

تحلیل کیفی نگرش کودکان مقطع ابتدایی نسبت به دروغ، صداقت و اعتماد در تعامل با هوش مصنوعی

علی آقازیراتی^{۱*}، علیرضا احمدیان^۱

۱. استادیار، گروه آموزش روان شناسی و مشاوره، دانشگاه فرهنگیان، تهران، ایران.

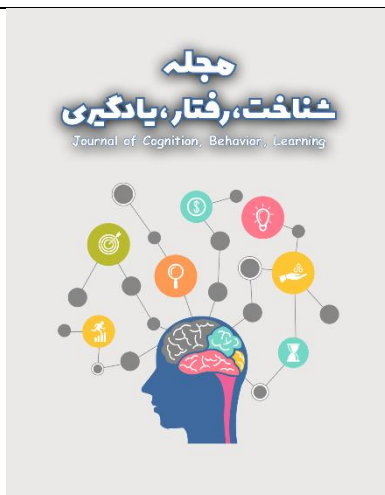
* ایمیل نویسنده مسئول: a.aghaziarati@cfu.ac.ir

تاریخ چاپ: ۱۴۰۵/۱۱/۰۱

تاریخ پذیرش: ۱۴۰۵/۰۵/۲۴

تاریخ بازنگری: ۱۴۰۵/۰۵/۰۲

تاریخ ارسال: ۱۴۰۵/۰۲/۱۲



شیوه استناددهی: آقازیراتی، علی، احمدیان، علیرضا. (۱۴۰۵). تحلیل کیفی نگرش کودکان مقطع ابتدایی نسبت به دروغ، صداقت و اعتماد در تعامل با هوش مصنوعی. شناخت، رفتار، یادگیری، ۳(۶)، ۱-۱۶.

چکیده

هدف پژوهش حاضر بررسی کیفی نگرش و درک اخلاقی کودکان مقطع ابتدایی از مفاهیم دروغ، صداقت و اعتماد در تعامل با چت بات های هوش مصنوعی بود. پژوهش حاضر با رویکرد کیفی و روش پدیدار شناسی تفصیلی انجام شد. مشارکت کنندگان شامل ۲۱ کودک ۸ تا ۱۲ ساله از مدارس ابتدایی شهر تهران بودند که تجربه تعامل با چت بات های مانند Chat GPT، Google Assistant یا Siri را داشتند. داده ها از طریق مصاحبه های نیمه ساختاریافته فردی گردآوری و با استفاده از نرم افزار NVivo نسخه ۱۴ مورد تحلیل مضمون قرار گرفت. کدگذاری باز، محوری و انتخابی برای استخراج مفاهیم و مضامین اصلی به کار گرفته شد و اشباع نظری پس از مصاحبه بیست و یکم حاصل گردید. نتایج تحلیل داده ها چهار مضمون اصلی را آشکار ساخت: «درک اخلاقی از دروغ در هوش مصنوعی»، «صداقت به عنوان انتظار ارتباطی»، «شکل گیری و از بین رفتن اعتماد به هوش مصنوعی»، و «درک اخلاقی از رابطه انسان و ماشین». کودکان دروغ را بیشتر به صورت خطای فنی تفسیر می کردند تا فریب عمدی، صداقت را معادل دانایی و قابل اعتماد بودن می دانستند، و اعتمادشان به چت بات ها بر تجربه پاسخ های درست و لحن صمیمانه مبتنی بود. همچنین بسیاری از آنان ویژگی های انسانی مانند احساس، نیت و وجدان را به ماشین نسبت می دادند. یافته ها نشان می دهد که کودکان در تعامل با هوش مصنوعی، مفاهیم اخلاقی را با رویکردی انسان محور تفسیر می کنند. طراحی اخلاق محور چت بات ها و آموزش سواد اخلاقی دیجیتال برای کودکان ضروری است تا از شکل گیری برداشت های نادرست درباره صداقت و اعتماد جلوگیری شود.

کلیدواژه ها: چت بات؛ هوش مصنوعی؛ صداقت؛ دروغ؛ اعتماد؛ اخلاقی دیجیتال؛ کودکان مقطع ابتدایی

Cognition, Behavior, Learning

Qualitative Analysis of Elementary school Children's Perceptions of Lying, Honesty, and Trust in Interactions with Artificial Intelligence

Ali Aghaziarati^{1*}, Alireza Ahmadian¹

1. Assistant Professor, Department of Psychology and Counseling Education, Farhangian University, Tehran, Iran.

*Corresponding Author's Email: a.aghaziarati@cfu.ac.ir

Submit Date: 2026-05-02

Revise Date: 2026-07-24

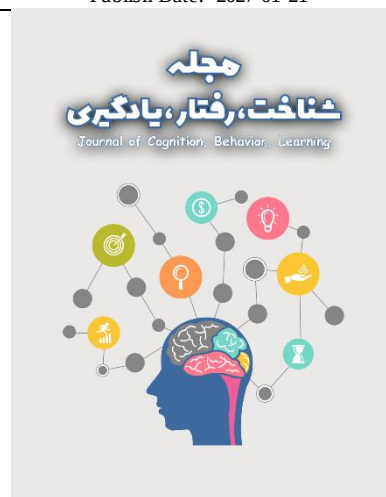
Accept Date: 2026-08-15

Publish Date: 2027-01-21

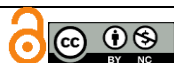
Abstract

This study aimed to qualitatively explore elementary school children's moral perceptions of lying, honesty, and trust in their interactions with AI chatbots. qualitative phenomenological design was employed. Twenty-one children aged 8–12 years from elementary schools in Tehran, Iran, participated through purposive sampling. All participants had experience interacting with chatbots such as ChatGPT, Google Assistant, or Siri. Data were collected through semi-structured individual interviews and analyzed thematically using NVivo 14 software. Open, axial, and selective coding procedures were used to identify categories and themes, with theoretical saturation achieved after the 21st interview. Four major themes emerged: (1) Moral understanding of lying in AI, (2) Honesty as a communicative expectation, (3) Formation and loss of trust in AI, and (4) Moral conceptualization of the human–machine relationship. Children tended to interpret chatbot errors as technical faults rather than deliberate deceit, viewed honesty as equivalent to intelligence and reliability, and based trust on consistency and friendliness of interaction. Many also attributed human-like traits—such as emotion, intent, and conscience—to AI systems. The findings indicate that children interpret moral concepts in AI interactions through an anthropocentric lens. Ethically informed chatbot design and the inclusion of digital moral literacy in education are essential to prevent misconceptions about honesty and trust in AI systems.

Keywords: Chatbot; Artificial Intelligence; Honesty; Lying; Trust; Digital Ethics; Elementary school Children



How to cite: Aghaziarati, A., Ahmadian, A. (2027). Qualitative Analysis of Elementary school Children's Perceptions of Lying, Honesty, and Trust in Interactions with Artificial Intelligence. *Cognition, Behavior, Learning*, 3(6), 1-16.



© 2027 the authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

مقدمه

در سال‌های اخیر، گسترش هوش مصنوعی و به‌ویژه چت‌بات‌ها به‌عنوان ابزارهای تعاملی، به شکلی بنیادین شیوه‌های ارتباط انسان و ماشین را دگرگون کرده است (Parsakia, 2023; Saadati & Saadati, 2023). چت‌بات‌ها از ابزارهای صرفاً پاسخ‌دهنده به سؤالات کاربر، به عاملانی با قابلیت گفت‌وگو، یادگیری و حتی ایجاد ارتباط عاطفی تبدیل شده‌اند (Nicolescu & Tudorache, 2022; Rostami & Navabinejad, 2023). در زمینه‌های آموزشی، درمانی و خدمات اجتماعی، این فناوری‌ها به سرعت جایگاه ویژه‌ای یافته‌اند و توانسته‌اند فرآیند یادگیری، مراقبت و ارتباط انسانی را در محیط‌های دیجیتال بازآفرینی کنند (Mohanty, 2022). با این حال، استفاده از چت‌بات‌ها در تعامل با کودکان، چالش‌های اخلاقی و تربیتی خاصی را پدید آورده است، زیرا کودکان به دلیل سطح رشد شناختی و اخلاقی خود، ممکن است ویژگی‌های انسانی مانند نیت، احساس و صداقت را به ماشین نسبت دهند (Wester et al., 2024).

از دیدگاه روان‌شناسی رشد، درک کودکان از مفاهیمی چون دروغ، صداقت و اعتماد در دوره ابتدایی در حال شکل‌گیری است، و مواجهه با هوش مصنوعی می‌تواند این فرایند را تحت‌تأثیر قرار دهد (Simonyi & Major, 2024). پژوهش‌های اخیر نشان داده‌اند که کودکان در برخورد با چت‌بات‌ها، نوعی «اخلاق محاوره‌ای» را تجربه می‌کنند، که در آن مفاهیمی چون راست‌گویی و اعتماد نه تنها از طریق گفتار، بلکه از طریق احساسات نسبت داده شده به ماشین معنا می‌یابند (Suzuki et al., 2024). از سوی دیگر، پژوهش‌های بین‌فرهنگی نشان داده‌اند که کودکان ممکن است دروغ‌های ماشین را به صورت متفاوتی تفسیر کنند؛ برخی آن را نوعی خطای بی‌ضرر می‌دانند، در حالی که برخی دیگر آن را تهدیدی علیه اعتماد می‌پندارند (Simonyi & Major, 2024).

