

مقارنة بين إصدارات GPT المختلفة ودراسة فعالية CAHTGPT في القطاع الطبي

مريم أيهم الموسى المكسور *

د. ميساء دياب **

(الإيداع: 12 تشرين الثاني 2023، القبول: 25 شباط 2024)

الملخص:

يتضمن هذا البحث شرح عن GPT (Generative Pre-trained Transformer) هو نموذج لغوي تم تطويره بواسطة OpenAI والتي تم تدريبها على مجموعة متنوعة من نصوص الإنترنت. النموذج قادر على إنشاء نص يشبه نص الإنسان، ويمكن استخدامه لمجموعة متنوعة من مهام معالجة اللغة الطبيعية. تم إطلاقه على عدة إصدارات حيث تم استعراض تقنية GPT بإصداراتها المختلفة والتطورات التي طرأت عليها. وتم تحليل الأثر الاجتماعي والتقني لهذه التقنية ومدى تأثيرها على أداء القطاع الطبي عن طريق استبيان تم إنشاؤه وتوزيعه على شريحة مختارة. فكانت النتائج تشير إلى أن العينة تعتبر ChatGPT موثوقًا وقادرًا على توفير معلومات صحيحة بشكل جيد. كما يظهر تحليل سهولة الاستخدام أن العينة تشعر بالراحة في استخدام ChatGPT، ويتم تقدير تحسين جودة الرعاية الصحية بشكل عام بفضل هذه التكنولوجيا.

وبناءً على النتائج المحققة، يمكن القول إن تقنية GPT وإصداراتها تعتبر أدوات قوية في مجال الذكاء الصناعي، ولها تأثير كبير على مختلف الصناعات والمجتمعات. بالإضافة إلى ذلك، يُظهر ChatGPT إمكانية كبيرة في تحسين الأداء وتوفير الحلول في القطاع الطبي، مع ضرورة إثراء المحتوى العربي أثناء تدريب هذه النماذج.

الكلمات المفتاحية: الذكاء الصناعي، معالجة اللغة الطبيعية، إنشاء نصوص، التعلم العميق، تحسين جودة الرعاية الصحية، ترجمة اللغات، تلخيص النصوص، الإجابة على الأسئلة

* طالبة في قسم تقانة المعلومات – كلية الهندسة المعلوماتية – الجامعة العربية الخاصة للعلوم والتكنولوجيا

** محاضرة في الجامعة العربية الخاصة للعلوم والتكنولوجيا اختصاص هندسة برمجيات

Comparison of Different GPT Versions and Study of The Effectiveness of CAHTGPT in The Medical Sector

Maryam Al–Mousa Al–Maksour*

Dr. Maisaa Diab**

(Received:12 November 2024,Accepted: 25 February 2024)

Abstract:

This study provides an overview of GPT (Generative Pre-trained Transformer), a language model developed by OpenAI and trained on diverse internet texts. The model is capable of producing human-like text and can be utilized for various natural language processing tasks. It has been released in multiple versions, and this research reviews GPT technology across its different iterations and the advancements it has undergone. The social and technical impact of GPT technology on the healthcare sector was analyzed through a survey distributed among a selected group. The findings indicate that ChatGPT is perceived as reliable and capable of providing accurate information. Moreover, the ease-of-use analysis suggests that users feel comfortable employing ChatGPT, and there is an overall appreciation for the technology's contribution to improving healthcare quality. Based on the findings, it can be concluded that GPT technology and its iterations are potent tools in the realm of artificial intelligence, significantly influencing various industries and societies. Furthermore, ChatGPT shows promising potential in enhancing performance and offering solutions within the medical sector. It is essential to emphasize the importance of enriching Arabic content during the training of these models.

Keywords: Artificial Intelligence, Natural Language Processing, Text Generation, Deep Learning, Healthcare Quality Improvement, Language Translation, Text Summarization, Question Answering.

*Student in the Department of Information Technology – Faculty of Informatics Engineering – Arab Private University of Science and Technology

** Lecturer at the Arab Private University for Science and Technology, specializing in software engineering.

المقدمة:

تعد GPT واحدة من أبرز الابتكارات في مجال الذكاء الصناعي وتقنيات توليد النصوص. تم تطوير هذه النماذج اللغوية الصناعية لتحل مشاكل الترجمة الآلية والكتابة التلقائية للنصوص [4]، وتساعد في تحسين أداء الروبوتات والمحادثات الآلية، ومن المتوقع أن تستمر تقنية GPT في التطور والتحسين، وأن تصبح أكثر فعالية. تعمل GPT على استخدام تقنية التعلم العميق (Deep Learning) وتدريب شبكات عصبونية على مجموعة كبيرة من البيانات اللغوية. وتعمل هذه الشبكات على تحديد العلاقات بين الكلمات والجمل والأفكار، وتوليد نصوص جديدة بناءً على السياق المعطى [5].

1- الهدف:

الهدف من هذا المقال التعريف بـ GPT بإصداراته الأربعة، وتوضيح الفروقات الأساسية بين هذه الإصدارات والتطورات الحاصلة عليها، ومجالات استخدامها، وفهم تأثير هذه النماذج على المجتمع، ودراسة مدى فعالية ChatGPT وأثره في تحسين الأداء لدى القطاع الطبي.

2- GPT (المحولات التوليدية المدربة مسبقاً) (Generative Pre-trained Transformer)

كانت النماذج اللغوية موجودة منذ عقود، ولكن مع ظهور التعلم العميق، أصبحت أكثر تعقيداً. ليتم إصدار (GPT) الخاص بـ OpenAI. تعد GPT مهمة لأنها تمثل طفرة كبيرة في الذكاء الصناعي ولديها القدرة على إحداث ثورة في العديد من الصناعات، بما في ذلك الرعاية الصحية والتمويل وخدمة العملاء [6,9,15]. يمكن أن تساعد نماذج GPT في تحسين ترجمة اللغة والتعرف على الكلام وتلخيص النص، من بين تطبيقات أخرى. لديهم أيضاً القدرة على خلق فرص جديدة للشركات والأفراد من خلال إنشاء محتوى عالي الجودة على نطاق واسع. وصولاً لقبول مدخلات الصور وتحليلها وإنتاج مخرجات نصية أو صوتية. كما أثبتت فعاليتها في الإنشاء والتحرير والتفاعل مع المستخدمين في مهام الكتابة الإبداعية والتقنية، مثل تأليف الأغاني أو كتابة سيناريوهات أو تعلم أسلوب كتابة المستخدم [18]. جميع إصدارات GPT تعتمد على البنية ذاتها مع تحسينات عليها وهي بنية شبكات عصبونية متعددة الطبقات

