the Linux operating system with 32GB of RAM and an AMD Ryzen-5 2600 six-core CPU. No specialised hardware was used. The response time was approximately five seconds when an image was supplied to describe. The sample images used to run the model were obtained from two sources with ethical approval (Children’s Health Queensland Human Ethics Research Committee (HREC) HREC/21/QCHQ/73157) with reciprocal approval from CSIRO Social Science (HREC 003/24)).

2.2. *Drawings From Children Who Have Experienced Abuse*

Drawings by children who have experienced abuse were examined to test the capabilities of the system, as there is a body of literature analysing the independent features of these drawings. Sample images were taken from various literature sources. Numerous features appear significantly more often in the drawings of children who have experienced abuse

than in those who have not. For young children, this includes houses with no doors, dark clouds or rain, lack of hands and arms, and human figures with enlarged mouths, hands and/or genitalia [1,2]. If generative AI can detect these red flags, it may apply to the screening of drawings to augment early identification of abuse. The ability to identify these features retrospectively indicated a potential use case in prospective identification and consequent discussion of salient features.

2.3. Drawings From People Living With Chronic Pain

Dolores [4], a chatbot designed to provide education and conduct a pain history in adolescents and adults, allowed the user to access a drawing tool. The tool could be used to indicate areas on a body outline where the pain was located, to draw sensations associated with pain, or simply for recreation to support engagement. Only drawings captured when Dolores was conducting a pain history were used for this study, comprising questions about the description, history and sensations of the person's specific pain. Analogous to the abuse dataset, several characteristic features were observed in the descriptions of pain, including fires to show heat and volcanoes to indicate pain flare-ups. The colour red was often chosen above other colours [4, 6], consistent with previous studies, particularly those conducted with minority groups with limited English proficiency [6]. It is postulated that red is a colour used universally to denote pain due to its similarity in hue to both inflammation and blood.

During the Dolores study [4], the chatbot could not *view* the image and only asked a generic question, "What did you draw?". The capacity of the chatbot was assessed by using the VQA component to determine if the system recognised features of clinical interest. This recognition is the first step towards engaging in a deeper interaction around the drawings, their pertinence to the individual and their suitability for inclusion in the broader body of work around the symbology of pain.

3. Results

The early version of the system was able to describe features of the drawings from children, adolescents and adults with varying levels of accuracy.

3.1. Identification Of Elements Is Generally Consistent With Drawing Features

In three drawings by children (approximately eight years of age) who were the victims of sexual abuse, (Figure 1a, c [2] and b [7]), red flags indicative of abuse [3] were depicted. For each of these drawings, the model was able to recognise specific features identifying the house (Figure 1a), a girl (Figure 1b), and a girl with a sports toy (Figure 1c). Using the VQA component, the model was then probed to provide further information on content, including questions on features indicative of abuse worthy of further exploration (Table 1). Questions included asking if the house has doors, if there is smoke coming from the chimney, or if there is rain. For the figures, pertinent questions included whether there was a presence of arms, genitalia, and/or tears.

The initial descriptions, while consistent with drawing features, were not necessarily an accurate interpretation. The drawing of the girl (Figure 1b) may have been intended to show rosy cheeks rather than a clown face, and it is likely that the toy (Figure 1c) is a skipping rope rather than a hula hoop. Nonetheless, it was able to describe features that,

used retrospectively, could flag drawings for the attention of a professional (for example, Q&A in bold italics in Table 1). The VQA model has correctly answered that a door is not present in Figure 1a, but a chimney with smoke is. It was correct to say that rain and storm clouds were absent. The dialogue for Figure 1b seems to identify the asymmetry of the face and the lack of arms. It recognised some words and provided the correct spelling (“Good morning”).

