論文とETIC 18

ケアの写真を、AIはどう読むか

馬介在学習の写真にみる活動分類の性能と、ケア評価に必要な検証



Daiki Morimoto / Unknown Lab
ORCID: 0009-0009-6444-750X
Version 1.0 / 公開日 2026-11-09
DOI: 10.5281/zenodo.24074799
未査読 / 専門職としての著者確認済み
独立専門レビュー未実施 / AI支援執筆
ライセンス: CC BY 4.0

30秒で読む

Demaらの研究が示したのは、馬介在学習プログラムYawardani Jan-gaのセッション写真について、既存の事前学習済みモデルに少数の参照例を組み合わせた構成が、写っている活動の種類と、子どもと馬の身体接触の有無を、専門家のラベルと約9割一致して分類できたことである[1]。写真は6〜14歳の子ども10人のセッションから得たもので、9種類の活動分類では試験画像374枚で正解率90.6%、接触の有無では264枚で88.6%であった。この結果から、セッション記録を活動や接触で整理する用途は検討に値する候補となる。姿勢の分類は193枚に正解ラベルがなく、出力の一貫性のみが評価された。実務上の利益、心理的な意味、臨床的な転帰は検証の対象外である。

一枚の写真から

仮の写真を一枚想定する。屋外の囲いで、子どもがブラシを手に馬の首筋を整えている。この写真に「手入れ（グルーミング）」「馬に触れている」と書き添えるのは、写っているものの記述である。この子どもが前回より落ち着いたのか、馬との関係に手応えを感じ始めたのかは、写真の外にある経緯と本人の語りがなければ決められない。

本稿は、AIがケアの写真をどこまで使える記録に変えられるか、そして意味ある変化の評価にどこまで近づけるかを問う。素材は二つある。一つは、西オーストラリア州キンバリー地域のアボリジナルの子ども・若者を対象とする馬介在学習（EAL）プログラム「Yawardani Jan-ga」の画像を視覚言語モデルで解析したDemaらの研究である[1]。もう一つは、同プログラムの評価計画を示したCoffinらのプロトコルである[2]。先に答えを示す。観察できる活動と接触の分類は、一つのデータセットの中で専門家ラベルと9割前後一致した。記録を整理する手段の候補にはなりうる。姿勢から自信を読む部分と、変化の意味を評価する部分は未検証のままである。

研究の場と資料

主論文は、DemaらがJournal of Medical Internet Researchに2026年10月30日に発表した開発・評価研究である[1]。投稿は2026年3月23日、改訂稿の受領は7月11日、採択は7月14日で、CC BY 4.0で公開されている。著者順は出版社の最終版の表記に従った。採択前のプレプリントは同じ原稿であり、別の研究として数えない。

研究の場は、西オーストラリア州キンバリー地方のアボリジナル主導プログラム「Yawardani Jan-ga（馬が助ける）」である。馬介在学習（EAL）は、馬の手入れや引き馬などの活動と、その後の振り返りを組み合わせる学習法である。社会的・情緒的ウェルビーイング（SEWB）を支えることを目的とする[1,2]。プログラムの対象は6〜25歳だが、分析した写真は10人の子ども（男子5人・6〜12歳、女子5人・6〜14歳）のセッションから得られた。撮影の期間や暦日は報告されていない。

画像の集合は課題ごとに作られた。活動分類では、2人の子どもの写真から各活動4枚、計36枚の見本を選んだ。残る464枚は、検証用90枚（3人）と評価用374枚（5人）に分けた。子どもは集合をまたがない。ただし馬は頭数が限られるため、複数の集合に写っている。評価用の内訳は、ボール遊び52、手入れ50、群れの観察40、引き馬48、障害物つき引き馬17、騎乗35、ペイント54、素手で触れる47、洗う31で、区分によって枚数が異なる。接触判定では、268枚（接触あり108、なし160）から各区分2枚を見本に回し、残る264枚で評価した。姿勢の課題では、姿勢が見て取れる193枚を選んだ。課題間で写真が重なりうるため、枚数の合計は独立な写真数にも人数にもならない。

