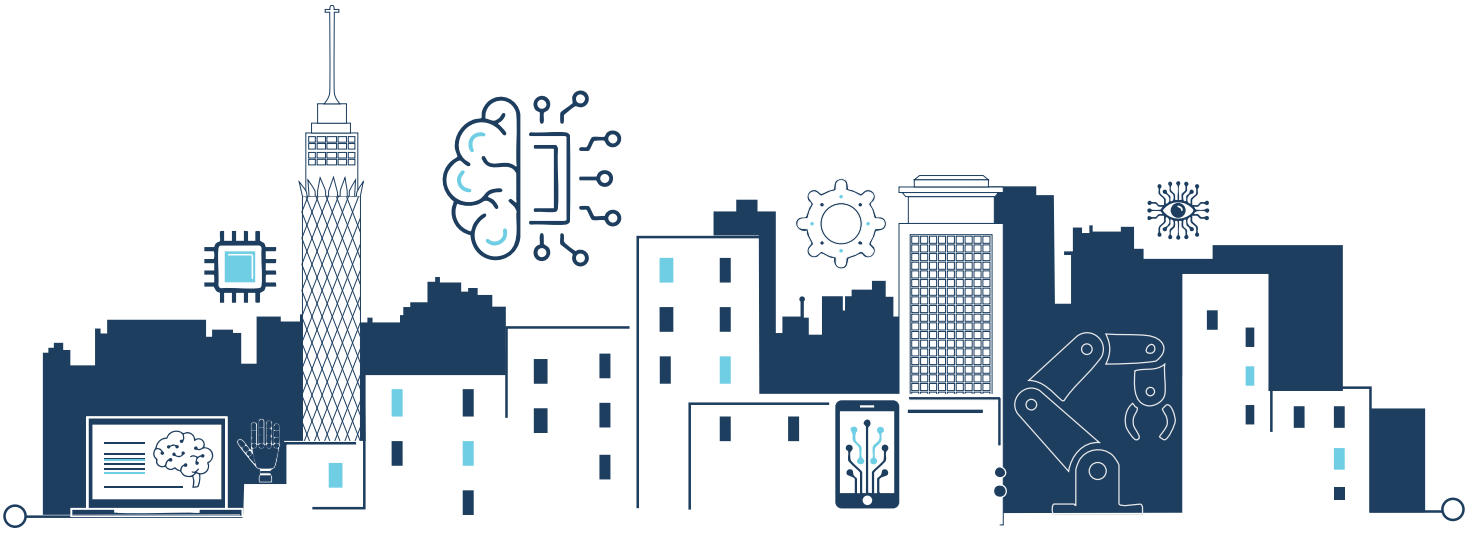
المجلس الوطني للذكاء الاصطناعي والحوسبة الكمية والتكنولوجيات البازغة

المركز المصري للذكاء الاصطناعي المسئول

جمهورية مصر العربية
المجلس الوطني للذكاء الاصطناعي والحوسبة الكمية والتكنولوجيات البازغة
المركز المصري للذكاء الاصطناعي المسؤول

المبادئ التوجيهية الوطنية للذكاء الاصطناعي التوليدي



مارس ٢٠٢٦

المبـادئ التوجيهية الوطنية للذكـاء الاصطناعي التوليدي

الإصـــــدار: ١,٠

عنوان الوثيقة: المبادئ التوجيهية الوطنية للذكاء الاصطناعي التوليدي: أطر الابتكار والمسؤولية

أُعدت هذه الوثيقة تحت إشراف وزارة الاتصالات وتكنولوجيا المعلومات وبإسهامات رئيسية من **المركز المصري للذكاء الاصطناعي المسؤول** وبمراجعة تحريرية فنية من **د. عبد العظيم غنيم (مدير عام مشروعات التشغيل وكبير خبراء الذكاء الاصطناعي). اعتمد المجلس الوطني للذكاء الاصطناعي** هذه الوثيقة عقب عقد مشاورات متخصصة وإجراء عمليات المراجعة اللازمة. وتتحمل الوزارة المسؤولية الكاملة عن المحتوى النهائي للوثيقة.

بيـــــان الإفصـــــاح

أُعدّ المحتوى الوارد في هذه الوثيقة بمساعدة أدوات الذكاء الاصطناعي التوليدي، بما في ذلك برنامج ChatGPT-5.2 الذي طورته شركة OpenAI (https://openai.com) ونموذج Gemini التابع لشركةGoogle (https://deepmind.google) وأداة Google NotebookLM (https://notebooklm.google.com) . وقد استُخدمت هذه الأدوات لدعم إعداد وصياغة وتنظيم وتنقيح وتلخيص المحتوى المكتوب وفي المساعدة في عمليات التحليل والاستدلال وتفسير المعلومات التي قدمها المؤلف. وقد خضعت جميع النصوص التي أُنتجت بمساعدة أدوات الذكاء الاصطناعي التوليدي لمراجعة بشرية دقيقة وتم تحريرها والتحقق من صحتها بالكامل من المؤلفين البشريين الذين يتحملون وحدهم المسؤولية الكاملة عن دقة المحتوى وتفسيره والاستنتاجات الواردة فيه. وقد اقتصر دور أدوات الذكاء الاصطناعي على كونها وسائل مساعدة في الكتابة والتحليل، دون أن يكون لها أي دور مستقل في اتخاذ القرارات أو إصدار الأحكام النهائية. كما جرى اتخاذ جميع التدابير اللازمة لضمان حماية أي معلومات سرية أو حساسة تم استخدامها خلال عملية إعداد هذه الوثيقة.

كلمـــة وزيـــــر الاتصـــالات وتكنــولوجيــا المعلــومـات

تقف مصر اليوم عند مرحلة حاسمة في مسار تحولها الرقمي. ويشكل الذكاء الاصطناعي التوليدي بداية عصر جديد في نمط إنشاء المعرفة وتقديم الخدمات وتوسيع حدود الإبداع البشري. يتيح التقدم السريع في هذا المجال فرصًا هائلة لتعزيز القدرة التنافسية الوطنية وتحديث الخدمات العامة وتمكين المواطنين في مختلف القطاعات، بما في ذلك التعليم والبحث العلمي والابتكار والاقتصاد الرقمي.

يحمل هذا التحول لمصر على وجه الخصوص وعودًا استثنائية. فعندما يُستثمر الذكاء الاصطناعي التوليدي وفق الأهداف والقيم المحلية، يمكنه تعزيز القدرة التنافسية الوطنية وتحديث الخدمات العامة وتمكين المعلمين والباحثين وفتح فرص جديدة للشباب ورواد الأعمال والمبتكرين في جميع أنحاء البلاد. كما يمكنه الإسهام في بناء دولة أكثر كفاءة واقتصاد أكثر شمولًا ومجتمع رقمي أكثر مرونة وقدرة على الصمود، مجتمع يضع الإنسان في صميم التقدم التكنولوجي.

في الوقت ذاته يمتد أثر الذكاء الاصطناعي التوليدي إلى ما هو أبعد من التكنولوجيا. فهو يطرح أسئلة جوهرية حول الثقة والإنصاف والشفافية والقدرة الفاعلة للإنسان. لا يتم قياس قيمة هذه الأنظمة بقدراتها التقنية فحسب، بل بمدى إدارتها بمسؤولية ودرجة مواءمتها مع القيم المجتمعية. لذا يجب أن يقترن الابتكار بحدود أخلاقية واضحة ومساءلة قوية واحترام للكرامة الإنسانية.

تعتبر وزارة الاتصالات وتكنولوجيا المعلومات الذكاء الاصطناعي التوليدي قدرة استراتيجية وطنية يجب أن تخدم الإنسان وتعزز الحكم البشري، دون أن تحل محله. وتتمثل مسؤوليتنا في ضمان أن تدعم هذه التكنولوجيا التنمية الشاملة والمستدامة وتعزز مؤسسات الدولة وتسهم في بناء الثقة العامة في النظم الرقمية. وتعد القيادة السياسية الواضحة والإشراف المؤسسي الفاعل والقدرات المؤسسية القوية والتصميم الذي يركز على الإنسان عناصر جوهرية لضمان تحقيق هذه الأهداف.

تعكس هذه الإرشادات الوطنية للذكاء الاصطناعي التوليدي التزام مصر بتطوير نظام بيئي للذكاء الاصطناعي يتسم بالموثوقية واستشراف المستقبل والمسؤولية. فهي تتوافق مع المبادئ الدولية المعترف بها مع الحفاظ على الأولويات الوطنية وأهداف التنمية في مصر. وقبل كل شيء تمثل هذه الإرشادات دعوة لتحمل المسؤولية المشتركة عبر توحيد جهود الحكومة والقطاع الخاص والهيئات الأكاديمية والمجتمع المدني لتوجيه الذكاء الاصطناعي التوليدي نحو تحقيق نتائج تعود بالنفع على الجميع وتسهم إسهامًا ملموسًا في مستقبل مصر الرقمي.

