جهوزية البيانات في مؤسسات الدولة للاستفادة من الذكاء الاصطناعي

د . براق عبد القادر عبد الكريم

الوكيل الفني/ وزارة الاتصالات

يسلط هذا البحث الضوء على أهمية جهوزية البيانات في مؤسسات الدولة والعناصر الرئيسة لهذه الجهوزية للاستفادة من تقنيات الذكاء الاصطناعي، ويقدم البحث دراسة حالة عن البيانات الحكومية التي تمت استضافتها في مركز البيانات الوطني لتقييم الفجوات الرقمية فيها والتحديات المرتبطة بها، وتشير النتائج إلى الحاجة لتحسين جودة البيانات، وتعزيز التكامل بين الأنظمة، وتطوير البنى التحتية والسياسات لتحقيق الاستفادة المثلى من الذكاء الاصطناعي، كما يقترح البحث حلولًا لتحسين جهوزية بيانات مؤسسات الدولة للذكاء الاصطناعي.

الكلمات المفتاحية: الذكاء الاصطناعي، جهوزية البيانات الرقمية، جودة البيانات، البنية التحتية التكنولوجية، حماية البيانات، تكامل الأنظمة، سياسة البيانات.

Data Readiness in State's Institutions to Leverage Artificial Intelligence

Dr. Buraq A. Abdulkareem
Technical Undersecretary
Ministry of Communications

This research highlights the importance of data readiness in the State's institutions and the key elements of such readiness to leverage artificial intelligence technologies. It presents a case study on the governmental data that are hosted in National Data Center to assess their digital gaps and associated challenges. Results indicate the need to improve data quality, enhance system integration, and develop infrastructure and policies to maximize the benefits of artificial intelligence. The research proposes solutions to improve the data readiness of state institutions for artificial intelligence.

Keywords: Artificial intelligence, digital data readiness, data quality, technical infrastructure, data protection, system integration, data policy.

الاستلام	الإرجاع	القبول
2025/ 4/14	2025/4/21	2025/ 4/28

المقدمة

يستفاد من الذكاء الاصطناعي في دوائر الدولة لتحسين كفاءة الخدمات الحكومية، والمساعدة في اتخاذ القرارات، وتعزيز الشفافية، كما أنه يسهم بشكل كبير في تحسين إدارة الموارد، ومن هنا فإن الذكاء الاصطناعي (Artificial Intelligence) والمعروف اختصارا بـ (AI) هو فرع من علوم الحاسوب يهدف إلى إنشاء أنظمة قادرة على أداء المهمات التي تتطلب عادة ذكاء بشريا، مثل التعلّم، والتفكير، والتعرف على الأنماط، وإيجاد الحلول للمشكلات والمعوقات، ويشمل الذكاء الاصطناعي مجموعة متنوعة من التقنيات بما في ذلك معالجة اللغة الطبيعية (Natural Language Processing)، والتعلم الآلي ، (Machine Learning) والتعلم العميق (Deep Learning) وغيرها، من التي تعتمد بصورة أساس على تحليل كميات كبيرة من البيانات لاستخلاص الأنماط والتنبؤات للتعلم وإنجاز المهم[1].

تعتمد المؤسسات على البيانات لإنجاز مهماتها، والعمل على تطوير أعمالها بالطريقة التي تسهم في زيادة الإنتاجية، وجودة الخدمات المقدمة من خلالها، وبالإمكان تعريف البيانات بأنها حقائق ومعلومات خام تجمع من مختلف المصادر، التي يمكن الاستفادة منها للحصول على معلومات ذات معنى، أو اتخاذ قرارات معينة، أو إيضاح حقائق أو غيرها من المهمات من خلال تحليل هذه البيانات التي تتنوع بين مهيكلة مثل قواعد البيانات التقليدية، وغير مهيكلة مثل النصوص والصور والفيديوهات، أو مزيج بين الاثنين، ويعد تحليل البيانات جزءا أساسا من عملية اتخاذ القرار في العديد من المجالات، بما في ذلك الأعمال التجارية، والرعاية الصحية، والعلوم وغيرها[2].