活動と接触の正解は、専門家パネルの2人が決めた。2人が全画像に独立してラベルを付け、意見が分かれた画像は協議で解消した。偶然の一致を差し引いた一致度であるコーエンのκは、活動0.938、接触0.955だった。この値はラベル付けの再現性を示すが、ラベルが臨床的・文化的に妥当かどうかまでは示さない。

表1 課題ごとの評価資料

課題	評価に使った画像	正解の作り方	測られたこと
活動分類（9区分）	評価374枚（見本36、検証90は除く）	専門家2人の独立評定と協議（κ=0.938）	活動名の正しさ
接触判定（2区分）	評価264枚（見本4は除く）	同上（κ=0.955）	接触の有無の正しさ
姿勢分類（3区分）	姿勢が明瞭な193枚	正解なし	繰り返したときの出力の揃い方

出典：[1]のDataset節から作成。解釈：三課題の画像は重なりうるため、合計を標本数として扱えない。

資料の範囲について付記する。本稿は、出版社のHTML全文（表1〜6、方法、考察、参考文献）に基づく。掲載された5つの数式画像（融合得点、全一致率、平均一致度、根拠の一貫性、投票エントロピー）は今回直接確認し、本文の変数説明と照合した。図1〜8の画像は確認しておらず、本文とキャプションを参照した。図にしかない数値は追加しておらず、区分別の数値は本文に記された範囲に限った。

この研究が加えたもの

EALの評価はこれまで、実践者の観察記録、チェックリスト、面接、質問紙に頼ってきた。著者らはその課題として、時間がかかること、観察者によるぶれ、標準化の難しさを挙げる[1]。心拍やコルチゾールの測定も考えられるが、機器が子どもと馬のやりとりを妨げうる。一方で、写真は振り返りや記録のためにすでに撮られている。

AIで馬を観察する先行研究は、厩舎での歩行や採食など、馬だけの行動が中心だった。この研究は、子どもと馬の関わりを読み取ろうとした点で新しい。既存の視覚言語モデルに説明文と少数の見本を与えるだけで、追加の学習（ファインチューニング）は報告されていない。枠組みは、プログラムの実践者や研究者と共同で設計されたという。モデルはローカルで動

く公開モデルから選ばれた。理由として、通信が限られる遠隔地での運用と、データの保護が挙げられている。

三つの仕組み

活動分類：説明文と見本の平均像を混ぜる

活動分類にはSigLIP2（google/siglip2-giant-opt-patch16-384）を使う[1,5]。このモデルは画像と文章を数値の並び（ベクトル）に変換する。意味が近いものほど、ベクトルの向きも近くなるよう作られている。向きの近さはコサイン類似度で測る。

処理は準備と分類の二段階に分かれる。準備段階では、9活動の説明文をそれぞれベクトルにする。各活動の見本写真もベクトルにして、活動ごとに平均する。この平均が視覚的な原型（プロトタイプ）である。分類段階では、判定する写真のベクトルを、各活動の説明文と原型の両方と比べる。説明文との近さと原型との近さを重みαで混ぜた得点を活動ごとに求め、最も得点の高い活動を答えとする。αが1なら説明文だけ、0なら見本だけで判定することになる。

nショットは、各活動に与える見本の枚数を指す。0ショットは説明文だけで判定する設定である。見本を増やしても、モデル自体を学習し直すことはなく、参加者が増えるわけでもない。

αは検証用90枚を使って決めた。0から1まで0.1刻みに試し、マクロF1が最大になる値を選んでいる。2ショットでは、SigLIP2がα=0.9、比較対象のCLIP-ViT-L-14が0.8だった。説明文は紛らわしい活動を書き分けている。手入れは道具を使うもの、触れるは素手で道具を使わないもの、引き馬はコーンや障害物がないもの、と定義された。α=0.9という値は、この設定で判定の大部分を説明文が担っていたことを示す。

接触判定：見本を会話として見せる

接触判定には、生成型の視覚言語モデルQwen2.5-VL-7B-Instructを使う[1,6]。著者らの予備実験では、類似度で判定するSigLIPやCLIPは、「触れているか」という空間関係を安定して区別できなかった。

