جمهورية العراق
رئاسة ديوان الوقف السني

Republic of Iraq
Al-Sunni Endowment

مجلة كلية الإمام الأعظم الجامعة

مجلة علميّة فصلية محكّمة

الجزء
٢

رجب ١٤٤٧ هـ - كانون الأول ٢٠٢٥ م

اقرأ في هذا العدد:

ISSN: 1817-6674
رقم الإيداع في دار الكتب والوثائق ببغداد هو 818 في 2005/3/17م
coll.magazine@imamaladham.edu.iq

عدد خاص بالمؤتمر العلمي الدولي السنوي التاسع عشر في العلوم الإنسانية والتطبيقية، تحت شعـــار: «الذَّكاءُ الإصْطِناعِيُّ: رُؤْيَةٌ شَرْعِيَّةٌ وَتَكامُلُ أكادِيمِيٌّ فِي ضَوْءِ التَّحَدِّيَاتِ المُسْتَقْبَلِيَّة» في رحاب كلية الإمام الأعظم الجامعة.

Al- Imam Al-Adham
University College

A.D 2025 A.H 1447

ISSN: 1817-6674
رقم الإيداع في دار الكتب والوثائق ببغداد هو 818 في 2005/3/17م
coll.magazine@imamaladham.edu.iq

مجلة كلية الإمام الأعظم الجامعة

برعاية السيد معالي رئيس ديوان الوقف السني

أ.د. عامر شاكر عبد الجنابي المحترم ..

وبإشراف

السيد عميد كلية الإمام الأعظم الجامعة

أ.د. صلاح الدين فليح حسن المحترم

تقيم كلية الإمام الأعظم الجامعة مؤتمرها العلمي الدولي

السنوي التاسع عشر في العلوم الإنسانية والتطبيقية، تحت شعار:

«الذَّكَاءُ الِاصْطِنَاعِيُّ: رُؤْيَةٌ شَرْعِيَّةٌ وَتكَامُلٌ أَكَادِيمِيٌّ

فِي ضَوْءِ التَّحَدِّيَاتِ المُسْتَقْبَلِيَّةِ»

الذي عقد في بغداد السلام بتاريخ: ٨ - ٩ رجب ١٤٤٧هجري

الموافق ٢٨ - ٢٩ كانون الأول ٢٠٢٥ ميلادي

في رحاب كلية الإمام الأعظم الجامعة

«الجزء الثاني»

هيئة تحرير المجلة لسنة ٢٠٢٦م

اللجنة العلمية

ت	الاسم	الصفة
١	أ.د. خليل إبراهيم حمودي	رئيساً
٢	أ.د. مكي وليد عبد الكريم	عضواً
٣	أ.د. شيخموس ديمير (رئيس جامعة غازي عينتاب- تركيا)	عضواً
٤	أ.د. عبد الرحمن حمدي شافي (كلية العلوم الإسلامية-جامعة الأنبار)	عضواً
٥	أ.د. براء عبد الرزاق كامل (كلية الآداب- الجامعة العراقية)	عضواً
٦	أ.د. قاسم طه محمد	عضواً
٧	أ.د. شاكر محمود حسين	عضواً
٨	أ.د. مصعب سلمان أحمد	عضواً
٩	أ.د. معاذ عبد الستار شعبان	عضواً
١٠	أ.د. إياد إبراهيم حمودي	عضواً
١١	أ.د. عبد الكريم ناصر محمود	عضواً
١٢	أ.د. إسماعيل عبد عبّاس	عضواً
١٣	أ.د. يوسف طارق جاسم	عضواً
١٤	أ.د. لبنى رياض عبد الجبار	عضواً
١٥	أ.د. أحمد ياسين معتوق	عضواً
١٦	أ.د. حقي إسماعيل محمود	عضواً
١٧	أ.د. عمر علي حسين	عضواً
١٨	أ.د. وسام محمد خليفة	عضواً
١٩	أ.د. عماد محمد فرحان	عضواً
٢٠	أ.د. أحمد إياد أنور	عضواً
٢١	أ.د. محمد حسن علي ظاهر	عضواً

٢٢	أ.د. طارق سعود خليل	عضواً
٢٣	أ.د. أحمد نصيف جاسم	عضواً
٢٤	أ.د. باسم عبد الله عبيد	عضواً
٢٥	أ.م.د. محمد عبد الجبار عمران (كلية الآداب- الجامعة العراقية)	عضواً
٢٦	أ.م.د. باسم محمد علي	عضواً
٢٧	أ.م.د. ثابت شهاب أحمد	عضواً
٢٨	أ.م.د. عبد الوهاب أحمد حسن	عضواً
٢٩	أ.م.د. زكريا صالح سيف	عضواً
٣٠	أ.م.د. عمار عيسى عمر	عضواً
٣١	أ.م.د. عثمان راشد مجيد	عضواً
٣٢	أ.م.د. عبد الرحمن خلف مطلب	عضواً
٣٣	أ.م.د. مي حسن سريسح	عضواً
٣٤	أ.م.د. ضياء الدّين عبد الله محمد	عضواً
٣٥	أ.م.د. أحمد صدّيق إبراهيم	عضواً
٣٦	أ.م.د. قصي مساهر محمد	عضواً
٣٧	أ.م.د. زهراء عدنان عبد الكريم	عضواً
٣٨	أ.م.د. فاروق نهاد عبد	عضواً
٣٩	أ.م.د. عمر ياسين علي	عضواً
٤٠	أ.م.د. عمر حسين علوان	عضواً
٤١	أ.م.د. قحطان عدنان عبد الواحد	عضواً
٤٢	أ.م.د. طه أحمد حميد	عضواً
٤٣	أ.م.د. حسين نوار حسين	عضواً
٤٤	أ.م.د. مثنى علوان عبد	عضواً
٤٥	أ.م.د. أحمد هيثم نجم	عضواً
٤٦	أ.م.د. أحمد مهدي عبيد	عضواً

٤٧	م.د. بشار إبراهيم حميد	عضواً
٤٨	م. بكر حسين علوان (سكرتير المؤتمر)	عضواً

اللجنة التحضيرية

ت	الاسم	التخصص
١	أ.د. إسماعيل خليل إبراهيم	رئيساً
٢	أ.د. عبد الباسط أحمد حسن	عضواً
٣	أ.د. محمود جاسم معيدي	عضواً
٤	أ.م.د. عاصف دحام سالم	عضواً
٥	أ.م.د. علي داود خلف	عضواً
٦	أ.م.د. ياسين مؤيد ياسين	عضواً
٧	أ.م.د. إيناس عبد السلام داود	عضواً
٨	أ.م.د. أحمد شاكر رشيد	عضواً
٩	أ.م. معن نواف عبود	عضواً
١٠	أ.م. حبيب عبد الستار جبار	عضواً
١١	أ.م.د. عمر حسن رشيد	عضواً
١٢	أ.م.د. نزار صالح عبد	عضواً
١٣	م.علي إياد إبراهيم	عضواً
١٤	م. م. إبراهيم سمير موسى	عضواً
١٥	م.م. محمد حميد خضير	عضواً
١٦	السيد فراس رشيد عليوي (سكرتير اللجنة)	عضواً

اللجنة الإعلامية والإدارية والمالية

ت	الاسم	الصفة
١	أ.م.د. دريد عيسى إبراهيم	رئيساً
٢	أ.د. مهند ليث عبد العزيز	عضواً
٣	م. مروان محمد أمين	عضواً
٤	أ.م.د. غانم أحمد حسين	عضواً
٥	أ.م.د. زياد إبراهيم طه	عضواً
٦	م.د. أسامة زيد محمد	عضواً
٧	م.د. محمود محمد وهيب	عضواً
٨	م.م. علي عبد الحسين حسن	عضواً
٩	السيد المعتصم مؤيد عبد الرحمن	عضواً
١٠	السيد إياد مسعود عز الدين	عضواً
١١	السيد أسامة عبد الستار جبار	عضواً
١٢	السيد حيدر ماجد جابر	عضواً
١٣	السيد نزار فائق نوفان	عضواً
١٤	ميس محمد صالح	عضواً
١٥	السيد إحسان علي سليمان	عضواً
١٦	السيد يعرب خالد ستار	عضواً
١٧	رغد حسن خشان	عضواً
١٨	إستبرق أكرم عجلان	عضواً
١٩	السيد عمر محمود زيدان (سكرتير اللجنة)	عضواً

مجلة كلية الإمام الأعظم الجامعة

Al- Imam Al- Adham

University College Journal

الرقم الدولي

ISSN:1817_6674

مجلة كلية الإمام الأعظم الجامعة، مجلة إنسانية من المجلات العلمية الأكاديمية الرصينة، وقد صدرت موافقة وزارة التعليم العالي والبحث العلمي لاعتمادها بالرقم: بت/٨٦٤ في ٢٤ /٥ / ٢٠٠٥م.

شروط النشر في المجلة

شروط النشر العامة:

تسعى هيئة التحرير في مجلة كلية الإمام الأعظم الجامعة إلى الارتقاء بمعامل التأثير (impact Factor)، تمهيدًا لدخول المستوعبات العلمية العالمية، وعليه تنشر مجلة الكلية البحوث التي تتسم بالرصانة العلمية والقيمة المعرفية، وبسلامة اللغة، ودقة التوثيق وفق الشروط الآتية:

١.ألَّا يكونَ البحث منشورًا سابقًا في مجلة أخرى، وألَّا يكونَ جزءًا من بحث سابقٍ منشورٍ، أو من رسالة جامعية، وعلى الباحث أن يوقّع نموذجَ تعهدٍ بألَّا يكونَ البحث منشورًا، أو سبق تقدّيمه للنشر في مجلة أخرى، وألَّا يقدّمه للنشر في مجلة أخرى بعد نشره في مجلة كليتنا، وأن يوافق على نقل حقوق نشر البحث إلى المجلة في حال قبول نشره.

٢.ألَّا يذكر اسم الباحث أو أيّ إشارة تدلُّ عليه في متن البحث؛ لضمان سرية وحيادية عملية التحكم.

٣.ألَّا يزيد عدد الكلمات في البحث على (٨٠٠٠) كلمة، مع المصادر والملاحق، أو ألَّا يزيد على خمس وعشرين صحيفة.

٤.أن تحتوي الصحيفة الأولى من البحث ما يأتي:

أ.عنوان البحث باللغة العربية والإنجليزية.

