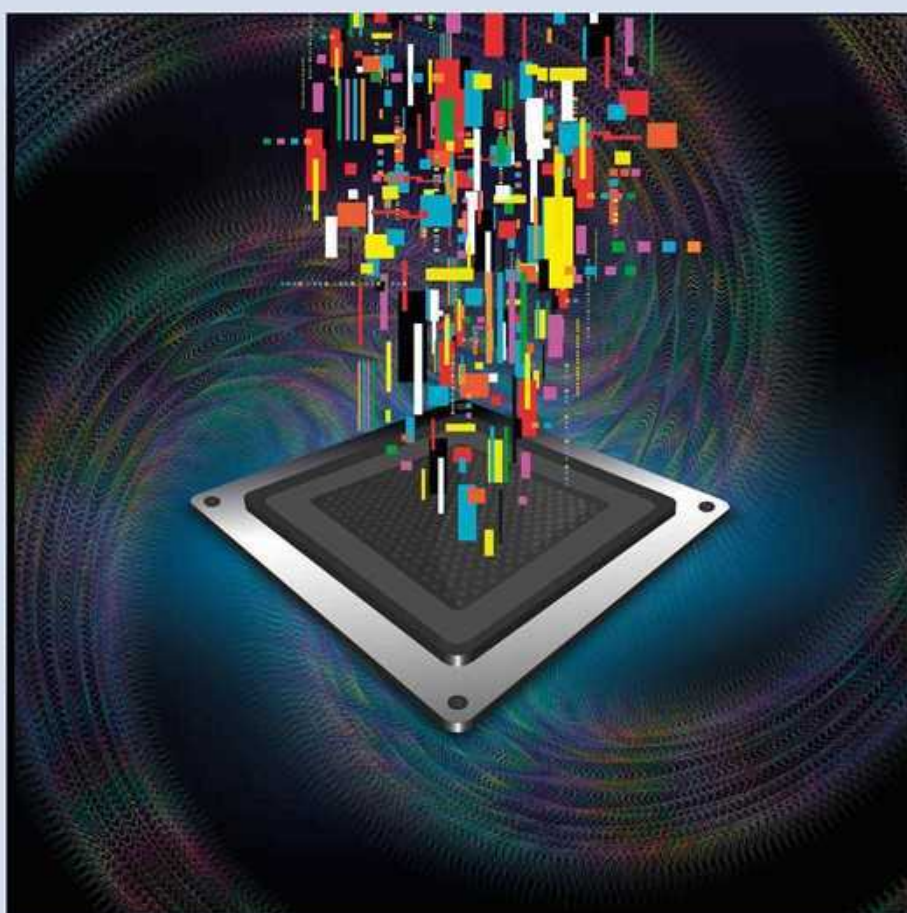


الهلوسة الخوارزمية

Algorithmic Hallucination

وانعكاسها على الثقة القانونية فى النظم الذكية



دار الفكر الجامعى

أمام كلية الحقوق - الاسكندرية

ت: ٤٨٤٣١٣٢

دكتور

باسم محمد فاضل مديولى

الهلوسة الخوارزمية

(Algorithmic Hallucination)

وانعكاسها على الثقة القانونية

في النظم الذكية

الدكتور

باسم محمد فاضل مدبولي

2026

دار الفكر الجامعي

30 شارع سوتير - الإسكندرية

ت : 4843132

المقدمة

شهدت تقنيات توليد اللغة الطبيعية (NLG) تطوراً هائلاً في السنوات الأخيرة بفضل تطوير تقنيات التعلم العميق من تسلسل إلى تسلسل، مثل نماذج اللغة القائمة على Transformer. وقد أدى هذا التقدم إلى توليد لغة طبيعية أكثر سلاسة وتماسكاً، مما ساهم في تحسين أداء المهام اللاحقة، مثل التلخيص التجريدي، وتوليد الحوار، وتحويل البيانات إلى نصوص. ومن الواضح أيضاً أن التوليد القائم على التعلم العميق عرضة لإنتاج نصوص غير مقصودة، مما يُضعف أداء النظام ويؤدي إلى عدم تلبية توقعات المستخدمين في العديد من سيناريوهات العالم الحقيقي⁽¹⁾.