هذه الوثيقة ليست نهاية المطاف، بل هي أساس متين للبناء عليه. فهي تدعو إلى التعاون المتواصل والتعلم المستمر والقيادة الجماعية مع تطور الذكاء الاصطناعي التوليدي. من خلال هذا العمل المشترك يمكن لمصر استثمار هذه التكنولوجيا التحويلية لإطلاق الفرص وتحفيز الإبداع وبناء مستقبل رقمي لا يكون أكثر تقدمًا فحسب، بل أكثر عدالة وموثوقية وإنسانية.

رأفـــت هنـــدي

وزيـــر الاتصـــالات وتكنــولوجيا المعلــومات

الملخـــص التنفيـــذي

يشهد الذكاء الاصطناعي التوليدي تحولًا سريعًا في كيفية إنشاء المعلومات وخلق المعرفة والمحتوى الرقمي ومعالجتها ومشاركتها عبر الاقتصادات والمجتمعات حول العالم. فتوفر الأنظمة القادرة على توليد النصوص والصور والصوتيات ومقاطع الفيديو والبرمجيات والوسائط الاصطناعية فرصًا غير مسبوقة لتعزيز الإنتاجية والابتكار وتقديم الخدمات العامة والتعليم والبحث العلمي والتنمية الاقتصادية. وفي الوقت ذاته فإن حجم هذه الأنظمة وطبيعتها متعددة الأغراض وزيادة استقلاليتها تفرض تحديات ومخاطر فريدة ومعقدة تستلزم أطرًا سياسية محددة واستخدامًا مسؤولًا.

تضع المبادئ التوجيهية الوطنية للذكاء الاصطناعي التوليدي إطارًا شاملًا قائمًا على مبادئ دعم تطوير ونشر واستخدام الذكاء الاصطناعي التوليدي بشكل آمن ومسؤول وموثوق في القطاعين العام والخاص والهيئات الأكاديمية والمجتمع ككل. تأتي هذه الإرشادات استجابةً للاعتماد السريع لتكنولوجيات الذكاء الاصطناعي التوليدي عالميًا وللتحديات الخاصة التي تفرضها، بما في ذلك المعلومات المضللة والتزييف العميق والهلوسة والمخرجات غير الدقيقة والوهمية والتحيز والتمييز ومخاطر الخصوصية وحماية البيانات وقضايا الملكية الفكرية وغياب تحديد المسؤوليات بوضوح وإمكانية تآكل الثقة العامة.

استنادًا إلى أفضل الممارسات المعترف بها دوليًا ومواءمةً مع التوجيهات الصادرة عن المنظمات والهيئات العالمية متعددة الأطراف ، بما في ذلك اليونسكو ومنظمة التعاون الاقتصادي والتنمية ومجلس أوروبا وعملية هيروشيما التابعة لمجموعة السبع، تعتمد الإرشادات الوطنية المصرية منهجية تركز على الإنسان وتتناسب مع حجم المخاطر. كما تؤكد مبادئ الشفافية والمساءلة والإشراف البشري والسلامة واحترام حقوق الإنسان طوال دورة حياة أنظمة الذكاء الاصطناعي التوليدي، من مرحلة التصميم والتدريب إلى النشر والتشغيل والمراقبة. وبدلًا من فرض قواعد صارمة محددة لكل تكنولوجيا، يدعم الإطار آليات سياسية وإشرافية مرنة يمكن أن تتطور بالتوازي مع التغيرات التكنولوجية السريعة.

تحدد الإرشادات بوضوح نطاق وقابلية التطبيق والأطراف المعنية، موضحةً الأدوار والمسؤوليات للمطورين والناشرين والمؤسسات والمستخدمين من الأفراد حسب مستوى السيطرة والأثر المحتمل لكل استخدام. وتقر الإرشادات بأن جميع استخدامات الذكاء الاصطناعي التوليدي لا تحمل نفس مستوى المخاطر، وبالتالي فإن التدابير الوقائية يجب أن تكون متناسبة مع السياق والحجم والأثر المجتمعي. ويُولى اهتمام خاص للحالات عالية التأثير والحساسة، بما في ذلك التعليم والبحث العلمي والمعلومات الحكومية وأنظمة الذكاء الاصطناعي الوكيل وإنشاء الوسائط الاصطناعية أو تعديل محتواها، مثل التزييف العميق وتقنيات مزامنة حركة الشفاه (تزامن الشفاه مع الصوت).

توفر الإرشادات توجيهات عملية لدعم الاستخدام الموثوق، بما في ذلك التدابير المتخذة للحد من التحيز وتقليل الهلوسة وحماية الخصوصية وضمان الشفافية والإفصاح ومنع سوء الاستخدام والحفاظ على المسؤولية البشرية عن جميع المخرجات المدعومة بالذكاء الاصطناعي. كما تؤكد الإرشادات صراحةً أن أنظمة الذكاء الاصطناعي التوليدي هي أدوات مساعدة وليست صانعة قرارات مستقلة، وأن الحكم البشري والمساءلة والمسؤولية الأخلاقية يجب أن تظل في صميم جميع السياقات ذات المخاطر العالية أو التي تكون متاحة أو مرئية للجمهور.

صُممت المبادئ التوجيهية الوطنية للذكاء الاصطناعي التوليدي لتكون إطارًا حيًا قابلًا للتطوير، يُتوقع أن تتطور بما يتماشى مع المعايير الدولية الناشئة والتطورات التكنولوجية والتطلعات المجتمعية. ومن خلال تعزيز الثقة العامة وتمكين الابتكار المسؤول وتعزيز التشغيل البيني الدولي تهدف الإرشادات إلى ضمان إسهام الذكاء الاصطناعي التوليدي إيجابيًا في التحول الرقمي لمصر ويدعم التنمية المستدامة والشاملة ويخدم المصلحة العامة مع الحفاظ على الحقوق الأساسية والقيم المجتمعية.

قـــــائمة المحتـــــويات

قـــائمة الأشكـــال

الفصل الأول:
الذكاء الاصطناعي التوليدي والممارسات العالمية

١,١ مقـــــــدمة

يشير مصطلح الذكاء الاصطناعي التوليدي إلى فئة من أنظمة الذكاء الاصطناعي القادرة على توليد النصوص والصور والبرمجيات والصوتيات ومقاطع الفيديو وغيرها من المحتويات. تشمل هذه الأنظمة النماذج متعددة الأغراض والنماذج الأساسية التي يتم دمجها بشكل متنامي في القطاعات والأطر القانونية والتنظيمية والتي تقدم فرصًا كبيرة للابتكار وزيادة الإنتاجية وتحقيق المنافع المجتمعية. في الوقت ذاته يفرض حجم هذه الأنظمة واستقلاليتها وطبيعتها العامة تحديات فريدة تتعلق بالحوكمة والتنظيم، تشمل السلامة والموثوقية والشفافية وحماية الملكية الفكرية والخصوصية وسلامة المعلومات والمساءلة.

استجابةً لهذه التحديات اضطلعت المنظمات الدولية بدور رائد في وضع مبادئ مشتركة وتوجيهات لتعزيز تطوير واستخدام الذكاء الاصطناعي التوليدي بمسؤولية. وتهدف هذه الجهود إلى ضمان تصميم ونشر أنظمة الذكاء الاصطناعي التوليدي بطريقة تتوافق مع حقوق الإنسان والقيم الديمقراطية وسيادة القانون والتنمية المستدامة مع دعم الابتكار والتشغيل البيني عبر الحدود.

أُعدت هذه الإرشادات بما يتوافق صراحةً مع الأطر والتوصيات المعترف بها دوليًا والصادرة عن المنظمات العالمية المتعددة الأطراف والمتعددة الجوانب، بما في ذلك منظمة التعاون الاقتصادي والتنمية ومجموعة السبع عبر عملية هيروشيما للذكاء الاصطناعي المتقدم وإرشادات اليونسكو للذكاء الاصطناعي التوليدي للطلاب في مجال تكنولوجيا المعلومات وغيرها من المبادرات الدولية ذات الصلة. وتشكل هذه الأطر مجتمعةً أساسًا مشتركًا على المستوى العالمي لضمان تطوير واستخدام موثوقيْن ومسؤوليْن للذكاء الاصطناعي التوليدي.