البيانات هي الوقود الأساس والمادة الخام للذكاء الاصطناعي، وتعتمد تقنياته على تحليل كميات ضخمة من البيانات لاستخلاص الأنماط، والتنبؤات، والوصول للنتائج المرجوة، والقيام بالمهمات الموكلة إليه، كلما زادت كمية البيانات المتاحة وجودتها، كان أداء نماذج الذكاء الاصطناعي أدق وأفضل، والذي يعتمد بالأساس على تحليل هذه البيانات، كما يشمل تحليل البيانات خطوات متعددة مثل جمع البيانات، وتنظيفها، وتخزينها، وتحليلها باستخدام أدوات وتقنيات متقدمة[3]، ومن الواضح هنا أهمية جهوزية البيانات على أداء منظومات الذكاء الاصطناعي من حيث ما يلي:

1. التوافرية والجودة.
2. الحماية الأمنية لها.
3. هيكليتها وتنظيمها.
4. توفير البنية التحتية التكنولوجية المطلوبة لها.
5. السياسات والإجراءات.

يسلط هذا البحث الضوء على العناصر الرئيسة المطلوبة في بيانات مؤسسات الدولة للاستفادة من تبني تقنيات الذكاء الاصطناعي لإنجاز الأعمال الموكلة بها، وتحليل انموذج من البيانات الرقمية لمؤسسات في الدولة العراقية، وتقييم الفجوات الرقمية والأساليب المستخدمة حاليا، وملائمتها لتبني الذكاء الاصطناعي في هذه المؤسسات، وهنا تختلف مؤسسات الدولة من حيث العناصر المذكورة آنفاً اختلافاً نسبيا، ومن الصعوبة تقييم مؤسسات الدولة الحالية جميعا في سياق هذا البحث وحدوده؛ وعليه سيتم اختيار مجموعة من البيانات تمت استضافتها في مركز البيانات الوطني كانموذج عن البيانات الرقمية في مؤسسات الدولة لأغراض هذا البحث، وتقييم نسب جهوزية هذه البيانات، واقتراح حلول لتحسين جهوزية البيانات، كما إن البحث لم يشتمل على عنصر الحماية الأمنية للبيانات، ذلك أن عملية تقييم هذا العنصر فنيا تتطلب وقتا وجهودا كبيرة وموارد وموافقات خاصة لا يمكن توفيرها في الوقت المطلوب لإنجاز هذا البحث.

المحور الأول: جهوزية البيانات للذكاء الاصطناعي

تعرف جهوزية البيانات للذكاء الاصطناعي بأنها مدى توافق البيانات مع متطلبات تطبيقات الذكاء الاصطناعي، وتشمل هذه الجهوزية على توافرية هذه البيانات وجودتها والحماية الأمنية لها، وهيكليتها، وتنظيمها، والبنية التحتية، والسياسات والإجراءات الخاصة بها، ويعرض هذا الجزء من البحث الإطار النظري لهذه العناصر، الذي يستفاد منه لدراسة التحديات والفجوات التي تواجه دوائر الدولة.

أولًا: التوافرية والجودة

التوافرية هي مدى إمكانية الوصول إلى البيانات المطلوبة من المستخدمين أو الأنظمة عند الحاجة إليها، وتتكون من المكونات الآتية:

1. توفر البيانات:

أ. يجب أن تكون البيانات موجودة وقابلة للاستخدام.

ب. تعتمد على البنية التحتية التقنية مثل الخوادم ومراكز البيانات.

2. سهولة الوصول:

أ. يجب أن تكون البيانات متاحة بشكل سهل وآمن للمستخدمين المصرح لهم.

ب. يشمل ذلك توفير واجهات برمجية (APIs) مناسبة للوصول إلى البيانات.

3. استمرارية التوافر:

أ. يعني الحفاظ على توفر البيانات حتى في حالات الطوارئ، مثل الكوارث الطبيعية، أو الانقطاعات الفنية.

ب. يتطلب ذلك وجود أنظمة احتياطية أو نسخ مكررة من البيانات.

إن التطبيقات التي تعتمد على الذكاء الاصطناعي، مثل التحليلات التنبؤية، وأنظمة اتخاذ القرار، وغيرها، تحتاج إلى بيانات فورية ودقيقة، كما إن ضعف التوافرية يؤدي إلى توقف الأنظمة أو الوصول إلى نتائج غير دقيقة.