تشير الهلوسة الخوارزمية في نظم الذكاء الاصطناعي التوليدي مثل Chat GPT إلى إنتاج معلومات غير دقيقة أو مضللة رغم أن النموذج قد يبدو مقنعاً وموثوقاً، وتحدث هذه الظاهرة عندما يُنشئ النموذج محتوى يبدو منطقياً أو متماسكاً لغوياً، لكنه لا يستند إلى مصادر صحيحة أو بيانات حقيقية، مما قد يؤدي إلى تضليل المستخدمين في سياقات معرفية أو قانونية أو علمية.

1- موضوع الدراسة

تعد الهلوسة أحد أبرز المخاطر المرتبطة بتقنيات الذكاء الاصطناعي التوليدي، إذ تظهر عندما تعجز هذه الأنظمة عن فهم الأسئلة أو تفسيرها

(1) Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Delong Chen, Wenliang Dai, Ho Shu Chan, Andrea Madotto, Pascale Fung: Survey of Hallucination in Natural Language Generation, Submitted on 8 Feb 2022 (v1), last revised 14 Jul 2024 (this version, v7)] <https://arxiv.org/abs/2202.03629>

تفسيراً دقيقاً، فنقوم بإنتاج إجابات مختلفة لا تستند إلى أساس واقعي أو معرفي صحيح. وفي هذه الحالة، وبسبب عدم قدرتها على توليد استجابات سليمة، تلجأ الخوارزميات إلى إنشاء مخرجات افتراضية، وهي الظاهرة التي يُشار إليها اصطلاحاً بـ«هلوسة الذكاء الاصطناعي»⁽¹⁾.

وتشير الهلوسة في الذكاء الاصطناعي إلى ظاهرة تنتج عن قيام النماذج اللغوية الكبيرة (Large Language Models – LLMs) ، أو أنظمة الرؤية الحاسوبية المعتمدة على الذكاء الاصطناعي التوليدي، بتوليد أنماط أو كيانات غير موجودة في الواقع، أو بإنتاج مخرجات غير دقيقة أو غير منطقية من منظور المستخدم البشري، وتظهر هذه الظاهرة عندما يقدم النظام إجابات أو تفسيرات لا تستند إلى وقائع صحيحة أو لا تعكس مضمون بيانات التدريب على نحو سليم⁽²⁾.

وبوجه عام، يتوقع المستخدم عند تفاعله مع أنظمة الذكاء الاصطناعي التوليدي الحصول على مخرجات متوافقة مع الطلب المقدم، سواء كانت إجابة صحيحة أو تحليلاً منطقياً للمعطيات، غير أن الخوارزميات قد تفشل في بعض الحالات في تحقيق هذا الهدف، فتنتج مخرجات لا تستند إلى بيانات تدريب موثوقة، أو تُفسّر بشكل خاطئ داخل بنية النموذج التحويلي، أو تفتقر إلى أي نمط معرفي يمكن التحقق منه. وفي هذا السياق، يُستخدم مصطلح «الهلوسة» مجازاً لوصف هذه الانحرافات في الاستجابة⁽³⁾.

(1) **Jose Antonio** :ChatGTP and the Generative AI Hallucinations Mar 15, 2023 <https://medium.com/chatgpt-learning/chatgtp-and-the-generative-ai-hallucinations-62feddc72369>

(2) ما هي هلاوس الذكاء الاصطناعي؟

<https://www.ibm.com/sa-ar/think/topics/ai-hallucinations>

(3) ورغم أن مفهوم الهلوسة يرتبط تقليدياً بالإدراك البشري أو الحيواني، فإن توظيفه في مجال الذكاء الاصطناعي يُعد توصيفاً مجازياً دقيقاً نسبياً، لاسيما في تطبيقات التعرف على الصور والأنماط، حيث قد تُنتج النماذج تفسيرات تبدو بعيدة تماماً عن=

وتجدر الإشارة إلى أن هلوسة الذكاء الاصطناعي لا تقتصر على نوع معين من البيانات، بل يمكن أن تظهر عند معالجة مختلف أشكال المحتوى التوليدي، بما في ذلك النصوص والصور والتسجيلات الصوتية ومقاطع الفيديو، بل وحتى الشفرات البرمجية، مما يجعلها إشكالية متعددة الأبعاد ذات آثار تقنية وقانونية ومعرفية⁽¹⁾.