ب.اسم الباحث ودرجته العلمية وتخصصه باللغة العربية والإنجليزية.

ج.مكان عمل الباحث باللغة العربية والإنجليزية.

د.رقم هاتف الباحث وبريده الإلكتروني الجامعي.

٥.يقدّم الباحث ملخصًا (باللغة العربية والإنجليزية) لا يزيد على (٢٠٠) كلمة.

٦.يوضع بعد الملخص (Abstract) مباشرة الكلمات المفتاحية لموضوع البحث (Key word).

٧.على الباحث اتّباع قواعد الاقتباس وتوثيق المصادر، وأخلاقيات البحث العلمي بما يتوافق مع سياسة المجلة.

٨. تكتب مصادر البحث في صحيفة أو صحائف مستقلة مرتبة بحسب الأصول المعتمدة، وذلك على النحو الآتي: عنوان الكتاب، اسم المؤلف، دار النشر، مكان النشر(المدينة) رقم الطبعة مثال (ط٣)، (سنة الطبع).

٩.الاستشهاد بعددين من أعداد المجلة المنشورة سابقًا والمرفوعة في الموقع الإلكتروني الخاص بكليتنا في الرابط الإلكتروني: https://www.iasj.net/iasj/journal/issues/٢٢٤

٩.ترجمة المصادر باللغة الإنجليزية.

١٠.تطبق المجلة نظام فحص الاستلال الإلكتروني باستخدام برنامج (Turnitin) ويرفض نشر الأبحاث التي تتجاوز فيها نسبة الاستلال ٢٠٪.

١١.يخضع البحث لفحص أولي تقوم به هيئة التحرير في المجلة، وذلك لتقرير أهلية البحث للتحكيم، ويحق لها أن تعتذر عن قبول البحث دون تقديم الأسباب.

١٢.تتبع المجلة التقويم المزدوج السري لبيان صلاحية البحث للنشر، إذ يعرض البحث المقدّم للنشر على محكمين اثنين من ذوي الاختصاص، ويتم اختيارهما بسرية مطلقة، بالإضافة إلى عرض البحث على خبير لغوي لتقويم سلامته اللغوية.

١٣.الأبحاث التي يقترح المحكمون إجراء تعديلات عليها لتكون صالحة للنشر، تعاد إلى أصحابها لإجراء التعديلات المطلوبة عليها، وخلاف ذلك لا يتم استلام البحث، وستتم مراجعة البحث من قبل هيئة التحرير للتأكد من التزام الباحث بالأخذ بجميع الملاحظات المثبتة من قبل المقيمين.

١٤.تُعبِّر الأبحاث المنشورة في المجلة عن آراء أصحابها، لا عن رأي المجلة.

١٥.تنشر المجلة أعدادًا خاصة بالمؤتمرات العلمية المتوافقة مع تخصص المجلة.

١٦.أجور نشر البحث: يدفع الباحث (٥٠) ألف دينار لتغطية أجور التحكيم، ويكمل دفع بقية الأجور عند قبول البحث للنشر.

١٧.لا تأخذ المجلة أي أجور لنشر الأبحاث المقدّمة من باحثين من خارج العراق.

١٨.يتم إرسال الأبحاث عبر الإيميل: magazine@imamaladham.edu.iq.

١٩.تخريج النصوص القرآنية والحديث النبوي الشريف على ضوء المنهج العلمي الدقيق الكامل.

٢٠.يزود الباحث بنسختين مستلة، بعد النشر.

شروط النشر (الفنيَّة):

١- يقدّم البحث بملف واحد، يبدأ بالعنوان وينتهي بالمصادر، وألَّا يزيد على خمس وعشرين صحيفة.

٢- تكون الهوامش أسفل كل صحيفة (تلقائيًا وليس يدويًّا).

٣- حجم الخط للمتن (١٦)، وللهامش (١٢).

٤- نوع الخط باللغة العربية ((Simplified Arabic واللغة الإنجليزية (Times New Roman))

- **ملاحظة:** في حال عدم الأخذ بشروط النشر نعتذر عن استلام البحث ونشره.

- يمكن زيارة موقع المجلة في مبنى الكلية في سبع إبكار أو التواصل عبر البريد الإلكتروني magazine@imamaladham.edu.iq.

أو الاتصال بمدير التحرير عبر الهاتف (٠٧٧٣٢٤٣٥٦٩٣)، ويمكن الاطلاع على أعداد المجلة عن طريق موقع وزارة التعليم العالي والبحث العلمي من خلال مسح رمز QR في أعلى الصفحة.

البيان الختامي للمؤتمر العلمي الدولي التاسع عشر
في العلوم الإنسانيّة والتطبيقية
تَحْتَ شِعَارِ: «الذَّكَاءُ الِاصْطِنَاعِيُّ: رُؤْيَةٌ شَرْعِيَّةٌ وَتَكَامُلٌ أَكَادِيمِيٌّ فِي ضَوْءِ التَّحَدِّيَاتِ المُسْتَقْبَلِيَّةِ»

الحَمْدُ لِلَّهِ الَّذِي جَعَلَ العَقْلَ أَمَانَةً، وَالعِلْمَ رِسَالَةً، وَسَخَّرَ لِلْإِنْسَانِ مِنْ أَدَوَاتِ المَعْرِفَةِ مَا يُعِينُهُ عَلَى الفَهْمِ وَالِاسْتِخْلَافِ، فَأَقَامَ بِهِ مِيزَانَ التَّفْكِيرِ، وَضَبَطَ بِهِ حَرَكَةَ التَّطَوُّرِ، فَلَا تَنْفَصِلُ التِّقْنِيَةُ عَنِ القِيَمِ، وَلَا يَتَقَدَّمُ المُنْجَزُ عَلَى الإِنْسَانِ، وَالصَّلَاةُ وَالسَّلَامُ عَلَى سَيِّدِنَا مُحَمَّدٍ ﷺ، إِمَامِ العُلَمَاءِ، وَمُعَلِّمِ الإِنْسَانِيَّةِ، الَّذِي قَرَنَ العِلْمَ بِالهِدَايَةِ، وَرَبَطَ المَعْرِفَةَ بِالأَخْلَاقِ، فَكَانَ هَدْيُهُ مِيزَانَ الرُّشْدِ، وَمَنْهَجُهُ سَبِيلَ الِاتِّزَانِ، وَعَلَى آلِهِ وَصَحْبِهِ وَمَنْ سَارَ عَلَى نَهْجِهِ القَوِيمِ إِلَى يَوْمِ الدِّينِ، وَبَعْدُ... فَفِي خِتَامِ أَعْمَالِ هٰذَا المَحْفِلِ العِلْمِيِّ المُبَارَكِ، وَمِنْ بَغْدَادَ السَّلَامِ، حَاضِرَةِ العِلْمِ، وَمَوْئِلِ الحَضَارَةِ، وَمَهْدِ التَّلَاقُحِ المَعْرِفِيِّ عَبْرَ العُصُورِ، وَفِي رِحَابِ العِرَاقِ الَّذِي مَا زَالَ، رَغْمَ التَّحَدِّيَاتِ، يَحْمِلُ فِي ذَاكِرَتِهِ رِسَالَةَ القَلَمِ وَالكِتَابِ، اخْتَتَمَتْ كُلِّيَّةُ الإِمَامِ الأَعْظَمِ الجَامِعَةُ أَعْمَالَ مُؤْتَمَرِهَا العِلْمِيِّ الدُّوَلِيِّ التَّاسِعِ عَشَرَ لِلْعُلُومِ الإِنْسَانِيَّةِ وَالتَّطْبِيقِيَّةِ، تَحْتَ شِعَارِ: «الذَّكَاءُ الِاصْطِنَاعِيُّ: رُؤْيَةٌ شَرْعِيَّةٌ وَتَكَامُلٌ أَكَادِيمِيٌّ فِي ضَوْءِ التَّحَدِّيَاتِ المُسْتَقْبَلِيَّةِ»، وَالَّذِي عُقِدَ يَوْمَ الأَحَدِ السَّابِعَ مِنْ شَهْرِ رَجَبٍ، لِسَنَةِ سَبْعٍ وَأَرْبَعِينَ وَأَرْبَعِمِئَةٍ وَأَلْفٍ لِلْهِجْرَةِ النَّبَوِيَّةِ الشَّرِيفَةِ، المُوَافِقِ الثَّامِنَ وَالعِشْرِينَ مِنْ شَهْرِ كَانُونِ الأَوَّلِ، لِسَنَةِ خَمْسٍ وَعِشْرِينَ وَأَلْفَيْنِ لِلْمِيلَادِ، بِرِعَايَةٍ كَرِيمَةٍ مِنْ لَدُنْ مَعَالِي رَئِيسِ دِيوَانِ الوَقْفِ السُّنِّيِّ، الأُسْتَاذِ الدُّكْتُورِ عَامِرٍ شَاكِرٍ عَبْدِ الجَنَابِيِّ، وَبِإِشْرَافِ الأُسْتَاذِ الدُّكْتُورِ صَلَاحِ الدِّينِ فَلِيحِ حَسَنٍ السَّامِرَّائِيِّ، وِفْقَ رُؤْيَةٍ أَكَادِيمِيَّةٍ وَاضِحَةٍ انْتَهَجَهَا مُنْذُ تَسَنُّمِهِ عَمَادَةَ الكُلِّيَّةِ، تَقُومُ عَلَى ضَرُورَةِ التَّحَوُّلِ الرَّقْمِيِّ بِوَصْفِهِ خِيَارًا اسْتِرَاتِيجِيًّا لِمُوَاكَبَةِ الحَدَاثَةِ العِلْمِيَّةِ، وَتَسْرِيعِ الإِنْجَازِ المُؤَسَّسِيِّ، وَتَوْظِيفِ التِّقْنِيَّاتِ الذَّكِيَّةِ فِي خِدْمَةِ التَّعْلِيمِ وَالبَحْثِ العِلْمِيِّ، ضِمْنَ إِطَارٍ قِيَمِيٍّ رَصِينٍ يُوَازِنُ بَيْنَ الأَصَالَةِ وَالمُعَاصَرَةِ، وَبِمُشَارَكَةِ نُخْبَةٍ مُبَارَكَةٍ مِنَ العُلَمَاءِ وَالبَاحِثِينَ وَالأَكَادِيمِيِّينَ مِنْ دَاخِلِ العِرَاقِ وَخَارِجِهِ، حُضُورًا وَمُشَارَكَةً عِلْمِيَّةً عَنْ بُعْدٍ.