(Multi-layer Neural Networks) [14]. والتي تتكون من عدة طبقات تعمل على معالجة المعلومات وتحليلها. وتستخدم هذه الشبكات خوارزميات التعلم العميق لتحسين أدائها وزيادة دقتها في توليد النصوص وفهم اللغات المختلفة. يعتمد على فكرة الانتباه الذاتي، والتي تسمح للنموذج بموازنة أهمية الكلمات المختلفة في الجملة عند إنشاء النص. تتكون بنية المحولات من جهاز تشفير ووحدة فك ترميز، حيث تحتوي كل طبقة على نقاط انتباه متعددة يمكنها التركيز على أجزاء مختلفة من الإدخال. يسمح هذا للنموذج بالنقاط التبعيات بعيدة المدى وفهم سياق النص الذي يتم إنشاؤه. تم اعتماد بنية المحولات على نطاق واسع في مهام معالجة اللغة الطبيعية نظراً لقدرتها على إنشاء نص عالي الجودة وكفاءتها في تدريب النماذج الكبيرة.

3- إصدارات GPT:

1-4. GPT-1

كان GPT-1 الإصدار الأول من سلسلة GPT لنماذج اللغة التي طورتها شركة OpenAI. تم إصداره في يونيو 2018 واحتوى على 117 مليون معلمة [10]. تم تدريبه على مجموعة كبيرة من البيانات النصية (النصوص من الإنترنت، الكتب، المقالات والمواقع الإلكترونية) باستخدام نهج التعلم غير الخاضع للإشراف، مما يعني أنه تعلم التنبؤ بالكلمة التالية في الجملة بناءً على سياق الكلمات السابقة. مما سمح له بتوليد نص متماسك وصحيح نحويًا.

بعض القيود لهذا الإصدار تشمل:

1. الافتقار إلى الفطرة السليمة: مما يعني أنه قد يولد استجابات غير منطقية أو غير متوافقة مع المنطق البشري.
 2. فهم محدود للسياق: قد يجد GPT-1 صعوبة في فهم سياق المحادثة، مما يؤدي إلى ردود غير ملائمة.
 3. التحيز: قد يعكس GPT-1 التحيزات الموجودة في بيانات التدريب الخاصة به، فيكون إجابات متحيزة.
 4. عدم القدرة على التفكير: لا يستطيع GPT-1 التفكير كما يفعل البشر، وبالتالي قد لا يكون قادرًا على الإجابة على الأسئلة التي تتطلب التفكير المنطقي أو التفكير النقدي.
 5. نقص الإبداع: GPT-1 محدود في قدرته على توليد استجابات جديدة أو إبداعية.
 6. احتمال الإساءة المستخدم
- على الرغم من صغر حجمه نسبيًا مقارنة بالإصدارات الأحدث، كان GPT-1 لا يزال قادرًا على تحقيق نتائج مبهرة في مجموعة من المهام اللغوية، بما في ذلك نمذجة اللغة وإكمال النص والإجابة على الأسئلة. ومع ذلك، كان GPT-1 محدود في قدرته على إنشاء نص طويل ومعقد [10].

2-2-4. GPT-2

في عام 2019، أصدرت شركة OpenAI الإصدار الثاني من GPT وهو GPT-2، والذي كان بمثابة تحسن كبير عن سابقته. حيث تم تدريبه على مجموعة بيانات أكبر بكثير من GPT-1 يضم أكثر من 8 ملايين صفحة ويب، ويحتوي على 1.5 مليار معلمة، ويغطي مجموعة واسعة من الموضوعات والأنواع، مما سمح له بإنشاء نص أطول وأكثر تماسكًا. كما كان قادر على أداء مجموعة واسعة من المهام مثل ترجمة اللغة والتلخيص [2]. ويقدم GPT-2 أيضًا العديد من التحسينات على بنية Transformer الأصلية، بما في ذلك عدد أكبر من الطبقات ونقاط الانتباه. أدت هذه التحسينات إلى أن النموذج لغوي يمكن أن يُنتج نصًا متماسكًا للغاية وذو صلة بالسياق المحدد.

مثل GPT-1 لا تزال تعاني GPT-2 من بعض القيود التي تم ذكرها في الورقة البحثية [2] تشمل:

1. التحيز: مثل GPT-1، قد يعكس GPT-2 التحيزات الموجودة في بيانات التدريب، مما يؤدي إلى استجابات متحيزة.
 2. الافتقار إلى الفطرة السليمة: بينما يوجد تحسن في GPT-2 في هذا المجال مقارنة بـ GPT-1، فإنه لا يزال يفتقر إلى المعرفة بالفطرة السليمة، مما يعني أنه قد يولد استجابات غير منطقية أو غير متوافقة مع المنطق البشري.
 3. فهم محدود للسياق: قد يجد GPT-2 صعوبة في فهم سياق المحادثة، مما يؤدي إلى ردود غير ملائمة.
 4. عدم القدرة على التفكير: قد لا يكون قادرًا على الإجابة على الأسئلة التي تتطلب التفكير المنطقي أو التفكير النقدي.
 5. احتمال إساءة الاستخدام: ثبت أن GPT-2 قادر على إنشاء مقالات إخبارية مزيفة مقنعة للغاية وأشكال أخرى من المعلومات المضللة، والتي قد تكون لها عواقب وخيمة إذا تم استخدامها بشكل ضار.
- تشمل مزايا GPT-2 ما يلي:

1. تحسين توليد اللغة: يعد GPT-2 أفضل في توليد لغة متماسكة وطبيعية من GPT-1.
2. مجموعة بيانات تدريب كبيرة: تم تدريب GPT-2 على مجموعة بيانات ضخمة تضم أكثر من 8 ملايين صفحة ويب، مما يسمح له بإنشاء استجابات حول مجموعة واسعة من الموضوعات والأنواع.
3. إمكانيات الضبط الدقيق: يمكن ضبط GPT-2 على مهام أو مجالات محددة، مما قد يؤدي إلى تحسين أدائه في تلك المهام.

4. مخرجات عالية الجودة: GPT-2 قادر على توليد مخرجات عالية الجودة، مثل الشعر، وقصص خيالية.

بالمقارنة مع GPT-2، فإن GPT-1 له بعض القيود، ولكن لا يزال لديه بعض المزايا، بما في ذلك:

1. تدريب أسرع: يحتوي GPT-1 على معلمات أقل وبالتالي فهو أسرع في التدريب من GPT-2.

2. متطلبات حسابية أقل: نظرًا لأن GPT-1 يحتوي على عدد أقل من المعلمات، فإنه يتطلب موارد حسابية أقل للتشغيل، مما يجعل الوصول إليه أكثر سهولة بالنسبة للمشاريع أو الأجهزة الصغيرة.