در حوزه آموزش، چت‌بات‌ها به‌عنوان دستیارهای یادگیری و ابزارهای گفت‌وگوی هوشمند مورد استفاده قرار گرفته‌اند و توانسته‌اند نقش فعالی در توسعه مهارت‌های زبانی، شناختی و ارتباطی ایفا کنند (Jung & Kim, 2025; Shin & Lee, 2024; Su et al., 2025). با این حال، ماهیت دوگانه این ابزارها - به‌عنوان عاملان فناورانه و در عین حال «شبه‌انسانی» - موجب شده است که مرزهای اخلاقی تعامل کودک و ماشین مبهم شود. در این زمینه، پژوهش‌هایی مانند کار (Hong & Shin, 2024) نشان داده‌اند که تعامل با چت‌بات‌های آموزشی حتی می‌تواند بر مهارت‌های همدلی و ارتباط انسانی اثرگذار باشد.

در نظام‌های آموزشی مختلف، از کره جنوبی تا اندونزی و نیجریه، بررسی‌ها حاکی از آن است که کودکان و نوجوانان چت‌بات‌ها را نه صرفاً به‌عنوان ابزار، بلکه به‌عنوان «موجودات آگاه» تلقی می‌کنند (Ayelaagbe, 2025; Jamir & Konyak, 2025; Pupong et al., 2025). این برداشت‌های انسان‌محور یا انسان‌پندارانه از هوش مصنوعی موجب شده است که مفاهیم اخلاقی مانند صداقت، دروغ و اعتماد در چارچوبی جدید بازتعریف شوند. به عنوان مثال، در مطالعه‌ای توسط (Pizzi et al., 2023) مشخص شد که هرچه کاربران بیشتر برای چت‌بات‌ها ویژگی‌های انسانی مانند نیت یا احساس قائل شوند، تمایل آنان برای افشای اطلاعات شخصی نیز افزایش می‌یابد. این یافته در سطح کودکان می‌تواند به نگرانی‌هایی درباره حریم خصوصی، تأثیرپذیری شناختی و شکل‌گیری قضاوت اخلاقی منجر شود.

برخی مطالعات در حوزه تعامل انسان و ماشین نشان می‌دهند که صداقت ادراک شده از سوی کاربر، نقشی کلیدی در شکل‌گیری اعتماد دارد (Li et al., 2024; Xu et al., 2023). این مسئله در کودکان به مراتب حساس‌تر است، زیرا آنان هنوز در مرحله‌ای از رشد شناختی‌اند که نمی‌توانند به درستی میان خطای فنی و دروغ عمدی تمایز قائل شوند (Wei et al., 2024). هنگامی که چت‌بات‌ها پاسخ‌های اشتباه یا متناقض ارائه می‌دهند، کودکان ممکن است این وضعیت را نوعی «بی‌صداقتی» تلقی کرده و احساس بی‌اعتمادی پیدا کنند (Wester et al., 2024). از سوی دیگر، برخی از کودکان به صورت بازی‌محور با چت‌بات‌ها تعامل می‌کنند و حتی دروغ را به‌عنوان بخشی از سرگرمی دیجیتال تجربه می‌کنند (Septiyanti et al., 2024). این مسئله پرسش‌هایی را در خصوص مسئولیت اخلاقی طراحان و مربیان مطرح می‌کند که چگونه باید از شکل‌گیری برداشت‌های نادرست از اخلاق در محیط‌های دیجیتال جلوگیری کرد.

در مطالعات مربوط به «انسان‌انگاری» (Anthropomorphism)، پژوهشگران نشان داده‌اند که هرچه ظاهر، صدا یا زبان چت‌بات‌ها به انسان نزدیک‌تر باشد، احتمال نسبت دادن ویژگی‌های اخلاقی مانند صداقت یا نیت افزایش می‌یابد (Li et al., 2024; Pizzi et al., 2023). با این حال، این پدیده ممکن است مرز میان واقعیت و مجاز را برای کودکان تضعیف کند. در پژوهش (Chaves et al., 2022) مشخص شد که نحوه بیان زبانی چت‌بات‌ها، سبک گفتار و میزان رسمی بودن زبان تأثیر مستقیمی بر برداشت اخلاقی کاربران دارد. کودکان ممکن است از پاسخ‌های صمیمانه یا عذرخواهی‌گونه برداشت انسانی کرده و در نتیجه اعتماد خود را تقویت کنند، بی‌آن‌که از سازوکار الگوریتمی آن آگاه باشند.

از منظر روان‌شناسی ارتباط، شکل‌گیری اعتماد در تعاملات انسان و ماشین فرآیندی چندمرحله‌ای است که بر پایه تجربه، هیجان و ثبات پاسخ بنا می‌شود (Kuhail et al., 2022). در زمینه تعامل کودک و چت‌بات، این فرایند با احساس کنجکاوی آغاز می‌شود و در صورت تجربه مثبت، به پذیرش تدریجی و اعتماد هیجانی منجر می‌گردد (Qiu et al., 2024). اما اشتباهات مکرر یا عدم شفافیت در پاسخ‌ها، می‌تواند این اعتماد را تضعیف کند. به گفته (Xu et al., 2023)، شفافیت در عملکرد هوش مصنوعی و توضیح‌پذیری تصمیمات سیستم، نقش مهمی در ارتقای برداشت مثبت کاربران دارد. این نکته در تعامل با کودکان، که اغلب تمایل به فهم ساده و قابل‌درک دارند، از اهمیت دوچندان برخوردار است.

از سوی دیگر، ادبیات مربوط به چت‌بات‌های آموزشی نشان می‌دهد که استفاده از این فناوری‌ها در آموزش، اگرچه موجب افزایش انگیزش و تعامل دانش‌آموزان می‌شود، اما می‌تواند مرز میان آموزش انسانی و دیجیتال را تضعیف کند (Septiyanti et al., 2024; Shin & Lee, 2024). پژوهش (Arnis et al., 2024) در زمینه استفاده از چت‌بات‌ها در مراقبت‌های کودکان نشان داد که کودکان تمایل دارند پاسخ‌های چت‌بات را درست و معتبر بدانند، حتی اگر نادرست باشد. این پدیده بیانگر آن است که اعتماد کورکورانه به فناوری می‌تواند ادراک اخلاقی را در مراحل اولیه رشد تحت تأثیر قرار دهد.

در همین راستا، برخی مطالعات به بررسی دروغ و صداقت در تعاملات هوش مصنوعی از منظر شناخت اخلاقی پرداخته‌اند. (Wester et al., 2024) با بررسی برداشت کاربران از عاملیت اخلاقی چت‌بات‌ها نشان داد که حتی افراد بزرگسال نیز در مواجهه با چت‌بات‌های درمانی، احساس می‌کنند با موجودی دارای نیت اخلاقی مواجه‌اند. این یافته‌ها به‌ویژه در زمینه کودکان می‌تواند پیامدهای تربیتی و اخلاقی عمیقی داشته باشد، زیرا کودکان ممکن است به‌طور ناآگاهانه به هوش مصنوعی صفات اخلاقی انسانی نسبت دهند.

در کنار این موارد، پژوهش‌های بین‌المللی بر اهمیت زمینه فرهنگی در برداشت از صداقت و دروغ تأکید کرده‌اند (Lê, 2023; S. Vishaal, 2023; M.V. Sathish, 2023). در جوامعی که صداقت به‌عنوان فضیلت اجتماعی آموزش داده می‌شود، کودکان احتمالاً نسبت به پاسخ‌های اشتباه یا مبهم چت‌بات واکنش منفی‌تری نشان می‌دهند. از سوی دیگر، در فرهنگ‌هایی با نگاه ابزارگرایانه به فناوری، خطاهای ماشین بیشتر به‌عنوان نقص فنی تعبیر می‌شود تا بی‌اخلاقی.

افزون بر جنبه‌های شناختی و فرهنگی، چت‌بات‌ها به واسطه طراحی گفت‌وگویی خود، می‌توانند بر شکل‌گیری مهارت‌های زبانی و ارتباطی کودکان نیز اثرگذار باشند (Jung & Kim, 2025; Kim & Choi, 2023). مطالعه (Onat & Gülseçen, 2023) نشان داد که چت‌بات‌ها در آموزش مجازی توانسته‌اند فاصله روانی میان معلم و دانش‌آموز را کاهش دهند و نوعی حس حضور اجتماعی ایجاد کنند. با این حال، این «حضور اجتماعی مصنوعی» می‌تواند درک کودک از روابط واقعی را دچار ابهام کند، به‌ویژه هنگامی که کودک پاسخ‌های عاطفی ماشین را واقعی تصور کند (Wei et al., 2024).