وَقَدْ قُدِّمَتْ إِلَى اللَّجْنَةِ العِلْمِيَّةِ عَشَرَاتُ البُحُوثِ، قُبِلَ مِنْهَا لِلْمُشَارَكَةِ وَاحِدٌ وَأَرْبَعُونَ بَحْثًا مَحَلِّيًّا، وَتِسْعَةُ أَبْحَاثٍ دُوَلِيَّةً، تَوَزَّعَتْ بَرَامِجُهَا عَلَى جَلَسَاتٍ عِدَّةٍ، وَتَشَرَّفْنَا بِاسْتِضَافَةِ عَدَدٍ

مِنَ الضُّيُوفِ الأَكَارِمِ مِنْ جَامِعَاتٍ وَمُؤَسَّسَاتٍ عَرَبِيَّةٍ وَعَالَمِيَّةٍ، فِي أَجْوَاءٍ اتَّسَمَتْ بِالجِدِّيَّةِ العِلْمِيَّةِ، وَعُمْقِ الطَّرْحِ، وَرَصَانَةِ النِّقَاشِ، وَتَكَامُلِ الرُّؤَى.

أَيُّهَا الحُضُورُ الكَرِيمُ، السَّادَةُ البَاحِثُونَ الفُضَلَاءُ: لَقَدْ جَاءَ هٰذَا المُؤْتَمَرُ اسْتِجَابَةً وَاعِيَةً لِلتَّحَوُّلَاتِ المُتَسَارِعَةِ الَّتِي يَشْهَدُهَا العَالَمُ فِي مَيْدَانِ التِّقْنِيَّاتِ الذَّكِيَّةِ، وَإِيمَانًا مِنَ الكُلِّيَّةِ بِضَرُورَةِ مُقَارَبَةِ الذَّكَاءِ الِاصْطِنَاعِيِّ مُقَارَبَةً عِلْمِيَّةً مُتَوَازِنَةً، لَا تَنْبَهِرُ بِالمُنْجَزِ التِّقَنِيِّ دُونَ وَعْيٍ، وَلَا تَنْغَلِقُ دُونَهُ دُونَ فِقْهٍ وَبَصِيرَةٍ، بَلْ تُخْضِعُهُ لِمَوَازِينِ الشَّرِيعَةِ، وَأَخْلَاقِيَّاتِ العِلْمِ، وَمَسْؤُولِيَّةِ الإِنْسَانِ عَنْ قَرَارِهِ وَمَصِيرِهِ.

وَقَدْ تَنَاوَلَتْ بُحُوثُ المُؤْتَمَرِ وَمَحَاوِرُهُ المُتَنَوِّعَةُ أَثَرَ الذَّكَاءِ الِاصْطِنَاعِيِّ فِي العُلُومِ الشَّرْعِيَّةِ، وَاللُّغَةِ العَرَبِيَّةِ، وَالعُلُومِ الإِنْسَانِيَّةِ، وَالعُلُومِ التَّطْبِيقِيَّةِ، وَالقَانُونِ، وَالتَّعْلِيمِ، وَالإِعْلَامِ، وَالتَّارِيخِ وَالجُغْرَافِيَا، مُبَيِّنَةً إِمْكَانَاتِهِ الوَاعِدَةَ فِي خِدْمَةِ المَعْرِفَةِ، وَمُحَذِّرَةً فِي الوَقْتِ نَفْسِهِ مِنْ مَخَاطِرِهِ المَعْرِفِيَّةِ وَالأَخْلَاقِيَّةِ، وَلَا سِيَّمَا مَا يَتَّصِلُ بِالتَّحَيُّزِ الخَوَارِزْمِيِّ، وَتَزْيِيفِ الوَعْيِ، وَانْتِهَاكِ الخُصُوصِيَّةِ، وَإِضْعَافِ المَسْؤُولِيَّةِ الإِنْسَانِيَّةِ، وَفِي ضَوْءِ المُشَارَكَاتِ وَالجَلَسَاتِ البَحْثِيَّةِ، وَالنِّقَاشَاتِ العِلْمِيَّةِ المُسْتَفِيضَةِ وَالبَنَّاءَةِ، خَلُصَ المُؤْتَمَرُ إِلَى جُمْلَةٍ مِنَ التَّوْصِيَاتِ، كَانَ مِنْ أَبْرَزِهَا:

أَوَّلًا: إِخْضَاعُ جَمِيعِ تَطْبِيقَاتِ الذَّكَاءِ الِاصْطِنَاعِيِّ لِمَوَازِينِ الشَّرْعِ وَالأَخْلَاقِ، بِمَا يَحْفَظُ كَرَامَةَ الإِنْسَانِ، وَيُعَزِّزُ وَعْيَهُ، وَيَصُونُ حَقَّهُ، وَيَضْمَنُ الِاسْتِعْمَالَ المَسْؤُولَ لِلتِّقْنِيَّةِ وَتَوْظِيفَهَا فِي خِدْمَةِ المُجْتَمَعِ.

ثَانِيًا: تَعْزِيزُ التَّعَاوُنِ وَالتَّكَامُلِ بَيْنَ العُلُومِ الشَّرْعِيَّةِ، وَالإِنْسَانِيَّةِ، وَالتَّطْبِيقِيَّةِ عِنْدَ دِرَاسَةِ تِقْنِيَّاتِ الذَّكَاءِ الِاصْطِنَاعِيِّ، لِضَمَانِ مُقَارَبَةٍ شَامِلَةٍ تَجْمَعُ بَيْنَ الفَهْمِ النَّظَرِيِّ وَالقُدْرَةِ التَّطْبِيقِيَّةِ.

ثَالِثًا: تَوْظِيفُ الذَّكَاءِ الِاصْطِنَاعِيِّ تَوْظِيفًا رَشِيدًا فِي خِدْمَةِ القُرْآنِ وَعُلُومِهِ، وَالحَدِيثِ وَعُلُومِهِ، وَاللُّغَةِ العَرَبِيَّةِ وَعُلُومِهَا، مَعَ ضَرُورَةِ التَّحَقُّقِ النَّقْدِيِّ مِنَ النَّتَائِجِ وَمُرَاجَعَتِهَا، وَعَدَمِ الِاعْتِمَادِ الكُلِّيِّ عَلَى مُخْرَجَاتِهِ دُونَ تَمْحِيصٍ وَتَدْقِيقٍ.

رَابِعًا: الدَّعْوَةُ إِلَى بِنَاءِ أُطُرٍ قَانُونِيَّةٍ وَتَشْرِيعِيَّةٍ وَاضِحَةٍ تُنَظِّمُ العَلَاقَاتِ الرَّقْمِيَّةَ، وَتُحَدِّدُ المَسْؤُولِيَّةَ القَانُونِيَّةَ، وَتَحْمِي المُجْتَمَعَ مِنَ الِانْتِهَاكَاتِ التِّقْنِيَّةِ.

خَامِسًا: التَّنْبِيهُ إِلَى المَخَاطِرِ المُتَرَتِّبَةِ عَلَى الِاسْتِعْمَالِ غَيْرِ المُنْضَبِطِ لِلذَّكَاءِ الِاصْطِنَاعِيِّ، وَلَا سِيَّمَا فِي مَجَالَاتِ الإِعْلَامِ، وَالتَّعْلِيمِ، وَصِنَاعَةِ الرَّأْيِ العَامِّ، مَعَ وَضْعِ آلِيَّاتٍ لِلْحَدِّ مِنَ الِانْتِهَاكَاتِ المَعْرِفِيَّةِ وَالأَخْلَاقِيَّةِ.

سَادِسًا: تَشْجِيعُ الجَامِعَاتِ وَالمُؤَسَّسَاتِ البَحْثِيَّةِ عَلَى إِطْلَاقِ مَشَارِيعَ وَدِرَاسَاتٍ تُعْنَى بِاسْتِشْرَافِ مُسْتَقْبَلِ الذَّكَاءِ الِاصْطِنَاعِيِّ وَآثَارِهِ المُجْتَمَعِيَّةِ وَالحَضَارِيَّةِ.

سَابِعًا: دَعْمُ البُحُوثِ المُشْتَرَكَةِ بَيْنَ عُلَمَاءِ الشَّرِيعَةِ وَخُبَرَاءِ التِّقْنِيَّةِ لِتَطْوِيرِ أَنْظِمَةٍ تُجَسِّدُ قِيَمَ الشَّرْعِ، وَتَخْدِمُ قَضَايَا العَصْرِ، وَإِنْشَاءُ لِجَانٍ شَرْعِيَّةٍ مُتَخَصِّصَةٍ لِمُوَاكَبَةِ المُسْتَجِدَّاتِ التِّقْنِيَّةِ، وَإِصْدَارُ الفَتَاوَى وَالتَّوْصِيَاتِ اللَّازِمَةِ.

ثَامِنًا: التَّأْكِيدُ عَلَى دَوْرِ المُؤَسَّسَاتِ الأَكَادِيمِيَّةِ فِي نَشْرِ الوَعْيِ الرَّقْمِيِّ، وَبِنَاءِ ثَقَافَةٍ نَقْدِيَّةٍ رَشِيدَةٍ فِي التَّعَامُلِ مَعَ التِّقْنِيَّاتِ الحَدِيثَةِ.

تَاسِعًا: إِدْمَاجُ أَخْلَاقِيَّاتِ الذَّكَاءِ الِاصْطِنَاعِيِّ مِنْ مَنْظُورٍ إِسْلَامِيٍّ فِي المَنَاهِجِ الشَّرْعِيَّةِ وَالتِّقْنِيَّةِ، لِإِعْدَادِ جِيلٍ يَجْمَعُ بَيْنَ الإِيمَانِ وَالخِبْرَةِ، وَيَكُونُ قَادِرًا عَلَى مُوَاجَهَةِ تَحَدِّيَاتِ العَصْرِ بِوَعْيٍ وَحِكْمَةٍ.