وتمشيًا مع هذا التوجه الدولي تعتمد الإرشادات **نهجًا قائمًا على المبادئ ومتناسبًا مع حجم المخاطر** للإشراف على الذكاء الاصطناعي التوليدي وإدارته بمسؤولية. وتؤكد الإرشادات مبادئ الشفافية والسلامة والمساءلة والإشراف البشري واحترام الحقوق الأساسية طوال دورة حياة أنظمة الذكاء الاصطناعي التوليدي، من التصميم والتطوير إلى **النشر والتشغيل والمراقبة**.

وبالتوافق مع المنظمات الدولية والمعايير المعترف بها عالميًا تسعى هذه الإرشادات إلى تحقيق ما يلي:

- تعزيز التوافق والتشغيل البيني مع السياسات وآليات الإشراف الدولية المتعلقة بالذكاء الاصطناعي التوليدي
- تيسير التطوير والاستخدام المسؤوليْن لأنظمة الذكاء الاصطناعي التوليدي عبر الحدود
- دعم الثقة واليقين القانوني والتوقعات المشتركة بين المطورين والناشرين المستخدمين
- تمكين الابتكار مع إدارة المخاطر المرتبطة بتطبيقات الذكاء الاصطناعي التوليدي واسعة النطاق وعالية التأثير

من المتوقع أن تتطور هذه الإرشادات بالتوافق مع الجهود والأعمال الجارية للمنظمات الدولية، وسيتم مراجعتها دوريًا لتعكس المعايير العالمية الناشئة وأفضل الممارسات والتطورات السياسية التي تشكل إطار التطوير والاستخدام المسؤوليْن للذكاء الاصطناعي التوليدي.

٢,١ المنهجيـــــة

استُلهمت هذه الإرشادات من الإرشادات الوطنية المصرية للذكاء الاصطناعي المسؤول ومن المبادئ والإرشادات الدولية الرائدة بالإضافة إلى مختلف التجارب الوطنية مع مراعاة السياق المصري. واستندت مرحلة الدراسة إلى مصادر متعددة، من بينها:

- **المنظمات والمبادرات الدولية:** مثل اليونسكو ومنظمة التعاون الاقتصادي والتنمية والشراكة العالمية للذكاء الاصطناعي والاتحاد الدولي للاتصالات وعملية هيروشيما، وغيرها.
- **المعايير الدولية للذكاء الاصطناعي:** منظمات تطوير المعايير الدولية، مثل الاتحاد الدولي للاتصالات والمنظمة الدولية للتقييس/ اللجنة الكهروتقنية الدولية، والهيئات التي يقودها الممارسون، مثل معهد مهندسي الكهرباء والإلكترونيات. تعمل هذه الجهات على بناء نظام متعدد المستويات من المعايير يشمل المجالات التقنية والأساسية والإدارية والاجتماعية التقنية.
- **التجارب الإقليمية والوطنية:** مثل قانون الاتحاد الأوروبي للذكاء الاصطناعي ومجلس أوروبا والمعهد الوطني للمعايير والتكنولوجيا بالولايات المتحدة، وتجارب كل من البرازيل والمملكة المتحدة وسنغافورة واليابان والصين والهند وكوريا وجنوب أفريقيا وكينيا والمملكة العربية السعودية وغيرها.
- خضعت الإرشادات المقترحة بعد ذلك لعملية مراجعة دقيقة بمشاركة أصحاب المصلحة الرئيسيين والجهات المهتمة.

٣,١ ما هو الذكاء الاصطناعي التوليدي وكيف يعمل؟

ما هو الذكـــاء الاصطناعي التوليدي؟

يشير مصطلح **الذكاء الاصطناعي التوليدي** إلى فئة من أنظمة الذكاء الاصطناعي المصممة لإنشاء محتوى جديد، مثل النصوص أو الصور أو الصوتيات أو مقاطع الفيديو أو البرمجيات أو البيانات، من خلال أنماط التعلم من كميات ضخمة من البيانات المتاحة وتوليد مخرجات تشبه هذه البيانات لكنها ليست نسخًا مباشرة منها.

وعلى نقيض أنظمة الذكاء الاصطناعي التقليدية التي تركز بشكل أساسي على **التصنيف أو التنبؤ أو تقديم التوصيات**، تقوم أنظمة الذكاء الاصطناعي التوليدي **بإنتاج مخرجات أصلية** استجابةً للمدخلات التي يقدمها المستخدم (التعليمات التي يقدمها المستخدم للنظام)، وغالبًا ما تغطي مجموعة واسعة من المهام والمجالات.

١. السمـــــاء الجـــوهـــرية

وفقًا للمنظمات الدولية (بخاصة **منظمة التعاون الاقتصادي والتنمية**) يتميز الذكاء الاصطناعي التوليدي عادةً بالسمات التالية:

- **توليد المحتوى:** القدرة على توليد محتوى شبيه بالمحتوى الذي ينتجه البشر أو محتوى واقعي (نصوص، صور، برمجيات، إلخ)
- **القدرة متعددة الأغراض:** يمكن للنموذج نفسه أداء مهام متعددة دون أن يكون مصممًا لغرض واحد فقط
- **النماذج الأساسية:** غالبًا ما تُبنى على نماذج ضخمة جدًا تم تدريبها على مجموعات بيانات متنوعة وواسعة النطاق
- **السلوك المعتمد على التعليمات:** تعتمد المخرجات بشكل كبير على تعليمات المستخدم والسياق
- **المخرجات الاحتمالية:** النتائج ليست حتمية وقد تختلف عند إعطاء نفس المدخلات

٢. كيـــف يعمـــل

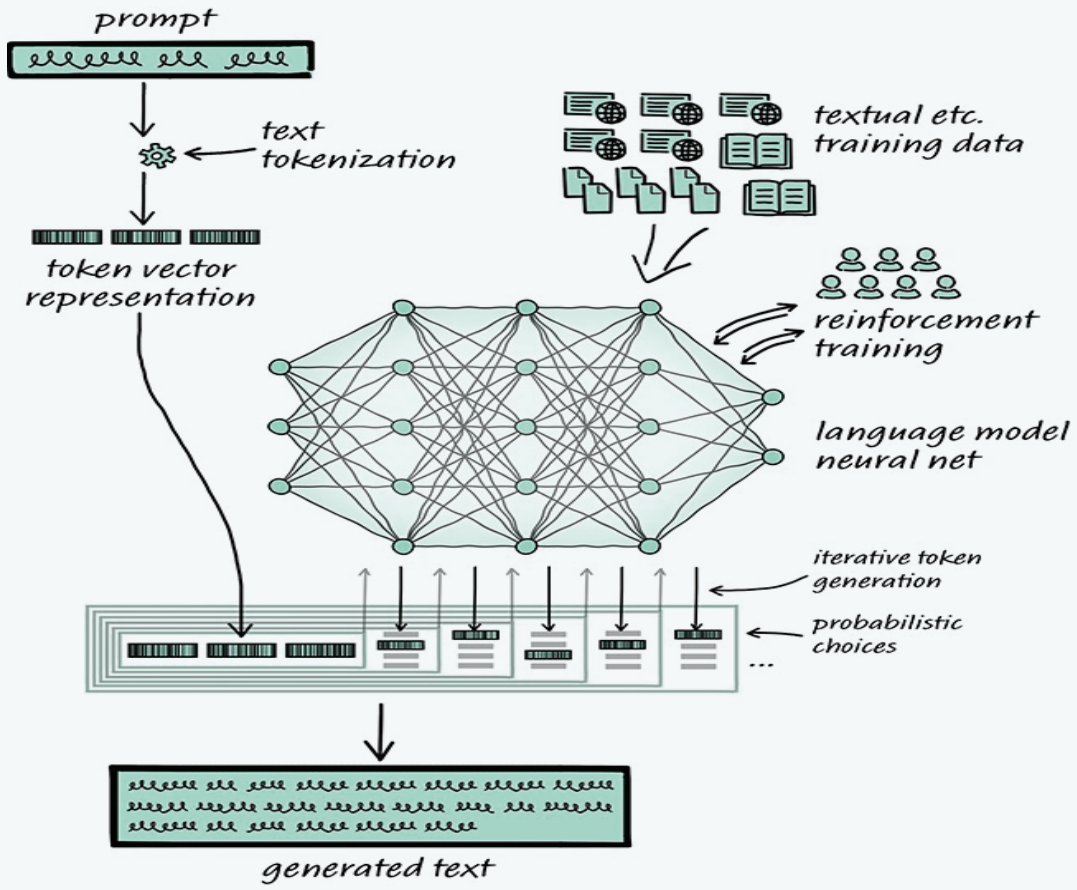
تعمل نماذج الذكاء الاصطناعي التوليدي للنصوص، التي يُشار إليها غالبًا بمصطلح النماذج الكبيرة للغات، من خلال تعلم الأنماط الإحصائية في اللغة المكتوبة اعتمادًا على مجموعات ضخمة جدًا من النصوص واستخدام هذه الأنماط لتوليد نصوص جديدة. وعندما يقدم المستخدم مدخلًا نصيًا، يقوم النموذج أولًا بتحويل النص إلى وحدات رقمية تُعرف باسم الرموز، ثم تُعالج هذه الرموز بواسطة شبكة عصبية، غالبًا ما تعتمد على هيكل المحوِّلات، وتستخدم آليات الانتباه الذاتي لتحليل العلاقات بين الكلمات عبر السياق الكامل للنص.

