

CLAUDE: РЕВОЛЮЦИОНИЗИРАНЕ НА ВЗАМОДЕЙСТВИЕТО МЕЖДУ ЧОВЕК И ИЗКУСТВЕН ИНТЕЛЕКТ ЧРЕЗ КОМПЮТЪРНО ЗРЕНИЕ

Деян Петков Софийков
Бургаски свободен университет

CLAUDE: COMPUTER VISION CREATES REVOLUTION IN HUMAN-AI INTERACTION

Deyan Petkov Sofiykov
Burgas Free University

Abstract: Claude AI, developed by Anthropic, is an advanced chatbot introduced in March 2023 and enhanced through several iterations, with Claude 3.5 released in October 2024. By enabling computer vision and control, Claude can autonomously execute tasks traditionally performed by humans, such as clicking, typing, and navigating applications. This innovation holds transformative potential for industries dependent on repetitive tasks, including finance, legal, and customer service. Despite current limitations and ethical considerations, Claude's capability to integrate directly into digital work environments represents a new paradigm in human-AI collaboration, advancing both efficiency and interactivity.

Key words: Claude, Artificial Intelligence, Human-AI Interaction.

1. Въведение към Claude AI

Claude AI е усъвършенстван чатбот, разработен от Anthropic и е предназначен да генерира отговори на различни входове, подобно на ChatGPT. Пуснат е през Март 2023 г. и оттогава е претърпял няколко подобрения, като последната версия е Claude 3, пусната през Март 2024 г. [1].

При тестове може да се види ясно способността на Claude за компютърно управление с пускането на Claude 3.5 Sonnet през Октомври 2024 г. [2]. Тази функция позволява на Claude да взаимодейства с компютрите автономно и да изпълнява различни задачи, които до скоро можеха да се изпълняват само от човек - оператор [2]. Claude може да мести курсора, да кликва върху бутони, да въвежда текст и да навигира между приложения без човешка намеса [3] За разлика от традиционните инструменти за автоматизация, Claude не се ограничава до конкретни работни процеси или софтуерни програми. Може да се адаптира към различни задачи и интерфейси. Тази възможност може да революционизира индустрии, които разчитат в голяма степен на повтарящи се и монотонни задачи, включващи множество приложения и раздели, като тези при финансите, правните услуги и поддръжка на клиенти. [3]

Въпреки това е важно да се отбележи, че тази способност е все още в ранен етап и е изправена пред няколко предизвикателства. Например моделът на Claude се среща с трудности при определени задачи, които хората намират за тривиални, като превъртане или мащабиране. [3]

Anthropic е въвела предпазни мерки за справяне с потенциални рискове, включително класификатори за откриване на злоупотреба и предотвратяване на взаимодействия с чувствителни уебсайтове. [3]

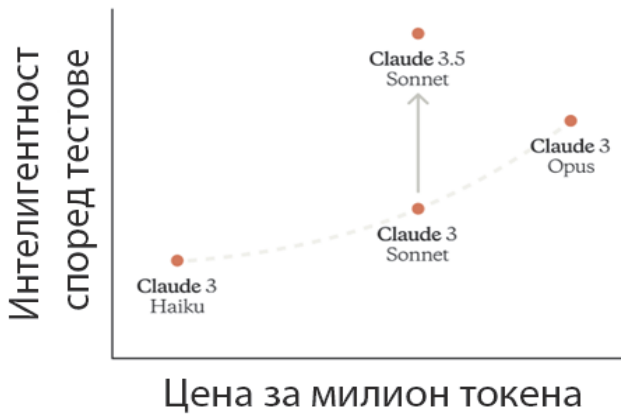
Въпреки тези предизвикателства, потенциалните ползи от възможностите за компютърно управление на Claude са значителни. С развитието на технологията, това може да доведе до значителни подобрения в анализа на данни, визуализацията и взаимодействията с потребителския интерфейс, като може дори потенциално подобри достъпността в различни сектори [3].