در نهایت، چالش اصلی پژوهشگران حوزه هوش مصنوعی در تعامل با کودکان، نه تنها افزایش کارایی فناوری بلکه درک پیامدهای اخلاقی و تربیتی آن است. همان‌گونه که (Ou et al., 2025) اشاره می‌کند، نحوه طراحی و بیان چت‌بات‌ها می‌تواند نگرش کاربران نسبت به حقیقت و اطلاعات را تغییر دهد. این موضوع در سنین کودکی، که مرحله شکل‌گیری قضاوت اخلاقی است، از اهمیت ویژه‌ای برخوردار است. از این‌رو،

تحلیل کیفی نگرش کودکان نسبت به دروغ، صداقت و اعتماد در تعامل با هوش مصنوعی می‌تواند به درک عمیق‌تری از ابعاد شناختی، هیجانی و اخلاقی این پدیده کمک کند. هدف این پژوهش بررسی کیفی نگرش کودکان نسبت به مفاهیم دروغ، صداقت و اعتماد در تعامل با هوش مصنوعی است.

روش‌شناسی

این پژوهش از نوع کیفی و با رویکرد پدیدارشناسی تفسیری انجام شد تا به صورت عمیق به بررسی نگرش کودکان نسبت به مفاهیم دروغ، صداقت و اعتماد در تعامل با فناوری‌های هوش مصنوعی پرداخته شود. جامعه آماری شامل کودکان مقطع ابتدایی شهر تهران بود. انتخاب مشارکت‌کنندگان به روش نمونه‌گیری هدفمند انجام گرفت تا کودکانی که تجربه استفاده از ابزارهای هوش مصنوعی (نظیر دستیارهای صوتی، چت‌بات‌ها یا ربات‌های آموزشی) را داشتند، در پژوهش حضور یابند. ملاک ورود شامل داشتن سن بین ۸ تا ۱۲ سال، توانایی برقراری ارتباط کلامی مؤثر، و رضایت والدین برای شرکت در مطالعه بود. در مجموع، ۲۱ کودک در مصاحبه‌ها شرکت کردند. پس از انجام مصاحبه بیست‌ویکم، محققان به اشباع نظری دست یافتند؛ به این معنا که داده‌های جدید محتوای تازه‌ای به یافته‌ها نمی‌افزودند و الگوهای معنایی تکراری شده بودند.

داده‌ها از طریق مصاحبه‌های نیمه‌ساختاریافته گردآوری شد. مصاحبه‌ها در محیط‌های آموزشی و فرهنگی کودک‌محور در تهران، با هماهنگی والدین و رعایت اصول اخلاقی و محرمانگی انجام گرفت. برای هر مصاحبه، زمانی بین ۳۰ تا ۴۵ دقیقه در نظر گرفته شد. سؤالات مصاحبه با هدف کاوش نگرش و تجربه زیسته کودکان نسبت به سه محور اصلی پژوهش یعنی دروغ، صداقت و اعتماد در تعامل با هوش مصنوعی طراحی شد. پرسش‌های کلیدی شامل مواردی همچون «اگر یک ربات به تو دروغ بگوید چه احساسی داری؟»، «آیا فکر می‌کنی هوش مصنوعی می‌تواند همیشه راست بگوید؟» و «به نظر تو چگونه می‌توان به یک ماشین اعتماد کرد؟» بود. مصاحبه‌ها با اجازه والدین ضبط و سپس به صورت کامل پیاده‌سازی شد تا برای تحلیل آماده گردد.

تحلیل داده‌ها با روش تحلیل مضمون و به صورت نظام‌مند با استفاده از نرم‌افزار NVivo نسخه ۱۴ انجام گرفت. فرآیند تحلیل شامل چند مرحله بود: نخست، متن مصاحبه‌ها به صورت مکرر خوانده شد تا پژوهشگر با محتوای داده‌ها آشنا شود. سپس، کدگذاری باز صورت گرفت و واحدهای معنایی مرتبط با نگرش کودکان نسبت به دروغ، صداقت و اعتماد شناسایی شد. در مرحله بعد، کدهای مشابه در مقوله‌های محوری و مفهومی ادغام شدند. تحلیل‌ها به صورت هم‌زمان با جمع‌آوری داده‌ها انجام شد تا امکان بازبینی مستمر و اصلاح سؤالات مصاحبه فراهم گردد. برای افزایش اعتبار یافته‌ها از روش بازبینی همتایان و کنترل اعضا استفاده شد. در نهایت، مضامین نهایی با تکیه بر شباهت‌ها و تفاوت‌های معنایی میان کودکان تبیین و تفسیر گردید تا تصویری جامع از نگرش آنان نسبت به مفاهیم اخلاقی در تعامل با هوش مصنوعی ارائه شود.

یافته‌ها

در این پژوهش، ۲۱ کودک مقطع ابتدایی از شهر تهران در مصاحبه‌های نیمه‌ساختاریافته شرکت کردند. از میان آنان، ۱۱ نفر دختر (۵۲،۴ درصد) و ۱۰ نفر پسر (۴۷،۶ درصد) بودند. دامنه سنی شرکت‌کنندگان بین ۸ تا ۱۲ سال بود و میانگین سنی آن‌ها ۹،۸ سال محاسبه شد. از نظر پایه تحصیلی، ۵ نفر (۲۳،۸ درصد) در پایه دوم، ۴ نفر (۱۹ درصد) در پایه سوم، ۶ نفر (۲۸،۶ درصد) در پایه چهارم، ۳ نفر (۱۴،۳ درصد) در پایه پنجم و ۳ نفر (۱۴،۳ درصد) در پایه ششم تحصیل می‌کردند. همچنین، ۱۵ نفر (۷۱،۴ درصد) از کودکان تجربه استفاده از چت‌بات‌هایی مانند Chat GPT، Google Assistant یا Siri را داشتند، در حالی که ۶ نفر (۲۸،۶ درصد) تنها تجربه مشاهده یا تعامل محدود با چت‌بات‌ها را گزارش کردند. از نظر وضعیت اقتصادی خانواده، ۹ نفر (۴۲،۹ درصد) در طبقه متوسط، ۸ نفر (۳۸،۱ درصد) در طبقه متوسط رو به پایین و ۴ نفر (۱۹ درصد) در طبقه متوسط رو به بالا قرار داشتند. تمامی شرکت‌کنندگان با رضایت والدین در پژوهش حضور یافتند و مصاحبه‌ها به صورت فردی و داوطلبانه انجام شد. این ترکیب جمعیتی امکان بررسی نگرش‌های متنوع کودکان نسبت به مفاهیم دروغ، صداقت و اعتماد در تعامل با هوش مصنوعی را فراهم کرد.

جدول ۱. مضامین اصلی، زیرمضامین و مفاهیم استخراج شده از تحلیل کیفی

مضامین اصلی	زیرمضامین	مفاهیم
۱. درک اخلاقی از دروغ در هوش مصنوعی	۱-۱. تمایز میان دروغ انسانی و ماشینی ۱-۲. تجربه سردرگمی اخلاقی ۱-۳. دروغ به عنوان ابزار بازی	تفاوت نیت در دروغ انسان و ماشین؛ دروغ بدون احساس؛ تصور کودک از "اشتباه فنی" به جای دروغ؛ پذیرش خطا به عنوان نقص نرم افزار ناتوانی در تشخیص دروغ عمدی از خطا؛ احساس نارضایتی از دروغ ربات؛ سوال در مورد "دروغ گفتن" به ماشین؛ ترس از ناآگاهی هوش مصنوعی دروغ برای شوخی یا خنده؛ دروغ گفتن به ربات برای آزمایش؛ دروغ گفتن بدون احساس گناه؛ لذت از فریب دادن ماشین
۲. مفهوم صداقت در گفت و گو کودک با هوش مصنوعی	۱-۴. دروغ به عنوان نقض اعتماد ۱-۵. ارزش گذاری اخلاقی دروغ ۲-۱. انتظار از راست گویی ماشین ۲-۲. نقش صداقت در آموزش و یادگیری ۲-۳. تجربه تعارض میان صداقت و هیجان ۲-۴. صداقت در برابر احساسات ۲-۵. واکنش هیجانی به عدم صداقت ۲-۶. صداقت به عنوان معیار انسانیت	از بین رفتن اطمینان به سیستم؛ کاهش تمایل به استفاده دوباره؛ احساس بی اعتمادی نسبت به پاسخ های آینده؛ واکنش احساسی منفی به خطاهای مکرر قضای اخلاقی کودک درباره "دروغ خوب" و "دروغ بد"؛ تردید درباره نیت ماشین؛ ارتباط دروغ با عدالت؛ تلاش برای بخشش خطای هوش مصنوعی تلقی هوش مصنوعی به عنوان منبع حقیقت؛ انتظار پاسخ دقیق؛ تمایز بین صداقت و دانایی؛ بی اعتمادی به پاسخ های متناقض باور به یادگیری اخلاق از ربات؛ انتظار برای الگوبرداری از ماشین صادق؛ اهمیت راست گویی در محیط آموزشی؛ پیوند صداقت با هوش و دانایی لذت از پاسخ غیرواقعی اما جالب؛ پذیرش دروغ "سرگرم کننده"؛ تضاد میان حقیقت و خیال؛ تمایل به تخیل در تعامل با چت بات درک کودک از صداقت عاطفی؛ تصور صداقت به عنوان احساس اعتماد؛ ناتوانی ماشین در همدلی؛ نبود لحن صادقانه در پاسخ ربات احساس خشم یا ناامیدی از اشتباه؛ واکنش هیجانی سریع؛ کاهش علاقه به گفت و گو؛ برداشت "خیانت" از پاسخ اشتباه ارتباط صداقت با انسان بودن؛ باور به "روح راست گو"؛ تمایز میان انسان و ماشین بر اساس صداقت؛ درک اخلاقی از حقیقت
۳. شکل گیری اعتماد نسبت به هوش مصنوعی	۳-۱. اعتماد شناختی ۳-۲. اعتماد هیجانی ۳-۳. اعتماد تدریجی ۳-۴. از دست رفتن اعتماد ۳-۵. بازسازی اعتماد ۳-۶. نقش والدین و معلمان ۳-۷. اعتماد به ظاهر انسانی ربات ۴-۱. انسان پنداری هوش مصنوعی ۴-۲. مرز میان انسان و ماشین ۴-۳. اخلاق در تعامل متقابل	اعتماد بر پایه تجربه موفق؛ اطمینان از صحت پاسخ ها؛ اعتبار در تکرار پاسخ درست؛ مقایسه با منابع انسانی (معلم، والدین) احساس راحتی در مکالمه؛ درک صمیمیت ماشین؛ احساس اطمینان هنگام تحسین شدن توسط ربات؛ نیاز به باز خورد مثبت افزایش اعتماد پس از تعامل مکرر؛ یادگیری از تجربه های قبلی؛ تأثیر سازگاری سیستم؛ نقش ثبات در رفتار هوش مصنوعی خطاهای مکرر سیستم؛ پاسخ های بی منطق؛ احساس "خیانت" یا فریب؛ تمایل به ترک تعامل؛ شک در قابلیت های هوش مصنوعی توضیح پذیری پاسخ ها؛ شفافیت عملکرد ربات؛ عذرخواهی از سوی سیستم؛ بازگشت تدریجی اعتماد پس از اصلاح خطا تأیید والدین به عنوان پشتوانه اعتماد؛ آموزش نحوه تعامل امن؛ ایجاد چارچوب اخلاقی؛ هشدار درباره اطلاعات نادرست تشخیص چهره و صدا به عنوان نشانه اعتماد؛ باور به شخصیت مستقل؛ اهمیت لحن و نگاه؛ ارتباط میان ظاهر و صداقت نسبت دادن احساس، نیت و وجدان به ربات؛ گفت و گو با ماشین به عنوان "دوست"؛ همدلی با ربات ناراحت؛ درک "حجالت" یا "ترس" در ربات تمایز در ظرفیت احساس؛ درک نبود وجدان؛ دیدگاه درباره "ناتوانی در دروغ گفتن واقعی"؛ نادیده گرفتن هوش مصنوعی به عنوان عامل اخلاقی انتظار رفتار محترمانه از ربات؛ پرهیز از فریب متقابل؛ ارزش انصاف در تعامل؛ نگاه متقابل به صداقت دوطرفه