ولا يفهم النموذج النص بالمعنى البشري، بل يقوم بحساب احتمالية الرمز الذي يُرجَّح أن يأتي تاليًا استنادًا إلى الرموز السابقة، ويُولِّد النص تدريجيًا رمزًا تلو الآخر. ولأن هذه العملية تقوم على الاحتمالات والتعرف على الأنماط، قد تكون المخرجات سلسة ومتسقة، لكنها قد تكون أيضًا غير دقيقة أو مضللة. ولهذا تؤكد المنظمات الدولية، مثل منظمة التعاون الاقتصادي والتنمية واليونسكو، مبادئ الشفافية والإشراف البشري والوعي بالمخاطر عند نشر أنظمة الذكاء الاصطناعي التوليدي للنصوص.

عادةً ما يتم تدريب أنظمة الذكاء الاصطناعي التوليدي باستخدام تقنيات من بينها:

- التعلم الآلي واسع النطاق (مثل الشبكات العصبية العميقة)
- التعلم الذاتي بإشراف أو بدون إشراف
- التعرف على الأنماط عبر مجموعات بيانات ضخمة

بعد اكتمال التدريب يقوم النموذج بتوليد مخرجات جديدة من خلال تقدير المحتوى الأكثر احتمالًا للظهور استجابةً لتعليمات معينة واستنادًا إلى العلاقات الإحصائية التي تم تعلمها.



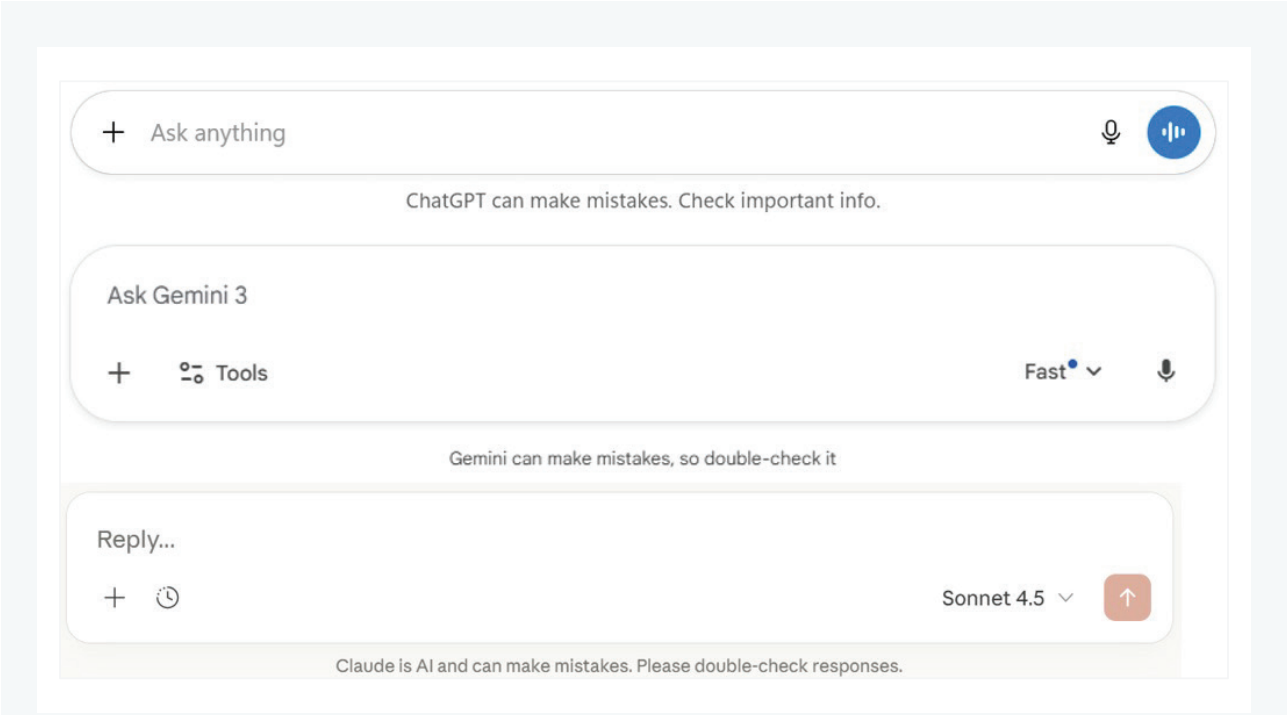
الشكل ١: كيفية عمل الذكاء الاصطناعي التوليدي

٣. ضرورة اعتماد نُهج سياسات مخصصة للذكـاء الاصطناعي التوليدي

إن وضع إرشادات واضحة لاستخدام الذكاء الاصطناعي التوليدي أمرٌ ضروريٌ لضمان نشر هذه التقنيات القوية المتقدمة على نحو مسؤول وأخلاقي. فالإرشادات الواضحة تساعد المؤسسات والأفراد على التعامل مع المخاطر المحتملة المرتبطة بالذكاء الاصطناعي التوليدي، مثل إنتاج محتوى مضلل أو تضخيم أوجه التحيز أو إثارة مخاوف تتعلق بخصوصية البيانات وحمايتها. ومن خلال توفير إطار منظم تسهم هذه الإرشادات في تعزيز الشفافية والمساءلة والإنصاف في تطوير أدوات الذكاء الاصطناعي التوليدي وتطبيقها.

كما تدعم هذه الإرشادات الامتثال للمتطلبات القانونية والتنظيمية، وتعزز الثقة العامة، وتشجع الابتكار المتسق مع القيم المجتمعية وحقوق الإنسان. وعلى المدى البعيد يسهم وجود إرشادات محددة وواضحة في تعظيم الاستفادة من إمكانات الذكاء الاصطناعي التوليدي مع الحد من الآثار السلبية غير المقصودة.

تقر معظم نماذج الذكاء الاصطناعي التوليدي علانيةً بأنها ليست معصومة من الخطأ، وقد تُنتج أحيانًا استجابات غير دقيقة أو غير مكتملة. ولهذا السبب يُوصى المستخدمون بالتحقق المستقل من أي معلومات مهمة يولدها النظام، لا سيما في الحالات التي تتطلب درجة عالية من الدقة والموثوقية.