من جانب آخر فإن جودة البيانات تشير إلى مدى ملاءمة البيانات للاستخدام في سياق معين، وهنا البيانات ذات الجودة العالية تكون دقيقة، ومكتملة، وموثوقة، وبحسب الأبعاد الآتية:

1. الدقة (Accuracy):

أ. خلو البيانات من الأخطاء والتشوهات.

ب. أن تكون البيانات متطابقة مع السجلات الأصلية.

2. الاكتمال (Completeness):

أ. يجب أن تحوي البيانات على العناصر الضرورية كافة.

ب. وجود الأعمدة المطلوبة جميعا في قاعدة البيانات.

3. الاتساق (Consistency):

أ. توافق البيانات عبر الأنظمة المختلفة.

ب. أن تكون البيانات متماثلة في الأنظمة جميعا.

4. التوقيت(Timeliness) :

أ. أن تكون البيانات حديثة ومحدثة بانتظام.

ب.البيانات المعتمدة على الوقت الحقيقي تحتاج إلى تحديث لحظي.

5. الصحة (Validity):

أ. توافق البيانات مع القواعد والمعايير المحددة.

ب.أن تكون البيانات ضمن نطاق منطقي.

الذكاء الاصطناعي يعتمد على تحليل كميات كبيرة من البيانات؛ لذا تؤثر جودة البيانات بشكل مباشر في دقة النماذج، وفاعلية التحليلات، وموثوقية النتائج.

ثانيًا: الحماية الأمنية للبيانات

الحماية الأمنية للبيانات هي مجموعة من السياسات، والأدوات، والممارسات التي تهدف إلى حماية البيانات من الوصول غير المصرح به، والتعديل، والسرقة، أو التدمير، وتعد حماية البيانات أمرا حيويا للمؤسسات، لا سيما مع تزايد الاعتماد على البيانات في اتخاذ القرارات، واستخدام تقنيات الذكاء الاصطناعي؛ لذا فإن للحماية الأمنية للبيانات فوائد عديدة منها الحفاظ على سلامة البيانات من التلاعب أو التلف، ومنع الخسائر الاقتصادية، وحماية خصوصية بيانات المؤسسات والمستخدمين، وتتكون الحماية الأمنية للبيانات من العناصر الآتية[4]، [5]، [6]:

1. السرية (Confidentiality):

أ. الحفاظ على سرية البيانات من خلال منع الوصول غير المصرح به.

ب. تحويل البيانات إلى صيغة غير مفهومة إلا للمصرح لهم.

ج. التحقق من هوية المستخدم قبل الوصول إلى البيانات.

2. السلامة (Integrity):

أ. ضمان أن البيانات لم تَعدل أو تفسد في أثناء النقل أو التخزين.

ب.للتحقق من أن البيانات لم تَعدل.

ج. التنبه عند حدوث تغييرات غير مصرح بها.

3. التكرار (Redundancy):

أ. الاحتفاظ بنسخ احتياطية من البيانات لتجنب فقدانها.

ب. توفير نسخ عديدة من البيانات عبر مواقع مختلفة.

ثالثًا: هيكلية وتنظيم البيانات

تمثل هيكلية البيانات وتنظيمها الأساس الذي تعتمد عليه المؤسسات لتحليل البيانات، واتخاذ القرارات، وتطبيق تقنيات الذكاء الاصطناعي، كما إن الهيكلية الجيدة والتنظيم الفعال يمكّنان المؤسسات من تحقيق أقصى استفادة من البيانات، إذ تمكن هيكلية البيانات من الإسراع في عملية التحليل، وزيادة الدقة في اتخاذ القرارات، فضلاً عن أنها تقلل تكاليف التخزين والمعالجة[7].

وتتكون هيكلية البيانات من العناصر الآتية:

1. مصادر البيانات (Data Sources):

أ. بيانات داخلية من الأنظمة والتطبيقات الداخلية مثل أنظمة الموارد البشرية.

ب. بيانات خارجية مثل بيانات العملاء أو بيانات السوق.