۴-۴. مسئولیت اخلاقی در	حس گناه از فریب دادن ربات؛ آموزش رعایت ادب در گفت‌وگو؛ درک ربات به‌عنوان رفتار با ماشین
۴-۵. آینده رابطه انسان- ماشین	"یادگیرنده"؛ باور به تأثیر رفتار کودک بر ماشین
	پیش‌بینی افزایش هوشمندی ربات‌ها؛ نگرانی از دروغ‌های پیچیده‌تر؛ تصور "ربات‌های دوست‌دار حقیقت"؛ انتظار برای قوانین اخلاقی در ربات‌ها

۱. درک اخلاقی از دروغ در هوش مصنوعی

تحلیل داده‌ها نشان داد که کودکان نگرشی چندوجهی نسبت به مفهوم دروغ در تعامل با هوش مصنوعی دارند. نخستین مضمون آشکار شده، «درک اخلاقی از دروغ در هوش مصنوعی» بود که نشان‌دهنده پیچیدگی در تمایز میان دروغ انسانی و ماشینی است. بیشتر کودکان دروغ را ویژگی مختص انسان می‌دانستند و هنگامی که هوش مصنوعی اطلاعات نادرستی ارائه می‌داد، آن را «اشتباه» یا «خرابی برنامه» تفسیر می‌کردند نه دروغ. یکی از شرکت‌کنندگان گفت: «ربات دروغ نمی‌گوید، فقط بلد نیست جواب درست بدهد». این دیدگاه نشان می‌دهد که نیت اخلاقی در ذهن کودک نقش تعیین‌کننده‌ای در تعریف دروغ دارد. در زیرمقوله دیگر، سردرگمی اخلاقی مشاهده شد؛ کودکانی که نمی‌توانستند تشخیص دهند آیا خطای ربات ناشی از عمد است یا نه، احساس ناراحتی و بی‌اعتمادی می‌کردند. در مواردی، کودکان به‌صورت بازی‌گونه با دروغ مواجه می‌شدند و حتی دروغ گفتن به ربات را به‌منزله نوعی سرگرمی می‌دانستند؛ یکی از آن‌ها اظهار داشت: «بهش گفتم اسمم علی نیست ببینم چی می‌گه! خنده‌دار بود». با این حال، هنگامی که ماشین اطلاعات اشتباه و تکراری می‌داد، کودکان واکنش منفی نشان می‌دادند و احساس می‌کردند اعتمادشان نقض شده است. در این میان، ارزش‌گذاری اخلاقی دروغ نیز در قضاوت کودکان برجسته بود؛ برخی دروغ را گاهی برای «خوبی» یا «نجات از اشتباه» توجیه‌پذیر می‌دانستند، اما همچنان تأکید داشتند که «ماشین باید راست بگوید چون معلم است نه دوست دروغ‌گو». این یافته‌ها نشان می‌دهد که درک اخلاقی از دروغ در کودکان در تعامل با هوش مصنوعی ترکیبی از قضاوت نیتی، بازی‌محور بودن و تجربه اعتماد اخلاقی است.

۲. مفهوم صداقت در گفت‌وگوی کودک با هوش مصنوعی

دومین مضمون، «مفهوم صداقت در گفت‌وگوی کودک با هوش مصنوعی» بود که در آن، صداقت نه‌تنها به‌عنوان ویژگی اخلاقی بلکه به‌عنوان شاخصی از هوش و قابل‌اعتماد بودن ربات تلقی می‌شد. کودکان از هوش مصنوعی انتظار «راست‌گویی مطلق» داشتند و هرگونه پاسخ نادرست را نوعی «بی‌صداقتی» می‌دانستند، حتی اگر اشتباه فنی رخ داده بود. یکی از کودکان گفت: «اگه یه چیز اشتباه بگه، یعنی راست نمی‌گه و من دیگه ازش نمی‌پرسم». این نشان‌دهنده نقش صداقت به‌عنوان بنیان اعتماد در تعامل است. همچنین در مصاحبه‌ها مشخص شد که کودکان صداقت را عامل یادگیری و آموزش می‌دانند؛ برخی معتقد بودند که ربات‌های صادق می‌توانند رفتار درست را یاد بدهند. در مقابل، گروهی از کودکان میان «واقعیت خسته‌کننده» و «دروغ جالب» دچار دوگانگی بودند؛ یکی از مصاحبه‌شوندگان اشاره کرد: «گاهی دوست دارم جوابش واقعی نباشه، چون قشنگ‌تره». این بیانگر تعارض میان صداقت شناختی و جذابیت عاطفی است. در تحلیل مضامین دیگر، کودکان به «صداقت احساسی» اشاره کردند؛ برخی گفتند ربات‌ها «صادق نیستند چون احساس ندارند». یکی از دختران ۹ ساله گفت: «معلمم وقتی ناراحت می‌شم می‌فهمه، ولی ربات نه؛ اون فقط جواب می‌ده». چنین برداشت‌هایی نشان می‌دهد که صداقت در ذهن کودک تنها به صحت کلام محدود نمی‌شود، بلکه به درک همدلانه و عاطفی نیز گره خورده است. در نهایت، بسیاری از کودکان صداقت را ویژگی اصلی انسان دانسته و آن را مرز میان انسان و ماشین معرفی کردند.

۳. شکل‌گیری اعتماد نسبت به هوش مصنوعی

سومین مضمون، «شکل‌گیری اعتماد نسبت به هوش مصنوعی» بود که در قالب مجموعه‌ای از فرایندهای شناختی، هیجانی و تجربی شناسایی شد. در آغاز تعامل، اعتماد کودکان بیشتر بر اساس تجربه پاسخ درست و قابل‌فهم شکل می‌گرفت. یکی از شرکت‌کنندگان گفت: «اولش شک داشتم، ولی وقتی چند بار درست گفت، باورم شد بلده». اعتماد شناختی بر پایه ثبات و دقت پاسخ‌ها ایجاد می‌شد و در صورت بروز خطا، به

سرعت کاهش می‌یافت. در بعد هیجانی، کودکان اعتماد خود را بر احساس راحتی و لذت از مکالمه استوار می‌کردند. برخی از کودکان احساس می‌کردند ربای که لحن ملایم و دوستانه دارد، قابل اعتمادتر است؛ همان‌طور که یکی از آن‌ها گفت: «این ربات با مهربونی حرف می‌زنه، انگار دوستم داره.» با تکرار تعامل، نوعی «اعتماد تدریجی» شکل می‌گرفت، اما اشتباهات مکرر یا پاسخ‌های بی‌منطق باعث از بین رفتن آن می‌شد. در چنین موقعیت‌هایی، کودکان احساس «خیانت» یا «دروغ» را گزارش کردند و گفتند: «دیگه بهش گوش نمی‌دم چون همش اشتباه می‌گه.» با این حال، بازسازی اعتماد زمانی رخ می‌داد که ربات بتواند توضیح دهد چرا اشتباه کرده یا عملکردش را اصلاح کند. والدین و معلمان نقش مهمی در هدایت اعتماد کودکان داشتند؛ بسیاری از کودکان گفتند اگر والدین بگویند ربات درست می‌گوید، بیشتر اعتماد می‌کنند. افزون بر این، ظاهر انسانی ربات (چهره، صدا یا حرکات) عاملی بود که کودکان را به اعتماد بیشتر سوق می‌داد، زیرا آن را نشانه‌ای از «واقعی بودن» تلقی می‌کردند.