そこで次の形式をとる。見本写真と正しい答えを、利用者とアシスタントの問答として並べる。その後に、判定する写真と指示を置く。モデルは文章で答え、プログラムがそこから接触あり・なしを取り出す。手の数や距離の目安も抜き出すが、評価の対象は接触の有無だけである。

比較にはLLaVA-OneVision-7Bを使った。条件は同じプロンプト、温度0.01（出力のばらつきを抑える設定）、最大512トークンである。入力画像の大きさは、QwenとLLaVAが1,024ピクセル、活動分類のSigLIP2が384ピクセル、CLIPが224ピクセルである。

姿勢分類：記述してから決め、5通り走らせる

姿勢分類にもQwen2.5-VL-7Bを使う。手法の土台は、答えの前に推論の過程を書かせる思考の連鎖（CoT）である。これを画像向けに二段階にしたものをMCoTと呼ぶ。

第1段階では、頭の位置、肩の並び、背骨の曲がり、立ち方を、ラベルを付けずに記述させる。温度は0.7で、記述にばらつきを持たせる設定である。第2段階では、画像と第1段階の記述をもとに、「自信がある / 自信がない / 無関心」の一つを選ばせる。温度は0.1である。さらに自己一貫性（SC）として、この二段階を1枚につき5通り（K=5）走らせ、多数決で最終ラベルを決める。

三つのラベルはモデルの出力区分の名前にすぎず、子どもの気持ちを測った値には当たらない。著者らは「組み合わせ確信度」も提案している。これは、多数派の得票割合に、5つの記述どうしの意味の近さを掛けた値である。意味の近さは、文章をベクトルにするSentenceBERTでのコサイン類似度で測る。この値は、人が見直すべき出力を拾う目安として示された。実際の正答率に合わせて較正された確率ではなく、正しさとの関係も検証されていない。実装は請求に応じて提供される形で、自由に入手できる形では公開されていない。本稿では、独立した再現の報告を確認していない。

どう確かめたか

見本写真は、実践者と相談して各区分の典型例を選んだ。評価用データでの成績を見て選んだものではないと報告されている。

活動と接触では、正解と照らして四つの指標を求めた。正解率は、全画像のうち当たった割合である。再現率は、ある区分の写真のうち、正しくその区分と判定された割合をいう。適合率は、その区分と判定したもののうち、実際にその区分だった割合をいう。F1は両者を一つにまとめた値で、マクロF1は区分ごとのF1を等しい重みで平均した値である。そのため、17枚しかない障害物つき引き馬の出来も、54枚あるペイントと同じ重さで反映される。

モデル間の差は、マクネマー検定で調べた。同じ写真に対する二つのモデルの当たり外れの組を比べる検定である。ただし、同じ子どもの写真が複数含まれるため、写真が互いに独立という前提は部分的に損なわれる可能性がある。著者らもこれを限界に挙げ、報告された有意性に影響しうるとしている。p値は写真単位の比較であり、子ども単位の検証にはならない。

姿勢では三つの条件を比べた。1回で推論と分類をするCoT、二段階のMCoT、それに多数決を加えたMCoT+SCである。各条件を独立に5回実験し、その平均と標準偏差を報告している。外側の5回の実験反復と、MCoT+SCの内側で1枚ごとに走らせる5経路は、数が同じでも役割が異なる。どちらも5つの独立したモデルを意味しない。

姿勢の指標は四つある。平均一致度は、各画像で多数派ラベルに一致した実行の割合を、全画像で平均したものである。全一致率は、すべての実行が同じラベルを出した画像の割合である。投票エントロピーはラベルの散らばりを表し、0なら完全に一致する。根拠の一貫性は、記述どうしの意味の近さの平均である。どれも安定性の指標であり、正しさは測らない。たとえば、5回とも同じ誤ったラベルを出した画像も、一致として数えられる。また、評価式のNと、各条件（CoT、MCoT、MCoT+SC）の反復手順との対応は詳しく記されておらず、同じ計算を再現するうえでの限界になる。