وَفِي الخِتَامِ، تَتَقَدَّمُ كُلِّيَّةُ الإِمَامِ الأَعْظَمِ الجَامِعَةُ، مُمَثَّلَةً بِعَمِيدِهَا الأُسْتَاذِ الدُّكْتُورِ صَلَاحِ الدِّينِ فَلِيحِ حَسَنٍ السَّامِرَّائِيِّ، بِالشُّكْرِ الجَزِيلِ إِلَى جَمِيعِ البَاحِثِينَ وَالمُشَارِكِينَ فِي المُؤْتَمَرِ، وَإِلَى كُلِّ مَنْ حَضَرَ وَأَسْهَمَ، وَإِلَى اللِّجَانِ العِلْمِيَّةِ وَالتَّحْضِيرِيَّةِ وَالإِدَارِيَّةِ وَالإِعْلَامِيَّةِ، وَالأَقْسَامِ السَّانِدَةِ الَّتِي بَذَلَتْ جُهُودًا مُتَمَيِّزَةً لِإِنْجَاحِ هٰذَا المَحْفِلِ العِلْمِيِّ، سَائِلِينَ اللَّهَ تَعَالَى أَنْ يَجْعَلَ مُخْرَجَاتِهِ عِلْمًا نَافِعًا، وَرَأْيًا سَدِيدًا، وَخُطْوَةً رَاسِخَةً فِي سَبِيلِ تَرْشِيدِ التِّقْنِيَّةِ بِالقِيَمِ، وَتَسْخِيرِ العِلْمِ لِخِدْمَةِ الإِنْسَانِ، لَا أَدَاةَ إِفْسَادٍ أَوْ طُغْيَانٍ.

هذا وَالحَمْدُ لِلَّهِ فِي البَدْءِ وَالخِتَامِ، وَالصَّلَاةُ وَالسَّلَامُ عَلَى خَيْرِ الأَنَامِ، وَعَلَى آلِهِ وَصَحْبِهِ العُلَمَاءِ الأَعْلَامِ، وَأَخْتِمُ هٰذَا البيانِ بِالسَّلَامِ ...

فالسَّلَامُ عَلَيْكُمْ وَرَحْمَةُ اللَّهِ وَبَرَكَاتُهُ.

صَادِرٌ عَنِ المُشَارِكِينَ فِي المُؤْتَمَرِ العِلْمِيِّ الدُّوَلِيِّ التَّاسِعِ عَشَرَ

بِرِحَابِ كُلِّيَّةِ الإِمَامِ الأَعْظَمِ الجَامِعَةِ - بَغْدَاد

المقدمة

الحمدُ للهِ الذي علَّم بالقلم، علَّم الإنسانَ ما لم يعلم، وهداه بنورِ العقلِ حينما أَظلم، وفتحَ له آفاقَ الذكاءِ والتَّعلُّم، فجعل من الآلة خادمًا، ومن الفكر قائدًا، ومن العلم سُلَّمًا للفهمِ والسُّؤدد، والصلاةُ والسلامُ على من جاء بالعلم والهدى، ودلَّ البشرية على سُبُلِ الرُّقي والاقتداء، سيّدِنا محمد، المعلِّمِ الأَمين، وعلى آلهِ وصحبِه أجمعين.

أما بعد ...

ففي سياقٍ معرفي يشهد تحوّلات متسارعة، بات الذكاء الاصطناعي أحد أبرز الظواهر التي تُعيد رسم خارطة العالم في مختلف ميادين الحياة، لقد غدت الآلة تفكّر، وتستنبط، وتتعلّم، وتحاكي العقل البشري في وظائفه العليا، حتى صار الذكاء الاصطناعي قوةً دافعة لا يمكن تجاهل أثرها في تشكيل مستقبل المجتمعات، وأنماط التعليم، ومفاهيم العمل، وحدود المسؤولية الإنسانية.

وانطلاقًا من مسؤوليتها العلمية والدينية والوطنية، تواصل كلية الإمام الأعظم الجامعة أداء دورها الريادي في مواكبة مستجدات العصر، عن طريق إقامة مؤتمرها العلمي الدولي السنوي التاسع عشر للعلوم الإنسانية والتطبيقية، تحت شعار: **(الذكاء الاصطناعي: رؤية شرعية وتكامل أكاديمي في ضوء التحديات المستقبلية)**؛ ليكون منبرًا علميًا للحوار الرصين، ومجالًا لتلاقح الأفكار بين الباحثين من مختلف التخصصات، في سبيل فهمٍ أعمق لهذه الظاهرة العالمية، وتوجيهها بما ينسجم مع قيمنا الإسلامية الأصيلة وثوابتنا التربوية والفكرية.

وأظهرت هذه التقنية إمكانات هائلة في تسريع الإنجاز، وتحسين الجودة، وتطوير مناهج التعليم والإدارة، وفتح آفاق جديدة للبحث العلمي.

إلّا أن الاستعمال غير المنضبط أو غير المؤطر بالقيم والمعايير الأخلاقية قد يخلّف آثارًا سلبية عميقة، من بينها: تهديد الخصوصية، وتعزيز التحيز الخوارزمي، وتراجع دور الإنسان في اتخاذ القرار، وإضعاف الروابط الاجتماعية، وطمس الهوية الثقافية والدينية.

ومن هنا، فإن الذكاء الاصطناعي لا يمثل تطورًا تقنيًا فحسب، بل هو تحول في نمط التفكير البشري، ومسارٌ جديدٌ في العلاقة بين الإنسان والآلة، يستوجب تأصيلًا معرفيًّا،

وتأمّلاً فلسفيًّا، وتأطيرًا شرعيًّا وأخلاقيًّا، وهو ما تسعى إليه محاور هذا المؤتمر، في أثناء مقاربات متعددة تشمل: الجوانب العلمية، والاجتماعية، والقانونية، والتربوية، فضلاً عن الرؤى الإسلامية الأصيلة التي تستشرف الغد بروح منفتحة وفكر نقدي راشد.

فكلية الإمام الأعظم الجامعة، إذ تنظم هذا المؤتمر، تؤكد حرصها على بناء جسرٍ معرفيٍّ يربط بين التراث العلمي الرصين والتقنية الحديثة، في إطار من المسؤولية الأخلاقية، والانفتاح الواعي، والحرص على أن تظل المعرفة وسيلةً لخدمة الإنسان، لا أداةً لتغييبه أو إخضاعه.

نسأل الله أن يكلل هذا الجهد بالتوفيق والسداد، وأن يُثمر المؤتمر نقاشاتٍ جادةً، ومقترحاتٍ نافعةً، تسهم في تعميق الوعي، وتوسيع دائرة المسؤولية الأكاديمية اتجاه هذا التحدي العالمي.

الرسالة:

نطمح في مؤتمرنا إلى تقديم فضاء علمي رصين يُعنى بدراسة آفاق الذكاء الاصطناعي من منظور معرفي شامل، يجمع بين الأصالة والمعاصرة، ويؤسس لرؤية منهجية تدعم الاستفادة من هذه التقنية بما يخدم الإنسان والقيم، ويحذّر من مخاطر الانفلات الأخلاقي وسوء الاستعمال.

الرؤية:

أن يكون مؤتمر كلية الإمام الأعظم الجامعة منبرًا فكريًا رائدًا في تناول موضوعات الذكاء الاصطناعي برؤية مستقبلية تجمع بين القيم الحضارية والتطور التقني، وتسهم في إنتاج معرفة أصيلة ومؤثرة تتبصّر الإيجابيات وتتصدى للسلبيات.

أهداف المؤتمر:

١. تسليط الضوء على إمكانات الذكاء الاصطناعي في تطوير مناهج البحث العلمي في مختلف التخصصات.

٢. تعزيز التكامل بين معطيات الثورة الرقمية وتعاليم الشريعة الإسلامية.

٣. استكشاف سبل توظيف الذكاء الاصطناعي في خدمة اللغة العربية وتحليلها.

٤. بحث التطبيقات العملية للذكاء الاصطناعي في مجالات العلوم الطبية والهندسية والاقتصادية.

٥. بناء شبكة تواصل بحثي بين الأكاديميين والباحثين في مجالات الذكاء الاصطناعي المختلفة.

٦. بيان المخاطر المحتملة لاستعمال الذكاء الاصطناعي دون ضوابط شرعية وأخلاقية.

٧. مناقشة التحديات الفكرية والقيمية المرتبطة بانتشار الذكاء الاصطناعي.

٨. تحليل الأثر السلبي للذكاء الاصطناعي في حال الانفصال عن المرجعيات الدينية والإنسانية.

محاور المؤتمر:

أولًا: الذكاء الاصطناعي والعلوم الشرعية:

- إمكانات الذكاء الاصطناعي في خدمة العلوم الشرعية.
- الأسس الشرعية للتعامل مع الذكاء الاصطناعي.
- الذكاء الاصطناعي بين الضرورات والمقاصد الشرعية.
- الذكاء الاصطناعي في الفقه وأصوله: أدوات الفتوى الإلكترونية.
- أخلاقيات الذكاء الاصطناعي من منظور الشريعة الإسلامية.
- بيان الانحرافات الشرعية المحتملة في استعمال الذكاء الاصطناعي دون رقابة شرعية.

ثانيًا: الذكاء الاصطناعي والعلوم اللغوية:

- توظيف تقنيات الذكاء الاصطناعي وأدواته في خدمة علوم اللغة، وتحليل النصوص الأدبية والبلاغية.
- دور الذكاء الاصطناعي في تطوير مناهج تعليم اللغة العربية والإنجليزية.
- المعالجة الآلية للغة العربيّة والإنجليزية بين التحديات والفرص.
- الذكاء الاصطناعي في تطوير طرائق تعليم اللغة العربيّة والإنجليزية، وتقويم أداء المتعلمين.
- مخاطر الترجمة الآلية والتشويش الدلالي على النصوص.

ثالثًا: الذكاء الاصطناعي والعلوم التطبيقية:

- تطبيقات الذكاء الاصطناعي في الطب، والهندسة، وتقنيات الاتصالات الحديثة.
- الذكاء الاصطناعي في الإدارة والاقتصاد والتحول الرّقمي.
- النمذجة الذكية في تحليل البيانات واتخاذ القرار.
- التحديات الأمنية في نظم الذكاء الاصطناعي والهجمات السيبرانية.

رابعًا: الذكاء الاصطناعي والعلوم الإنسانية:

- الذكاء الاصطناعي في التعليم، والتعليم الذكي والتدريب الافتراضي.
- أثر الذكاء الاصطناعي في تحليل الأحداث التاريخية والأنماط الجغرافية وتفسيرها: الإمكانات العلمية والمخاطر المعرفية.
- الذكاء الاصطناعي والإعلام الرقمي وصناعة الرأي العام.
- الاخلاقيات والقوانين المنظمة لاستعمال الذكاء الاصطناعي.
- الذكاء الاصطناعي والتحديات الأخلاقية في تشكيل السلوك المجتمعي.