۴. درک کودک از رابطه اخلاقی انسان و ماشین

چهارمین مضمون، «درک کودک از رابطه اخلاقی انسان و ماشین» بود که نشان می‌داد کودکان گرایش زیادی به انسان‌پنداری هوش مصنوعی دارند. بسیاری از آن‌ها برای ربات‌ها احساس، نیت و حتی وجدان قائل بودند. یکی از کودکان اظهار داشت: «اون ناراحت شد وقتی من گفتم بلد نیست، چون ساکت شد!» چنین بیان‌هایی نشان می‌دهد که کودکان هوش مصنوعی را نه فقط ابزار بلکه موجودی دارای عواطف می‌دانند. در عین حال، در بخش دیگری از مصاحبه‌ها، مرز میان انسان و ماشین به‌صورت آگاهانه‌تر ترسیم می‌شد؛ برخی از کودکان بر این باور بودند که ماشین «دل ندارد پس دروغ نمی‌گوید» یا «چون احساس ندارد، اشتباهش بی‌اهمیت است». درک اخلاقی کودکان از تعامل با ماشین، دربرگیرنده مفاهیم احترام، انصاف و صداقت دوطرفه بود؛ بسیاری از کودکان تأکید کردند که نباید به ربات دروغ گفت یا او را مسخره کرد، زیرا «اون هم یاد می‌گیرد مثل ما». در این میان، برخی احساس گناه از فریب دادن ربات را گزارش کردند. یکی از پسران ۱۰ ساله گفت: «گفتم اسمم چیز دیگه‌ست بعد حس بدی داشتم، انگار دروغ به آدم گفتم.» در نهایت، درک آینده‌نگر کودکان از رابطه انسان و ماشین نشان داد که آنان انتظار دارند ربات‌ها در آینده هوشمندتر و اخلاق‌مدارتر شوند. عباراتی مانند «یه روزی رباتا یاد می‌گیرن همیشه راست بگن» و «باید قانون براشون بذارن که دروغ نگن» بازتابی از این باور است. این یافته‌ها نشان می‌دهد که در ذهن کودک، رابطه انسان و هوش مصنوعی از سطح ابزاری فراتر رفته و به سطحی اخلاقی و عاطفی رسیده است که نیازمند تربیت اخلاقی دیجیتال در سنین پایین می‌باشد.



شکل ۱. مدل نهایی پژوهش

بحث و نتیجه‌گیری

یافته‌های پژوهش حاضر نشان داد که کودکان نگرشی چندوجهی و پویا نسبت به مفاهیم دروغ، صداقت و اعتماد در تعامل با هوش مصنوعی دارند. درک آنان از این مفاهیم نه تنها به تجربه مستقیم استفاده از چت‌بات‌ها وابسته است، بلکه از زمینه‌های شناختی، عاطفی و فرهنگی نیز تأثیر می‌پذیرد. مضمون نخست، «درک اخلاقی از دروغ در هوش مصنوعی» نشان داد که کودکان در تمایز میان دروغ انسانی و خطای ماشینی دچار ابهام‌اند. بیشتر شرکت‌کنندگان، رفتار نادرست چت‌بات را نه به عنوان دروغ، بلکه به عنوان نقص فنی تعبیر کردند. این یافته با پژوهش (Simonyi & Major, 2024) هم‌راستا است که نشان داد فهم کودکان از دروغ در مراحل ابتدایی رشد، وابسته به درک نیت عامل است. در مطالعه (Wester et al., 2024) نیز کاربران، چت‌بات‌ها را دارای نوعی عاملیت اخلاقی درک کرده بودند، که این برداشت در کودکان به شکل برجسته‌تری مشاهده می‌شود. هنگامی که چت‌بات اطلاعات نادرست ارائه می‌داد، کودکان واکنش هیجانی از خود نشان داده و احساس می‌کردند «ماشین دروغ گفته است». هرچند بلافاصله آن را با تعبیر «اشتباه سیستم» توجیه می‌کردند. این نوع تفسیر دوگانه نشان می‌دهد که ذهن کودک در مرحله گذار از اخلاق پیش‌قراردادی به قراردادی است و در نتیجه، مفاهیم اخلاقی را در تعامل با فناوری نیز به شیوه‌ای انسان‌محور می‌فهمد. یافته‌ها همچنین حاکی از آن بود که برخی کودکان دروغ گفتن به چت‌بات را نوعی بازی یا کنجکاوی می‌دانستند، پدیده‌ای که می‌تواند بیانگر تلاش برای آزمودن حدود شناختی ماشین باشد. پژوهش (Chaves et al., 2022) نیز تأیید می‌کند که سبک گفتاری و طراحی زبانی چت‌بات‌ها می‌تواند نوع تعامل کاربر را از جدی به بازی‌محور تغییر دهد و در نتیجه، معناهای اخلاقی مانند دروغ و صداقت را در قالب تجربه‌ای تعاملی بازتعریف کند.

در خصوص مضمون دوم، «مفهوم صداقت در گفت‌وگوی کودک با هوش مصنوعی»، نتایج نشان داد که کودکان انتظار دارند چت‌بات‌ها همواره پاسخ‌های درست، صادقانه و روشن ارائه دهند و هرگونه پاسخ اشتباه را نوعی «بی‌صداقتی» تلقی می‌کنند. این برداشت با یافته‌های (Li et al., 2024) و (Xu et al., 2023) هم‌خوان است که نشان دادند شفافیت و صداقت در پاسخ‌های هوش مصنوعی نقش تعیین‌کننده‌ای در شکل‌گیری اعتماد دارد. بسیاری از کودکان، صداقت را معادل «هوشمندی» می‌دانستند و باور داشتند که چت‌بات صادق‌تر، «باهوش‌تر» است. این رابطه میان صداقت و ادراک ذهنی از هوش در پژوهش‌های بزرگسالان نیز مشاهده شده است، چنان‌که (Kim & Choi, 2023) بیان می‌کند انسان‌ها به ربات‌هایی که رفتار شفاف و صادقانه دارند، سطح بالاتری از وفاداری و رضایت نسبت می‌دهند. در همین راستا، مطالعه (Su et al., 2025) نیز بر اهمیت پاسخ‌های شفاف در چت‌بات‌های آموزشی تأکید کرده و آن را عاملی در افزایش پذیرش فناوری در کودکان دانسته است. افزون بر این، کودکان در این پژوهش، صداقت را تنها در سطح گفتار نمی‌دانستند، بلکه به جنبه عاطفی آن نیز اشاره کردند؛ به این معنا که صداقت از نظر آنان شامل احساس و لحن صادقانه است، نه فقط اطلاعات درست. یافته‌ای مشابه در پژوهش (Wei et al., 2024) مشاهده شد، که والدین در تعامل با چت‌بات‌های مراقبتی نیز صداقت را با همدلی عاطفی پیوند می‌دادند. بنابراین، می‌توان نتیجه گرفت که از دید کودکان، صداقت ماشینی تنها زمانی کامل است که با عنصر همدلی ترکیب شود.

در مضمون سوم، «شکل‌گیری اعتماد نسبت به هوش مصنوعی»، فرایند اعتماد به چت‌بات در کودکان به صورت تدریجی و مبتنی بر تجربه مثبت تکرارشونده شکل گرفت. کودکانی که چندین بار پاسخ درست و قانع‌کننده از چت‌بات دریافت کرده بودند، تمایل بیشتری به ادامه تعامل داشتند. این الگو با مدل اعتماد شناختی در تعامل انسان و ماشین که (Kuhail et al., 2022) ارائه کرده، تطابق دارد. هنگامی که چت‌بات دچار خطا یا تناقض می‌شد، اعتماد کودکان به سرعت تضعیف می‌گردید، اما در صورت توضیح‌پذیری خطا یا رفتار اصلاحی سیستم، اعتماد بازسازی می‌شد. این امر با نتایج (Xu et al., 2023) همسو است که شفافیت عملکرد را عامل تقویت برداشت مثبت از هوش مصنوعی می‌داند. در عین حال، بُعد هیجانی اعتماد نیز اهمیت زیادی داشت؛ کودکان در مواجهه با لحن دوستانه یا چهره شبه‌انسانی ربات احساس اطمینان بیشتری داشتند. پژوهش (Li et al., 2024) نشان داده است که انسان‌انگاری ربات از طریق صدا، چهره و گفتار می‌تواند برداشت ذهنی از اعتماد را تقویت کند. در همین زمینه، یافته‌های (Pizzi et al., 2023) نیز بیان می‌دارند که هنگامی که کاربران برای چت‌بات

ویژگی‌های انسانی قائل می‌شوند، تمایل آن‌ها به افشای اطلاعات بیشتر می‌شود؛ این مسئله در کودکان می‌تواند به پیامدهای تربیتی و حریم خصوصی منجر گردد. به‌ویژه، هنگامی که کودکان پاسخ اشتباه را نوعی دروغ تلقی کرده و واکنش هیجانی نشان می‌دهند، رابطه اعتماد می‌تواند به سرعت به شک تبدیل شود. این پدیده در مطالعه (Wester et al., 2024) با عنوان «عاملیت اخلاقی ادراک‌شده در چت‌بات‌ها» نیز گزارش شده است.