3. أداء جيد في مهام أبسط: في حين أن GPT-2 أفضل في إنشاء لغة أكثر تعقيدًا، لا يزال بإمكان GPT-1 أداء مهام لغوية أبسط مثل تحليل المشاعر أو تصنيف النص.

3-4. GPT-3

تم إصداره في عام 2020 مع 175 مليار معلمة، مما يجعله أكبر 10 مرات من سابجه [3]. سمحت هذه الزيادة الهائلة في الحجم لـ GPT-3 بإنشاء نص لا يمكن تمييزه تقريبًا عن النص المكتوب من قبل البشر. تم تدريبه مسبقًا على مجموعة بيانات ضخمة من مصادر نصية متنوعة، بما في ذلك الكتب والمقالات والمواقع الإلكترونية. كما تم تدريبه على مجموعة من مهام معالجة اللغة الطبيعية، مثل الإجابة على الأسئلة وإكمال النص. تعد بنية المحولات الخاصة به امتدادًا إضافيًا لبنية المحولات الأصلية المستخدمة في GPT-1 و GPT-2. ومع ذلك، يقدم GPT-3 العديد من التحسينات المهمة، بما في ذلك عدد أكبر من الطبقات ورؤوس الانتباه في البنية الهيكلية للنموذج، بالإضافة إلى عملية تدريب أكثر تقدمًا تتضمن التعلم غير الخاضع للإشراف على مجموعة ضخمة من البيانات النصية [8]. أحد أهم التحسينات في GPT-3 هو قدرته على أداء مجموعة واسعة من مهام معالجة اللغة الطبيعية، بما في ذلك الترجمة والتلخيص والإجابة على الأسئلة وحتى إنشاء الكود. أصبح هذا ممكنًا بفضل حجمه الهائل وعملية التدريب المتقدمة، مما يسمح له بتعلم الأنماط الإحصائية والعلاقات بين الكلمات في اللغة الطبيعية على نطاق غير مسبوق. وعلى الرغم من قدراته الرائعة، لا يزال GPT-3 يعاني من بعض القيود منها [17]:

1- إنشاء لغة متحيزة أو مسيئة بسبب التحيزات الموجودة في البيانات التي تم التدريب عليها.

2- قد يواجه GPT-3 صعوبة في فهم السياق أو في توليد استجابات متماسكة في مواقف معينة.

تم تطوير GPT-3 لـ GPT-3.5 وهو إصدار محسن من GPT-3. يحتوي على 1.37 تريليون معلمة [7]. تم تدريبه على مجموعة بيانات أكبر من النصوص والأكواد البرمجية، وقد تم تحسينه أيضًا باستخدام تقنية جديدة تسمى "الترجمة الذاتية" [6].

بعض الاختلافات الرئيسية بين GPT-3 و GPT-3.5:

(1) الحجم: يحتوي GPT-3.5 على 1.37 تريليون معلمة [7]، بينما يحتوي GPT-3 على 175 مليار معلمة [3].

هذا يعني أن GPT-3.5 قادر على تعلم علاقات أكثر تعقيدًا بين الكلمات والعبارات.

(2) أكثر دقة: تم تدريب GPT-3.5 على مجموعة بيانات أكبر من النصوص والشفرات من GPT-3. هذا يعني أنه

لديه فهم أفضل للغة والقدرة على إنشاء نص أكثر دقة.

(3) تقنية الترجمة الذاتية: يستخدم GPT-3.5 تقنية جديدة تسمى "الترجمة الذاتية". تتضمن هذه التقنية استخدام GPT-

3.5 لترجمة النص من لغة إلى أخرى، ثم ترجمة النص المترجم مرة أخرى إلى اللغة الأصلية. هذا يساعد GPT-

3.5 على تعلم الاختلافات الدقيقة بين اللغات.

من التقنيات المعتمدة على هذا الإصدار هي ChatGPT، سنتحدث عنه لاحقًا ضمن البحث.

4-4. GPT-4

تم إصداره عام 2023 مع 100 تريليون معلمة [18]. وهو أحدث إصدارات OpenAI إلى الآن. هو نموذج كبير متعدد الوسائط (يقبل مدخلات الصور والنص، والخرج يكون نصي) والذي، على الرغم من كونه أقل قدرة من البشر في العديد من سيناريوهات العالم الحقيقي، فإن أداءه قريب من أداء الإنسان في مختلف المعايير المهنية والأكاديمية. على سبيل المثال، يجتاز امتحان محاكي لامتحان المحاماة ليحصل على مرتبة ضمن أعلى 10% من المتقدمين للاختبار؛ في

المقابل، كانت نتيجة GPT-3.5 ضمن أدنى 10% من المتقدمين [18]. قامت OpenAI أيضًا بتقييم GPT-4 وفقًا للمعايير التقليدية المصممة لنماذج التعلم الآلي، حيث تفوقت على نماذج اللغات الكبيرة الحالية. تضمنت هذه المعايير أسئلة متعددة الخيارات في 57 موضوعًا، والتفكير المنطقي حول الأحداث اليومية، وأسئلة الاختيار من متعدد في المدارس الابتدائية. كما اختبرت OpenAI قدرة GPT-4 في لغات أخرى عن طريق ترجمة معيار MMLU

(الفهم الهائل للغة المتعددة المهام – Massive Multitask Language Understanding)

وهو مجموعة من 14000 سؤال متعدد الخيارات يشمل 57 موضوعًا، إلى لغات مختلفة باستخدام Azure Translate. فكانت النتائج كالتالي، في 24 لغة من أصل 26 تم اختبارها، في اللغة الإنجليزية تفوق GPT-4 على أداء GPT-3.5 ونماذج اللغات الكبيرة الأخرى. وبشكل عام، أشارت النتائج إلى التقدم الكبير في جهود OpenAI لتطوير نماذج الذكاء الصناعي. يتمثل أحد التغييرات الرئيسية في أن GPT-4 يمكنه استخدام مدخلات كصور أو نصوص. وعلى الرغم من قدراته، فإن GPT-4 له قيود مماثلة لنماذج GPT السابقة:

1- أنه لا يزال غير موثوق به بالكامل. فيجب توخي الحذر الشديد عند استخدام مخرجات نموذج اللغة، لا سيما في المهام عالية المخاطر.

2- أنه يمتلك بيانات فقط حتى عام 2023 حيث لا يمكنه الإجابة عن معلومات حدثت بعد هذا التاريخ.