結果

活動分類

2ショットのSigLIP2は、評価用374枚で正解率90.6%、マクロF1 91.1%だった。CLIPは正解率62.0%で、差は28.6ポイントある。説明文だけの0ショットでも、SigLIP2は89.3%に達した。見本2枚による上積みは1.3ポイント（マクロF1で1.5ポイント）である。マクロF1が最も高かったのは4ショットの91.2%だった。2ショットは、同じ正解率をより少ない見本で得られるとして採用された。

表2 活動分類の成績（評価用374枚、%）

見本数	SigLIP2 正解率 / マクロF1	CLIP 正解率 / マクロF1
0	89.3 / 89.6	57.5 / 52.5
1	90.4 / 90.7	61.8 / 60.6
2（採用）	90.6 / 91.1	62.0 / 61.6
3	89.8 / 90.0	62.3 / 61.3
4	90.6 / 91.2	61.5 / 60.9

出典：[1]の表1から作成。解釈：見本数による変化は小さく、見本が多いほど成績が上がる関係も見られない。成績差の大半はモデルの選択に伴って生じている。なお、αの重みは検証用データで探索されたが、説明文の書き方や見本の選び方を変えて各要素の寄与を切り分けた実験は報告されていない。

マクネマー検定では、全設定でp<.001だった。2ショットでは、SigLIP2だけが当てた写真が124枚、CLIPだけが当てた写真が17枚である（$\chi^2=79.69$）。本文によると、SigLIP2の区分別の正答割合は0.74〜1.00で、ボール遊びと洗う作業はすべて正解だった。CLIPは引き馬で0.00と大きく崩れた。

接触判定

Qwen2.5-VLは評価用264枚で、1ショット時に正解率88.6%、マクロF1 88.5%だった。0ショットの87.5%から改善した分は、主に「接触あり」の取りこぼしが減ったことによる。接触ありの再現率は、72%から85%に上がった（2ショットでは86%）。

LLaVA-OneVisionは、見本を増やすと成績が下がった。正解率は0ショットの80.3%から、1ショットで68.2%、2ショットで65.5%に落ちた。接触ありの再現率も、59.4%、26.4%、19.0%と下がった。一方、接触なしの再現率は94.3%、96%、97%と高いままだった。見本を見せるほど、答えが枚数の多い「接触なし」へ偏ったことになる。

表3 接触判定の成績（評価用264枚、%）

見本数	Qwen2.5-VL 正解率 / マクロF1 / 接触あり再現率	LLaVA-OneVision 正解率 / マクロF1 / 接触あり再現率
0	87.5 / 86.4 / 72.0	80.3 / 78.3 / 59.4
1（採用）	88.6 / 88.5 / 85.0	68.2 / 59.1 / 26.4
2	87.9 / 87.3 / 86.0	65.5 / 53.9 / 19.0

出典：[1]の表3・表5から作成。解釈：見本の効果の向きは、モデルによって異なった。LLaVAの低下はこのモデルとこの課題での観察であり、少数見本の与え方一般に広げられる結果ではない。

両モデルの差のマクネマー検定は、0ショットでp=.01、1・2ショットでp<.001だった。ここでも、写真の独立性は保証されない。

姿勢分類

表4 姿勢分類の出力の安定性（193枚、5回の実験反復の平均（標準偏差））

指標	CoT	MCoT	MCoT+SC
平均一致度	0.675（0.006）	0.853（0.004）	0.916（0.010）
全一致率	0.104（0.004）	0.528（0.019）	0.705（0.024）
投票エントロピー（低いほど一致）	0.653（0.009）	0.300（0.010）	0.174（0.027）
根拠の一貫性	0.913（0.001）	0.931（0.001）	0.857（0.002）

出典：[1]の表6から作成。解釈：いずれの値も正解との照合を含まない。標準偏差は反復間のばらつきを表す。信頼区間は報告されていない。

記述の段階を挟み、多数決を加えるほど、ラベルは揃った。すべての実行が一致した画像は、CoTでは約1割、MCoT+SCでは約7割である。ただし、0.916は正解率に相当する値ではない。70.5%も、支援の効果を表す値ではない。