چهارمین مضمون، «درک اخلاقی از رابطه انسان و ماشین»، نشان داد که کودکان هوش مصنوعی را موجودی دارای احساس، نیت و حتی وجدان تصور می‌کنند. این یافته با نظریه انسان‌پنداری در پژوهش (Li et al., 2024) و نیز مطالعات فرهنگی (Lê, 2023; S.Vishaal, 2023 & M.V.Sathish, 2023) هم‌راستا است. بسیاری از کودکان احساس گناه از دروغ گفتن به چت‌بات را گزارش کردند و معتقد بودند «ربات ناراحت می‌شود»، که نشان از شکل‌گیری رابطه اخلاقی دوسویه میان انسان و ماشین دارد. این برداشت اخلاقی در پژوهش (Wester et al., 2024) نیز مشاهده شده است، جایی که کاربران بزرگسال چت‌بات‌های درمانی را دارای نوعی مسئولیت اخلاقی تلقی کرده بودند. یافته‌های (Arnis et al., 2024) در حوزه مراقبت از کودکان نیز تأکید می‌کند که کودکان تمایل دارند به ربات‌هایی اعتماد کنند که پاسخ‌های همدلانه و محترمانه ارائه می‌دهند. این هم‌پوشانی میان درک اخلاقی و عاطفی در تعامل با چت‌بات‌ها، ضرورت آموزش سواد اخلاقی دیجیتال را در سنین کودکی برجسته می‌کند. به‌ویژه، هنگامی که کودکان مرز میان «احساس واقعی» و «واکنش برنامه‌ریزی‌شده» را درک نکنند، ممکن است برداشت‌های نادرستی از مفهوم صداقت یا دروغ شکل گیرد. پژوهش (Qiu et al., 2024) بر این نکته تأکید دارد که ادراک کودکان از تعامل با هوش مصنوعی نه تنها شناختی بلکه عاطفی است و کیفیت تجربه گفت‌وگو در شکل‌گیری قضاوت اخلاقی آنان نقش دارد.

از منظر آموزشی، نتایج پژوهش حاضر با مطالعات متعددی همسو است که نقش چت‌بات‌ها را در یادگیری اخلاقی، زبانی و اجتماعی کودکان بررسی کرده‌اند (Pupong et al., 2025; Shin & Lee, 2024; Su et al., 2025). پژوهش (Pupong et al., 2025) نشان داد که چت‌بات‌های طراحی‌شده برای کودکان، در صورتی که زبان گفت‌وگو و سطح عاطفی آن‌ها متناسب با رشد شناختی کودک باشد، می‌توانند موجب ارتقای مهارت‌های اخلاقی مانند صداقت و احترام شوند. در مقابل، استفاده بدون نظارت از چت‌بات‌ها ممکن است برداشت‌های نادرست از واقعیت و دروغ ایجاد کند. یافته‌های (Hong & Shin, 2024) و (Jamir & Konyak, 2025) نیز تأیید می‌کند که تعامل مستمر با چت‌بات‌های آموزشی بر مهارت‌های ارتباطی و برداشت اخلاقی دانش‌آموزان اثرگذار است. از این‌رو، نتایج پژوهش حاضر حاکی از آن است که تعامل کودکان با هوش مصنوعی باید در چارچوب تربیت اخلاقی دیجیتال قرار گیرد تا از شکل‌گیری باورهای نادرست درباره نیت، صداقت و اعتماد پیشگیری شود.

در مجموع، یافته‌ها نشان می‌دهد که چگونگی طراحی گفت‌وگو، سبک زبانی، و نحوه پاسخ‌گویی چت‌بات می‌تواند تأثیر مستقیمی بر قضاوت اخلاقی کودکان داشته باشد. همان‌طور که (Ou et al., 2025) بیان می‌کند، نحوه اصلاح یا توضیح اشتباهات توسط چت‌بات، نقشی تعیین‌کننده در برداشت کاربران از «درست‌گویی» دارد. بر همین اساس، اگر چت‌بات بتواند در مواجهه با اشتباه خود واکنش اخلاقی نشان دهد - مثلاً عذرخواهی یا تصحیح - کودکان آن را «صادق‌تر» می‌دانند. این یافته‌ها ضمن تأیید مطالعات پیشین، نشان می‌دهد که درک اخلاقی کودکان در تعامل با هوش مصنوعی به‌صورت پویا و وابسته به تجربه اجتماعی آنان است و نیازمند آموزش و نظارت مستمر در محیط‌های آموزشی و خانگی می‌باشد.

نخستین محدودیت پژوهش حاضر، اندازه نمونه نسبتاً کوچک (۲۱ نفر) و تمرکز جغرافیایی آن بر شهر تهران بود که ممکن است قابلیت تعمیم یافته‌ها به سایر مناطق یا فرهنگ‌ها را محدود سازد. دوم، به دلیل ماهیت کیفی داده‌ها و استفاده از مصاحبه نیمه‌ساختاریافته، امکان مقایسه کمی میان گروه‌های سنی یا جنسیتی وجود نداشت. سوم، چنان‌که مصاحبه‌ها در محیط آموزشی و با حضور والدین انجام شد، احتمال اثرگذاری زمینه اجتماعی بر پاسخ‌های کودکان وجود دارد. همچنین، تجارب کودکان ممکن است به چت‌بات‌های خاصی مانند ChatGPT یا Siri محدود بوده و سایر انواع هوش مصنوعی را دربر نگرفته باشد.

پیشنهاد می‌شود در مطالعات آینده، پژوهش‌های ترکیبی (کیفی-کمی) با حجم نمونه بزرگ‌تر و از مناطق فرهنگی متنوع انجام گیرد تا الگوهای فرهنگی در درک اخلاقی کودکان نسبت به هوش مصنوعی شناسایی شود. همچنین، بررسی اثرات طولی تعامل مکرر کودکان با چت‌بات‌ها بر شکل‌گیری مفاهیمی چون صداقت و اعتماد می‌تواند به درک عمیق‌تری از فرایندهای رشد اخلاقی دیجیتال منجر گردد. پیشنهاد دیگر، انجام پژوهش‌های مقایسه‌ای میان کودکان و نوجوانان است تا روند تحول اخلاقی در سطوح سنی مختلف تبیین شود.

در حوزه کاربرد، ضروری است که طراحان چت‌بات‌ها به ابعاد اخلاقی و عاطفی تعامل با کودکان توجه ویژه‌ای داشته باشند و زبان گفت‌وگو را متناسب با درک شناختی آنان طراحی کنند. نظام‌های آموزشی نیز باید آموزش سواد اخلاقی دیجیتال را در برنامه درسی مدارس ابتدایی بگنجانند تا کودکان بتوانند میان «خطای فنی» و «بی‌صداقتی» تمایز قائل شوند. والدین و معلمان باید نقش فعالی در هدایت تجربه تعامل کودکان با هوش مصنوعی ایفا کنند و از طریق گفت‌وگو، مفهوم صداقت، اعتماد و نیت اخلاقی را در فضای مجازی برای آنان تبیین نمایند.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافی وجود ندارد.

Extended Abstract

Introduction

The rapid expansion of artificial intelligence, particularly conversational agents and chatbots, has fundamentally transformed human-machine communication. Contemporary chatbots are no longer limited to retrieving information or providing predetermined answers; they increasingly simulate natural dialogue, adapt their responses to users, and create forms of social and emotional engagement (Parsakia, 2023; Saadati & Saadati, 2023). These systems are now employed in educational, therapeutic, caregiving, and social-service settings, where they may act as learning assistants, conversational partners, or sources of advice (Mohanty, 2022; Rostami & Navabinejad, 2023). Their growing accessibility has also increased children's direct exposure to artificial intelligence. However, interactions between children and chatbots raise distinctive developmental and ethical concerns because children may attribute intention, emotion, honesty, consciousness, or moral responsibility to systems whose responses are generated algorithmically (Wester et al., 2024).