تُعد الهلوسة الخوارزمية (Algorithmic Hallucination) من الظواهر التقنية المرتبطة بأنظمة الذكاء الاصطناعي التوليدي، حيث يقوم النظام بإنتاج مخرجات أو بيانات لا تستند إلى واقع فعلي أو مصدر صحيح، رغم اتسامها بالتماسك الشكلي والمنطقي، ويرجع ذلك إلى أن هذه الأنظمة لا تعمل وفق منطق الإدراك أو الفهم، وإنما تعتمد على نماذج رياضية إحصائية تهدف إلى توليد أكثر المخرجات احتمالاً من حيث البنية، لا من حيث الصدق الواقعي⁽²⁾.

وتنشأ هذه الظاهرة - المعروفة اصطلاحاً بالهلوسة الخوارزمية - عندما يملأ النظام الفجوات المعرفية استناداً إلى أنماط لغوية وليس إلى حقائق موضوعية، فيولد محتوى غير صحيح أو غير موثوق، وهو ما قد يؤدي إلى تضليل المستخدم غير المتخصص.

وتزداد خطورة الهلوسة الخوارزمية في الأنظمة التي تُستخدم في بيئات عالية الحساسية، كالنظم القانونية أو الطبية أو الأمنية، حيث يؤدي

=الواقع. ويُشبه هذا السلوك ما يحدث لدى البشر عند تفسير أشكال عشوائية، مثل رؤية وجوه أو رموز في السحب أو على سطح القمر، نتيجة الميل الطبيعي إلى البحث عن أنماط ذات معنى.

(1) **Jose Antonio** :ChatGTP and the Generative AI Hallucinations Mar 15, 2023 <https://medium.com/chatgpt-learning/chatgtp-and-the-generative-ai-hallucinations-62feddc72369>

(2) Bender, E. M., et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of FAccT.

غياب آليات التحقق الخارجي (External Verification) إلى تضخيم أثر المخرجات الوهمية. ولهذا اتجهت الدراسات التقنية الحديثة إلى تعزيز نماذج "الإنسان في الحلقة" (Human-in-the-loop) للحد من الاعتماد المطلق على المخرجات الآلية⁽¹⁾.

أدى الانتشار الواسع لاستخدام نماذج الذكاء الاصطناعي التوليدي، وعلى رأسها Chat GPT، إلى انتقال ظاهرة الهلوسة الخوارزمية من نطاقها النظري إلى حيز التطبيق العملي، فلم تعد الهلوسة مجرد احتمال تقني، بل أصبحت واقعاً ملموساً يظهر في مخرجات تتضمن وقائع غير صحيحة، أو نصوصاً قانونية مختلفة، أو إحالات وهمية إلى أحكام قضائية وتشريعات لا وجود لها، وتكمن خطورة هذه الظاهرة في أن Chat GPT يقدم هذه المخرجات بصياغة لغوية متماسكة ومنطقية، بما قد يدفع المستخدم إلى الوثوق بها دون تحقق، خاصة في البيئات القانونية والأكاديمية التي تقوم أساساً على دقة المعلومة وصحة المصدر.

وتبرز التطبيقات العملية للهلوسة الخوارزمية في اعتماد بعض المستخدمين على مخرجات Chat GPT في إعداد مذكرات قانونية أو بحوث علمية أو تقارير مهنية، دون مراجعة بشرية كافية، مما يؤدي إلى تضليل القرار القانوني أو العلمي. وقد يترتب على ذلك آثار قانونية مباشرة، مثل بطلان الإجراءات، أو قيام المسؤولية المهنية، أو الإخلال بواجب العناية. ومن ثم، فإن دراسة هذه التطبيقات العملية تمثل مدخلاً أساسياً لفهم الأبعاد الواقعية للهلوسة الخوارزمية، وتقييم مدى كفاية الأطر القانونية القائمة في مواجهة مخاطرها والحد من آثارها السلبية.

(1) Bommasani, R., et al. (2022). On the Opportunities and Risks of Foundation Models. Stanford University. Amodei, D., et al. (2016). Concrete Problems in AI Safety. arXiv.

