الذكاء الاصطناعي التوليدي في هندسة البيانات

أ.د. محمد محمد الهادي

أكاديمية السادات للعلوم الإدارية

E-mail: mohamed.m.elhadi@gmail.com

١. المقدمة

الذكاء الاصطناعي التوليدي في هندسة البرمجيات يتضمن كل من موضوعات مثل:

التوليد أي إنشاء مجموعة البيانات الاصطناعية؛امرتبطة بتصميم مجموعة البيانات غير المتوازنة وبيانات اختبار واقعية لاختبار التطبيق المعين

الإستهلاك / استيعاب كل من مدخل البيانات الآلية، إثراء البيانات، تطبيع البيانات، محادثة بيانات صوتية، إلي نصية، محادثة بيانات نصية إلي صوتية

التخزين: يتضمن ضغط البيانات، التكرار، التخزين المتدرج التالي، توليد البيانات الصناعية، إعادة بناء البيانات

التحويل: يتكون من توحيد نوع البيانات، توجيه التحويل والنقل المخطط، تنفيذ قاعدة الأعمال، تنقية / تنظيف البيانات الآلية، التوحيد لإثراء البيانات

الخدمة: تتضمن لوحات المعلومات التفاعلية، إعداد آلية التقارير، عملية إستخراج وتحويل وتحميل ETL عكسية، تقرير امتثال البيانات الآلي، تكامل اللغة الطبيعية

• في المشهد المتطور لهندسة البيانات، لم يعد تكامل الذكاء الاصطناي التوليدي إلي مفهوما مستقبليا وإن ذلك صار حقيقة في الوقت الحالي. ومع وقوف البيانات باعتبارها شريان حياة للبتكارصار توليدها ومعالجتها وإدارتها أكثر أهمي من أي وقت مضي. كما أن إدخال براعة التوليد المدعوم بالطور في نماذج اللغة الكبيرة LLMs مثل نموذج GPT أي محول التدريب المسبق التوليدي يتضح أن هذه التكنولوجيا لا تعمل فقط علي تقرير الأطر لقائمة، بل إنهاتدث أيضا ثورة في درة حياة البيانات بأكملها.

• **إعادة اختراع دورة حياة هندسة البيانات**

تتضمن هندسة البيانات تقليديا نقل البيانات وإدارتها من خلال عدو مراحل تتمثل في التوليد، التخزين، الاستيعاب، التحويل، الحويل والخدمة (سابق عرضهاعاليه) . أي إنها عملية دقيقة تعتمد عل دقة البيانات. وتوافر كل ذلك وجاهزيته للتحليل. ولكل مرحلة من هذه المراحل الخمسة تحدياتها ومتطلباتها ونماذج اللغة الكبيرة صارت أدوات لا غني عنها التي توفر الحلول الذكية.

وفيما يلي شكل يوضح دورة حياة هندسة البيانات هذه:

شكل (رقم 1) دورة حياة هندسة البيانات

والعرض التالي لاستكشاف هذا التآزر عبر كل مرحلة حيث تتعمق في كيف يمكن الذكاء الاصطناعي التوليدي أن يكون المايسترو في سيمفونية البيانات.ظهر ومع تزايد ندرة مجموعات البيانات االفعلنة ومخاوف خصوصية البيانات.

(١) التوليد: من مرحلة إنشاء البيانات

تعد عملية إنشاء دورة حياة هندسة البيانات مرحلة تأسيسية، حيث يتم جمع البيانات الأولية من موارد

متنوعة مثل قواعد بيانات المعاملات، الكثير من أجهزة إنترنت الأشياء IoT، وخدمات الويب. ونظرا

دورة الحياة بأكملها بدءا من الاستيعاب وحتي التحليلات.

ومع تزايد ندرة مجموعات البيانات الفعلية ومخاوف خصوصية البيانات ظهر الذكاء الاصطناعي التوليدي كأداة قوية لإنشاء بيانات تركيبية. علي سبيل المثال، المؤسسات المالية صارت تتبني هذه التكنووجيا المتقدمة بشكل متزايد، التي منها شبكات الخصومة التوليدية GANs لخاصة لإنتاج معاملات مالية تحاكي البيانات الأصلية بضوح، كما تستخدم المالية صارت تتبني هذه التكنووجيا المتقدمة بشكل متزايد، التي منها شبكات الخصومة التوليدية GANs بنية شبكة مزدوجة تضمن مولد يصنع بيانات جديدة وتمييز يقيمها من خلال عملياتها التكررية ويولدون بيانات تركيبية تحافظ علي الفروق الاحصائية للسلوك المالي الحقيقي دون المساس بخوصية العملاء. ويمند هذا الابتكار إلي ما هو أبع من التمويل. ويصحح الذكاء الاسطناعي التوليدي بإختلالات البياناتا يضمن تحليل المشاعر العادلة علي منصات التجارة الإلكترونية. كما يوفر مجموعات بينات اختبار واقعية لتطوير البرمجيات ويثري بيانات التدريب كمهام معالجة للغة الطبيعيةNLP التي علاوة علي ذلك ، فإنه تقدم المساعدة في التحسين اللوجستي. وفي الأساس، مجموعات يحدث الذكاء الاصطناعي التوليدي ثورة في إنشاء مجموعات بيانات متعددة لاستخدامات وواقعية عبر مجالات مختلفة وإعطاء الأولية لأمن البيانات والخصوصية.