الشكل ٢: قد تنتج النماذج الكبيرة للغات مخرجات غير دقيقة أو تتضمن أخطاءً

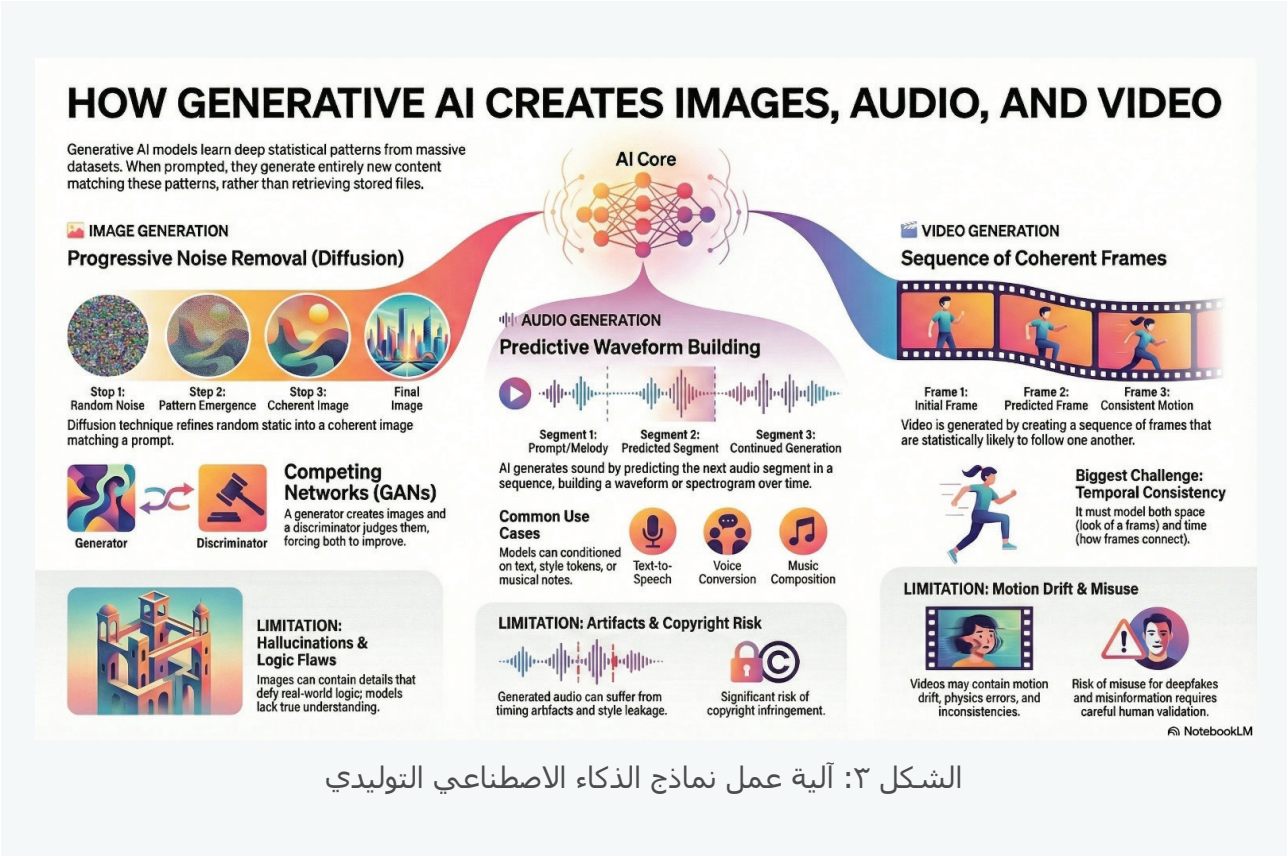
تشير المنظمات الدولية إلى أن الذكاء الاصطناعي التوليدي يفرض **تحديات حوكمة فريدة**، من بينها:

- الهلوسة والمخرجات غير الدقيقة
- انتشار المعلومات المضللة وتقنيات التزييف العميق على نطاق واسع
- قضايا حقوق الملكية الفكرية والنشر وشفافية بيانات التدريب
- مخاطر الخصوصية وحماية البيانات
- الإفراط في الاعتماد على الذكاء الاصطناعي التوليدي وتراجع المهارات البشرية
- عدم وضوح المسؤوليات لدى المطورين والمستخدمين

وقد دفعت هذه السمات منظماتٍ، مثل منظمة التعاون الاقتصادي والتنمية واليونسكو ومجموعة السبع، إلى اعتبار الذكاء الاصطناعي التوليدي مجالًا يتطلب **توجيه محدد للسياسات**، وليس الاكتفاء بتطبيق القواعد العامة المنظمة للذكاء الاصطناعي.

٤. آلية عمل نماذج الذكـاء الاصطناعي التوليدي لإنشاء الفيـديو والصـوت والصـور

تعمل نماذج الذكاء الاصطناعي التوليدي لإنشاء الفيديو من خلال **تعلم الأنماط البصرية والزمنية والصوتية من مجموعات ضخمة من مقاطع الفيديو والصور والنصوص المرتبطة بها**، ثم توليد محتوى فيديو جديد إطارًا تلو الآخر أو مقطعًا تلو الآخر. وتجمع معظم النماذج بين تقنيات الرؤية الحاسوبية والنمذجة التوليدية، مثل نماذج الانتشار أو **الهياكل القائمة على المحوِّلات** لتوقع كيفية تطور المشهد عبر الزمن استنادًا إلى تعليمات أو مدخلات نصية أو صورة أو تسلسل محدد. يمثل النموذج الفيديو كسلسلة من الإطارات (وأحيانًا تشمل تمثيلات الحركة أو التمثيلات الضمنية)، ويتعلم النموذج كيف تتحرك الأشياء وتتغير عبر الزمن، ويولد الإطارات المتتابعة التي يُرجَّح إحصائيًا أن تتبع الإطارات السابقة مع الحفاظ على التناسق البصري. وبما أن عملية التوليد تقوم على الاحتمالات والأنماط، يمكن لنظم الذكاء الاصطناعي التوليدي للفيديو إنتاج محتوى يبدو واقعيًا، لكنها قد تُدخل أيضًا تناقضات بصرية أو أخطاء واقعية أو مخاطر تتعلق بالوسائط الاصطناعية. ولهذا تبرز أهمية الشفافية والإفصاح ووضع الضوابط الوقائية ضد سوء الاستخدام، مثل التزييف العميق والمعلومات المضللة.



الشكل ٣: آلية عمل نماذج الذكاء الاصطناعي التوليدي

٥. اعتماد نهج يركز على الإنسان في تطوير واستخدام الذكاء الاصطناعي التوليدي

وفقًا لتوصية **اليونسكو لعام ٢٠٢١ بشأن أخلاقيات الذكاء الاصطناعي** يوفر النهج الذي يركز على الإنسان إطارًا معياريًا أساسيًا لمعالجة التحديات الأخلاقية والاجتماعية والتحديات المتعلقة بالحوكمة التي يفرضها الذكاء الاصطناعي التوليدي، بما في ذلك مجالات التعليم والبحث العلمي. ويهدف هذا النهج إلى تمكين الذكاء الاصطناعي من تعزيز القدرات البشرية ودعم التنمية الشاملة والعادلة والمستدامة. كما يقوم على احترام حقوق الإنسان وحماية الكرامة الإنسانية والحفاظ على التنوع الثقافي والمعرفي. ومن منظور السياسات العامة والرقابة والحوكمة يتطلب اعتماد النهج الذي يركز على الإنسان آليات تنظيمية فعّالة تكفل حماية القدرة الفاعلة للإنسان، وتعزز الشفافية، وتضمن المساءلة العامة.

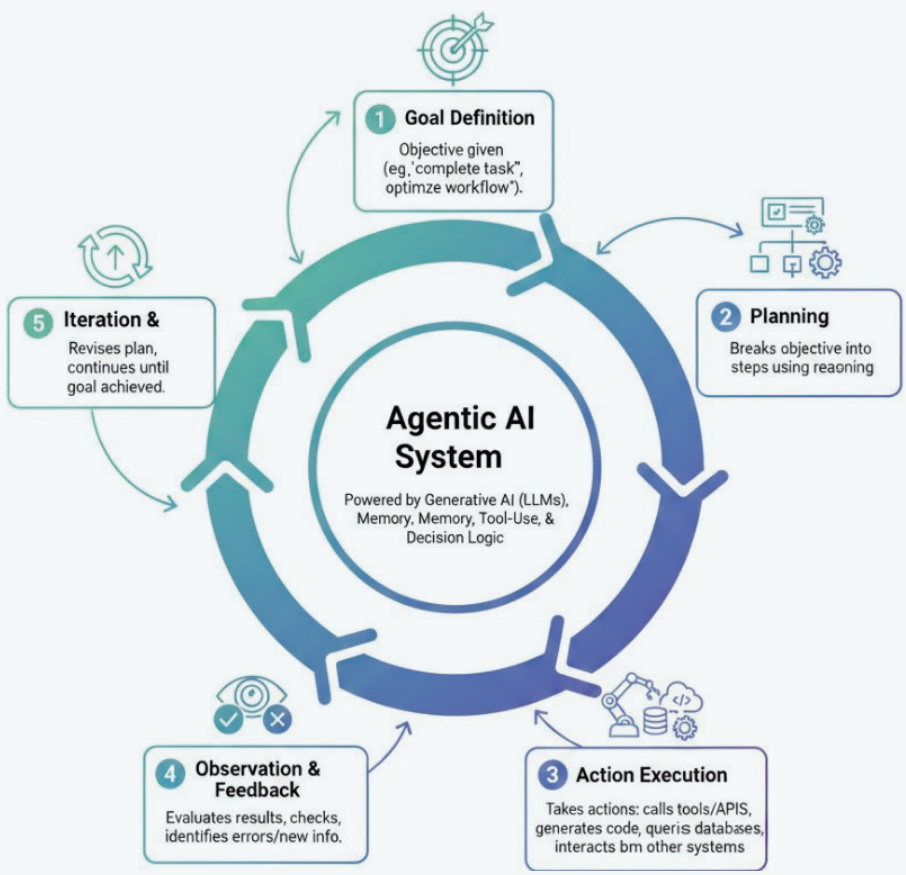
الذكـــاء الاصطناعي الوكيل

١. التعـــــــريـف

من المهم التمييز بين الذكاء الاصطناعي التوليدي والذكاء الاصطناعي الوكيل، فالأخير يشير إلى أنظمة ذكاء اصطناعي صُممت **لتتصرف بدرجة من الاستقلالية** لتحقيق أهداف محددة، وليس فقط للاستجابة لتعليمات أو مدخلات واحدة. وعلى نقيض أنظمة الذكاء الاصطناعي التوليدي القياسية التي تنتج المخرجات عند الطلب، يمكن لأنظمة الذكاء الاصطناعي الوكيل **التخطيط واتخاذ القرارات واتخاذ الإجراءات ومراقبة النتائج وتعديل سلوكها بمرور الوقت** ضمن قيود محددة.