ومن زاوية المسؤولية القانونية، تبرز إشكالية جوهرية تتمثل في صعوبة تطبيق قواعد المسؤولية التقليدية القائمة على الخطأ الشخصي أو الإهمال البشري. فالهلوسة الخوارزمية لا يمكن إسنادها بسهولة إلى مستخدم النظام، ولا إلى المبرمج بشكل مباشر، طالما أن النظام يعمل وفق ما صُمم له. وهو ما يخلق فجوة قانونية بين وقوع الضرر وغياب الفاعل البشري الواضح، مما يضعف من فعالية نظم التعويض التقليدية.

وفي هذا السياق، اتجه الفقه القانوني المقارن إلى البحث عن بدائل نظرية لمعالجة هذه الإشكالية، من بينها تبني المسؤولية الموضوعية القائمة على مجرد تحقق الضرر، أو توزيع المسؤولية بين عدة أطراف (المطور، والمشغل، والمالك)، أو إدماج آليات التأمين الإجباري لتغطية المخاطر الناشئة عن أخطاء الأنظمة الذكية. وتعكس هذه الاتجاهات إدراكاً متزايداً لخصوصية الخطأ الخوارزمي بوصفه خطأً تقنياً مستقلاً عن الإرادة الإنسانية.

وعليه، فإن الهلوسة الخوارزمية تفرض إعادة صياغة المفاهيم القانونية التقليدية للخطأ والمسؤولية، بما يحقق التوازن بين تشجيع الابتكار التكنولوجي من جهة، وضمان حماية المتضررين وعدم تركهم دون تعويض من جهة أخرى. ويستلزم ذلك تطوير أطر تشريعية مرنة تستوعب الطبيعة الاحتمالية للذكاء الاصطناعي، وتعيد تعريف العلاقة بين الإنسان والآلة في مجال المسؤولية القانونية.

2- أهمية الدراسة:

تتبع أهمية دراسة الهلوسة الخوارزمية بوصفها خطراً تقنياً من كونها إحدى أبرز الإشكاليات المعاصرة التي أفرزها الاعتماد المتزايد على أنظمة الذكاء الاصطناعي التوليدي في المجالات القانونية والطبية والإدارية والاقتصادية. فهذه الأنظمة لم تعد أدوات مساعدة فحسب، بل

أصبحت تؤثر تأثيراً مباشراً في صناعة القرار، الأمر الذي يجعل المخرجات الوهمية أو غير الدقيقة مصدر تهديد حقيقي للحقوق والمراكز القانونية للأفراد.

وتتجلى أهمية الموضوع كذلك في قصور القواعد القانونية التقليدية عن استيعاب طبيعة المخاطر الناشئة عن الهلوسة الخوارزمية، إذ إن هذه المخاطر لا تتدرج بسهولة ضمن مفاهيم الخطأ المادي أو العيب التقني المعروفين في نظرية المسؤولية المدنية، وهو ما يفرض ضرورة إعادة تقييم الأسس القانونية للمساءلة، والانتقال من منطق الخطأ الفردي إلى منطق إدارة المخاطر التقنية وتوزيع المسؤولية بين أطراف المنظومة الرقمية.

كما تبرز أهمية الدراسة من زاوية حماية الثقة العامة في التقنيات الذكية، حيث يؤدي انتشار المخرجات الوهمية إلى تقويض الثقة في أنظمة الذكاء الاصطناعي، ويُضعف من مشروعية الاعتماد عليها في المرافق العامة والأنشطة المهنية. ومن ثم، فإن التنظيم القانوني الفعّال للهلوسة الخوارزمية يُعد شرطاً لازماً لتحقيق التوازن بين تشجيع الابتكار التقني وضمان الأمن القانوني وحماية المتعاملين.

وتكتسب الدراسة أهميتها أيضاً من الناحية التشريعية المقارنة، في ظل اتجاه عدد من الأنظمة القانونية الحديثة إلى وضع أطر تنظيمية خاصة للذكاء الاصطناعي، كما هو الحال في الاتحاد الأوروبي، دون أن تُفرد تنظيمًا صريحًا ومتكاملاً لظاهرة الهلوسة الخوارزمية. الأمر الذي يجعل البحث في هذا الموضوع مساهمة علمية ضرورية لسد فراغ تشريعي قائم واقتراح حلول قانونية قابلة للتطبيق.