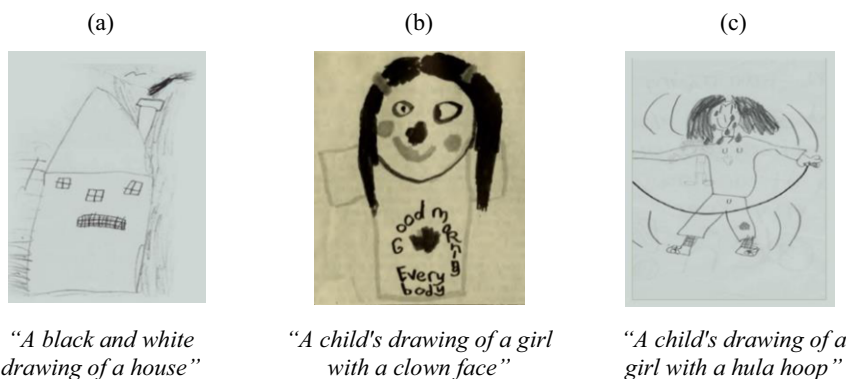


Figure 1. Generated text captions of drawings by children who were the victims of sexual abuse [2,7].

Table 1. Visual question and answer dialogue from the drawings in Figure 1 (a) and (b).

	Figure 1a: House	Figure 1b: Girl
PROMPT:	<i>Does the house have a door?</i>	What is wrong with the girl?
LAVIS:	<i>No</i>	It's her face
PROMPT:	Is there a chimney?	<i>Does the girl have arms?</i>
LAVIS:	Yes	<i>No</i>
PROMPT:	<i>Does the chimney have smoke?</i>	Are there any letters
LAVIS:	<i>Yes</i>	Yes
PROMPT:	Is it raining?	What do the letters say
LAVIS:	No	Good morning
PROMPT:	Are there storm clouds?	Do the eyes look normal
LAVIS:	No	No

3.2. The System Shows Promise For Prospective Applications

While the use case for the drawings completed by children who had experienced abuse was envisaged to be a retrospective analysis by a health professional, the use case for the pain drawings was designed to be prospective – that is, opening a conversation with the user about what the drawing represented and how they reflected the physical experiences of the person with chronic pain. This was important as symbolism was commonly used; hence, it was necessary to ascertain if the system could detect features salient to pain.

The model was first provided with drawings from the Dolores study to generate text descriptions (Figure 2). All were accurate descriptions of the images, which ranged from various shapes (Figure 2a) to erupting volcanos (Figure 2b) to anatomical drawings (Figure 2c). A dialogue with LAVIS about features of the drawings was entered to see if it could correctly determine colours, shapes and anatomical anomalies. While the LAVIS framework was able to answer questions about the colours correctly, it could not

identify the swirl shapes in Figure 2a, referring to them as rectangles and circles (Table 2). Participant discussions indicated that the spirals reflected cyclic pain, and the jagged lines represented instant pain sensations.

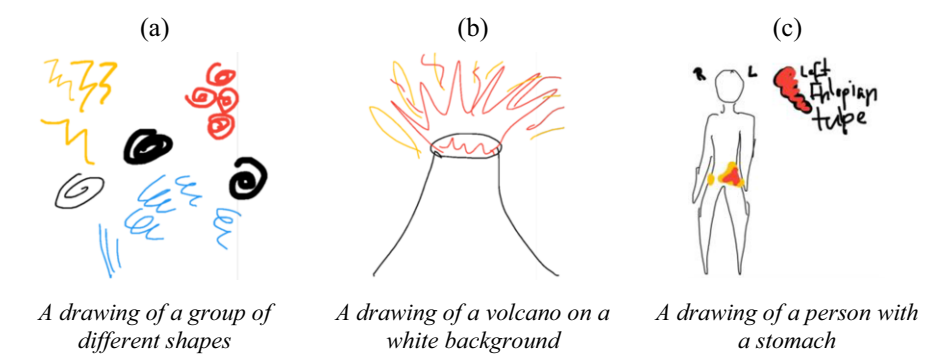


Figure 2. Generated text captions for each drawing captured during the Dolores chatbot pilot when the user was asked to describe their pain [4,6].

The system detected anatomical drawings (Figure 2c, Table 2) and that the participant was referring to the abdomen. The system also detected the use of red and yellow colours and associated them with ‘fire’, a common symbol for pain.

Table 2. Visual question and answer dialogue from the drawings in Figure 2 1 (a) and (c).