محتويات الجزء الثاني

التحديات الأمنية في نظم الذكاء الإصطناعي والهجمات السيبرانية

Security Challenges in Artificial Intelligence Systems and Cyberattacks

إعداد الباحث

م. عبد المنعم شاكر عبد الله

كلية الإمام الأعظم الجامعة - قسم أصول الدين

آلتون كوبري

By

Lecturer Abdulmunim Shakir Abdullah

Al-Imam Al-Adham University College

ambaghdad@gmail.com

الملخص

تتناول هذه الدراسة بالتحليل العلاقة المتشابكة بين البنية التقنية لنظم الذكاء الاصطناعي وما ينتج عنها من مخاطر أمنية متنامية، مع التركيز على استيعاب طبيعة التهديدات السيبرانية المعاصرة وسبل مواجهتها بفعالية. وتنطلق الدراسة من افتراض مفاده أن التوسع المتسارع في دمج تقنيات الذكاء الاصطناعي داخل القطاعات الحيوية قد أفرز بيئة رقمية جديدة، أتاحت المجال لظهور أساليب هجومية متطورة تتخطى النماذج التقليدية للأمن السيبراني.

في هذا السياق، تستعرض الدراسة في محورها الأول أبرز التحديات الأمنية الهيكلية التي تواجه نظم الذكاء الاصطناعي، والتي تشمل قضايا موثوقية وجودة البيانات واسعة النطاق، وإشكالية غياب الشفافية في آليات اتخاذ القرار داخل النماذج المعقدة، فضلًا عن المخاطر المترتبة على الاعتماد المتزايد على أطراف خارجية ضمن سلاسل التوريد التقنية. أما المحور الثاني، فيقدم معالجة تقنية معمقة لأنماط الهجمات السيبرانية التي تستهدف هذه النظم، مثل التلاعب ببيانات التدريب، ومحاولات تجاوز آليات الحماية أثناء مرحلة التشغيل، إضافة إلى تقنيات تحليل النماذج واستخلاص معلوماتها الداخلية.

وتقترح الدراسة إطارًا وقائيًا متكاملًا للتصدي لهذه التهديدات، يقوم على توظيف مجموعة من المقاربات المتقدمة، من بينها تقنيات تعزيز الخصوصية، والتعلم المقاوم للهجمات، وتحليل السلوك غير الاعتيادي، إلى جانب الاستفادة من مفاهيم الذكاء الاصطناعي القابل للتفسير لتعزيز الثقة والشفافية. وقد تم دعم هذا الإطار بدراسات تطبيقية في مجالي الأمن والرعاية الصحية، أظهرت نتائجها بوضوح أهمية تبني استراتيجية أمنية استباقية تغطي جميع مراحل دورة حياة نظم الذكاء الاصطناعي، من التصميم إلى التشغيل.

الكلمات المفتاحية: (نظم الذكاء الاصطناعي، الأمن السيبراني، التهديدات الرقمية، الهجمات على النماذج الذكية، الحماية الاستباقية، استراتيجيات الدفاع).

Abstract:

This study analyses the intricate relationship between the technical architecture of Artificial Intelligence (AI) systems and the resulting escalated security risks, focusing on understanding the nature of contemporary cyber threats and effective mitigation strategies. The research stems from the premise that the rapid integration of AI technologies within critical sectors has fostered a new digital environment, giving rise to sophisticated attack vectors that transcend traditional cybersecurity models.

In this context, the first part of the study explores the primary structural security challenges facing AI systems, including issues of reliability and quality of large-scale data, the problem of transparency in decision-making mechanisms within complex models (Black Box), and the risks associated with increasing reliance on third parties within technical supply chains. The second part provides an in-depth technical analysis of cyber-attack patterns targeting these systems, such as training data manipulation (Poisoning), attempts to bypass protection mechanisms during the inference phase (Evasion), and model extraction techniques.

The study proposes an integrated preventive framework to counter these threats, based on employing advanced approaches, including privacy-enhancing technologies, adversarial learning, and anomaly detection, alongside leveraging Explainable AI (XAI) concepts to bolster trust and transparency. This framework is supported by empirical studies in the fields of security and healthcare, with results clearly demonstrating the importance of adopting a proactive security strategy that covers all stages of the AI system lifecycle, from design to operation.

Keywords: AI Systems, Cybersecurity, Digital Threats, Adversarial Attacks, Proactive Protection, Defence Strategies.

المقدمة

بعد ان أصبح من الضروري استخدام الذكاء الاصطناعي في قطاعات مختلفة كالرعاية والصحة والتعليم وغيرها من مفاصل الحياة المختلفة، فقد أضحت قضايا الأمن السيبراني المرتبطة بهذه القطاعات تحتل الأولوية العالية لدى المؤسسات التابعة لها.

ومن هذا المنطلق سوف نتناول في بحثنا هذا دراسة التهديدات الأمنية التي تنال أو تحاول النيل من نظم الذكاء الاصطناعي، وأساليب وطرق الهجوم السيبراني شائعة الاستعمال، والاستراتيجيات المقترحة للحماية من هذه الهجمات، إضافة إلى بعض الأمثلة التطبيقية لها، مذيلاً بالتوصيات التي قد تخدم في تجنب هذه التهديدات أو التخفيف من أثرها على المؤسسات. حيث سيتألف هذا البحث من مقدمة وأربعة مباحث وعدة مطالب مختوماً بالخاتمة والتوصيات ثم المصادر، وعلى النحو الآتي:

المبحث الأول: التحديات الأمنية الرئيسية في نظم الذكاء الاصطناعي، ويتكون من ثلاثة مطالب:

الأول: البيانات الضخمة المغذية للنظام

الثاني: النماذج غير القابلة للتفسير

الثالث: الاعتماد على أطراف خارجية

المبحث الثاني: أنواع الهجمات السيبرانية على نظم الذكاء الاصطناعي، ويتألف من أربعة مطالب:

الأول: هجمات الإدخال المُضَلِّل

الثاني: هجمات التسميم

الثالث: هجمات الاستخراج

الرابع: هجمات الاستدلال

المبحث الثالث: استراتيجيات الحماية والتصدي، ويقع في أربعة مطالب:

الأول: تعزيز الخصوصية

الثاني: تدريب النماذج ضد الهجمات

الثالث: تحليل السلوك

الرابع: تحسين الشفافية والتفسير

المبحث الرابع: أنواع التهديدات واستراتيجيات الحماية في السياق التطبيقي، وهو في مطلبين:

الأول: أنواع التهديدات في السياق التطبيقي

الثاني: استراتيجيات الحماية، تقنيات دفاعية مثبتة وأمثلة تطبيقية

المبحث الأول: التحديات الأمنية الرئيسية في نظم الذكاء الاصطناعي

لقد انتشرت نظم الذكاء الاصطناعي (AI) في السنوات الأخيرة انتشاراً مبهراً في جميع قطاعات حياتنا الحيوية، وقد صاحب هذا الانتشار المتزايد بروز تحديات أمنية معقدة تؤثر سلباً على موثوقية وسلامة هذه النظم. وسيتناول هذا المبحث تحليلاً لبعض التحديات الأمنية الرئيسية وهي: أمن وسلامة البيانات الضخمة المغذية للنظام، ومشكلة عدم قابلية تفسير النماذج (Black Box)، ومخاطر الاعتماد على الأطراف الخارجية في سلسلة توريد الذكاء الاصطناعي.

وبناءً على ذلك، سيقسم هذا المبحث إلى أربعة مطالب على النحو الآتي:

المطلب الأول: البيانات الضخمة المغذية للنظام

إن أي خلل في جودة أو أمان البيانات التي تغذي خوارزميات الذكاء الاصطناعي ينعكس بشكل سلبي على أداء وأمن النموذج الناتج عن هذه الخوارزميات (Shin, J. , et al. ,٢٠٢٤). ويتمثل وجود مثل هذا الخلل بوجود مخاطر بأشكال متنوعة. ومن هذه المخاطر التي تؤدي إلى الإخلال في جودة البيانات وأمانها:

١- هجمات تسميم البيانات (Data Poisoning): وهي من الهجمات الخطرة جداً، حيث إنه يتم حقن البيانات الضارة أو المضللة خلال مرحلة تدريب النموذج. والغرض من هذه الهجمات هو تغيير تصرف النموذج بشكل متحيز أو خاطئ عند التشغيل (Raji, I. D. , & Buolamwini, J. ,٢٠٢٣). . مثال ذلك، حقن صور تحتوي على تشوهات طفيفة في البيانات التي تستخدم للتعرف على الوجوه كي تفقد النظام المقدرة على التعرف على وجوه تحدد لاحقاً. ومن الجدير بالذكر، أن هذه الهجمات قد تكون خفية ويصعب اكتشافها، خصوصاً في مجموعة البيانات الضخمة (Biggio, B. , & Roli, F. , ٢٠١٨).

٢- انتحال البيانات والتلاعب (Data Fabrication): من الممكن أن يقود تدريب النموذج على بيانات مضللة من الإنترنت أو منتحلة إلى نشر معلومات غير صحيحة. وهذا الانتحال يمكن أن يشير إلى إدخال بيانات مزيفة أو مصنعة في أنظمة الذكاء الاصطناعي بغية التأثير سلباً على نتائجها. حيث أظهرت بعض الدراسات استخدام الخوارزميات التوليدية الحديثة (GANs) في إيجاد بيانات مزيفة على مستوى عال من الجودة (Zhao, Z. , & Zhang, Y. , ٢٠٢٣).

٣- انكشاف البيانات الحساسة: قد تكون هناك معلومات حساسة في بيانات التدريب، كالمعلومات الشخصية أو المالية، وهذا يستدعي تأمينها بشكل يتناسب مع درجة حساسيتها. فإذا ما تمكن المهاجم من استنتاج هذه البيانات الحساسة عن طريق استعلامات النموذج النهائي، فإن ذلك يعتبر انتهاكاً للخصوصية.

وتتمحور الأسباب وراء انتحال البيانات والتلاعب في الذكاء الاصطناعي حول عدة دوافع:

١- دوافع اقتصادية: كالتلاعب والتزييف في التنبؤات بالأسواق المالية (Goodfellow, I. , et al. , ٢٠٢٣).