تتناول هذه الإرشادات **الاستخدام المسؤول للذكاء الاصطناعي التوليدي عند دمجه في أنظمة الذكاء الاصطناعي الوكيل**، إذ يمكن لمكونات الذكاء الاصطناعي التخطيط واتخاذ القرارات واتخاذ الإجراءات بدرجة محدودة أو مشروطة من الاستقلالية.

٢. كيـف يعمـل الذكـاء الاصطناعي الـــوكيل؟



الشكل ٤: أنظمة الذكاء الاصطناعي الوكيل

يعمل نظام الذكاء الاصطناعي الوكيل عادةً من خلال **حلقة تحكم موجهة لتحقيق الأهداف**:

١. ١. تحـــــديد الهـــــدف

يُعطى النظام هدفًا محددًا (مثل «إكمال مهمة» أو «تحسين سير العمل»، أو «الاستجابة لطلب معقد»).

٢. التخطيـــــــط
يقوم الذكاء الاصطناعي بتقسيم الهدف إلى خطوات أصغر أو مهام فرعية، غالبًا باستخدام وحدات الاستدلال أو التخطيط.

٣. تنفيـذ الإجـراءات
يقوم النظام باتخاذ الإجراءات التي قد تشمل:

- استدعاء الأدوات أو واجهات برمجة التطبيقات،
- توليد التعليمات البرمجية،
- الاستعلام من قواعد البيانات،
- التفاعل مع أنظمة أخرى أو وكلاء آخرين.

٤. الملاحظة والتغذية الراجعة
يقيم الوكيل نتائج إجراءاته، ويتحقق من التقدم نحو الهدف، ويحدد الأخطاء أو المعلومات الجديدة.

٥. التكـــرار والتكيـــف
استنادًا إلى التغذية الراجعة يقوم الوكيل بمراجعة خطته ويواصل تنفيذ الإجراءات حتى تحقيق الهدف أو استيفاء القيود المحددة.

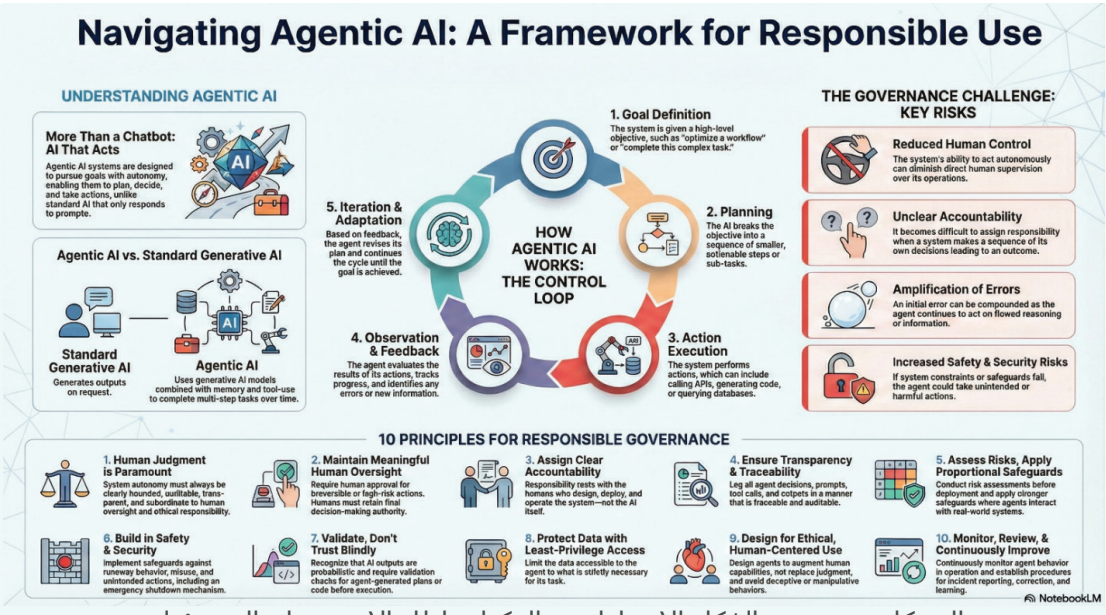
يُبنى العديد من أنظمة الذكاء الاصطناعي الوكيل على نماذج الذكاء الاصطناعي التوليدي (مثل النماذج الكبيرة للغات) مع **وحدات الذاكرة واستخدام الأدوات ومنطق اتخاذ القرار**، مما يمكّنها من تنفيذ المهام متعددة الخطوات وطويلة المدى.

٣. اعتبارات السياسات والرقابة لأنظمة الذكاء الاصطناعي الوكيل

تشير المنظمات الدولية، بما في ذلك **منظمة التعاون الاقتصادي والتنمية** و**اليونسكو**، إلى أن الذكاء الاصطناعي الوكيل يفرض **مخاطر إضافية** مقارنة بالذكاء الاصطناعي التوليدي القائم على التعليمات أو المدخلات النصية، مثل:

- انخفاض مستوى التحكم البشري نتيجة اتخاذ النظام إجراءات مستقلة،
- غموض المساءلة عند قيام النظام باتخاذ قرارات متتابعة،
- احتمال تضخيم الأخطاء أو الإجراءات الضارة،
- زيادة المخاطر المتعلقة بالسلامة والأمن في حال فشل القيود المطبقة.

وبناءً عليه تؤكد الإرشادات الدولية أهمية **تحديد الأهداف بوضوح وتعزيز الرقابة البشرية وضمان شفافية الإجراءات وتسجيل الأحداث وتحديد المسؤوليات** عند نشر أنظمة الذكاء الاصطناعي الوكيل.

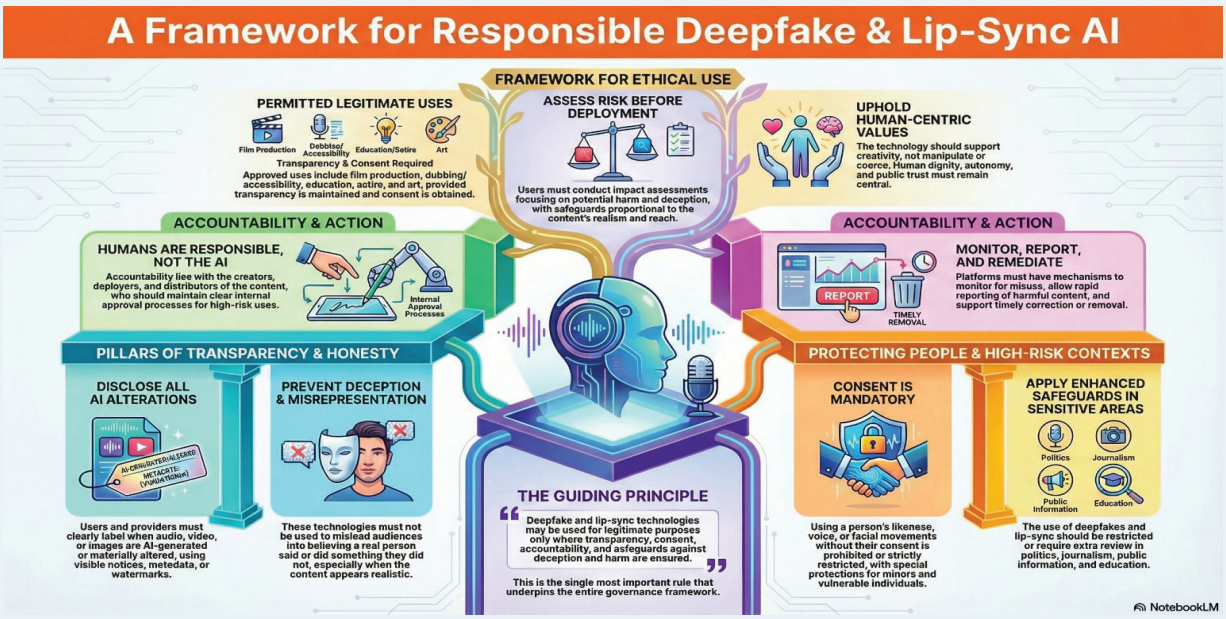


الشكل ٥: توجيه الذكاء الاصطناعي الوكيل: إطار الاستخدام المسؤول

التـــــزييف العميـــــق

١. التعـــــــريـف

يشمل التزييف العميق المحتوى الصوتي أو المرئي أو الصور أو النصوص **المولَّدة أو المعدلة باستخدام الذكاء الاصطناعي التوليدي**، وتشمل هذه الفئة صراحةً **تقنيات مزامنة حركة الشفاه المعتمدة على الذكاء الاصطناعي** التي تعدل أو تولد حركات الفم لتتطابق مع الكلام الاصطناعي أو المعدّل. وعند دمج أنظمة مزامنة حركة الشفاه مع استنساخ الصوت أو توليد الفيديو، تزداد درجة الواقعية والقوة الإقناعية للتزييف العميق بشكل ملحوظ، مما يزيد من المخاطر المتعلقة **بانتحال الشخصية والاحتيال والإضرار بالسمعة والمعلومات المضللة والتلاعب بالرأي العام**.