وأخيراً، فإن تناول الهلوسة الخوارزمية باعتبارها خطراً تقنياً مستقلاً يُسهم في تطوير الفقه القانوني العربي في مجال الذكاء الاصطناعي،

ويدعم الجهود الرامية إلى بناء نظرية قانونية حديثة للمخاطر الرقمية، بما يتوافق مع التحولات التقنية المتسارعة ومتطلبات العدالة وحماية الحقوق في العصر الرقمي.

3- أهداف الدراسة:

يهدف هذا البحث إلى تحليل ظاهرة الهلوسة الخوارزمية بوصفها أحد المخاطر التقنية الملازمة لأنظمة الذكاء الاصطناعي التوليدي، وبيان آثارها القانونية في ضوء القواعد العامة للمسؤولية والتنظيم القانوني الحديث، وذلك من خلال تحقيق الأهداف الآتية:

- تحديد المفهوم القانوني الدقيق للهلوسة الخوارزمية وبيان خصائصها المميزة لها عن الخطأ البرمجي التقليدي أو الخطأ البشري.
- تحليل الطبيعة القانونية للهلوسة الخوارزمية وتكييفها بوصفها خطرًا تقنيًا يستوجب تنظيمًا خاصًا.
- بيان أوجه القصور في القواعد القانونية التقليدية عند تطبيقها على الأضرار الناشئة عن المخرجات الوهمية للأنظمة الذكية.
- تحديد نطاق المسؤولية المدنية المترتبة على الهلوسة الخوارزمية، مع بيان شروط قيامها وأركانها.
- دراسة توزيع المسؤولية القانونية بين أطراف منظومة الذكاء الاصطناعي (المصمم، مزود الخدمة، المستخدم).
- تحليل دور واجب الإعلام والشفافية في الحد من مخاطر الهلوسة الخوارزمية وحماية المستخدمين.
- استعراض الاتجاهات التشريعية المقارنة في تنظيم مخاطر الذكاء الاصطناعي ومدى كفايتها لمعالجة الهلوسة الخوارزمية.
- تقييم الدور القضائي في مواجهة الأضرار الناشئة عن المخرجات الوهمية وإشكاليات الإثبات المرتبطة بها.

- اقتراح إطار قانوني وتنظيمي متكامل يوازن بين تشجيع الابتكار التقني وضمان الحماية القانونية للأفراد.

4- إشكالية الدراسة:

تتمحور إشكالية هذه الدراسة حول مدى كفاية القواعد القانونية التقليدية في مواجهة المخاطر الناشئة عن الهلوسة الخوارزمية التي تصدر عن أنظمة الذكاء الاصطناعي التوليدي، ولا سيما في ظل اعتماد هذه الأنظمة على آليات تنبؤية احتمالية قد تُنتج مخرجات وهمية أو غير مطابقة للواقع، مع ما يترتب على ذلك من آثار قانونية تمس الحقوق والمراكز القانونية للأفراد.

وتتفرع عن هذه الإشكالية الرئيسة تساؤلات جوهرية، من أهمها:

- هل يمكن تكييف الهلوسة الخوارزمية بوصفها خطأ قانونياً أم خطراً تقنياً؟
- ومن هو المسؤول قانوناً عن الأضرار الناشئة عنها في ظل غياب الشخصية القانونية للنظام الذكي؟
- وإلى أي مدى تكفي قواعد المسؤولية المدنية التقليدية، القائمة على الخطأ والضرر وعلاقة السببية، لمعالجة هذه الظاهرة؟
- وهل يحقق التنظيم القانوني الحالي التوازن المطلوب بين تشجيع الابتكار التقني وضمان الحماية القانونية للمستخدمين والغير؟

كما تثير الدراسة إشكالية عملية تتعلق بصعوبة الإثبات في منازعات الهلوسة الخوارزمية، سواء من حيث تحديد مصدر المخرج الوهمي، أو إثبات العلاقة السببية بين النظام والضرر، في ظل تعقيد البنية التقنية للأنظمة الذكية وطابعها غير القابل للتفسير. وهو ما يفرض التساؤل حول الحاجة إلى تبني آليات قانونية خاصة، كافتراض الخطأ أو تخفيف عبء الإثبات أو إعادة توزيع المسؤولية.