一方、根拠の一貫性はMCoT+SCで0.857に下がった。ラベルが揃うにつれ、そこに至る記述どうしは似なくなっている。著者らは、第1段階の高い温度が記述を多様にした可能性を挙げる。記述が多様なこと自体は欠点とは限らず、同じ姿勢の異なる側面を書いているのかもしれない。それでも、異なる記述が同じラベルに集まる理由は、正解のないこの表からは決められない。記述が判断の根拠として働いたのか、記述とあまり関係なくラベルが決まったのかは分からない。それを見分けるには、記述とラベルの対応を、合意した基準や独立した評定に照らして調べる追加の検証が必要になる。

結果をどう読むか

著者らは、少数の例を示す学習法を使えば、ラベル付きデータが限られていてもEALで観察できる活動と空間的な関わりを分類できる可能性があり、実践者が主導する記録を補えると結論した[1]。用途としては、活動の種類ごとにセッション記録を整理することを想定している。同時に、この結果を治療的進歩や臨床アウトカムを評価できる証拠と解釈すべきではないと明記し、実践者の評価、参加者の目標、標準化されたウェルビーイング・治療アウトカム指標との照合を今後の条件に挙げた。実践者の負担軽減も期待として述べているが、時間や労力は測定していない。

数値から読める技術上の要点は、既存の公開モデルに活動の説明文を与えるだけで、分類の大半が成り立ったことである。SigLIP2は例を与えない0-shotで89.3%に達しており、各活動2枚の例を加えた上昇は1.3ポイントにとどまった。検証データで選ばれたテキスト重みα=0.9も、説明文との照合が判定の主軸だったことを示す。Qwen2.5-VL-7Bでは、接触ありの画像を正しく拾えた割合（再現率）が、参照例なしの72.0%から、各カテゴリに参照例を1枚ずつ示した条件で85.0%となった。LLaVA-OneVisionでは例を足すほど同じ再現率が59.4%から19.0%へ下がった。例示の効果は、モデルによって上がる場合も下がる場合もある。姿勢の一致率0.705は、反復実行でラベルが揃った画像の割合を示す。「自信あり」という判定が正しいかどうかは、別に作る正解ラベルがなければ分からない。

解釈の範囲を決める主な制約は次のとおりである。

- 画像は子ども10人、一つの地域・プログラムに由来する。子どもはデータの分割間で重複しないが、馬の頭数が限られるため同じ馬が複数の分割に登場する。
- 姿勢の193枚は姿勢が明瞭な画像として選ばれたもので、正解ラベルがない。
- 屋外の複雑な背景や、体が隠れることは、成績を下げうる。
- 静止画のため、姿勢の移り変わりや関わりの展開を捉えられない。

- 1枚に1ラベルを付けるため、同時に起きる活動や移行の途中を表現できない。
- 馬の反応はごく一部しか扱っていない。また、較正されていない単一画像から距離を推定するのは難しい。
- 別の施設やデータでの外部検証は行われていない。プロンプトの言い回しを変えたときの頑健性も体系的には試されていない。
- 姿勢モジュールは1画像につき5回推論するため計算負荷が大きく、設備や通信の限られた地域で運用する際の制約になりうる。
- 大規模な一般データで学習したモデルが、身体言語について特定の文化の前提を持ち込むおそれがある。
- 記録時間、費用、業務量、社会的・情緒的ウェルビーイング（SEWB）は測定されていない。

本研究はAIを使う記録と使わない記録を比べる設計をとっていない。そのため、AIによる記録が負担を減らすとも、子どもの状態を改善するとも、因果的には主張できない。

姿勢の意味は誰が決めるのか

Yawardani Jan-gaの研究計画（プロトコル）[2]は、2024年12月に公表された。実施と評価の計画を示した文書であり、結果は含まない。それでも、この問いに対する計画上の答えを示している。

プログラムは、地域の長老（エルダー）、若者、実践者と共同で設計された。統括するのは、五つの文化圏の長老を含む文化諮問グループである。若者のフィードバックグループや、地域のサービス機関も解釈に関わる。