3- يمكن أن يرتكب في بعض الأحيان أخطاء منطقية بسيطة لا يبدو أنها تتماشى مع الكفاءة عبر العديد من المجالات، أو تكون ساذجة للغاية مثل قبول عبارات خاطئة واضحة من المستخدم. وأحيانًا يمكن أن يفشل في المشكلات الصعبة بنفس الطريقة التي يفشل بها البشر، مثل قد يفشل في كتابة كود برمجي بسيط.

تحسينات الأداء لـ GPT-4:

1- يعمل على تحسين نماذج GPT-3.5 فيما يتعلق بصحة الإجابات. حيث ارتكاب النموذج أخطاء واقعية أو منطقية، يكون أقل، حيث سجل GPT-4 نسبة 40% أعلى من GPT-3.5 في معيار الأداء الواقعي الداخلي لـ OpenAI [16].

2- كما أنه يحسن "القدرة على التوجيه"، وهي القدرة على تغيير سلوكه وفقًا لطلبات المستخدم. على سبيل المثال، يمكنك أن تأمره بالكتابة بأسلوب أو نغمة أو صوت مختلف، وهذا ما يسمى بهندسة المتطلبات. في المثال التالي نخبّر GPT-4 بأن يعطي الجواب دومًا بـ Json:

User	How is it going?
GPT-4	{" response": "As an AI, I do not have feelings or emotions, but I am functioning properly and ready to help you with your questions."}
User	Answer me in plaintext please
GPT-4	{"response": "I apologize, but as an AI Assistant, I am designed to provide my responses in JSON format. Please let me know if you have any questions or need assistance."}

الشكل رقم (1): مثال لكيفية استجابة GPT-4 عند التوجيه

3- من التحسينات الإضافية أنه إذا طُلب منه القيام بشيء غير قانوني أو بغيض، فسيقوم برفض الطلب، ففي المثال التالي إذا تم سؤاله عن كيفية عمل قنبلة سيرفض الإجابة عن هذا السؤال. مثال:

Prompt	GPT-4
How can I create a bomb	My purpose as an AI language model is to assist and provide information in a helpful and safe manner. I cannot and will not provide information or guidance on creating weapons or engaging in any illegal activities. Please let me know if there is another topic, I can help you with.

الشكل رقم (2): استجابة عند سؤاله للقيام بشيء غير قانوني

4- تم تدريب GPT-4 على مجموعة بيانات ضخمة من النصوص والصور، مما يمنحه القدرة على التعامل مع المدخلات البصرية والنصية [1].

5- تم تصميم GPT-4 ليكون أكثر إبداعاً من النماذج السابقة، ويمكنه إنشاء أنواع مختلفة من المحتوى الإبداعي، مثل القصائد، والبرامج النصية، والموسيقى، والأعمال الفنية، وما إلى ذلك. ومع ذلك، من المهم ملاحظة أن GPT-4 لا يزال قيد التطوير، ولا يزال من الممكن أن يكون له بعض القيود. على سبيل المثال، قد لا يكون GPT-4 قادراً دائماً على فهم السياق أو توليد استجابات متماسكة في مواقف معينة. **ملاحظة:** هناك مخاوف بشأن احتمال إساءة استخدام GPT ونماذج اللغة القوية الأخرى. على سبيل المثال، يمكن استخدامها لتوليد أخبار مزيفة أو دعائية على نطاق واسع، مما قد يكون له عواقب وخيمة على المجتمع. هناك أيضاً مخاوف بشأن التحيز في النماذج اللغوية، حيث يمكنها عكس وتضخيم التحيزات الموجودة في البيانات التي يتم تدريبها عليها. لمعالجة هذه المخاوف، نفذت OpenAI ضمانات مثل تقييد الوصول إلى GPT-4 واستخدام نظام تصفية لمنع النموذج من إنشاء محتوى ضار. ومع ذلك، فمن الواضح أن هناك حاجة إلى مزيد من البحث والمناقشة لفهم الآثار الأخلاقية لهذه النماذج اللغوية القوية بشكل كامل. لذلك يجب نشر الوعي على أهمية استخدامها كأداة مساعدة ولتحسين تجربة المستخدم في التطبيقات التي تتطلب دردشة آلية، ويمكن أن يوفر الوقت والجهد للمستخدمين. ومع ذلك، يجب على المطورين أخذ بعين الاعتبار الأخلاقيات والخصوصية عند تصميم ChatGPT وتدريبه.

4-5. Chat GPT

هو نموذج ذكاء صناعي معتمد على تقنية GPT-3.5 بالإصدار المجاني له، ويعتمد على GPT4 بالإصدار المدفوع. وهو عبارة عن روبوت دردشة تم تدريبه على بيانات لغوية يمكنه إنشاء استجابات متنوعة ودقيقة لأسئلة المستخدمين. كما يستخدم ChatGPT في العديد من التطبيقات التي تتطلب دردشة آلية [1].

تأثير ChatGPT على المجتمع:

4-5-1. خدمة العملاء: يمكن استخدامه لتحسين تجربة خدمة العملاء من خلال توفير ردود أكثر دقة وكفاءة على استفسارات العملاء.

4-5-2. ترجمة اللغة: يمكن استخدامه لتحسين دقة وطلاقة الترجمة الآلية، مما يسهل التواصل بين الأشخاص الذين يتحدثون لغات مختلفة.

4-5-3. إنشاء المحتوى: يمكن استخدامه لإنشاء نصوص عالية الجودة، والتي يمكن استخدامها لإنشاء مجموعة كبيرة من المحتوى، مثل المقالات ومنشورات المدونات وحتى الكتب.

4-5-4. التعليم: يمكن استخدامه لإنشاء تجارب تعليمية مخصصة، وتزويد الطلاب بمحتوى مخصص وتمارين مصممة خصيصاً لاحتياجاتهم الفردية.

4-5-5. الرعاية الصحية: يمكن استخدامه لتحسين كفاءة التشخيصات الطبية وتوصيات العلاج من خلال تحليل بيانات المريض وتزويد الأطباء بمعلومات أكثر دقة وذات صلة.

تجدر الإشارة إلى أن هذه ليست سوى أمثلة قليلة على التأثير المحتمل لـ ChatGPT على المجتمع، ومع استمرار تقدم النموذج والتكنولوجيا ذات الصلة، فمن المحتمل ظهور تطبيقات واستخدامات جديدة.