(٢) الاستهلاك / الاستيعاب: من بداية استيعاب البيانات

في عملية هندسة البيانات، تعتبر مرحلة استهلاك أي استيعاب البيات ضرورية جدا أيضا حيث أنها أيضا تجمع البيانات من مصادر متنوعة معالجة إخراجها.وهذه المرحلة يمكن أن تشكل تحديات كبيرة بسبب متغيرات البيانات ومصادرها وتدفقها. ويعد الاختيار بعناية بين العرض المجمع أو البث المباشر أمرا بالغ الأهمية من أجل بناء علي المتطلبات وحجم البيانات وقدرة المؤسسة علي معالجة البيانات في الوقت الفعلي أو في الوقت القريب.

وأحد التحديات التي تواجهها البنوك عند تحويل معلومات الطلبات المكتوبة بخط اليد إلي سجلات رقمية هو محدودية تكنولوجيا التعرف الضوئي علي حروف ORC في معالجة الكتابة المكتوبة يدويا وغيرالمقروءة. للتخفيف من هذا يأتي الذكاء الاصطناعي التوليدي ونماذج اللغة الكبيرة LLMs للعمل، وذلك باستخدام السياق من الأجـزاء الواضحة من النص لاستنتاج معاني الأجزاء غير الواضحة. وبالاعتماد علي بيانات التدريب المكتوبة، تتمتع هذه النماذج بمهارة في استنتاج النص وإعادة بنائها مما يضمن أن المستند الرقمي يعكس المادة الأصلية المكتوبة بخط اليد. كما تستخدم هذه التكنولوجيا المتقدمة أيضا في للأغراض التحليلية، وتحويل الصور إلي نص لتبسيط العمليات اللوجستية.قوعلي ذلك، يخدم الذكاء الاصطناعي التوليدي ونماذج اللغة الكبيرة كأدوات تحليلية حيوية في تعزيز دقة البيانات وجدواها، وتحويل تحديات الاستيعاب المعقدة إلي فرص للإبتكار والكفاءة.

(٣) التخزين: قبو الأصول الرقمية

في هندسة البيانات، يعد التحزين الفعال أمرا بالغ الأهمية، حيث يحقق التوازن بين توافر البيانات في الكفاءة التشغيلية، وتتوقف هذه المرحلة علي عدة عوامل التي منها ضمان إمكانية المقارنة مع متطلبات القراءة / الكتابة، منع الاختناقات، إتخاذ قرار بشأن المخاذن الأولية، ومع مراعاة قابلية التوسع، وإلتقاط بيانات الميتاداتا التي يتم الوصول إليها بشكل متكرر، والبيانات الباردة الأقل نشاطا.

ومع الزيادة الهائلة في إنشاء البيانات، وتحسين كفاءة التخزين التي تعتبر بالغة الأهمية، علي سبيل المثال خدمات بث الفيديو، والتي يمكنها الاستفادة من من الذكاء الاصطناعي التوليدي من أجل تقليص أحجام بيانات الفيديو تتعلم نماذج اللغة الكبيرة كيفية تشفير مقاطع الفيديو بإيجازوتحفق توازن دقيق بين الحفاظ علي الجودة وتقليل مساحة التخزين. وقد حددت تكنولوجيا الذكاء الاصطناعي التوليدي هذه البيانات القابلة للتوسع التي تتعلق فقط بالسبب الضروري للتخزين وإعادة ضغط مثير للإعجاب دون الإضرار بخبرة وتجربة المستخدم.

وإلي جانب ضغط الفيديو، تشمل حالات الاستخدام الأخري التي أحدثت ثورة في إدارة التخزين وتحسين التخزين الساحبي من خلال النسخ الذكي، واستخدام الطبقات التنبؤية لتوفير التكاليف وإنشاء مجموعات بيانات تركيبية للإعمال الجديدة واستعادة المتندات القديمة.

كما أنه من خلال هذه الابتكارات يؤديلذكاء الاصطناعي

التوليدي دورا محوريا في تحويل أساليب ومداخل التخزين توفير فعالية الكلفة والوظائف الضروية لعمليات البيانات المتطورة.