評価は質的な方法が中心である。記述的現象学（Colaizziの方法）で分析し、少なくとも2人のアボリジナルの研究者がコード化を担う予定である。参加者、保護者、紹介元、実践者の語りを、写真、動画、セッション記録

と突き合わせる。さらに、参加者が使う言葉と行動の表れを用語集にまとめる。これをもとに、SEWBをとらえる評価手段を地域と相談して作る計画である。

写真については、次の手順が計画されている。写真と動画から重要な関わりを特定し、それを数えるコード化の枠組みを試す。その厳密さと効率を高めるため、人工知能や機械学習の利用を探る。主論文は、この流れの技術的な一歩と読める。同時に計画は、SEWBの変化をどう読むかを、言葉や身振りの対応表で決めようとはしていない。共同設計、解釈、統治の問題として扱っている。計画には、実践者30人を15日間・120時間で養成するといった規模も記されている。ただし、これは予定であり、実際の費用や負担の測定値ではない。

このプロトコルは、写真に写った場面が参加者にとって何を意味したかを評価する手がかりを、本人、家族や養育者、紹介元の語りと、地域との協議に置いている。質的分析は初期コホートでは帰納的に始め、後続コホートでは帰納と演繹を組み合わせて比較を重ねる計画であり、社会的・情緒的ウェルビーイングを捉える尺度や、馬との相互作用を記録する方式も、写真・映像の分析と協議を通じて形成され、見直されていくものとして示されている。Unknown Labは、次に検証すべき点を、AIの活動ラベルがこうした語りと記録の中から関連する場面を探し出す助けになるかどうかにあると解釈する。

ETIC 2.0からの論考

ETIC 2.0は、意味のある活動や参加を、人・環境・技術・データの関係として捉えるためにUnknown Labが提案している概念的枠組みであり、臨床的に検証されたものではない[3]。この枠組みから本研究を読むと、活動ラベルの使い道として考えやすいのは、記録の中から特定の場面を探し出す索引である。

索引を使う際の起点は、参加者本人の目標に置かれる。紹介時の懸念は支援の理由にはなっても、そのまま本人の目標になるとは限らない。架空の例として、ある子どもが自分から馬の世話に加わりたいと望んでいるとす

る。手入れや洗う活動のラベルは、この目標に関わりうる場面を記録の中から見つける手がかりになる。ただし、その活動を誰が始めたのか、本人にとって何を意味したのかは、写真に添えられた記録と、本人や周囲の語りによって確かめる必要がある。また、撮影も選択も標準化されていない写真では、ラベルの付いた枚数から活動の頻度やその増加は推定できない。経時的な比較には撮影と選択の条件をそろえることが前提となり、馬との接触写真が増えたことを社会的関係の改善と読むこともできない。

そこでUnknown Labは、同じセッションについて、通常の記録（メモと写真）と、AIの活動・接触ラベルを加えた記録を比べることを提案する。評価するのは、関連する場面を探し出す正確さ、見落とした場面、ラベルの確認と修正にかかる時間である。比較に先立って、何を関連する場面とみなすかの基準を参加者と実践者とともに定めておく必要がある。語りも記憶や解釈を含むため、誤りのない正解として扱うことはできないからである。この比較は未検証の提案であり、臨床での導入を勧めるものではない。

次に読む

Wangらは、作業療法学生の面接をChatGPTを用いた採点法で評価した[4]。対象は、台湾北部の大学で学生が模擬脳卒中患者に行った8分間の主訴面接で、その逐語記録58件である。この研究は、同じ記録を繰り返し採点したときの再現性と、専門家の合議による評定との一致を、別々に検討している。安定しているかと基準に合っているかを分けて測る設計であり、再現性だけを測った本稿の姿勢分類と対照的である。参照できたのはPubMedの書誌と抄録のみで、本文、モデルの版、プロンプトは確認していない。