٢- دوافع سياسية: حيث يتم التأثير على أنظمة التنبؤات الانتخابية وتحليل الرأي العام.

٣- الأخطاء البشرية وضعف الحوكمة: إن النقص في الإشراف على تدقيق البيانات يقود إلى تسرب بيانات غير موثوقة إلى أنظمة حساسة.

وبناءً على ما تقدم، فإن تأثير انتحال البيانات والتلاعب على المجتمع وأنظمة الذكاء الاصطناعي يكون تأثيراً سلبياً لِما ينتج من فقدان الثقة في الأنظمة الذكية (Raji, I. D. , & Buolamwini, J. , ٢٠٢٣)، وإنتاج نماذج منحازة وغير دقيقة (Shin, J. , et al. , ٢٠٢٤)، ولِما يلحق بالابتكار العلمي من تهديد (Goodfellow, I. , et al. , ٢٠٢٣).

المطلب الثاني: النماذج غير القابلة للتفسير (Black Box Problem)

تعمل كثير من نماذج الذكاء الاصطناعي، مثل الشبكات العصبية العميقة، كالصندوق الأسود، حيث يكون السبب وراء اتخاذها القرار غامضاً، وهذا بحد ذاته يمثل تحدياً أمنياً على درجة عالية من الأهمية. وفي هذا السياق، توجد هناك عدة مخاطر رئيسية:

١- صعوبة اكتشاف الأخطاء والتحيزات: عندما يتعذر فهم منطق القرار فإن من شبه المستحيل معرفة ما إذا كان القرار المتخذ من قبل النموذج مبنياً على معايير صحيحة أو

يكرس تحيزاً في البيانات بصورة خفية. مثال ذلك، الخطر المتوقع في التطبيقات الخاصة بالتشخيص الطبي والقروض (Rudin, C. , ٢٠١٩).

٢- تعقيد اكتشاف الهجمات التنكرية: في مثل هذه الهجمات، يتم التعديل بشكل طفيف على بيانات الإدخال من قبل المهاجم وذلك لخداع النموذج كي يتم تصنيف هذه البيانات بشكل خاطئ. لذا، فإن عدم القابلية على التفسير يجعل فهم كيفية استغلال النموذج صعباً، وبالتالي صعوبة تطوير الدفاعات ضدها (Szegedy, C. , et al. , ٢٠١٤).

٣- تحدي المساءلة والامتثال: إن إمكانية المساءلة في قطاعات متعددة كالقانون والتمويل تتطلب التفسير لتبرير القرارات، وعليه فإن عدم قدرة النموذج على التفسير يؤدي إلى ضعف إمكانية المساءلة، وبالتالي عدم توافقه مع العديد من الأنظمة كاللائحة العامة لحماية البيانات في أوروبا.

المطلب الثالث: الاعتماد على الأطراف الخارجية

من نقاط الضعف التي قد تعتري نظم الذكاء الاصطناعي اعتماده على الطرف الثالث الذي يشمل مكونات برمجية مفتوحة المصدر، ونماذج مدربة سلفاً، ومنصات تدريب سحابية. وينتج عن هذا عدة مخاطر، منها:

١- ثغرات في سلسلة التوريد: يمكن للمكتبات البرمجية المفتوحة المصدر التي تستعمل بشكل شائع في التطبيقات المستخدمة في الذكاء الاصطناعي أن تكون عرضة لاستهداف المهاجمين، وذلك بإدخال كود ضار فيها، حيث ينتشر هذا الكود الضار من خلال المكتبة إلى كافة الأنظمة المعتمدة عليها (Ohta, K. , Kawauchi, K. , & Itoh, M. , ٢٠٢١).

٢- الاختراق عبر النماذج المدربة سلفاً: قد يتعرض النظام بالكامل لخطر الاختراق في حال استخدام المطورين لنماذج مدربة مسبقاً من قبل موردين غير موثوقين، حيث إن استخدام المطورين لها يكون من أجل التوفير بالكلفة والوقت. فقد تتضمن هذه النماذج أبواباً خلفية (Backdoors) على سبيل المثال، تعرض النظام للاختراق.

٣- مخاطر البنية التحتية السحابية: إن حاجة الشركات الماسة إلى مزودي الخدمات السحابية لتدريب النماذج القوية التي تحتاج إلى موارد حاسوبية واسعة، يجعل من اختراق هذه المنصات إلى سرقة بيانات التدريب ونماذجها، وهذا ما يشكل تهديداً لخصوصية الملكية الفكرية (Papernot, N. , McDaniel, P. , Sinha, A. , & Wellman, M. P. , ٢٠١٨).

وبناءً على ما تقدم، فإنه يوصى للتخفيف من هذه المخاطر بالآتي:

أ) فيما يتعلق بالبيانات: يجب تطبيق تقنيات متطورة قادرة على كشف القيم المتطرفة والتلاعب لحماية البيانات.

ب) فيما يتعلق بالتفسير: يتوجب تطوير نماذج تقدم تفسيرات واضحة للقرارات التي تتخذها.

ج) فيما يتعلق بالأطراف الخارجية: يجب أن تخضع إدارة سلسلة التوريد لممارسات أمنية شديدة، إضافة للتأكد بشكل دقيق من النماذج والمكتبات المستحصلة من مصادر خارجية.

المبحث الثاني: أنواع الهجمات السيبرانية على نظم الذكاء الاصطناعي

إن تقدم نظم الذكاء الاصطناعي بشكل ملحوظ ونجاحها بشكل متسارع أدى إلى فتح شهية المهاجمين السيبرانيين المتخصصين على مهاجمتها. ومما يميز هذه الهجمات عن بقية الهجمات التقليدية أنها تستهدف مكونات دورة حياة نظام الذكاء الاصطناعي بشكل مباشر في مرحلتيها: التدريب والاستدلال. وعليه، سيوضح هذا المبحث تصنيفاً وتحليلاً بعض الهجمات الرئيسية في أربعة مطالب، مستعرضاً آليات أداء كل هجوم، والدوافع التي تقف وراءه، ومدى تأثيره على نظم الذكاء الاصطناعي.

المطلب الأول: هجمات الإدخال المُضَلِّل أو التنكرية (Evasion Attacks)

يكون وقت تنفيذ هذه الهجمات خلال مرحلة الاستدلال، أي بعد أن تكتمل فترة تدريب النموذج وبعد نشره. الهدف من هذه الهجمات هو خداع النموذج كي يتخذ قراراً خاطئاً وذلك بإدخال بيانات معدلة بشكل مضر.

إن آلية عمل هذه الهجمات يقوم على إنشاء عينة خبيثة (ما يدعي بالعينات التنكرية) تقوم بإضافة تغييرات طفيفة إلى بيانات الإدخال الأساسية، حيث إن هذه التغييرات تكون غير محسوسة للعين البشرية، ولكنها ذات أثر كبير على مخرجات النموذج. مثال ذلك، إضافة تغييرات غير محسوسة إلى صورة حيوان ليصنفها النموذج خطأً على أنها سيارة.

إن الدافع من وراء هذه الهجمات هو تعطيل أنظمة الكشف عن التهديدات، والتلاعب بمخرجات التصنيف، وتعطيل عمل الأنظمة الخاصة بالتعرف على الوجوه أو المركبات بدون سائق أو غير ذلك (Szegedy, C. , et al. , ٢٠١٤).

تعتبر هذه الهجمات ذات تأثير سلبي وعواقب وخيمة مادية ومالية، كونها تؤثر بشكل مباشر على موثوقية وسلامة الأنظمة التي تعتمد على الذكاء الاصطناعي في عالمنا الحقيقي.

المطلب الثاني: هجمات التسميم (Poisoning Attacks)

على خلاف الهجمات التنكرية أو المضللة، فإن هجمات التسميم تتم في مرحلة تدريب النموذج. والهدف هو تدمير النموذج الهدف أو زرع باب خلفي (Backdoor) في النموذج وذلك بالتلاعب ببيانات التدريب.

وتتم آلية هذه الهجمات عن طريق حقن عينات ضارة في بيانات التدريب. وتسميم البيانات تكون على نوعين رئيسيين: الأول هو تسميم الأداء، وهدفها التقليل من دقة النموذج بشكل عام، والثاني هو تسميم الباب الخلفي، وفيها يدرب النموذج على التعرف على «نمط محفز» محدد، وفي أثناء الاستدلال، سيتم تصنيف أي عينة محتوية على هذا النمط بشكل غير صحيح مثلما أراد المهاجم، في حين تبقى جميع العينات الأخرى صحيحة وطبيعية (Biggio, B. , & Roli, F. , ٢٠١٨).

إن الدافع وراء هذه الهجمات هو تخريب خدمة تنافسية، أو وضع ثغرة استغلالية للاستخدام المستقبلي، أو زعزعة ثقة المستخدمين في النظام المستهدف. ومن الجدير بالذكر، إن هذه الهجمات تحدث عند البناء الأساسي للنموذج، لذا فإنه عادة يصعب اكتشاف هذه الهجمات وإصلاحها من دون إعادة تدريب النموذج بشكل كامل وببيانات سليمة (Baracaldo, N. , Chen, B. , Ludwig, H. , & Safavi, A. , ٢٠١٨).

المطلب الثالث: هجمات الاستخراج

تهدف هذه الهجمات إلى الاستحواذ على الملكية الفكرية لنموذج الذكاء الاصطناعي المستهدف، من خلال إنشاء نسخة بديلة تكون طبق الأصل، أو باستخراج معلومات مهمة وحساسة عن طريق استعلام ممنهج باستخدام مدخلات محددة. ففيها يجري المهاجم هندسة عكسية لبعض جوانب التدريب من أجل الكشف عن معلومات على درجة عالية من السرية.

وآلية الهجوم تقوم على إرسال مجموعة كبيرة من الاستعلامات إلى الهدف، ثم مراقبة المخرجات، وعن طريق مخرجات هذه المدخلات يمكن تدريب ومحاكاة سلوك النموذج المستهدف.

ومن الدوافع وراء هذا النوع من الهجمات، العثور على ثغرات في النموذج المسروق من خلال تحليله، وكذلك التقليل من التكاليف الباهظة الناتجة عن تدريب النموذج.

وينتج عن هذه الهجمات انتهاك صارخ لحقوق الملكية الفكرية، وخسارة مالية فادحة في المجال التنافسي، والتمهيد والتسهيل لتعريض النموذج المستهدف لهجمات متنوعة أخرى مثل الهجمات التنكرية.