2. التخزين (Data Storage):

أ. تخزن البيانات في قواعد بيانات تقليدية أو مستودعات البيانات (Data Warehousesأو بحيرات البيانات (Data Lakes).

ب. أمثلة على ذلك قواعد البيانات العلائقية مثل MySQL و PostgreSQL وبحيرات البيانات مثل Hadoop وAmazon S3.

3. معالجة البيانات (Data Processing):

أ. تتضمن تنظيم البيانات وإعدادها للتحليل.

ب. تقنيات معالجة البيانات تشمل:

- المعالجة في الوقت الحقيقي (Real-Time Processing).
- المعالجة بالدفعات(Batch Processing).

4. النقل والتكامل (Data Integration):

أ.نقل البيانات بين الأنظمة المختلفة وضمان تكاملها.

ب.توحيد البيانات من مصادر متعددة باستخدام أدوات مثل ETL Extract, Transform) (Load.

5. إدارة البيانات (Data Governance):

أ.وضع سياسات وقواعد لإدارة البيانات وضمان جودتها.

ب.تحديد المسؤوليات للأفراد داخل المؤسسة.

تسهم هيكلية البيانات بشكل فعال في تحليل البيانات بصورة أسرع، وتقوم على تحسين دقة اتخاذ القرارات، كما إنها تقلل من تكاليف التخزين والإدارة والمعالجة لهذه البيانات[7].

يشير تنظيم البيانات إلى كيفية تصنيف وترتيب البيانات داخل الأنظمة لجعلها قابلة للاستخدام بسهولة، وتكون أساليب تنظيم البيانات على النحو الآتي[8]:

1. البيانات المهيكلة (Structured Data):

تكون البيانات منظمة في جداول وقواعد بيانات، مثل البيانات المالية، أو بيانات المبيعات SQL.

2. البيانات غير المهيكلة (Unstructured Data):

أ.بيانات غير منظمة مثل النصوص، والصور، والفيديوهات.

ب. تُخزن في أنظمة تخزين مرنة مثل NoSQL .

3. البيانات شبه المهيكلة (Semi-Structured Data):

أ.تحوي على هيكل معين ولكنه غير صارم.

ب.مثال على ذلك البيانات بصيغ XML وJSON.

رابعًا: البنية التحتية التكنولوجية

البنية التحتية التكنولوجية تمثل الأساس الذي تعتمد عليه المؤسسات لتخزين البيانات ومعالجتها وتحليلها، وهي أساسية في التطبيق الذي يستخدم تقنيات الذكاء الاصطناعي، ومن هنا فإن بنية تحتية قوية وفعالة تعد عاملاً حاسما في نجاح مبادرات التحول الرقمي، وتحقيق

الاستفادة القصوى من البيانات، وتدعم هذه البنية التحتية العمليات الرقمية من خلال تخزين البيانات الضخمة، ومعالجتها بكفاءة عالية، كما إنها تقوم بتسريع التطبيقات الذكية من خلال تمكين تقنيات مثل التعلم الآلي، والرؤية الحاسوبية من العمل بفاعلية، فضلاً عن أنها تعزز المرونة من خلال تمكين المؤسسات من التكيف مع التغيرات التكنولوجية بسهولة[9].

تتكون البنى التحتية التكنولوجية الخاصة ببيانات الذكاء الاصطناعي من العناصر الآتية:

1. التخزين (Storage):

يجب أن تكون أنظمة التخزين قادرة على التعامل مع كميات كبيرة من البيانات التي تتنوع بين:

أ. البيانات المهيكلة مثل قواعد البيانات العلائقية.

ب. البيانات غير المهيكلة مثل النصوص والصور والفيديوهات.

تتكون أنظمة التخزين من الأنواع الآتية[10]:

- مستودعات البيانات (Data Warehouses) التي تستخدم لتخزين البيانات القديمة مثال ذلك Amazon Redshift.
- بحيرات البيانات (Data Lakes)، التي تستخدم لتخزين البيانات غير المهيكلة مثالHadoop Azure Data Lake .
- التخزين السحابي (Cloud Storage)، الذي يوفر سعة تخزين مرنة وقابلة للتوسع مثال ذلك AWS S3 وGoogle Cloud Storage.