結論

AIはケアの写真をどう読むか。Demaらの研究の条件の下では、既存のモデルは写真に写る活動の種類と、子どもと馬の接触の有無を、専門家のラベルとおおむね9割一致して分類できた。活動分類は参照例なしでもすでに高い水準にあり、少数の参照例による上積みは小さい。セッション記録を活動や接触で整理する用途は、具体的に検討する価値がある。一方、測られたのは専門家ラベルとの一致であり、記録作業の負担、参加者にとっての意味、臨床的な転帰は測られていない。姿勢の分類は正解ラベルとの照合がまだ行われておらず、文化的知識に基づくカテゴリーの定義と基準による検証を経るまで、その可否は判断できない。Unknown Labは次の段階として、同じセッションの通常の記録とAIラベルを加えた記録を比べ、関連場面を探す正確さ、見落とし、確認と修正の時間を評価することを提案する。この比較はまだ行われていない。

補足

表5 運用と根拠の範囲

論点	報告されたこと	報告されていないこと
計算環境	RTX 4090（24GB VRAM）、Core i9-14900F、32GB RAMの1台で公開モデルをローカル実行	処理時間、費用、職員の作業時間
データとコード	データは非公開で、合理的な請求に対し、関連する倫理承認、アボリジナル主導の統治手続き、適切なアクセス合意を条件に提供。推論スクリプト、プロンプトのテンプレート、データ処理パイプラインは、合理的な請求に応じて著者から入手可能	自由にダウンロードできるデータセットとコード
基盤モデル	SigLIP2とQwen2.5-VL-7Bのモデルカードの表示はApache 2.0（2026-11-09確認）	子どもの画像に使ってよいかの許可
倫理と資金	西オーストラリア州アボリジナル保健倫理委員会（WAAHEC、HREC926）承認。元の収集時の同意・アセントと、識別情報を除いたデータの二次利用。Healthway助成（G-202509-88731）	新たな地域での検証、臨床での導入

出典：[1,5,6]から作成。解釈：モデルの重みの利用許諾は、子どもの写真を扱う同意や地域の統治に代わるものではない。本表は法的助言ではない。

著者らは利益相反なしと申告している。また、原稿の言語編集と校正にClaude（Anthropic）を使ったことを開示している。

参考文献

1. Dema KD, Coffin J, Vaz S, Wang G, Olsson C. Multimodal Analysis of Child-Horse Therapeutic Interactions Using Vision-Language Models in an Australian Aboriginal Equine-Assisted Learning Program for Social and Emotional Well-Being: Development and Evaluation Study. J Med Internet Res. 2026;28:e95887. https://doi.org/10.2196/95887（ https://www.jmir.org/2026/1/e95887/ 、2026-11-09参照）

2. Coffin J, Vaz S, Kickett-Tucker C, Milroy H, Olsson C, Kirby M, et al. Protocol to implement and evaluate a culturally secure, strength-based, equine-assisted learning program, "Yawardani Jan-ga" (horses helping), to support the social and emotional wellbeing of Australian aboriginal children and young people. PLoS ONE. 2024;19(12):e0312389. https://doi.org/10.1371/journal.pone.0312389

3. Morimoto D. エコテクノロジカル・インテグラティブ・ケア(ETIC)2.0, Version 2.1改訂版 - AI時代の作業療法とケア生態系設計. Unknown Lab; 2026. https://doi.org/10.5281/zenodo.21848456

4. Wang YC, Chen AJ, How MJ, Yang CW, Lee YC, Huang SL, Hsieh CL. Development and Validation of a ChatGPT-Based Assessment for Conversation-Related Clinical Skills in Occupational Therapy OSCEs. American Journal of Occupational Therapy. 2026 (online 2026-11-03). https://doi.org/10.1177/19437676261500028 PMID: 42974814（PubMedの書誌・抄録のみ参照）

5. google. siglip2-giant-opt-patch16-384 (model card). Hugging Face. https://huggingface.co/google/siglip2-giant-opt-patch16-384（2026-11-09参照）

6. Qwen. Qwen2.5-VL-7B-Instruct (model card). Hugging Face. https://huggingface.co/Qwen/Qwen2.5-VL-7B-Instruct（2026-11-09参照）