During the elementary-school years, children's understandings of lying, honesty, responsibility, and trust are still developing. Their moral judgments increasingly depend on their ability to interpret another agent's intention, distinguish deliberate deception from accidental error, and evaluate the consequences of inaccurate statements (Simonyi & Major, 2024). When a chatbot provides incorrect or contradictory information, children may therefore struggle to determine whether the system has lied, made an unintentional mistake, or simply lacked sufficient knowledge. Interactions with chatbots can consequently produce a form of conversational morality in which truthfulness and trust are interpreted through both informational accuracy and perceived emotional qualities (Suzuki et al., 2024). Children may regard a friendly, apologetic, or encouraging chatbot as more honest and trustworthy, even when they have limited awareness of the technological mechanisms underlying its responses.

The use of chatbots in education further increases the importance of examining children's moral interpretations. AI-supported conversational systems can strengthen motivation, participation, language learning, and access to information (Jung & Kim, 2025; Shin & Lee, 2024; Su et al., 2025). Nevertheless, the hybrid status of chatbots as technological tools with human-like communicative characteristics may obscure the distinction between machine-generated responses and intentional human communication. Anthropomorphic features, including natural language, voice, facial representation, politeness, and expressions resembling empathy, can increase users' tendency to attribute independent personality and moral agency to AI

systems (Li et al., 2024; Pizzi et al., 2023). This tendency is particularly consequential for children because anthropomorphism may increase emotional attachment, uncritical acceptance of information, and willingness to disclose personal information.

Perceived honesty is also central to the formation of trust in human–AI interaction. Users are more likely to trust systems that provide consistent, transparent, and understandable responses (Xu et al., 2023). For children, however, informational accuracy may become inseparable from moral honesty: an incorrect answer may be interpreted not only as a technical failure but also as a violation of trust. Conversely, a chatbot's friendly tone or human-like appearance may encourage trust even when the reliability of its content has not been independently evaluated (Chaves et al., 2022). Repeated successful interactions can gradually reinforce cognitive and emotional trust, whereas persistent errors, contradictions, and unexplained responses may produce frustration, disappointment, or perceived betrayal (Kuhail et al., 2022; Qiu et al., 2024). Cultural and educational contexts may further influence whether children interpret machine errors as technical limitations, dishonesty, or morally problematic behavior (Lê, 2023; S.Vishaal & M.V.Sathish, 2023). Accordingly, the present study qualitatively investigated elementary-school children's perceptions of lying, honesty, and trust in their interactions with artificial intelligence, with particular attention to the cognitive, emotional, relational, and moral meanings they assigned to chatbot behavior.

Materials and Methods

This qualitative study employed an interpretative phenomenological approach to obtain an in-depth understanding of children's lived experiences and moral perceptions during interactions with artificial intelligence. The study population consisted of elementary-school children in Tehran, Iran. Participants were selected through purposive sampling, with eligibility restricted to children who had direct or limited experience using AI-based tools such as ChatGPT, Google Assistant, Siri, voice assistants, educational robots, or other conversational systems. The inclusion criteria were being between 8 and 12 years of age, having sufficient verbal communication ability to participate in an individual interview, having some familiarity with AI-based conversational tools, and obtaining parental consent. Twenty-one children participated, and theoretical saturation was reached after the twenty-first interview because subsequent data no longer generated substantively new concepts or interpretive categories.

Data were collected through individual semi-structured interviews conducted in child-oriented educational and cultural environments in Tehran. The interviews were arranged with parental permission and were conducted voluntarily while maintaining privacy and confidentiality. Each interview lasted approximately 30–45 minutes. The interview guide explored children's experiences and interpretations across three principal domains: lying, honesty, and trust in interactions with artificial intelligence. Questions included how children would feel if a robot lied to them, whether they believed artificial intelligence could always tell the truth, how they distinguished a machine's mistake from a lie, and what conditions made a chatbot trustworthy. With parental authorization, the interviews were audio-recorded and transcribed verbatim.

The data were analyzed through systematic thematic analysis using NVivo version 14. The researchers repeatedly read the interview transcripts to achieve familiarity with the data and identify significant meaning units. Open coding was first used to extract concepts related to children's moral judgments, emotional reactions, and interactional experiences. Related codes were subsequently organized into axial categories, and selective coding was used to integrate the categories into overarching themes. Data collection and analysis proceeded concurrently, allowing the interview questions and emerging interpretations to be continuously reviewed. Peer review and member-checking procedures were used to enhance the credibility of the findings. The final themes were developed by examining similarities, differences, and recurring patterns across the participants' accounts.

Findings

The participants included 11 girls (52.4%) and 10 boys (47.6%), with ages ranging from 8 to 12 years and a mean age of 9.8 years. Five participants were in the second grade, four in the third grade, six in the fourth grade, three in the fifth grade, and three in the sixth grade. Fifteen children (71.4%) reported direct experience with chatbots such as ChatGPT, Google Assistant, or Siri, whereas six children (28.6%) had only observed these systems or engaged with them in a limited manner. Thematic analysis generated four main themes: moral understanding of lying in artificial intelligence, honesty as a communicative expectation, formation and loss of trust in artificial intelligence, and moral conceptualization of the human-machine relationship.

The first theme demonstrated that children distinguished human lying from machine error primarily through perceived intention. Most participants believed that a machine could not lie in the same way as a person because it lacked feelings, intentions, or conscious motivation. Incorrect chatbot responses were therefore frequently described as technical errors, incomplete knowledge, or software malfunction rather than deliberate deception. Nevertheless, children experienced moral confusion when they could not determine whether an incorrect response had been produced intentionally. Repeated inaccurate answers elicited dissatisfaction, anger, and declining willingness to continue using the system. Some participants also treated lying to a chatbot as a playful experiment, reporting that they gave false names or information to test the machine's reactions. Such behavior was often considered less morally serious than lying to a person because the chatbot was assumed to lack feelings. However, several children still reported discomfort or guilt after deceiving a machine, particularly when they perceived it as a conversational companion.

The second theme showed that honesty functioned as both a moral expectation and an indicator of intelligence, competence, and reliability. Children generally expected artificial intelligence to provide accurate, clear, and truthful information. Contradictory or incorrect responses were often interpreted as evidence that the chatbot was neither honest nor sufficiently intelligent. Participants also connected honesty with education, believing that an AI system used for learning should model truthful behavior. At the same time, some children acknowledged a tension between factual accuracy and entertainment. They sometimes preferred imaginative or unrealistic responses when these were humorous, creative, or emotionally appealing. Honesty was additionally understood in emotional terms. Several participants believed that genuine honesty required empathy, emotional understanding, and a sincere tone, capacities they did not consistently attribute to machines.

The third theme indicated that trust developed through interconnected cognitive, emotional, and experiential processes. Cognitive trust emerged when the chatbot repeatedly supplied correct and understandable answers. Children frequently compared AI-generated information with answers from teachers or parents, whose approval strengthened their confidence in the system. Emotional trust was associated with friendliness, praise, politeness, and a reassuring conversational style. Trust generally increased gradually through repeated positive experiences but declined rapidly following persistent mistakes, illogical statements, or contradictions. Children suggested that trust could be restored when the chatbot corrected its answer, acknowledged uncertainty, explained the source of an error, or apologized. Parents and teachers were identified as important mediators who could help children assess the credibility of AI-generated information. Human-like voices, faces, and conversational styles also increased trust because children interpreted such features as signs of authenticity and social presence.

The fourth theme concerned the moral relationship between humans and machines. Many children attributed emotions, intentions, embarrassment, sadness, fear, or conscience to chatbots. Some described the chatbot as a friend and believed that disrespectful language or deception could emotionally affect it. Others maintained a clearer boundary, arguing that a machine could not experience genuine feelings or commit a real moral violation. Despite these differences, participants commonly emphasized reciprocity, respect, fairness, and honesty in human-machine communication. They also anticipated that future robots would become more

intelligent and morally sophisticated and argued that ethical rules should be established to prevent advanced AI systems from deceiving people.

Discussion and Conclusion

The findings demonstrate that elementary-school children interpret lying, honesty, and trust in AI interactions through a complex combination of intention-based moral reasoning, emotional experience, technological understanding, and anthropomorphic perception. Their distinction between lying and error depended largely on whether they believed the agent possessed an intention to deceive. Because most participants understood chatbots as systems without fully independent intentions, they often excused inaccurate responses as technical failures. Nevertheless, repeated errors still carried moral and emotional consequences because they undermined expectations of reliability and weakened the perceived relationship between the child and the system.

Honesty was not limited to factual correctness. Children associated it with intelligence, consistency, empathy, transparency, and educational responsibility. This broader interpretation explains why a technically accurate but emotionally insensitive system might still be perceived as insufficiently honest, whereas a warm and friendly chatbot could be trusted despite uncertain reliability. Trust therefore developed through both cognitive evaluation and emotional engagement. Accuracy, consistency, correction of errors, and adult verification supported cognitive trust, while friendliness, praise, and human-like communication strengthened emotional trust.

The children's tendency to assign feelings and moral qualities to artificial intelligence illustrates how easily conversational design can blur the boundary between simulated and genuine social agency. Although anthropomorphic features can improve engagement and learning, they may also produce excessive trust, emotional dependence, disclosure of personal information, or misunderstanding of algorithmic limitations. Children should therefore be taught that fluent, confident, or friendly responses do not necessarily guarantee accuracy, intention, understanding, or moral responsibility.