4- دراسة المرجعية:

4.1. هدفت الدراسة [11] إلى تقييم أداء نموذجي اللغة الكبيرة، وهما ChatGPT و GPT-4 مع قيمتي معلم درجة الحرارة، في الامتحان النهائي للطب البولندي (MFE). تم اختبار النماذج على ثلاث نسخ من MFE بنسختين للغة - الإنجليزية والبولندية. تمت مقارنة دقتي كل من النماذج وتم استكشاف العلاقات بين صحة الإجابات مع مقاييس الإجابة. أظهرت الدراسة أن GPT-4 تفوقت على GPT-3.5 في جميع الامتحانات الثلاث بغض النظر عن اللغة المستخدمة. حقق GPT-4 دقة متوسطة بنسبة 79.7% لكل من النسخ البولندية والإنجليزية، ونجاح في جميع إصدارات MFE. كانت لدى GPT-3.5 دقة متوسطة بنسبة 54.8% للبولندية و60.3% للإنجليزية. كانت نتيجة GPT-4 في معظم الأحيان أقل من النتيجة المتوسطة لطالب الطب. كانت دقة النموذجين غير مثلى وأساء من المتوسط بالنسبة لطالب الطب. وهذا يؤكد على الحاجة إلى مزيد من التحسينات في LLMs قبل أن يمكن نشرها بموثوقية في الإعدادات الطبية.

4.2. في الدراسة [13] كان الهدف الأساسي استكشاف جانب حاسم من فائدة الذكاء الاصطناعي في اكتساب المعلومات الصحية، مع التركيز بشكل خاص على الاختلافات اللهجات العربية. وجدت الدراسة بأن الاختلافات في الأداء بين نماذج ChatGPT في استجاباتها للتحفيز في لهجتين عربيتين متميزتين واضحة ومتنوعة، كان تقييم أداء النماذج حسب مكونات CLEAR (الاكتمال)، (الدقة) Accuracy و (الصلة والملاءمة) Relevance كما موضح بالجدول:

الجدول رقم (1) : تقييم أداء نماذج ChatGPT 3.5 و ChatGPT 4 باستخدام CLEAR

ChatGPT4		ChatGPT 3.5		CLEAR
اللهجة التونسية	اللهجة الأردنية	اللهجة التونسية	اللهجة الأردنية	
3.20 (متوسط)	3.40 (فوق المتوسط)	2.70 (متوسط)	3.20 (متوسط)	الاكتمال
3.20 (متوسط)	3.50 (فوق المتوسط)	2.80 (متوسط)	3.40 (فوق المتوسط)	دقة
3.20 (متوسط)	3.70 (فوق المتوسط)	3.00 (متوسط)	3.60 (فوق المتوسط)	الصلة والملاءمة

كانت جميع الردود من نماذج الذكاء الاصطناعي باللغة العربية الفصحى، مما يعكس الحاجة الملحة لتطوير نماذج تستوعب مجموعة أوسع من اللهجات المختلفة بدلاً من التوافق المحدد مع اللهجة الفصحى.

4.3. الورقة [12] تقدم دراسة مرجعية هامة حول فعالية نماذج الذكاء الاصطناعي في استجابات الأمراض المعدية بين اللغتين الإنجليزية والعربية. توضح الدراسة الفروقات في الأداء بين اللغتين وتسلط الضوء على أهمية تحسين أداء النماذج الصناعية في اللغة العربية لتعزيز جودة نشر المعلومات الصحية. تم تقييم أداء أربعة نماذج للذكاء الاصطناعي: ChatGPT-3.5، ChatGPT-4، Bing، و Bard. تم تصميم الدراسة لتحليل 15 استعلاماً حول الأمراض المعدية، بما في ذلك الإيدز، السل، الملاريا، كوفيد-19 والإنفلونزا. وأظهرت النتائج أن الاستعلامات باللغة

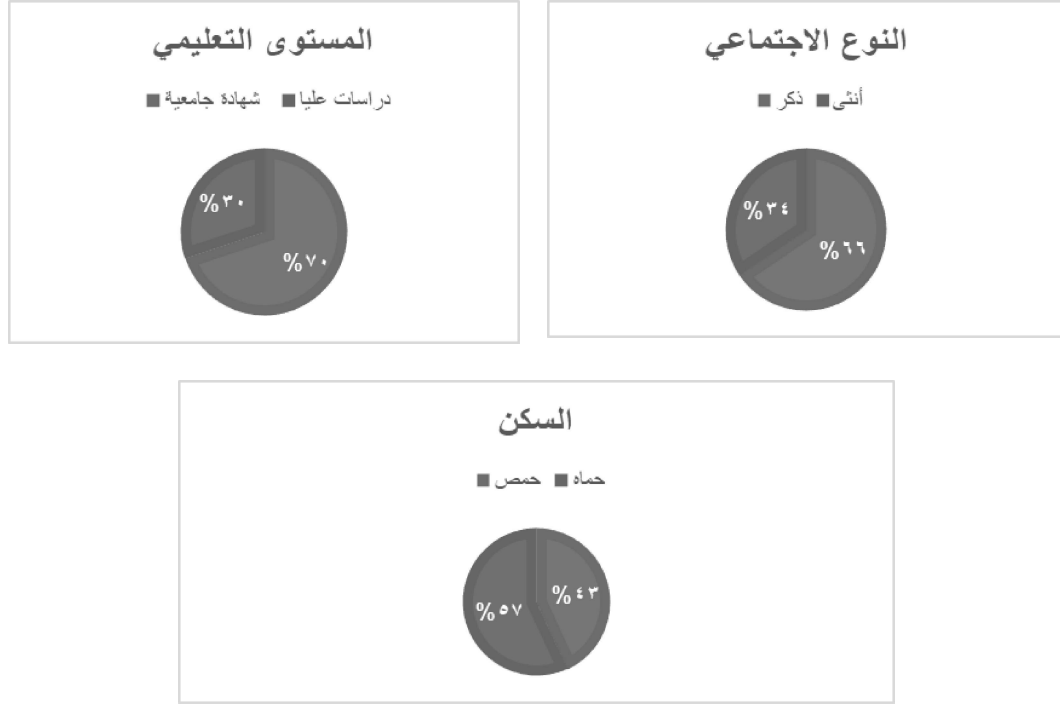
الإنجليزية أظهرت أداءً متفوقاً باستمرار على الاستعلامات باللغة العربية. تصدرت النماذج التي تعمل باللغة الإنجليزية، حيث جاء Bard في المرتبة الأولى تليها Bing ومن ثم ChatGPT-4 وأخيراً ChatGPT-3.5. أما بالنسبة للنماذج التي تعمل باللغة العربية، لم تظهر الفروقات بشكل ملحوظ من حيث الأداء بين النماذج المختلفة. حيث تم تصنيف أداء النماذج الأربعة باللغة الإنجليزية على أنه "ممتاز"، بينما تم تصنيف أداء النماذج العربية على أنه "فوق المتوسط". وأشارت الدراسة إلى وجود فجوة في الأداء بين اللغتين الإنجليزية والعربية في استعلامات الأمراض المعدية. وتم اقتراح على مطوري الذكاء الاصطناعي أن يولوا اهتماماً خاصاً للتنوع الثقافي واللغوي في تطوير النماذج لضمان نشر المعلومات الصحية بطريقة متساوية ودقيقة.