وعليه، تسعى هذه الدراسة إلى معالجة هذه الإشكاليات من خلال تحليل الإطار المفاهيمي والتقني للهلوسة الخوارزمية، ودراسة تنظيمها القانوني في ضوء التشريعات المقارنة، وصولاً إلى اقتراح حلول قانونية وتنظيمية قادرة على استيعاب المخاطر التقنية المستحدثة وضمان الحماية الفعالة للحقوق في البيئة الرقمية.

5- خطة الدراسة:

المبحث الأول: تقنية Chat GPT كمصدر للهلوسة الخوارزمية.

المطلب الأول: البنية التقنية لنماذج اللغة وعلاقتها بالهلوسة الخوارزمية.

المطلب الثاني: ماهية ChatGPT وجذورها التقنية.

المبحث الثاني: ماهية الهلوسة الخوارزمية.

المطلب الأول: تعريف الهلوسة الخوارزمية.

المطلب الثاني: صور الهلوسة الخوارزمية.

المطلب الثالث: الأسباب التقنية للهلوسة الخوارزمية

المبحث الثالث: المسؤولية المدنية عن أضرار الهلوسة الخوارزمية.

المطلب الأول: المسؤولية التقصيرية عن أضرار الهلوسة الخوارزمية

المطلب الثاني: المسؤولية العقدية عن أضرار الهلوسة الخوارزمية



الفهرس

الصفحة

الموضوع

.....	الاهداء
.....	المقدمة
.....	موضوع الدراسة
.....	اهمية الدراسة
.....	أهداف الدراسة
.....	اشكالية الدراسة
.....	منهج الدراسة
.....	خطة الدراسة

المبحث الأول

تقنية Chat GPT

كمصدر للهلوسة الخوارزمية

.....	المطلب الأول: البنية التقنية لنماذج اللغة وعلاقتها بالهلوسة
.....	الخوارزمية
.....	الفرع الأول: الطبيعة الاحتمالية لنماذج اللغة الكبيرة ودورها
.....	في توليد الهلوسة
.....	الفرع الثاني: التدريب المسبق وحدود التحقق من صحة
.....	المعلومات
.....	المطلب الثاني: ماهية Chat GPT وجذورها التقنية
.....	الفرع الأول: الجذور التاريخية لتقنية الشات جي بي تي
.....	(Chat GPT)
.....	الفرع الثاني: المقصود بتقنية Chat GPT

المبحث الثاني

الاطار المفاهيمي والتقني

للهلوسة الخوارزمية

المطلب الأول: تعريف الهلوسة الخوارزمية

الفرع الأول: التعريف التقني للهلوسة الخوارزمية

الفرع الثاني: التعريف القانوني للهلوسة الخوارزمية

الفرع الثالث: التمييز بين الهلوسة وما يتشابه معها

المطلب الثاني: صور الهلوسة الخوارزمية

الفرع الأول: الهلوسة المعلوماتية (اختلاق وقائع أو مصادر)

الفرع الثاني: الهلوسة القانونية (نصوص أو أحكام قضائية

وهمية)

الفرع الثالث: الهلوسة التنبؤية والقرارات الآلية الخاطئة

المطلب الثالث: الأسباب التقنية للهلوسة الخوارزمية

المطلب الرابع: أثر الهلوسة الخوارزمية بين المخاطر النظرية

والتطبيقات العملية

الفرع الأول: المخاطر والتحديات القانونية للهلوسة

الخوارزمية

الفرع الثاني: تطبيقات عملية على الهلوسة الخوارزمية

لمخرجات Chat GPT

المبحث الثالث

المسؤولية المدنية عن أضرار الهلوسة الخوارزمية

المطلب الأول: المسؤولية التقصيرية عن أضرار الهلوسة

الخوارزمية

الفرع الأول: قيام الخطأ القانوني في الهلوسة الخوارزمية	
الفرع الثاني: الضرر الناجم عن الهلوسة الخوارزمية	
الفرع الثالث: علاقة السببية وصعوباتها التقنية في الهلوسة الخوارزمية	
المطلب الثاني: المسؤولية العقدية عن أضرار الهلوسة الخوارزمية	
الفرع الأول: الأساس العقدي لاستخدام أنظمة الذكاء الاصطناعي	
الفرع الثاني: أركان العقد بين مستخدم Chat GPT ومزوّد الخدمة	
الخاتمة	
النتائج	
التوصيات	