2. قدرة الحوسبة (Computing Power):

تعتمد تطبيقات الذكاء الاصطناعي على قدرات حوسبية عالية لمعالجة البيانات بسرعة وكفاءة، وعرض النتائج المطلوبة، ويمكن إيجاز الخيارات الحوسبة بما يلي[11]:

أ. الخوادم المحلية (On-Premise Servers)، وتكون مناسبة للمؤسسات التي تحتاج إلى سيطرة كاملة على بياناتها.

ب. الحوسبة السحابية (Cloud Computing) تقدم مرونة عالية لتوسيع القدرات حسب الحاجة، مثال ذلك Amazon AWS, Microsoft Azure, Google Cloud.

ج. الحوسبة الهجينة (Hybrid Computing) وتمثل مزيجا بين السحابة والخوادم المحلية لتلبية الاحتياجات المختلفة.

3. الشبكات (Networking):

بنية تحتية شبكية قوية تضمن نقل البيانات بسرعة وأمان بين الأنظمة، وتكون على أن تحقق المتطلبات الآتية[12]:

أ. شبكات عالية السرعة مثل شبكات 5G لدعم نقل البيانات الفوري.

ب. أنظمة التوصيل بين الأنظمة (Interconnectivity) مثل بروتوكولات REST APIs.

ج. أمنية الشبكات (Network Security) العالية باستخدام تقنيات مثل جدران الحماية (Firewalls)، وأنظمة الكشف عن التطفل (IDS).

4. أطر معالجة البيانات (Data Processing Frameworks):

تعتمد تطبيقات الذكاء الاصطناعي على معالجة البيانات الضخمة باستخدام أطر عمل قوية على سبيل المثال الأطر الآتية[13]:

أ. Apache Hadoop لتحليل البيانات الضخمة.

ب. Apache Spark لمعالجة البيانات في الوقت الحقيقي.

ج. Google TensorFlow لتطوير نماذج التعلم الآلي.

5. أدوات إدارة البيانات (Data Management Tools)

تساعد على تنظيم البيانات، وضمان جودتها وتكاملها، ومن أشهر هذه الأدوات[7]:

أ. أدوات ETL (Extract, Transform, Load) مثل Talend و Informatica.

ب. أدوات Master Data Management (MDM) مثل Profisee و SAP Governance.Master Data

6. تقنيات الذكاء الاصطناعي (AI Technologies):

تحتاج المؤسسات إلى بنية تحتية تدعم تقنيات الذكاء الاصطناعي مثل[14]:

أ. الشبكات العصبية الاصطناعية (Neural Networks) التي تعمل على منصات مثل NVIDIA CUDA.

ب. أدوات التحليل التنبؤي (Predictive Analytics) وتعتمد على Python و R.

7. الأمان السيبراني (Cybersecurity):

حماية البيانات والتطبيقات من الهجمات السيبرانية، وذلك من خلال تطبيق استراتيجيات الأمان الآتية[15]:

أ. التشفير (Encryption) لحماية البيانات في أثناء النقل والتخزين.

ب. إدارة الوصول (Access Management) باستخدام المصادقة متعددة العوامل.

ج. تقييم المخاطر (Threats Assessment) لمعرفة الثغرات وإيجاد الحلول لغلقها.

د. الاستجابة للأحداث (Incidents Response) للتقليل من الأضرار الناتجة عن الهجمات السيبرانية وصدها.

هـ. أنظمة النسخ الاحتياطي (Backup Systems) لضمان استعادة البيانات في حالة الفقدان.

خامسًا: السياسات والإجراءات

تشير السياسات إلى مجموعة من القواعد والإرشادات التي تضعها المؤسسة لتحديد كيفية إدارة البيانات، وتخزينها، وحمايتها، واستخدامها، وتعد السياسات بمثابة الإطار العام الذي يوجه القرارات اليومية المتعلقة بالبيانات، بينما الإجراءات تمثل الخطوات أو العمليات التنفيذية، التي يجب اتباعها لضمان الالتزام بالسياسات، التي تركز على الجوانب العملية مثل جمع البيانات، ومعالجتها، وتوفيرها للمستخدمين[7].