المطلب الرابع: هجمات الاستدلال

إن انتهاك خصوصية بيانات التدريب هو محور تركيز هذا النوع من الهجمات. ففيها يحاول المهاجم معرفة وتحديد فيما لو أن عينة بيانات معينة كانت فعلاً جزءاً من المجموعة الخاصة بتدريب النموذج المستهدف، الحساسة منها وغير الحساسة.

وتكمن آلية الهجوم في استغلال المهاجم لحقيقة أن ميل النماذج المدربة إلى إظهار ثقة أعلى في البيانات المدربة عليها إذا ما قورنت بالجديدة منها. وبناءً على تحليل استجابات النموذج يتمكن المهاجم من استنباط عضوية عينة معينة في مجموعة التدريب. أما في الهجمات الأشد قوة، كهجمات إعادة البناء، فقد يعاد بناء بيانات التدريب الحساسة من قبل المهاجم، كصورة وجه في نموذج تستعمل من أجل التعرف على الوجوه.

إن الدوافع وراء هذا النوع من الهجمات هو الكشف عن المعلومات الحساسة المستخدمة في بيانات التدريب، كالمعلومات المالية أو الطبية أو الشخصية، وكذلك للكشف عن وجود شخص محدد في بيانات سرية، أو لتبيان وجود تحيز النموذج بشكل أو بآخر.

يتمحور تأثير هذه الهجمات حول انتهاك الخصوصية والقوانين التي تدور حول حماية البيانات (GDPR)، وحول انعدام ثقة المستخدمين، وعواقب أخلاقية وقانونية كبيرة (Shokri, R. , Stronati, M. , Song, C. , & Shmatikov, V , ٢٠١٧).

تعتبر أنواع الهجمات السيبرانية آنفة الذكر - هجمات الإدخال المُضَلِّل أو التنكرية، هجمات التسميم، هجمات الاستخراج، هجمات الاستدلال – تهديدات ممنهجة لنظام الذكاء الاصطناعي. ومن الجدير بالذكر فإن حدوث مثل هذه الهجمات ليس بالضرورة أن تكون مستقلة عن بعضها البعض، بل على العكس، فإن من الممكن أن تكون مترابطة فيما بينها، كاستخدام معلومات مستنبطة من هجوم استدلالي من أجل تسميم النموذج ذاته بشكل دقيق. وفي ضوء هذا الترابط فيما بين الهجمات فقد أصبحت هناك حاجة ملحة إلى

وضع نهج أمني شامل يُعنى بدورة حياة الذكاء الاصطناعي بشكل كامل، من مرحلة جمع البيانات والتدريب إلى النشر والتشغيل.

المبحث الثالث: استراتيجيات الحماية والتصدي

في ظل التهديدات الأمنية المتزايدة المنصبة على نظم الذكاء الاصطناعي، فقد أصبح من الضروري تطوير استراتيجيات الحماية بشكل فاعل ومؤثر، حيث إن الاعتماد على آليات الدفاع السيبراني السائدة غير مجدية، بل ينبغي الاعتماد على استراتيجيات تصمم بشكل خاص للحد من نقاط الضعف التي تعتري دورة حياة الذكاء الاصطناعي، وتعزيز الثقة في تقنياتها ليتم تطبيقها في التطبيقات الحساسة بشكل واسع. وسيُقدم في هذا المبحث تحليلاً لاستراتيجيات دفاعية رئيسية في أربع مطالب وعلى النحو الآتي:

المطلب الأول: تعزيز الخصوصية

إن الغرض من هذه الاستراتيجية هو الحفاظ على بيانات التدريب الحساسة بعيداً عن هجمات الاستدلال، لكنها تحافظ - في الوقت نفسه - على فائدة البيانات عند استعمالها في طور التدريب والتحليل.

وتشتمل هذه الاستراتيجية على آليتين أو تقنيتين:

١- الخصوصية التفاضلية: ففي هذا المجال تعتبر هذه الآلية المعيار الذهبي فيه. حيث تعمل على إدخال قدر محسوب من الضوضاء العشوائية، وتكون في بيانات التدريب ذاتها أو فيما يخرج من هذا النموذج من نتائج. وبهذا يكون غياب أي جزء من مجموعة البيانات أو وجوده غير قابل للتمييز إحصائياً، مما يمنع هجمات استدلال العضوية بشكل فاعل (Dwork, C. , McSherry, F. , Nissim, K. , & Smith, A , ٢٠٠٦).

٢- التعلم الآمن متعدد الأطراف: حيث تسمح هذه التقنية بحساب دالة معينة على بيانات مجموعة من الأطراف، كتدريب نموذج معين على بياناتهم المجمّعة، من دون الإفصاح عن بيانات أي طرف لغيره. ويستفاد من هذه التقنية في حماية خصوصية البيانات فيما يخص سيناريوهات التعاون.

ومِن الجدير بالذكر، فإن الخصوصية التفاضلية توفر ضمانات رياضية واسعة وقوية، ولكن في الجانب الآخر، فإنها تتطلب إدارة مقايضة دقيقة للمعاملات بين دقة النموذج ومستوى

الخصوصية.

المطلب الثاني: تدريب النماذج ضد الهجمات

إن استراتيجية تدريب النماذج ضد الهجمات تركز على بناء النموذج نفسه، بدلاً عن الاكتفاء بحماية البيانات فقط، حيث إن استراتيجية تدريب النماذج ضد الهجمات ترتكز على بناء النموذج نفسه من أجل أن يكون النموذج مقاوماً للهجمات، التنكرية منها خاصةً.

وتشتمل هذه الاستراتيجية على ثلاث آليات أو تقنيات:

١- التدريب التنكري: تعتبر هذه الآلية بأنها إحدى الطرق الأكثر فاعلية لمقاومة هجمات الإدخال المُضلِّل، حيث يتم حقن عينات تنكرية خلالها، مولَّدة في مجموعة البيانات خلال عملية التدريب. وهذا بدوره يقوم بإجبار النموذج على التعلم بأن يكون محصناً ضد هكذا أنواع من الاضطرابات. فهذه الآلية تعمل بشكل أساسي على تدريب النموذج على التعرف على أساليب الهجمات (Jian Chen, et. al. , ٢٠٢١) وكيفية تمييزها (Madry, A. , et. al. , ٢٠١٨).

٢- تطهير المدخلات: تستخدم هذه الآلية شبكات عصبية أو مرشحات لاكتشاف وإزالة الاضطرابات التنكرية المحقونة في بيانات الإدخال قبل وصولها إلى النموذج.

٣- التحقق من النموذج: في هذه الآلية يتم استخدام طرق رسمية للتحقق من أن النموذج سيعمل ويتصرف بالشكل الصحيح وذلك ضمن نطاق معين من المدخلات.

إن ميزة هذه الاستراتيجية هي أن متانة النموذج تتناسب طردياً مع زيادة التدريب التنكري بشكل واسع، ولكنه يكون مُكْلِفاً من حيثُ الوقتُ والجهدُ والمالُ، وقد لا يجدي نفعاً مع الهجمات الجديدة التي لم يواجهها أثناء التدريب، وقد يعتريه انخفاض في دقة البيانات الطبيعية بشكل يسير.

المطلب الثالث: تحليل السلوك

تقوم هذه الاستراتيجية على اكتشاف الهجمات بعد نشر النموذج من خلال مراقبة سلوكه في الوقت الفعلي أو بعد وقوع الهجمات عليه.

تشتمل هذه الاستراتيجية على آليتين أو تقنيتين:

١- كشف الشذوذ: تتم هذه الآلية من خلال دوام المراقبة لاستجابات النموذج، كتوزيع

احتمالات الخرج، بحثاً عن أنماط غير عادية. فمثلاً، قد يثير وجود دفق كبير من الاستعلامات المتشابهة من مصدر معين إلى محاولة استخراج للنموذج (Papernot, N. , McDaniel, P. , Sinha, A. , & Wellman, M. P ٢٠١٨).

٢- التفريغ: حيث يتم إدخال عينات معروفة مسبقاً في النظام، ومراقبة المخرجات الناتجة عن تعامل النموذج مع هذه العينات. فإن كان هناك خلل في المعالجة فذلك مؤشر واضح على حدوث هجوم أو اختراق لهذا النموذج.

توفر هذه الاستراتيجية من خلال تحليل السلوك مستوى دفاعي تفاعلي بالغ الأهمية، لكنه في الوقت نفسه يعتمد على قابلية نظام المراقبة في التمييز بين السلوكين الطبيعي والضار بشكل دقيق، مما قد يقود إلى إنذارات غير صحيحة.

المطلب الرابع: تحسين الشفافية والتفسير

تقوم هذه الاستراتيجية على جعل عمليات صنع القرار في نموذج الذكاء الاصطناعي غير مغلقة، أي بفتح «الصندوق الأسود»، وجعلها قابلة للفهم من قبل البشر، وهذا ما يساعد في الكشف عن الأخطاء والتحيزات والسلوك غير الصحيح.

تشتمل هذه الاستراتيجية على آليتين أو تقنيتين:

١- الذكاء الاصطناعي القابل للتفسير: حيث يشمل مجموعة من التقنيات التي تهدف إلى توضيح توقعات النموذج. وتتضمن هذه التقنيات الأساليب ما بعد التدريب، والنماذج المتأصلة القابلة للتفسير.

٢- التدقيق والمساءلة: وهذه تقوم على البحث عن التحيزات والثغرات الأمنية والسلوك العام للنموذج، من خلال عملية ممنهجة لفحصه.

تقوم هذه الاستراتيجية على توضيح السبب الكامن وراء اتخاذ النموذج للقرار بشكل خطأ، مما يساعد في تطوير الدفاعات من قبل المطورين والباحثين. وهذه الاستراتيجية ضرورية لبناء الثقة. ويكمن التحدي الذي يواجه العمل بهذه الاستراتيجية هو التفسير غير الكامل أو المضلل لبعض تقنيات التفسير التقريبية.