2. Технически черти и възможности

Claude използва API, който позволява на разработчиците да интегрират функционалност за използване на компютър в своите приложения. Той има достъп до набор от функции, които могат да манипулират среда на работния плот на компютъра. Тези функции се наричат „инструменти за използване на компютър“ респективно. Claude взаимодейства с компютрите чрез агентен цикъл, който представлява повторение на стъпки, при които Claude отговаря с искане за използване на инструмент, а приложението отговаря с резултатите от оценката на това искане. Когато използва инструменти за използване на компютър, Claude получава специална системна подкана, която позволява тази функционалност [4].

Claude 3.5 Sonnet постави нов стандарт в производителността на модели за изкуствен интелект, надминавайки своите предшественици, включително модела от най-високо ниво на Anthropic, Claude 3 Opus, в различни показатели за оценка. По-специално, той постига това, като същевременно запазва скоростта и рентабилността на оригиналния модел Claude 3 Sonnet. Значителното предимство на Claude 3.5 се крие в съотношението цена-качество. Потребителите вече могат да придобият интелигентност и скорост, сравними с тези на Claude 3 Opus на малка част от цената, като Claude 3.5 е приблизително 80% по-евтин от своя предшественик. Тази последна версия подчертава ангажимента на Anthropic да предоставя решения за изкуствен интелект, пригодени за различни потребителски нужди, предлагайки баланс между разширени възможности, скорост на обработка и достъпност [5]. Отличието на Claude 3.5 може да се види на фигура 1, сравнен с останалите свои версии.

Според оценките на Anthropic, този най-нов модел превъзхожда конкуренти като GPT-4o на OpenAI и Gemini 1.5 Pro на Google в почти всеки бенчмарк.



Фигура 1. Графика за сравнение с конкуриращите версии на Claude

По-конкретно, Claude 3.5 постигна най-високо класиране в няколко ключови области:

- Той поставя нови стандарти за експертни познания на основно ниво (MMLU) и на експертно разсъждение (GPQA).
- Моделът се отличава с умения за кодиране, задминавайки други модели, включително своя предшественик Claude 3.
- Claude 3.5 показва забележими подобрения в схващането на нюанси, хумор и сложни инструкции.
- Той демонстрира изключителна способност за писане на висококачествено съдържание с естествен, релативен тон.
- Моделът работи с два пъти по-висока скорост от Claude 3, като същевременно запазва ефективността на разходите.
- При вътрешни оценки на кодирането, Claude 3.5 решава 64% от проблемите, значително превъзхождайки Claude 3, който решава само 38%.
- Той обработва кодови преводи с лекота и показва особена ефективност за актуализиране на наследени приложения и мигриране на кодови бази.

Claude 3.5 също надминава своя предшественик по отношение на стандартните показатели за визуализация, особено забележимо за задачи, изискващи визуално разсъждение като интерпретиране на диаграми и графики. Резултатите от сравненията на Claude 3.5 с останалите модели може да се види на таблица 1.

Claude има достъп до набор от предварително дефинирани инструменти, както беше споменато, предоставени от Anthropic, като например:

- 1) { "type": "computer_20241022", "name": "computer" }
- 2) { "type": "text_editor_20241022", "name": "str_replace_editor" }
- 3) { "type": "bash_20241022", "name": "bash" } ¹

Първата команда се отнася до общия инструмент за използване на компютъра, предоставен от Claude. Тя позволява на Клод да взаимодейства с цялата работна среда, включително преместване на курсора на мишката, клавиатурата, отваряне и

затваряне на прозорци, превключване между приложения и изпълнение на основните навигационни задачи.