السياسات تضمن أن المؤسسة تتوافق مع القوانين المحلية والدولية مثل اللائحة العامة لحماية البيانات وغيرها، فضلاً عن أن وضع سياسات واضحة يساعد في تحقيق دقة البيانات واكتمالها. كما إنها تضع قواعد لحماية البيانات الحساسة من التهديدات الأمنية، ومن هنا فإن وجود سياسات شفافة يعزز ثقة العملاء والشركاء في المؤسسة، هذا فضلاً عن أن السياسات الواضحة تسهل الوصول إلى البيانات وتحليلها مما يعزز من الابتكار[15]، [16]، [17]، [18].

يمكن تلخيص عناصر السياسات والإجراءات بما يلي:

1. سياسات إدارة البيانات (Data Management Policies):

أ. تحدد كيفية جمع، وتخزين، ومعالجة البيانات.

ب. تشمل سياسة الخصوصية (Privacy Policy) وسياسة التخزين (Storage Policy) وسياسة النسخ الاحتياطي (Backup Policy).

2. **سياسات أمن البيانات (Data Security Policies):**

أ. تهدف إلى حماية البيانات من التهديدات.

ب. تشمل سياسة الوصول (Access Policy) التي تحدد من يمكنه الوصول إلى البيانات، وسياسة التشفير (Encryption Policy) المسؤولة عن تحديد كيفية حماية البيانات في أثناء النقل والتخزين.

3. **إجراءات جودة البيانات (Data Quality Procedures):**

أ. تركز على ضمان دقة واكتمال البيانات.

ب. تشمل خطوات لتنظيف البيانات (Data Cleaning) وعمليات التحقق من الاتساق (Consistency Checks).

4. **إجراءات الامتثال (Compliance Procedures)** [7]:

أ. تضمن الامتثال للقوانين والسياسات التنظيمية.

ب. تشمل مراجعات دورية للبيانات وإعداد تقارير بشأن استخدام البيانات.

المحور الثاني: تحليل نماذج من البيانات الرقمية في دوائر الدولة

تم الاعتماد على البيانات الرقمية الخاصة بالتطبيقات الحكومية التي تمت استضافتها لدى مركز البيانات الوطني، وتمثل 10 قواعد بيانات لوزارات وهيئات، كما هو موضح في الجدول رقم(1) في أدناه:

جدول (1): عدد القيود الكلية لبيانات بعض مؤسسات الدولة ونسبة الجهوزية.

ت	اسم الوزارة او الهيئة	عدد القيود الكلي	عدد القيود المتلكئة	
1	وزارة التخطيط	2507240	328789	75%
2	هيئة التقاعد الوطنية	5986024	4971272	81%
3	وزارة الصحة	6791215	6437480	70%
4	وزارة الصحة	995221	1116	%86
5	وزارة الصناعة والمعادن	124850	89998	%88
6	وزارة البيئة	2403	2403	%90
7	وزارة الكهرباء	15722	15690	%89
8	دائرة التسجيل العقاري	14679	14190	%88
9	جهاز مكافحة الإرهاب	883846	قيود مكررة	%93
10	هيئة الحج والعمرة	2832119	قيود مكررة	%91

لم يتم التطرق في هذا البحث عن أنواع البيانات والتطبيقات التي تم استضافتها في مركز البيانات الوطني، وذلك حفاظاً على سرية هذه المعلومات، وبعد تحليل التقارير الصادرة من مركز البيانات الوطني عن البيانات الرقمية للتطبيقات التي تمت استضافتها في مركز البيانات الوطني، شخصت النقاط في أدناه:

1. الجهوزية الرقمية

يوضح الجدول أن نسب الجهوزية الرقمية تختلف بين الوزارات والهيئات، إلا أنها بمستويات جيدة، إذ لوحظ ما يلي:

أ. تحقق وزارة البيئة أعلى نسبة جاهزية رقمية، إذ تمتلك ما يقارب (90%) من القيود المتكاملة والمكتملة.

ب. وزارة الصحة أقل نسبة جاهزية، إذ بلغت ما يقارب (70%) بسبب غياب الرقم الوطني في معظم القيود.