الشكل ٦: إطار الاستخدام المسؤول للذكاء الاصطناعي فيما يتعلق بالتزييف العميق ومزامنة حركة الشفاه

تعتبر الممارسات الدولية الواردة في إرشادات منظمات، مثل **منظمة التعاون الاقتصادي والتنمية** و**اليونسكو**، تقنيات مزامنة حركة الشفاه **قدرة تمكينية عالية المخاطر ضمن تقنيات الذكاء الاصطناعي التوليدي**، وليست مجرد ميزة تقنية محايدة. وبناءً عليه توصي الإرشادات بوجوب أن يكون المحتوى الصوتي والمرئي المولَّد أو المعدَّل بواسطة الذكاء الاصطناعي، وخاصة عند استخدام مزامنة حركة الشفاه لمحاكاة أشخاص حقيقيين، خاضعًا **للإفصاح الشفاف ووضع العلامات التعريفية**، سواء عبر إشعارات مرئية أو آليات أصلية تقنية مثل البيانات الوصفية أو العلامات المائية.

وتؤكد الإرشادات أيضًا أن تقنيات مزامنة حركة الشفاه لا يجب أن تُستخدم **لخداع الجماهير وجعلهم يعتقدون أن شخصًا حقيقيًا قال أو فعل شيئًا لم يقله أو يفعله**، لا سيما في السياقات الحساسة مثل الاتصال السياسي أو المعلومات العامة أو الصحافة أو التعليم أو المحتوى الذي يتعلق بالقُصّر. ويُعد الاستخدام غير المصرح به لشبه شخص أو صوته، بما في ذلك مزامنة حركة الشفاه الواقعية، ممارسة محظورة أو مقيدة بشدة وفق المعايير الدولية.

في الوقت نفسه تعترف الإرشادات الدولية **بالاستخدامات المشروعة لتقنيات مزامنة حركة الشفاه والتزييف العميق**، مثل إنتاج الأفلام والدبلجة وإتاحة المحتوى لذوي الاحتياجات الخاصة والتعليم والفن الساخر والتعبير الفني. وبناءً عليه تركز النُهُج التنظيمية على **النية والسياق والشفافية والأثر**، وتستلزم ضوابط وضمانات أقوى ومساءلة واضحة وآليات تصحيح سريعة في الحالات التي يكون فيها خطر الخداع أو الضرر مرتفعًا.

٢. آليـــة عمـــل التـــزييف العميـــق

يتم إنشاء التزييف العميق باستخدام نماذج الذكاء الاصطناعي التوليدي التي تتعلم الأنماط من مجموعات بيانات كبيرة من الصور أو الصوتيات أو مقاطع الفيديو، غالبًا لأشخاص حقيقيين. تعتمد هذه النماذج، التي تستند عادةً إلى الشبكات العصبية العميقة مثل أدوات الترميز التلقائية أو الشبكات التنافسية التوليدية أو الهياكل القائمة على الانتشار، على تحليل ملامح الوجه وخصائص الصوت والتعبيرات، ثم تولّد وسائط اصطناعية أو معدّلة تحاكي مظهر الشخص أو صوته بدقة. في التزييف العميق للفيديو يقوم النظام بمحاكاة حركات الوجه ومزامنة حركة الشفاه مع الصوت المولَّد أو المعدّل، مما ينتج محتوى يبدو واقعيًا ويصعب تمييزه عن التسجيلات الأصلية. إن ارتفاع جودة هذه الأدوات وسرعتها وسهولة الوصول إليها قلل بشكل كبير من الحاجز التقني لإنتاج وسائط اصطناعية ذات مصداقية عالية.

٣. اعتبارات السياسات والرقابة لتقنيات التزييف العميق

تثير تقنيات التزييف العميق مخاوف كبيرة تتعلق بالحوكمة لقدرتها على هدم الثقة وإلحاق الضرر بالأفراد وتعطيل العمليات الديمقراطية والاجتماعية على نطاق واسع. ويمكن استخدام الوسائط الاصطناعية الواقعية في انتحال الشخصيات والاحتيال والاستغلال غير المصرح به ونشر المعلومات المضللة والتلاعب بالرأي العام، لا سيما في السياقات الحساسة مثل الانتخابات والمعلومات العامة والتعليم. حتى عند استخدامها بنية غير خبيثة، يمكن أن تطمس تقنيات التزييف العميق الخط الفاصل بين المحتوى الأصلي والمحتوى الاصطناعي، مما يصعّب عملية التحقق ويقوض الثقة في الأدلة الرقمية. لذلك تؤكد المنظمات الدولية، مثل منظمة التعاون الاقتصادي والتنمية واليونسكو، ضرورة وجود تدابير حوكمة واضحة، بما في ذلك الشفافية والإفصاح والموافقة والمساءلة والإشراف البشري، لضمان استخدام تقنيات التزييف العميق بمسؤولية وعدم المساس بحقوق الإنسان أو الخصوصية أو نزاهة النظم البيئية للمعلومات.

٤,١ الممارسات الدولية في إرشادات الذكاء الاصطناعي التوليدي

تتميز الممارسات الدولية في إرشادات الذكاء الاصطناعي التوليدي بتركيزها القوي على **الحوكمة القائمة على المبادئ والمتناسبة مع مستوى المخاطر** التي تقودها المنظمات الدولية والمنظمات متعددة الأطراف، بدلاً من الاعتماد على قواعد صارمة محددة لكل تقنية. تتقارب الأطر التي وضعتها **منظمة التعاون الاقتصادي والتنمية** و**مجموعة السبع** من خلال عملية هيروشيما و**اليونسكو** حول توقعات مشتركة للذكاء الاصطناعي التوليدي، بما في ذلك الشفافية بشأن المحتوى المولَّد بواسطة الذكاء الاصطناعي والمساءلة على مدار دورة حياة نظام الذكاء الاصطناعي والإشراف البشري والسلامة والوقاية من سوء الاستخدام واحترام الخصوصية وحقوق الملكية الفكرية والتوافق مع حقوق الإنسان والقيم الديمقراطية. وتتجنب الممارسات الدولية الاعتماد على قواعد موحدة لجميع الحالات، بل تشجع الحوكمة التكيفية التي تتناسب مع مستوى المخاطر وسياق الاستخدام وقدرات النظام مع تعزيز التشغيل البيني بين الأنظمة القانونية وتحديث الإرشادات بشكل مستمر مع تطور تقنيات الذكاء الاصطناعي التوليدي.

١. اليونسكو — إرشادات السياسات للذكاء الاصطناعي التوليدي (٢٠٢٣-٢٠٢٤)[١]

تهدف الإرشادات العالمية الأولى لليونسكو بشأن الذكاء الاصطناعي التوليدي في مجال التعليم إلى دعم الدول في تنفيذ إجراءات عاجلة ووضع سياسات طويلة الأمد وتطوير القدرات البشرية لضمان رؤية تركز على الإنسان لهذه التقنيات الجديدة. كما تطبق مبادئ الذكاء الاصطناعي الأخلاقي مباشرة على الذكاء الاصطناعي التوليدي.