In conclusion, children's interactions with artificial intelligence have already moved beyond purely instrumental technology use and entered a moral and relational domain. Age-appropriate digital moral literacy should be incorporated into elementary education so that children can distinguish technical error from intentional deception, evaluate AI-generated information critically, protect personal information, and understand the difference between programmed responsiveness and genuine emotion. Parents and teachers should actively mediate children's use of conversational AI, while designers should employ transparent uncertainty statements, understandable explanations, corrective mechanisms, and developmentally appropriate language. Ethical chatbot design and guided education are essential for ensuring that children benefit from artificial intelligence without developing misleading assumptions about truthfulness, trust, emotion, and moral agency.

References

- Arnis, A., Rustina, Y., Allenidekania, A., & Darari, F. (2024). The Role of Chatbots in Paediatric Chronic Disease Management: Trends, Findings, and Future Recommendations. *Malahayati International Journal of Nursing and Health Science*, 7(10), 1227-1235. <https://doi.org/10.33024/minh.v7i10.564>
- Ayelaagbe, S. O. (2025). Adoption of Chatbots in Nigerian Higher Education: Perceptions and Academic Utilization at Adeyemi Federal University of Education, Ondo, Ondo State, Nigeria. *Amjd*, 5(1), 170-177. <https://doi.org/10.59568/amjd-2025-5-1-16>
- Chaves, A. P., Egbert, J., Hocking, T. D., Doerry, E., & Gerosa, M. A. (2022). Chatbots Language Design: The Influence of Language Variation on User Experience With Tourist Assistant Chatbots. *Acm Transactions on Computer-Human Interaction*, 29(2), 1-38. <https://doi.org/10.1145/3487193>
- Hong, H., & Shin, S. (2024). Effects of the Use of a Conversational Artificial Intelligence Chatbot on Medical Students' Patient-Centered Communication Skill Development in a Metaverse Environment. *Journal of Medicine and Life Science*, 21(3), 92-101. <https://doi.org/10.22730/jmls.2024.21.3.92>

- Jamir, I., & Konyak, H. (2025). Perceptions of Undergraduate Arts Students of Nagaland Towards the Use of Chatbots in Education. *International Journal for Multidisciplinary Research*, 7(2). <https://doi.org/10.36948/ijfmr.2025.v07i02.39372>
- Jung, J., & Kim, H. S. (2025). Instructors' Perceptions and Utilization of Chatbots for Korean Language Education. *Korean Association for Learner-Centered Curriculum and Instruction*, 25(6), 481-497. <https://doi.org/10.22251/jlcci.2025.25.6.481>
- Kimll, J. T., & Choi, D. Y. (2023). The Impact of Chatbot Human Characteristics on User Satisfaction and Loyalty: Focusing on the Moderating Effect of Social Presence. *Asia-pacific Journal of Convergent Research Interchange*, 9(11), 271-289. <https://doi.org/10.47116/apjcri.2023.11.23>
- Kuhail, M. A., Thomas, J., Alamlawi, S., Shah, S. J. H., & Thornquist, E. (2022). Interacting With a Chatbot-Based Advising System: Understanding the Effect of Chatbot Personality and User Gender on Behavior. *Informatics*, 9(4), 81. <https://doi.org/10.3390/informatics9040081>
- Lê, X. C. (2023). Nghiên Cứu Các Yếu Tố Ảnh Hưởng Đến Ý Định Sử Dụng Chatbot Của Khách Hàng: Góc Nhìn Tại Việt Nam. *Tạp Chí Khoa Học Đại Học Mở Thành Phố Hồ Chí Minh - Kinh Tế Và Quản Trị Kinh Doanh*, 18(5), 33-47. <https://doi.org/10.46223/hcmcoujs.econ.vi.18.5.2407.2023>
- Li, B., Yao, R., & Nan, Y. (2024). How Does Anthropomorphism Promote Consumer Responses to Social Chatbots: Mind Perception Perspective. *Internet Research*. <https://doi.org/10.1108/intr-04-2024-0583>
- Mohanty, N. (2022). CHEC-UP: A Digital Intervention to Reduce Disparities in Well-Child and Immunization Completion in Community Health. *Telehealth and Medicine Today*, 7(5). <https://doi.org/10.30953/thmt.v7.375>
- Nicolescu, L., & Tudorache, M. T. (2022). Human-Computer Interaction in Customer Service: The Experience With AI Chatbots—A Systematic Literature Review. *Electronics*, 11(10), 1579. <https://doi.org/10.3390/electronics11101579>
- Onat, O., & Gülseçen, S. (2023). Reducing Perceived Transactional Distance in Distance Education: The Impact of the Chatbot. *International Journal of Technology in Education and Science*, 7(4), 483-499. <https://doi.org/10.46328/ijtes.511>
- Ou, M., Ho, S. S., & Wijaya, S. A. (2025). Harnessing AI to Address Misinformation on Cultivated Meat: The Impact of Chatbot Expertise and Correction Sidedness. *Science Communication*. <https://doi.org/10.1177/10755470251315097>
- Parsakia, K. (2023). The Effect of Chatbots and AI on The Self-Efficacy, Self-Esteem, Problem-Solving and Critical Thinking of Students. *Health Nexus*, 1(1), 71-76. <https://doi.org/10.61838/hn.1.1.14>
- Pizzi, G., Vannucci, V., Mazzoli, V., & Donvito, R. (2023). I, Chatbot! The Impact of Anthropomorphism and Gaze Direction on Willingness to Disclose Personal Information and Behavioral Intentions. *Psychology and Marketing*, 40(7), 1372-1387. <https://doi.org/10.1002/mar.21813>
- Pupong, K., Hunsrisakhun, J., Pithpornchaiyakul, S., & Naorungroj, S. (2025). Development of Chatbot-Based Oral Health Care for Young Children and Evaluation of Its Effectiveness, Usability, and Acceptability: Mixed Methods Study. *Jmir Pediatrics and Parenting*, 8, e62738-e62738. <https://doi.org/10.2196/62738>
- Qiu, W., Chit, L. S., Jamil, N. B., Thway, M., Hwee, S. N. S., Zhang, L., Lim, F. S., & Lai, J. W. (2024). A Systematic Approach to Evaluate the Use of Chatbots in Educational Contexts: Learning Gains, Engagements and Perceptions. <https://doi.org/10.35542/osf.io/7yga3>
- Rostami, M., & Navabinejad, S. (2023). Artificial Empathy: User Experiences with Emotionally Intelligent Chatbots. *AI and Tech in Behavioral and Social Sciences*, 1(3), 19-27. <https://doi.org/10.61838/kman.aitech.1.3.4>
- S.Vishaal, & M.V.Sathish. (2023). Impact on Consumer Perception Towards AI on Chatbot: A Study With Reference to FourWheelers in Chennai City. *JDMS*, 11(19), 183-192. <https://doi.org/10.53422/jdms.2024.111920>
- Saadati, S. A., & Saadati, S. M. (2023). The Role of Chatbots in Mental Health Interventions: User Experiences. *Aitechbesosci*, 1(2), 19-25. <https://doi.org/10.61838/kman.aitech.1.2.4>
- Septiyanti, N. D., Luthfi, M. I., & Darmawansah, D. (2024). Effect of Chatbot-Assisted Learning on Students' Learning Motivation and Its Pedagogical Approaches. *Khazanah Informatika Jurnal Ilmu Komputer Dan Informatika*, 10(1), 69-77. <https://doi.org/10.23917/khif.v10i1.4246>
- Shin, D., & Lee, J. H. (2024). Moving From Off-the-Shelf Chatbots to a User-Designed Bespoke L2 Chatbot. *LLT*, 28(1), 1-14. <https://doi.org/10.64152/10125/73581>
- Simonyi, G., & Major, L. (2024). Who Is Lying and Who Is Telling the Truth? The Role of Education in the Development of Understanding of Lies. *Nastava I Vaspitanje*, 73(3), 363-376. <https://doi.org/10.5937/nasvas2403363s>
- Su, Y., Luo, M., & Zhong, C. (2025). To Chat or Not: Pre-Service English Teachers' Perceptions of and Needs in Chatbot's Educational Application. *Sage Open*, 15(1). <https://doi.org/10.1177/21582440251321853>
- Suzuki, H., Zhang, J., Kim, D. D., Sagae, K., DeVon, H. A., & Fukuoka, Y. (2024). AI or Human? Message Humanness Predicts Perceiving AI as Human: A Secondary Data Analysis of the HeartBot Study (Preprint). <https://doi.org/10.2196/preprints.67717>

- Wei, X., Chen, X., Yue, L., & Liao, P. (2024). Parents' Information Needs and Perceptions of Chatbots for Self-Medicating Their Children: A Descriptive Qualitative Study. <https://doi.org/10.21203/rs.3.rs-5266454/v1>
- Wester, J., Pohl, H., Hosio, S., & Berkel, N. v. (2024). "This Chatbot Would Never...": Perceived Moral Agency of Mental Health Chatbots. *Proceedings of the Acm on Human-Computer Interaction*, 8(CSCW1), 1-28. <https://doi.org/10.1145/3637410>
- Xu, Y., Bradford, N., & Garg, R. (2023). Transparency Enhances Positive Perceptions of Social Artificial Intelligence. *Human Behavior and Emerging Technologies*, 2023, 1-15. <https://doi.org/10.1155/2023/5550418>