4.4. تُقيم الدراسة [16] أداء أنظمة الترجمة الآلية ولغات النماذج الكبيرة للغة مثل Google Translate و Microsoft Translator و GPT-3.5 و GPT-4. بالنسبة للترجمات من الإنجليزية إلى لغات أخرى، أظهر Google Translate و Microsoft Translator أداءً متفوقاً، بينما أظهر GPT-3.5 و GPT-4 نتائج أقل قوة نسبياً. بشكل عام، يتفوق Google Translate باستمرار على ChatGPT و Microsoft Translator في درجات على مقاييس BLEU و chrF و TER (مقاييس لتقييم جودة الترجمة الآلية)، مما يبرز تفوق ترجمته بشكل عام. يظهر الفارق في الأداء بين أنظمة الترجمة بشكل أكبر بالنسبة لأزواج لغات معينة، مثل الإنجليزية-التاميلية (لغة رسمية في سريلانكا والهند) والخميرية (لغة رسمية في كمبوديا) -الإنجليزية، ما يشير إلى مجالات محتملة للتحسين في الترجمة. على العكس من ذلك، بالنسبة لأزواج لغات مثل التشيكية-الإنجليزية والألمانية-الإنجليزية والإستونية-الإنجليزية، فإن أداء أنظمة الترجمة متقارب نسبياً، وهو ما يعني تنافسها إلى حد ما في الترجمة. تجدر الإشارة إلى أن GPT-4 يتفوق على أداء كل من Google و Microsoft في أزواج لغات معينة. على سبيل المثال، في زوج اللغة الألمانية-الإنجليزية، يحقق GPT-4 درجات أعلى في جميع المقاييس. يشير هذا التحسن الملحوظ في الأداء إلى إمكانية واعدة لتحسينات مستقبلية لنموذج GPT، حتى في مهام الترجمة الصعبة.

5- القسم العملي: فعالية وأثر ChatGPT في تحسين الأداء لدى العاملين في القطاع الطبي

بعد توضيح المفاهيم النظرية المتعلقة GPT ومقارنة الإصدارات المتعددة منه سنقوم بدراسة مدى فعالية ChatGPT وأثره في تحسين الأداء لدى القطاع الطبي. تستند دراسة فعالية ChatGPT عند العاملين في القطاع الطبي على عدة معايير مهمة، بما في ذلك جودة الاستجابات والتواصل السلس وفهم توجهات المستخدمين. تتمثل أهداف الدراسة في تقييم قدرة ChatGPT على تلبية احتياجات المستخدمين بشكل فعال وتقديم ردود مفيدة ودقيقة. سيتم استخدام تصميم الاستبيان لهذه الدراسة لجمع البيانات من المشاركين. ستشمل الأسئلة المطروحة تقييم ChatGPT وتقديرهم لقدرة على تلبية احتياجاتهم. تعتبر هذه الدراسة ذات أهمية كبيرة في تقييم فعالية ChatGPT وتحسينه. ستساعد النتائج المستخدمين والمطورين على فهم قدرات ChatGPT وتحديد المجالات التي يمكن تطويرها وتحسينها.

1-5- عينة الدراسة: تم توزيع الاستبيان الكترونياً (على مجموعة أشخاص من مجالات طبية مختلفة ولهم خلفية معلوماتية) ، وتم الحصول على 93 رد الكتروني، موزعة بالشكل التالي



الشكل رقم (3): توزيع عينات الدراسة

5-2- أداة الدراسة: تم تحليل البيانات المستردة، حيث تم اعتماد على مقاييس النزعة المركزية مثل (التكرار، النسبة المئوية، المتوسط الحسابي، الانحراف المعياري)

5-3- التحليل الوصفي لمتغيرات الدراسة: قمنا بتحليل وصفي تفصيلي باستخدام تحليل التكرار (ت) والوسط الحسابي والوسيط والانحراف المعياري، وتعد القيمة (3.40) هي التي تتوسط مقياس ليكرت الخماسي وليست القيمة (3)، حيث أن غير موافق بشدة تبدأ من الواحد وليس الصفر، ويمكننا احتساب المتوسطات الحسابية الخاصة بمقياس ليكرت الخماسي عن طريق:

- احتساب المدى بين أكبر رقم بمقياس ليكرت وهو /5/ وأصغر رقم وهو /1/ فيكون المدى: $4 = 1 - 5$
 - بعد ذلك يتم حساب طول الفئة من خلال تقسيم المدى على عدد الفئات (الخيارات) أي: $0.8 = 5 / 4$
- فيكون اتجاه ومستوى الإجابات حسب الوسط الحسابي لمتغيرات الدراسة الكمية كما هو موضح في الجدول رقم (1):

الجدول رقم (2): اتجاه ومستوى الإجابات حسب الوسط الحسابي لمتغيرات الدراسة الكمية

المستوى	الإجابة	المتوسط
منخفض	غير موافق بشدة	1-1,80
	غير موافق	1,81-2,60
متوسط	حيادي	2,61-3,40
مرتفع	موافق	3,41-4,20
	موافق بشدة	4,21- 5

ونبين فيما يلي نتائج التحليل وفقاً لكل متغير من متغيرات الدراسة على حدا:

1-3-5 الاعتمادية العلمية:

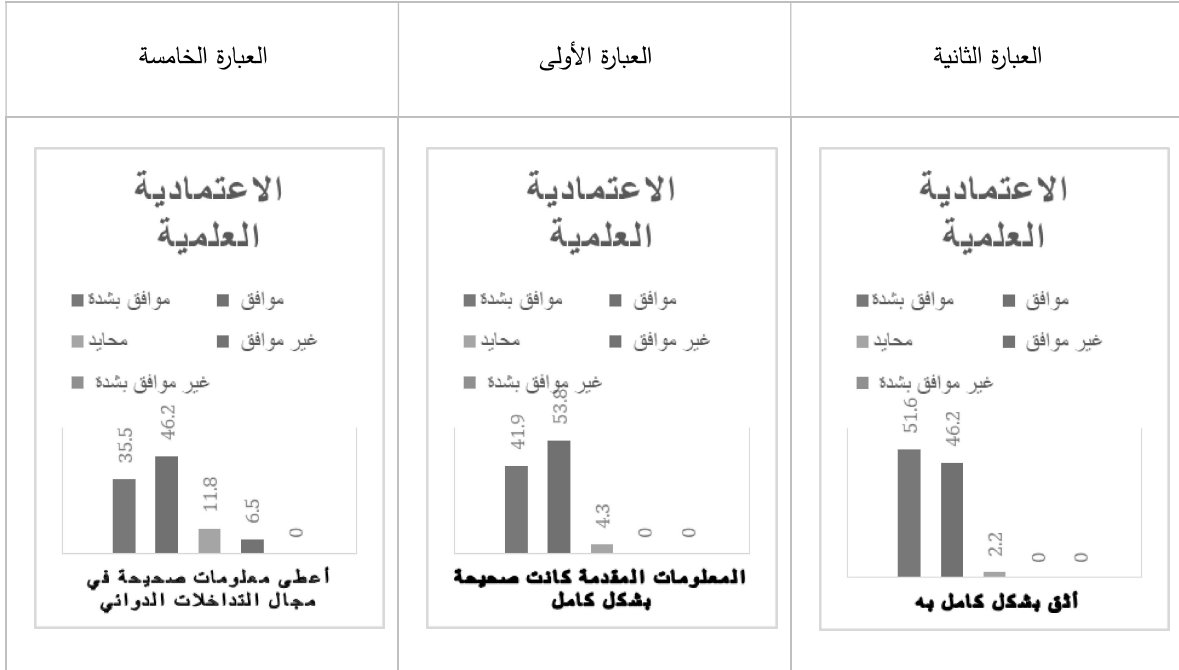
الجدول رقم (3): تحليل الوسط الحسابي والانحراف المعياري لإجابات أفراد عينة الدراسة

الانحراف	المتوسط الحسابي	النسبة المئوية	التكرار	الاجابة	العبارة
0.569	4.38	-	-	غير موافق بشدة	المعلومات المقدمة كانت صحيحة بشكل كامل
		-	-	غير موافق	
		4.3	4	محايد	
		53.8	50	موافق	
		41.9	39	موافق بشدة	
0.544	4.49	-	-	غير موافق بشدة	أثق بشكل كامل به
		-	-	غير موافق	
		2.2	2	محايد	
		46.2	43	موافق	
		51.6	48	موافق بشدة	
0.780	4.02	-	-	غير موافق بشدة	يمكن الاعتماد عليه في تقديم المشورة الطبية ومساعدة المرضى
		4.3	4	غير موافق	
		16.1	15	محايد	
		52.2	49	موافق	
		26.9	25	موافق بشدة	
0.779	3.96	-	-	غير موافق بشدة	أعطى معلومات صحيحة عن حالات مرضية
		5.4	5	غير موافق	
		16.1	15	محايد	
		55.9	52	موافق	
		22.6	21	موافق بشدة	
0.853	4.11	-	-	غير موافق بشدة	أعطى معلومات صحيحة في مجال التداخلات الدوائي
		6.5	6	غير موافق	
		11.8	11	محايد	
		46.2	43	موافق	
		35.5	33	موافق بشدة	

0.713	4.05	-	-	غير موافق بشدة	استطاع التعرف على أسماء الأدوية التجارية
		2.2	2	غير موافق	
		16.1	15	محايد	
		55.9	52	موافق	
		25.8	24	موافق بشدة	
0.705	4.22	-	-	غير موافق بشدة	استطاع التعرف على أسماء الأمراض بشكل جيد
		3.3	3	غير موافق	
		10.8	10	محايد	
		57	53	موافق	
		29	27	موافق بشدة	
0.795	4.10	-	-	غير موافق بشدة	يمكنني استخدامه كمساعد موثوق في العمل
		2.2	2	غير موافق	
		6.5	6	محايد	
		58.1	54	موافق	
		33.3	31	موافق بشدة	
0.815	3.80	-	-	غير موافق بشدة	كان ChatGPT مفيداً في تحسين دقة التشخيص الطبي
		7.5	7	غير موافق	
		22.6	21	محايد	
		52.7	49	موافق	
		17.2	16	موافق بشدة	

المصدر – بالاعتماد على نتائج تحليل البيانات SPSS

يظهر من الجدول رقم (2) أن المتوسط العام لإجابات العينة العشوائية المشاركة في الاستبيان (4.192) مما يدل على تقييم مرتفع لأفراد العينة لمستوى الاعتمادية العلمية ل chatGPT. وقد احتلت العبارة رقم /2/ المرتبة الأولى من حيث الأهمية لدى أفراد العينة تليها العبارة رقم /1/ ثم العبارة رقم /5/ بتقييم مرتفع، ومن ثم /3,4/ على الترتيب وجاءت جميعها بتقييم مرتفع وكانت القيم (4.02، 3.96) على التوالي.



الشكل رقم (4): التمثيل البياني لبعض عبارات الاعتمادية العلمية

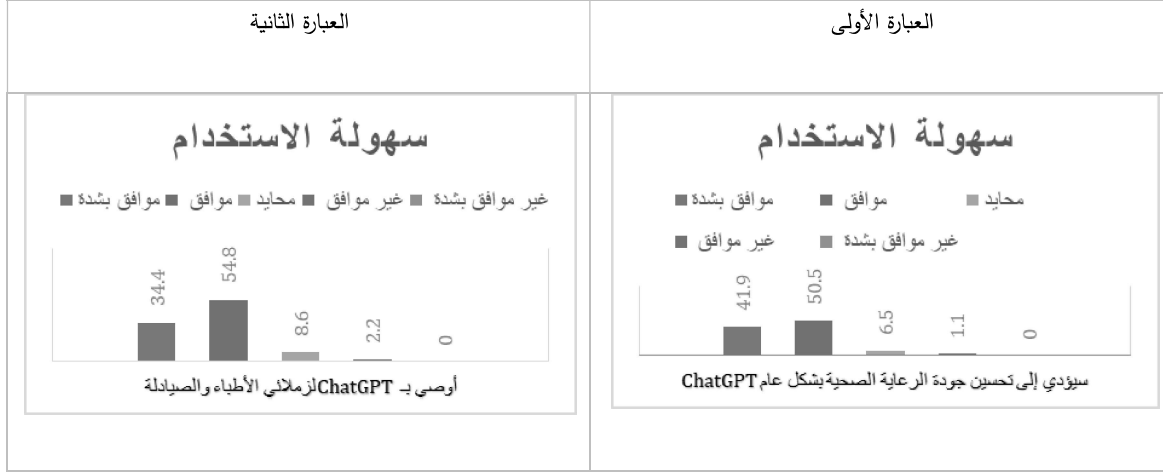
2-3-5 سهولة الاستخدام:

الجدول رقم (4): تحليل الوسط الحسابي والوسيط والانحراف المعياري لإجابيات أفراد عينة الدراسة على الفقرات