ج. هيئة الحج والعمرة نسبة عالية من القيود المكررة، التي بلغت 4,055 قيدا مكررا.

2. مصادر الضعف في الجاهزية

أ. البيانات المكررة مثال على ذلك قاعدة بيانات هيئة الحج والعمرة لديها عدد كبير من القيود المكررة (4,055 قيدا)، إذ إن التكرارات تضعف الجهوزية لأنها تؤثر في دقة التحليل.

ب. غياب البيانات الأساسية، ومثال ذلك الأرقام الوطنية، إذ إن وزارة التخطيط وهيئة التقاعد الوطنية تعانيان من غياب الأرقام الوطنية في معظم القيود، وكذلك تواريخ الميلاد، حيث إن قاعدة البيانات الخاصة بوزارة التخطيط تحوي على 92,513 قيدا من دون تواريخ ميلاد، و892,691 قيدا في قاعدة بيانات هيئة التقاعد الوطنية تفتقر لتواريخ الميلاد.

ج. تعدد الأنظمة وقواعد البيانات، التي تحوي اختلافاً في أنظمة تخزين البيانات بين الجهات، ما يتطلب إلى جهد إضافي لعملية التكامل بين الأنظمة.

مما تقدم آنفاً؛ يمكن إيجاز التحديات المرتبطة بجاهزية البيانات الرقمية في مؤسسات الدولة العراقية بما يلي:

1. جودة البيانات

تكمن تحديات جودة البيانات بأنها غير كاملة، أو تحوي على أخطاء تقلل من دقة التحليل، فضلاً عن أن وجود تكرارات في القيود يحدث تضاربا في النتائج، وتؤدي هذه التحديات إلى انخفاض موثوقية التحليل المستند إلى البيانات، وإهدار الوقت والموارد في تنظيف البيانات.

2. تكامل البيانات بين الأنظمة

إن استخدام قواعد بيانات وأنظمة مختلفة بين الجهات الحكومية، وعدم وجود توحيد في طريقة تخزين وتنظيم البيانات، وغياب التخطيط المسبق للتكامل مع الأنظمة الأخرى، يؤثر بصورة مباشرة في التحول الرقمي الشامل، إذ تصعب مشاركة البيانات بين الوزارات، والأمر الذي يؤدي إلى تأخر في اتخاذ القرارات المستندة إلى البيانات.

3. البنية التحتية التكنولوجية

نقص في البنية التحتية التكنولوجية لدعم التخزين الضخم، ومعالجة البيانات الضخمة المطلوبة في عمليات الذكاء الاصطناعي، والقدرة الاستيعابية للبيانات المتزايدة وتحليلها في الوقت المناسب.

4. السياسات والحوكمة

غياب سياسات موحدة لإدارة البيانات وتصنيفها وضمان جودتها يؤدي إلى غياب معايير واضحة لجمع وتنظيم البيانات للاستفادة منها في عمليات التحول الرقمي، فضلاً عن أنها تؤثر في جهوزية مؤسسات الدولة للذكاء الاصطناعي.

المحور الرابع: الحلول المقترحة لتحسين جاهزية البيانات

إن جهوزية البيانات تعد الأساس لاستخدام تقنيات الذكاء الاصطناعي، ومن هنا فإن من دون بيانات متوفرة وجاهزة للذكاء الاصطناعي لا يمكن اعتماد هذه التقنيات من أجل زيادة الإنتاجية، وتحقيق التنمية المستدامة، بل بالعكس فإن استخدام هذه التقنيات في حالة عدم جهوزية البيانات يؤدي إلى نتائج خاطئة تؤثر سلبا في عمل المؤسسات، وبعد دراسة وتحليل البيانات المتوفرة في مركز البيانات الوطني، وعكسها على ما تم إيضاحه في الإطار العلمي والنظري لجهوزية البيانات للذكاء الاصطناعي، فإنه بالإمكان تحسين هذه الجهوزية في مؤسسات الدولة من خلال ما يلي:

1. بناء قاعدة بيانات وطنية موحدة من خلال إنشاء نظام وطني مركزي يجمع البيانات من الوزارات والهيئات جميعا، وليس بالضرورة أن تكون بيانات الدولة جميعا موجودة بصورة مركزية في مكان واحد، ولكن ضرورة إنشاء نظام يوحد ويفهرس البيانات، وهيكليتها، وأماكن تواجدها، ستسهم هذه العملية بصورة فعالة في تحسين تكامل البيانات بين الجهات.
2. استخدام الرقم الوطني كمعرف رئيس لكل فرد لتجنب التكرارات، وتقليل الأخطاء، والحصول على نتائج صحيحة في عمليات تحليل البيانات، وتقليل الجهد المبذول فيها.
3. تنظيف البيانات باستخدام أدوات متقدمة مثل Talend و Informatica التي تؤدي إلى زيادة في دقة التحليل.
4. تطبيق عمليات تحقق منتظمة للكشف عن الأخطاء والتكرارات، التي تسهم في تحسين جاهزية البيانات لتطبيقات الذكاء الاصطناعي.
5. تطوير منصات تكامل البيانات مثل ETL (Extract, Transform, Load) لتوحيد البيانات التي تؤدي إلى سهولة نقل البيانات بين الأنظمة.
6. استخدام خدمات الحوسبة السحابية مثل AWS و Azure للبيانات غير الحساسة مع تشفيرها قبل خزنها لضمان أمان أكثر، وتؤدي هذه العملية إلى تحسين الوصول للبيانات.

7. تطوير سياسات لإدارة البيانات تتضمن معايير لجمع البيانات، وقواعد لحماية الخصوصية، وإجراءات لتنظيف البيانات، التي تسهم في زيادة الشفافية في إدارة البيانات، وتعزيز عملية الامتثال للقوانين الوطنية والدولية.
8. استخدام تقنيات التعلم الآلي لتحليل البيانات غير المكتملة، وملء الفجوات لتحسين جهوزية البيانات بشكل كبير.
9. تطبيق أدوات تحليل البيانات للكشف عن الأنماط، وتصحيح الأخطاء لتوفير الوقت والجهد.
10. الاستثمار في مراكز البيانات الحديثة، التي تؤدي إلى تعزيز قدرة المؤسسات على معالجة البيانات الضخمة التي تعد أساسا باستخدام تقنيات الذكاء الاصطناعي.
11. تبني الحوسبة السحابية (Cloud Computing) لدعم تخزين البيانات، وتحليلها لتحسين الأداء العام لأنظمة البيانات.

المصادر

[1] أي بي أم. (2021). جودة البيانات من أجل جاهزية الذكاء الاصطناعي.
[2] مركز علوم البيانات. (2020). عام محوري لعلماء البيانات.
[3] مار، ب. (2018). استراتيجية البيانات: كيفية الربح من عالم البيانات الضخمة.
[4] ستالينجز، و. (2020). مبادئ التشفير وأمن الشبكات.
[5] سيمانتك. (2021). حلول تكامل البيانات.
[6] غارتنر. (2021). جاهزية البنية التحتية لتوافر البيانات.
[7] جمعية إدارة البيانات الدولية. (2021). إدارة البيانات: دليل المعرفة.
[8] كيمبال، ر.، وروس، م. (2013). دليل أدوات مستودع البيانات.
[9] غارتنر. (2021). بناء بنية تحتية قابلة للتوسع للبيانات.
[10] إنمون، ب. (2005). بناء مستودع البيانات.
[11] أمازون ويب سيرفيسز. (2022). شرح الحوسبة السحابية.
[12] سيسكو. (2021). البنية التحتية للشبكات لتطبيقات الذكاء الاصطناعي.
[13] زاهاريا، م. (2016). دليل شامل لـ Apache Spark.
[14] راسل، س.، ونورفيغ، ب. (2021). الذكاء الاصطناعي: نهج حديث.
[15] المعهد الوطني للمعايير والتكنولوجيا. (2022). إطار الأمن السيبراني للبنية التحتية للبيانات.
[16] أي بي أم. (2021). جودة البيانات لجاهزية الذكاء الاصطناعي.
[17] ماكينزي وشركاه. (2021). بناء الثقة من خلال حوكمة البيانات.
[18] غارتنر. (2021). الابتكار في سياسات إدارة البيانات.