	Figure 2a: Shapes	Figure 2c: Anatomy
PROMPT:	What are the colours?	What are the colours?
LAVIS:	Black, red, yellow and white	Black, red, yellow and white
PROMPT:	What are the different shapes?	What's wrong with the stomach?
LAVIS:	Circles and rectangles	It's too big
PROMPT:	Are there any characters?	What's wrong with the body?
LAVIS:	Yes	Fire
PROMPT:	What are the characters?	Is the stomach red?
LAVIS:	Letters and numbers	Yes
PROMPT:	What letters are in the drawing?	Is the stomach yellow?
LAVIS:	Ey	Yes
PROMPT:	What numbers are in the drawing	What's wrong with the person?
LAVIS:	7 7	He's smoking

4. Discussion

The VQA dialogues, while providing descriptive information, had evident inconsistencies, suggesting that information obtained from this feature has to be considered carefully. Nonetheless, the LAVIS system offers promise as a practical solution for incorporating a visual intelligence system into the analysis of drawings obtained by or uploaded to the chatbot.

In the retrospective use case, it is expected that questions typically applied to clinical drawings will be programmed into the chatbot brain to be used in screening submitted

drawings. This will allow the AI to identify images requiring further investigation. In the prospective use case, an internal dialogue between two AI systems is envisioned, where the chatbot has an internal dialogue with the image intelligence system, deriving information it then provides to the user to engage in conversation on content. Equipping a chatbot with this feature allows it to demonstrate a degree of insight and offers the ability to ask relevant follow-up questions to solicit further information. With this ability, there is enormous potential for multi-modal interactions between humans and chatbots.

This research to date has explored content related to trauma – abuse or pain, however, a more positive application of the drawing tool conversation could be in children’s wellbeing. In this application, developing a library of symbols associated with wellbeing could contribute to the design of age and culture-appropriate wellbeing measures for children.

5. Conclusions

The ultimate goal is to provide an interactive system, such as a chatbot, that offers both a drawing modality and a conversation interface. Chatbot conversations can contribute to a library of symbols and images related to pain and suffering that will provide greater insight into the experience of non-verbal people with chronic pain.

References

- [1] McInnes E. Young children’s drawings after sexual abuse: disclosure and recovery. In: McInnes E, Mason AD, editors. *Where to from here? Examining conflict-related and relational interaction trauma*. Koninklijke: Brill | Rodopi; 2019. p. 47-69, doi: 10.1163/9789004397576_005.
- [2] Neubauer S, Mahan-Strupp R, Davis I, Neubauer E, Neubauer R. Identifying sexually abused children through their art. Child Abuse Education Foundation of WI, Inc. Available from: <https://www.insafehands.net.au/sites/default/files/u32/Identifying%20Sexually%20Abused%20Children%20Though%20Their%20Art.pdf>.
- [3] Kissos L, Goldner L, Butman M, Eliyahu N, Lev-Wiesel R. Can artificial intelligence achieve human-level performance? A pilot study of childhood sexual abuse detection in self-figure drawings. *Child Abuse Neglect*. 2020 Nov;104755, doi: 10.1016/j.chiabu.2020.104755.
- [4] Ireland D, Vijayakumar P, Andrews N. Dolores: a mobile chatbot for people living with chronic pain, studies in health technology and informatics. *Stud Health Technol Inform*. 2024 Jan;310:564-568, doi: 10.3233/SHTI231028.
- [5] Li D, Li J, Le H, Wang G, Savarese S, Hoi S. LAVIS: A one-stop library for language-vision intelligence. In: Bollegala D, Huang R, Ritter A, editors. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*; 2023 Jul 10-12; Toronto, CAN. ACL; c2023. p. 31-41, doi: 10.18653/v1/2023.acl-demo.3.
- [6] Andrews N, Ireland D, Vijayakumar P, Burvill L, Hay E, Westerman D, Rose T, Schlumpf M, Strong J, Claus A. Acceptability of a pain history assessment and education chatbot (Dolores) across age groups in populations with chronic pain: development and pilot testing. *JMIR Form Res*. 2023 Oct;7:e47267, doi: 10.2196/47267.
- [7] Malchiodi C. *Breaking the silence: art therapy with children from violent homes*. New York: Brunner/Mazel; 1990. 230 p, ISBN: 978-0876305782.