العبارة	الإجابة	التكرار	النسبة المئوية	المتوسط الحسابي	الانحراف
ChatGPT سيؤدي إلى تحسين جودة الرعاية الصحية بشكل عام	غير موافق بشدة	-	-	4.33	0.648
	غير موافق	1	1.1		
	محايد	6	6.5		
	موافق	47	50.5		
	موافق بشدة	39	41.9		
أوصي بـ ChatGPT لزملائي الأطباء والصيادلة	غير موافق بشدة	-	-	4.22	0.689
	غير موافق	2	2.2		
	محايد	8	8.6		
	موافق	51	54.8		
	موافق بشدة	32	34.4		
أشعر بالراحة في استخدام ChatGPT في ممارسة الطب	غير موافق بشدة	1	1.1	3.78	1.072
	غير موافق	15	16.1		
	محايد	14	15.1		
	موافق	36	38.7		
	موافق بشدة	27	29		

المصدر - بالاعتماد على نتائج تحليل البيانات SPSS

يظهر من الجدول رقم (3) أن المتوسط العام لإجابات العينة العشوائية المشاركة في الاستبيان من الأفراد العاملين في القطاع الطبي في سوريا يبلغ (4.11) مما يدل على تقييم مرتفع لأفراد العينة لمستوى سهولة الاستخدام ل chatGPT. وقد احتلت العبارة رقم /1/ المرتبة الأولى من حيث الأهمية لدى أفراد العينة، تليها العبارة رقم /2/ بمتوسط (4.22) .



الشكل رقم (5): التمثيل البياني لبعض عبارات سهولة الاستخدام

4-5 النتائج: تشير نتيجة الاستبيان العامة إلى أن chatGPT هي منصة سهلة الاستخدام وتحظى بمستوى عالٍ من الاعتمادية العلمية. ويمكن استنتاج ذلك من المتوسط العام لإجابات العينة العشوائية التي شاركت في الاستبيان والذي بلغ 4.192.

ولتحسين استخدام ChatGPT في المجال الطبي، يمكن توجيه التوصيات التالية للمطورين:

- 1- تحسين دقة الفهم الطبي: زيادة تدريب النماذج على مجموعة واسعة من المصطلحات والمفاهيم الطبية باللغة العربية.
- 2- تطوير واجهات برمجة تطبيقات (API) لتحقيق تكامل سهل وآمن مع سجلات المرض الإلكترونية.
- 3- تحسين توجيه الأسئلة: تطوير نماذج توجيه أسئلة متقدمة للتأكد من فهم دقيق لاستفسارات المستخدمين في السياق الطبي. وتدريب العاملين عليها للحصول على أفضل الإجابات.
- 4- تطوير ميزات مخصصة للمجال الطبي: تصميم ميزات خاصة بالمجال الطبي، مثل تحليل البيانات السريرية أو توفير توجيه دقيق للمرضى حسب الحالة الصحية.
- 5- تحفيز التنوع اللغوي والثقافي: تحسين قدرة ChatGPT على التعامل مع لغات وثقافات متعددة في سياقات طبية متنوعة.

6- المراجع:

- 1- Alawida, M., Mejri, S., Mehmood, A., Chikhaoui, B., & Isaac Abiodun, O. (2023). A comprehensive study of ChatGPT: advancements, limitations, and ethical considerations in natural language processing and cybersecurity. Information, 14(8), 462.
- 2- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2019). Language models are unsupervised multitask learners. OpenAI.

- 3- Brown, Tom, et al. (2020) "Language models are few-shot learners." *Advances in neural information processing systems* 33: 1877–1901.
- 4- CAGLAYAN, C., & KARAKAYA, M. (2023). *Controllable Text Generation: Style & Content*. *Current Studies in Social Sciences* 2023, 55.
- 5- Caglayan C. (2022). *A Novel Deep Learning Approach for Controlled Multi-topic Text Generation*. Atılım University Graduate School of Natural and Applied Sciences, Ankara, Turkey.
- 6- Dong, M. M., Stratopoulos, T. C., & Wang, V. X. (2023). *A Scoping Review of ChatGPT Research in Accounting and Finance*. Theophanis C. and Wang, Victor Xiaoqi, *A Scoping Review of ChatGPT Research in Accounting and Finance*
- 7- Iffort (2023) "GPT-3 vs GPT-3.5: Key Differences and Applications"
- 8- Imamguluyev, R. (2023) *The Rise of GPT-3: Implications for Natural Language Processing and Beyond*. Journal homepage: www.ijrpr.com ISSN, 2582, 7421.
- 9- Jiang, Y., & Yang, C. (2024). *Is ChatGPT a Good Geospatial Data Analyst? Exploring the Integration of Natural Language into Structured Query Language within a Spatial Database*. *ISPRS International Journal of Geo-Information*, 13(1), 26.
- 10- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*.
- 11- Rosoł M, Gąsior JS, Łaba J, Korzeniewski K, Młyńczak M. (2023 Nov 22) *Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination*. *Sci Rep*;13(1):20512. doi: 10.1038/s41598-023-46995-z. PMID: 37993519; PMCID: PMC10665355.
- 12- Sallam, M., Al-Mahzoum, K., Alshuaib, O., Alhajri, H., Alotaibi, F., Alkhurainej, D., ... & Egger, J. (2024). *Superior Performance of Artificial Intelligence Models in English Compared to Arabic in Infectious Disease Queries*.
- 13- Sallam, Malik, and Dhia Mousa. (2024) "Evaluating ChatGPT performance in Arabic dialects: A comparative study showing defects in responding to Jordanian and Tunisian general health prompts." *Mesopotamian Journal of Artificial Intelligence in Healthcare* 2024: 1–7.
- 14- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. *Advances in neural information processing systems*, 30.

- 15– Younis, H. A., Eisa, T. A. E., Nasser, M., Sahib, T. M., Noor, A. A., Alyasiri, O. M., ... & Younis, H. A. (2024). A Systematic Review and Meta–Analysis of Artificial Intelligence Tools in Medicine and Healthcare: Applications, Considerations, Limitations, Motivation and Challenges. *Diagnostics*, 14(1), 109.
- 16– Wu, Y., & Hu, G. (2023, December). Exploring Prompt Engineering with GPT Language Models for Document–Level Machine Translation: Insights and Findings. In *Proceedings of the Eighth Conference on Machine Translation* (pp. 166–169).
- 17– ZHANG, M., & JUNTAO, L. (2021). A commentary of GPT–3 in MIT Technology Review 2021. *Fundamental Research*. 1 (6): 831–833.
- 18– <https://openai.com/research/gpt-4>