Задача	Claude 3.5 Sonnet	Claude 3 Opus	GTP-4o	Gemini 1.5. Pro
Разсъждаване на висше ниво (GPAQ, Diamond)	59.4% (0-shot CoT)	50.4% (0-shot CoT)	53.6% (0-shot CoT)	-
Основно ниво на знания (MMLU)	88.7% (5-shot) 88.3% (0-shot CoT)	86.8% (5-shot) 85.7% (0-shot CoT)	88.7% (0-shot CoT)	85.9% (5-shot)
Код (HumanEval)	92.0% (0-shot)	84.9% (0-shot)	90.2% (0-shot)	84.1% (0-shot)
Многоезична математика (MGSM)	91.6% (0-shot CoT)	90.7% (0-shot CoT)	90.5% (0-shot CoT)	87.5% (8-shot)
Разсъждаване върху текст (DROP, F1 резултати)	87.1% (3-shot)	83.1% (3-shot)	83.4% (3-shot)	74.9% (Променилив брой опити)
Смесени оценки (BIG-Bench-Hard)	93.1% (3-shot CoT)	86.8% (3-shot CoT)	-	89.2% (3-shot CoT)
Решаване на математически задачи (MATH)	71.1% (0-shot CoT)	60.1% (0-shot CoT)	76.6% (0-shot CoT)	67.7% (4-shot)
Математика на начално ниво (GSM8K)	96.4% (0-shot CoT)	95.0% (0-shot CoT)	-	90.8% (11-shot)

Таблица 1. Сравнение по различни показатели с останалите водещи модели за изкуствен интелект

Втората команда се отнася до функционалността на текстови редактор. По-конкретно е специализиран инструмент за замяна на текст в текстов редактор. Някои възможни функции могат да включват намиране и заместване на текст в документ, замяна на текст в конкретен файл и боравене с текстови операции в конкретно приложение за текстов редактор. Името „str_replace_editor“ подсказва, че е фокусиран върху операции за заместване на текст в контекст на редактор.

Последната команда се отнася за инструмент на Bash shell. Bash (Bourne Again Shell) е Unix shell и интерпретатор на команди.

За разлика от традиционните инструменти за автоматизация, Claude не се ограничава до конкретни работни процеси или софтуерни програми. Може да се адаптира към различни задачи и интерфейси. Разработчиците, обаче, трябва изрично да изпълнят инструментите, които са под формата на команди, и да върнат резултатите на

Claude. Моделът не изпълнява автоматично инструментите, като предпазна мярка. Относно самата виртуална среда, Claude работи в така нареченият пясъчник (sandbox), отделна от външни ресурси. Разбира се Claude не е безплатен, заявките за използване на компютър се таксуват подобно на други заявки за Claude API услуги, с допълнителни токени за въвеждане, необходими за конкретни инструменти [4].

Claude е достъпен към момента, за всеки, който иска да го изпробва. За да се използва е нужен акаунт в сайта на Anthropic, от където може да се получи API ключ, който играе ролята на парола за услугите на Claude. Към момента има 4 модела на Claude, които могат да се използват, а именно Claude 1, Claude 2, Claude 3 и последната версия Claude 3.5. Версиите се различават със скорост и възможности респективно. За да се използва, който и да е от моделите, трябва да се използва JavaScript или Python по избор на потребителя и да се приложи съответният API ключ. От тази стъпка нататък модела може да се използва както потребителят иска, чрез API команди в кода. Препоръчително е да се започне с по-прости и леки заявки като за начало, за да се тества стабилността на модела.

3. Предиизвикателства и ограничения

Claude AI показва впечатляващи възможности, особено при генериране на код и технически задачи, но все още е изправен пред някои ограничения. Подобно на други модели за изкуствен интелект, Прекъсването на знанията на Claude обикновено е на няколко месеца, за разлика от последните версии на ChatGPT, който може да търси в интернет актуална информация. Подобно на генеративните модели за изкуствен интелект, Claude може да произведе халюцинации или неправилна информация, особено в теми, които са извън своите данни за обучение. Също така безплатната версия на Claude има лимит на токените за използване, след което напълно блокира, за разлика от ChatGPT, който преминава към по-нисък модел. Въпреки че Claude се отличава в специализирани области като генериране на код, това може да не съответства на гъвкавостта на ChatGPT в различни. Някои потребители съобщават за затруднения с определени действия като превъртане или мащабиране, което показва ограничения в сложните визуални взаимодействия на модела. [6]

Claude работи в компютърна среда, която е отделна от външни ресурси, което може да ограничи неговите възможности в сравнение с пълния достъп до системата. Anthropic са въвели предпазни мерки, включително класификатори за откриване на злоупотреби и предотвратяване на взаимодействия с чувствителни уебсайтове, които понякога могат да попречат на законните употреби. За задачи с голям риск, като тези, включващи чувствителна информация, Anthropic препоръчват човешки надзор поради ограниченията на технологията. [6]

Въпреки тези ограничения Claude AI изглежда обещаващ, особено в технически области като генерирането на код. Неговият фокус върху качеството и иновацията го отличава от някои конкуренти, което вешае, че текущото му развитие ще се справи с много от горе описаните предиизвикателства в бъдеще [6].