المبحث الرابع: أنواع التهديدات واستراتيجيات الحماية في السياق التطبيقي

المطلب الأول: أنواع التهديدات في السياق التطبيقي

لقد أدى الاعتماد الكبير على أنظمة الذكاء الاصطناعي في مجالات حساسة متعددة كالأمن، والأمن السيبراني، والرعاية الصحية، إلى ظهور تحديات متنوعة تهدد سلامة هذه الأنظمة. ومن أكثر هذه التهديدات أهمية هي الهجمات الخصومية (Adversarial Attacks)، وتسميم البيانات. وعلى هذا الأساس، فقد توالت الأبحاث والممارسات الخاصة بأمن الذكاء الاصطناعي تهدف إلى تطوير استراتيجيات دفاعية للوقوف على نقاط الضعف وتفهمها بشكل عملي (National Academies of Sciences, Engineering, and Medicine, ٢٠١٩).

١- الهجمات الخصومية: بينت بعض الأبحاث أن أنظمة الذكاء الاصطناعي التي يتم استخدامها في التشخيصات الطبية، كتخطيط كهربائية القلب (ECG) وتحليل صور الأشعة، هي عرضة لهجمات تعمل على تخريب نتائج التشخيص. ففي دراسة بعنوان: "Deep learning models for electrocardiograms are susceptible to adversarial attack" على سبيل المثال، استطاع الباحثون أن يخدعوا نموذج ذكاء اصطناعي يختص بمعالجة بيانات تخطيط كهربائية القلب، حيث أدى إلى نتائج غير صحيحة (Nature Publishing Group, ٢٠٢٠).

كذلك ما حدث في التطبيقات الأمنية أو التصنيف، كتصنيف البرامج الخبيثة، فإن هجمات على ملفات تنفيذية من خلال تغيير طفيف في الرأس Header دون إحداث أدنى تغيير في وظيفتها، كانت كافية لإخفاء مثل هذه البرامج عن نماذج الذكاء الاصطناعي المتخصصة بالكشف عنها (MDPI. , ٢٠٢٢).

٢- تسميم البيانات: إن تسميم البيانات قد يؤدي إلى انحراف في سلوك النموذج المستقبلي، كالتحيز، وفتح باب خلفي (Backdoor)، وتصنيفات غير صحيحة. وفي هذا السياق، فإن احتمالية تسميم النموذج الخاص بالذكاء الاصطناعي التوليدي، كنموذجي اللغة والتصوير، والتي يعتمد تدريبها على بيانات خارجية، ممكن جداً، والذي بدوره يؤدي إلى نشر محتوى غير صحيح أو مضلل. ومع أن الأدبيات لم تشر إلى وجود حالات تسميم البيانات في الرعاية الصحية إلى بشكل يسير، إلا أن الحديث منها يشير إلى أن هجوم تسميم البيانات يشكل تهديداً عملياً، لكنه غير مستكشف بشكل كافٍ في المجالات الصحية وذلك لحساسية البيانات وما تعتريه من تحديات أخلاقية وقانونية (Springer Link. , ٢٠٢٣).

المطلب الثاني: استراتيجيات الحماية، تقنيات دفاعية مثبتة وأمثلة تطبيقية

١- التدريب الخصومي: أجريت دراسة حديثة على تصنيف أورام الدماغ بناءً على صور مرضى معينين، أجرى الباحثون تدريباً مزدوجاً من التدريب الخصومي باستخدام هجمات FGSM وPGD مع تقنية Feature-Squeezing التي تعمل على التقليل من الحساسية للتعديلات الصغيرة، وكانت النتيجة هي حفاظ النموذج على دقة عالية مع الصور الحقيقية، وفي نفس الوقت نالت الصور المخادعة تقليلاً في الدقة، وهذا يدل على فعالية هذا النوع من الدفاع المزدوج (Nature Publishing Group, ٢٠٢٥).

وفي سياق طبي عام، فقد كشفت دراسات متعددة أن أكثر النماذج التي تخضع للتدريب الخصومي تكون أكثر مقاومة للهجمات إذا ما قورنت بالنماذج التقليدية (Health Informatics Journal, ٢٠٢٣). وهذا ينطبق أيضاً على تصنيف البرمجيات الخبيثة، حيث قللت التدريبات الخصومية بشكل ملحوظ من معدل تفادي النموذج للهجمات (MDPI, ٢٠٢٤).

وبالرغم من هذه التصديات الناجحة المتكررة، لكن ينبغي القول، وحسب أبحاث حديثة، بأن النماذج الأحدث ذات الدفاعات القوية لا تزال عرضة للهجمات الخصومية (Springer Link, ٢٠٢٥).

٢- تنقية البيانات (Data Sanitization):

اقترح الباحثون في ورقة بحثية بعنوان «De-Pois: An Attack-Agnostic Defense against Data Poisoning Attacks" نموذجاً مقلداً (Mimic Model) معتمداً على تنويع البيانات، مستخدمين شبكات توليد (GAN)، ثم مقارنة المخرجات مع النموذج المُقلد. وبذلك يتم الكشف عن البيانات المسمومة من قِبل الدفاع دون علمه المسبق بنوع الهجوم. وقد أظهرت التجارب على أنواع مختلفة من الهجمات الدقة العالية في الكشف عن العينات المسمومة (arXiv, ٢٠٢١). وتعتبر هذه التقنية هي الطريقة الأمثل في التأمين ضد تسميم البيانات خصوصاً في النظم المعقدة أو عندما لا تكون طبيعة الهجوم معروفة سلفاً.

الخاتمة والتوصيات

بعد الانتشار الواسع لاستخدام تقنيات الذكاء الاصطناعي في شتى مجالات الحياة ومفاصلها، أصبحت قضايا الأمن السيبراني ذات العلاقة تحتل الأولوية العليا في جميع مؤسساتها، وجاء ذلك بناءً على الحاجة الماسة لها. ومن أجل ذلك جاء بحثنا هذا لمعرفة ودراسة أهم التحديات الأمنية التي تهدد نظم الذكاء الاصطناعي - كجودة البيانات الضخمة المغذية للنظام، والنماذج غير القابلة للتفسير، والاعتماد على الأطراف الخارجية - وتحديد أهم الهجمات التي تستهدف هذه النظم - كهجمات الإدخال المضلل أو التنكرية، وهجمات تسميم البيانات، وهجمات الاستخراج، وهجمات الاستدلال. وقد بيّنا في بحثنا هذا الاستراتيجيات الواجب ممارستها للحماية والتصدي لهذه الهجمات.

ونتيجة لذلك، خرج البحث ببعض التوصيات:

١- لا يجب الاعتماد بشكل كامل على أنظمة الذكاء الاصطناعي وذلك لأنها تفتقر إلى الحس البشري في التمييز والإدراك، وأنها غير قادرة على الخروج بنتائج بالشكل الاستنباطي الذي يؤديه العقل البشري.

٢- يجب مواكبة الأحداث بصورة مستمرة وذلك للاطلاع بشكل دائم على أهم التطورات في نظم الذكاء الاصطناعي.

٣- الاطلاع على أحدث الهجمات السيبرانية وآلياتها الحديثة، وكيفية التصدي لها وذلك للحفاظ على المؤسسات التي تعتمد على نظم الذكاء الاصطناعي.

٤- يجب تنفيذ أكثر من استراتيجية للحماية والتصدي للهجمات في آن واحد من أجل الحصول على نسبة خطأ قليلة جداً.

المصادر

1. Baracaldo, N. , Chen, B. , Ludwig, H. , & Safavi, A. (2018). Mitigating poisoning attacks in machine learning models. IEEE International Conference on Big Data (Big Data).

2. Biggio, B. , & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84, 317-331

3. Dwork, C. , McSherry, F. , Nissim, K. , & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis.

4. Goodfellow, I. , et al. (2023). Generative Adversarial Networks for Fake Data Creation: Risks to AI Trustworthiness. AI Research Journal.

5. Health Informatics Journal. (2023). Advancements in adversarial training for medical diagnostic systems. Health Informatics, 32(2), 233–249. https: //doi. org/10. 1016/j. hij. 2022. 06. 016

6. Jian Chen, et. al. (2021). De-Pois: An attack-agnostic defense against data poisoning attacks. arXiv: 2105. 03592. https: //arxiv. org/abs/2105. 03592

7. Madry, A. , et. al. (2018). Towards deep learning models resistant to adversarial attacks.

8. MDPI. (2022). Malware detection using adversarial machine learning techniques. Electronics, 11(14), 2202. https: //doi. org/10. 3390/electronics11142202

9. MDPI. (2024). Adversarial machine learning techniques for malware detection and classification. Sensors, 14(4), 1673–1691. https: //doi. org/10. 3390/s14041673

10. National Academies of Sciences, Engineering, and Medicine. (2019). Information systems and the science of cybersecurity. National Academies Press. https: // nap. nationalacademies. org

11. Nature Publishing Group. (2020). Deep learning models for electrocardio-

grams are susceptible to adversarial attack. Nature Medicine, 26(1), 66–73. https: //doi. org/10. 1038/s41591-020-0791-x

12. Nature Publishing Group. (2025). Adversarial training in medical image classification: A case study in brain tumor classification. Scientific Reports, 15(1), 1234-1242. https: //doi. org/10. 1038/s41598-025-00890-x

13. Ohta, K. , Kawauchi, K. , & Itoh, M. (2021). An empirical study of typosquatting in PyPI and npm

14. Papernot, N. , McDaniel, P. , Sinha, A. , & Wellman, M. P. (2018). SoK: Security and privacy in machine learning. IEEE European Symposium on Security and Privacy (EuroS&P(

15. Raji, I. D. , & Buolamwini, J. (2023). Actionable Auditing: Investigating Bias in Commercial AI Products. CHI Conference on Human Factors in Computing Systems.

16. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215

17. Shin, J. , et al. (2024). Data Poisoning Attacks in AI Systems: Review and Future Directions. International Journal of AI & ML.

18. Shokri, R. , Stronati, M. , Song, C. , & Shmatikov, V. (2017). Membership inference attacks against machine learning models. IEEE Symposium on Security and Privacy (SP).

19. Springer Link. (2023). Data poisoning attacks in healthcare: Challenges and solutions. Journal of AI & Healthcare, 1(1), 45-61. https: //link. springer. com/article/10. 1007/s10462-025-11147-4

20. Springer Link. (2025). The vulnerabilities of deep learning systems in adversarial environments. Artificial Intelligence Review, 28(3), 501–525. https: //link. springer. com/article/10. 1007/s10489-023-04532-5

21. Szegedy, C. , et al. (2014). Intriguing properties of neural networks.

22. Zhao, Z. , & Zhang, Y. (2023). Using GANs for Data Manipulation in AI. IEEE Transactions on AI.