(٤) التحويل: تشكيل البيانات للمستقبل

في هندسة البيانات، تعتبر مرحلة التحويل او التحول حاسمة أيضا، خيث يتم تحسين البيانات لإطلاق إمكانياتها القصوي في توجيه رؤي وبصائر الأعمال. وتضمنت هذه المرحلة عمليات مختلفة منها التالي: محادثات الكتابة، التوحيد القياسي الرسمي، المعقد لمنطق الأعمال في نماذج البيانات، ومواءمة قواعد البيانات مع الحقائق الوظيفية لعمال مع تفوق شهادات نماذج اللغة الكبيرة مثل نموذج GPT-3 في المجال. حيث تستفيد من تعلمها الباهظ التكلفة لمعالجة مهام مثل توحيد تنسيقات التطبيقات باستخدام الاستنتاجية للتعرف علي الأنماط لإنشاء نصوص برمجية أو إعادة إنشائها وتحويل البيانات المتباينة إلي تنسيق موحد. وبالتالي تدفق المسار إلي البيانات النظرية للتحويل وتطبيقات التعلم الآلي.

وأبعد من مجرد التنسيق، تسهل نماذج اللغة الكبيرة ترجمة الهياكل التنظيمية المعقدة إلي تصميم قاعدة بيانات منظقية وتعريف قواعد العمل وألية عملية تنظيف تنظيف البيانات واقتراح إدراج البيانات الخارجية للحصول غلي غرض تحليلي أكثر إكتمالا.

كما تؤدي نماذج اللغة الكبيرة أيضا دورا تحويليا في هندسة البرمجيات ليس فقط من خلال تخسين جودة البيانات وتوحيدها، ولكم أيضا من خلال تسريع عملية إعداد البيانات مما يمهد الطريق لاتخاذ قرارات عمل قوية تتمحور حول البيانات.

(٥) الخدمة: إمداد بيانات بدقة متناهية

في هندسة البيانات، تعتبر مرحلة الخدمة هي المكان الذس يتم فيه تسليم ثمار العمل إلي المساهمين من خلال ثلاث طرق رئيسية هي: التحليلات، الحصول علي رؤي وبصائر عبر التقارير ولوحات المعلومات، والتعلم الألي لدعم التنبؤات والقرارات وتعكس استخلاص تحويل وتعميل الاستخلاس والتحويل والتحميل ETL للبيانات المحولة مرأخري إلي نظم الأعمال.

وأن تراكم دورة حياة البيانات هو في خدمة البيانات المعالجة للمستخمين النهائيين أو التطبيفات المستخدمة هنا التي تنثل لوحات المعلومات التفاعلية قمة سهولة الاستخدام. وتحدث نماذ الغة الكبيرة ثورة في تفاعل المستخدم مع تحليلات البيانات من خلال معالجة اللغة الظبيعية NLPعند دمجها في لوحات المعلومات التفاعلية، وتخدم نماذج اللغة الكبيرة كوسطاء أو وكلاء بين قواعد البيانات المعقدة والمستخدمين.

ويمكن للمستخدم كتابة استعلام أو التحدث به بلفة المحادثة وتقوم نماذج اللغة الكبيرة بعد ذلك بتحليل الإستعلام، باستخدام تدريب مكثف علي كمية هائلة من البيانات النصية لفهم نية المستخدم وتحديد الفروق الدقيقة في طلب الاستعلام. وبعد ذلك، يقوم النموذج بترجكة هذه الي استعلامات منظمة يمكن لنظام قاعدة البيانات لأساسي تنفيذها. فهو يسترد البيانات الدقيقة المطلوبة ويقدمها بنمسيق مفهوم. وتعمل هذه العملية السلسة علي تحسين خبرة وتجربة المستخدم بشكل كبيرما يسمح باستكشاف البيانات واتخاذ القراربشكل بديهي دون الحاجة إلي معرفة التكنولوجا بلغة الاستعلام. وتشمل حالات الاستخدام الأخري تبسيط أعداد التقارير الآلية من خلال تلخيص مجموعات بيانات معقدة. كما تشمل أيضا استخلاص وتحويل وتحميل ETL العكسية باستخدام التعيينات الذكية وضمان الامتثال التنظيمي كم خلال تقارير البيانات التي يتم إنشاؤها تلقائبا وتحويل تعقيد ذكاء الأعمال إلي سر مفهوم لاتخا القرارات التنفيذية.

وبذلك تعتبر نماذج اللغة الكبيرة في مرحلة الخدمة ككا يضمن أن مرحلة تحويل البيانات المتطورة تبلغ ذروتها في استخلاص قيمة استراتيجية مباشرة لمستخدمي الأعمال مما يعزز اتخاذ القرارات المستنيرة في جميع أنحاء المنظمة.

الاستنتاج

الذكاء الاصطناعي التوليدي، مع ١١وخاصة من خلال استخدام نماذج اللغة الكبيرة LLMs يبشر بنهضة في هندسة البيانات. فهو يؤدي لتحول التحديات إلي فرصا للإستثمار والتعقيدات إلي التبسيطات، والبيانات الأولية إلي سرود ثاقبة مع كل مرحلة من مراحل دورة حياة البيانات المعززة بالذكاء الاصطناعي التوليدي وبلك تصبح إمكانية الابتكار لا حدود لها.

وبينما نقف علي أعتابعصر جديد في هندسة البيانات، لم يعد السؤال هو ما إذا كان يجب الاعتماد علي الذكاء الاصطناعي التوليدي، ولكن ما مدي سرعة ذلك؟