4. Приложения в индустриите

Claude 3.5 Sonnet бележи важна опорна точка в технологията с изкуствен интелект, позволявайки на ѝ да взаимодейства с компютрите по начин, подобен на човека, като кликане, въвеждане на текст и изпълнение на задачи въз основа на екранни снимки. Изкуственият интелект преминава отвъд специализираните инструменти или

среди, за да взаимодейства директно със софтуера, имитирайки човешки асистенти в ежедневните работни процеси [7]. Потенциалните приложения включват:

- Обслужване на клиенти
- Работа с финанси
- Управление на проекти

5. Заключение

Тъй като изкуственият интелект става по-автономен, той може значително да повиши ефективността в различни сектори чрез справяне с рутинни задачи. Трансформация на динамиката на работата се представя силно в ангажираността на изкуственият интелект към софтуера подобно на нивото на човек – потребител, като без персонализирани интеграции променя начина, по който мислим за помощта на изкуственият интелект, преминавайки към активно дигитално сътрудничество. Съответно щом като изкуственият интелект поема повече отговорности, то респективно ще има опасения относно надеждността, сигурността и човешкото взаимодействие с все по-автономни системи. Промяната може да доведе до промени в организационните структури и да повдигне въпроси относно отчетността и етиката на изкуственият интелект. Въпреки че съществуват текущи ограничения, траекторията предполага, че Claude ще става по-умен, по-бърз и по-способен да действа като човек в цифровото работно пространство. [7]

Успехът на Claude може да зависи от ефективните стратегии за разпространение, подобно на случая с Microsoft Teams, който надмина Slack чрез силната си и стабилна интеграция. [7]

Източници:

- [1] Grammarly, „Claude AI 101: What It Is and How It Works“, Grammarly, 11 04 2024. [Онлайн]. Available: <https://www.grammarly.com/blog/ai/what-is-claude-ai/>. [Отваряно на 04 11 2024].
- [2] S. Mohanty, „Claude 3.5 Sonnet (New): Pioneering the Future of AI with Computer Control Capabilities“, Medium, 24 10 2024. [Онлайн]. Available: <https://medium.com/@subhraj07/claude-3-5-sonnet-new-pioneering-the-future-of-ai-with-computer-control-capabilities-37a6ff9f9033>. [Отваряно на 04 11 2024].
- [3] Anthropic, „Introducing computer use, a new Claude 3.5 Sonnet, and Claude 3.5 Haiku“, Anthropic, 22 10 2024. [Онлайн]. Available: <https://www.anthropic.com/news/3-5-models-and-computer-use>. [Отваряно на 04 11 2024].
- [4] Anthropic, „Computer use (beta)“, 22 10 2024. [Онлайн]. Available: <https://docs.anthropic.com/en/docs/build-with-claude/computer-use>. [Отваряно на 04 11 2024].
- [5] D. Poccia, „AWS“, Amazon, 20 06 2024. [Онлайн]. Available: <https://aws.amazon.com/blogs/aws/anthropics-claude-3-5-sonnet-model-now-available-in-amazon-bedrock-the-most-intelligent-claude-model-yet/>. [Отваряно на 13 11 2024].
- [6] L. Clairet, „Claude AI vs ChatGPT: advantages, limitations and feedback“, 23 08 2024. [Онлайн]. Available: <https://www.digidop.fr/en/blog/claude-vs-chatgpt>. [Отваряно на 05 11 2024].
- [7] Spearhead, „AI Takes Control – The Game-Changing Evolution of Anthropic’s Claude“, 24 10 2024. [Онлайн]. Available: <https://spearhead.so/ai-takes-control-the-game-changing-evolution-of-anthropics-claude/>. [Отваряно на 05 11 2024].