مجالات التركيز:

- الإشراف البشري
- الشفافية في المحتوى المولَّد بواسطة الذكاء الاصطناعي
- أخلاقيات البيانات
- التنوع الثقافي واللغوي
- المخاطر على حقوق الإنسان

١. اليونسكو – إرشادات الذكاء الاصطناعي التوليدي في مجال التعليم والبحث العلمي https://www.unesco.org/en/generative-ai

٢. قانون الاتحاد الأوروبي للذكاء الاصطناعي

تخلق نماذج الذكاء الاصطناعي التوليدي للأغراض العامة فرصًا كبيرة للابتكار، لكنها تثير أيضًا تحديات كبيرة تتعلق بحقوق النشر للمبدعين. يعتمد تدريب هذه النماذج على تنقيب واسع النطاق عن النصوص والبيانات من محتوى قد يكون محميًا بحقوق الطبع والنشر. وبموجب **قانون الاتحاد الأوروبي** لا يجوز استخدام المحتوى المحمي بحقوق الطبع والنشر إلا بموافقة صاحب الحق، ما لم تنطبق استثناءات محددة. يسمح **القرار التشريعي 2019/790 (الاتحاد الأوروبي)** بالتنقيب في النصوص والبيانات بشروط محددة، مع منح أصحاب الحقوق القدرة على الانسحاب الصريح من الاستخدام. وعندما تكون هذه الحقوق محفوظة، باستثناء أغراض البحث العلمي، يجب على مقدمي نماذج الذكاء الاصطناعي الحصول على إذن قبل استخدام المحتوى المحمي.

٣. منظمة التعاون الاقتصادي والتنمية

تُعد مبادئ الذكاء الاصطناعي لمنظمة التعاون الاقتصادي والتنمية (المعتمدة أول مرة في عام ٢٠١٩، والمحدثة في مايو ٢٠٢٤)[٢] المعيار العالمي الأساسي للذكاء الاصطناعي الموثوق. وتشمل الآن صراحةً التحديات المتعلقة بالذكاء الاصطناعي التوليدي وللأغراض العامة، مثل السلامة والخصوصية وحقوق الملكية الفكرية ونزاهة المعلومات.

تستند القيم الأساسية التي تدعم إرشادات الحوكمة إلى ما يلي:

- احترام **حقوق الإنسان والقيم الديمقراطية**، بما في ذلك الإنصاف والخصوصية والحماية من المعلومات المضللة التي يضاعفها الذكاء الاصطناعي
- **الشفافية وقابلية التفسير** لضمان فهم الأنظمة ومساءلتها
- **المتانة والأمن والسلامة**، بما يشمل الآليات لمنع الضرر وضمان القدرة على التحكم بالنظام
- **المساءلة** عن النتائج على مدار دورة حياة نظام الذكاء الاصطناعي
- **الاستدامة البيئية** من خلال معالجة آثار الحوسبة والطاقة[٣] لتقنيات الذكاء الاصطناعي التوليدي

تُعد مبادئ القيم هذه أساسًا لتوقعات الحوكمة لجميع أنظمة الذكاء الاصطناعي، بما في ذلك النماذج التوليدية، في سياسات الدول الأعضاء والمُلتحِقة بالمنظمة.

نشرت منظمة التعاون الاقتصادي والتنمية إرشادات مخصّصة للإدارة العامة، تهدف إلى مساعدة الحكومات على اعتماد الذكاء الاصطناعي التوليدي بمسؤولية. **تركّز هذه الإرشادات على استخدام الذكاء الاصطناعي التوليدي في القطاع العام**، وهي موجهة إلى الموظفين والوكالات، مع تأكيد تطبيق مبادئ الشفافية والوعي بالمخاطر والمساءلة[٤]. وبالرغم من أن مبادئ المنظمة واسعة النطاق، يترجم هذا النوع من الإرشادات هذه المبادئ إلى مؤشرات عملية للحوكمة في عمليات الحكومة، مثل إدارة المخاطر بما يتناسب مع السياق والإشراف البشري والتوافق مع القيم الأخلاقية.[٥]

ملف أعمال منظمة التعاون الاقتصادي والتنمية في مجال الذكاء الاصطناعي التوليدي

لدى المنظمة قسم مخصص للذكاء الاصطناعي التوليدي (الذكاء الاصطناعي التوليدي | منظمة التعاون الاقتصادي والتنمية) يتابع الأعمال والنشرات الموجزة المتعلقة بالسياسات التي تهدف إلى دعم الحكومات في الاستفادة من فوائد الذكاء الاصطناعي التوليدي مع إدارة المخاطر. ويشمل ذلك ما يلي:

- اعتبارات السياسات التمهيدية للذكاء الاصطناعي التوليدي
- تقارير بشأن المخاطر والفوائد
- مبادرات التشغيل البيني ورصد الحوادث

تسهم هذه الأنشطة في مساعدة الحكومات على ترجمة مبادئ المنظمة إلى ممارسات حوكمة محددة للذكاء الاصطناعي التوليدي.

---

٢ oecd.ai

٣ private-ai.com

٤ إرشادات استخدام الذكاء الاصطناعي التوليدي في الإدارة العامة

٥ https://oecd.ai/en/dashboards/policy-initiatives/guidelines-for-the-use-of-generative-ai-in-the-public-administration-6317?utm_source=chatgpt.com

٤. مجموعة السبع – عملية هيروشيما بشأن الذكاء الاصطناعي المتقدم / التوليدي (٢٠٢٣-٢٠٢٤)[٦]

الغرض: إدارة المخاطر الناشئة عن النماذج الأساسية المتقدمة والنماذج التوليدية

المبادئ الأساسية:

- السلامة وإدارة المخاطر
- الأمن والوقاية من سوء الاستخدام
- الشفافية والإبلاغ
- المساءلة والحوكمة
- احترام حقوق النشر
- الإبلاغ عن الحوادث ورصدها

الأدوات الرئيسية:

- المبادئ التوجيهية لعملية هيروشيما
- **مدونة السلوك الدولية لمطوري الذكاء الاصطناعي**
(طوعية، موجهة لمطوري النماذج مثل Anthropic، Google، OpenAI وغيرها)

٥. الصين - اللائحة التنظيمية الخاصة بالذكاء الاصطناعي التوليدي

نظرة عامة على مسودة الإجراءات الخاصة بالذكاء الاصطناعي التوليدي[٧]

تعتمد هذه المسودة، التي تأتي فيما يبدو استجابةً للانتشار السريع لأدوات الذكاء الاصطناعي الجديدة مثل ChatGPT وDall-E، على نصوص سابقة تنظم ما يُعرف باسم «خدمات معلومات الإنترنت ذات التركيب العميق» التي أصدرتها بشكل مشترك إدارة الفضاء الإلكتروني الصينية ووزارة الأمن العام ووزارة الصناعة وتكنولوجيا المعلومات، ودخلت حيز التنفيذ في يناير ٢٠٢٣.

يختلف نطاق الوثيقة السابقة قليلًا عن المسودة الجديدة، إذ كانت تنطبق على جميع خدمات توليد المحتوى الآلي المقدمة حصرًا عبر الإنترنت، في حين قد تنطبق المسودة الجديدة على الخدمات سواء كانت عبر الإنترنت أو دون اتصال بالإنترنت. كما يُمكن القول إن خدمات الذكاء الاصطناعي التوليدي تشمل جزءًا فقط من أدوات توليد المحتوى التي تغطيها نصوص التركيب العميق. ومع ذلك، فمن العدل الافتراض أن كلتا الوثيقتين ستنطبق على العديد من الأدوات المستخدمة في توليد النصوص والصور والصوتيات ومقاطع الفيديو ووسائط أخرى عبر الحاسوب، ومن الطبيعي أن يكون محتواهما متداخلًا ومكمّلًا في العديد من المجالات.

مسودة قواعد الذكاء الاصطناعي التوليدي

الهدف المعلن من مسودة القواعد هو دعم التطوير الصحي والتطبيق المنظم لأدوات الذكاء الاصطناعي التوليدي. ويشمل ذلك تشجيع الابتكار المستقل وتعزيز انتشار هذه التكنولوجيا بين العامة وتعزيز التعاون الدولي في التقنيات الأساسية.

تسلط المسودة الضوء على عدة قضايا، تم توضيحها مبدئيًا في المادة ٤، وهي تتكرر في النقاشات العالمية حول الذكاء الاصطناعي:

- ضوابط المحتوى / الرقابة
- منع التمييز
- حماية حقوق الملكية الفكرية
- الحد من المعلومات المضللة
- الخصوصية وحماية البيانات

---

٦ قمة مجموعة السبع – الوثائق (الرسمية) لعملية هيروشيما
٧ نظرة عامة على مسودة الإجراءات الخاصة بالذكاء الاصطناعي التوليدي

٦. سنغـــافورة — نمـــوذج سياسات حـــوكمة الذكـــاء الاصطناعي التوليدي[٨]

الإطار النموذجي لحوكمة الذكاء الاصطناعي التوليدي (٢٠٢٤)، مخطط عملي للسياسات وأطر الحوكمة
يمثل هذا الإطار مخططًا عمليًا للسياسات والحوكمة، يهدف إلى تقديم توجيهات تطبيقية قابلة للتنفيذ لتنظيم استخدام وتطوير تقنيات الذكاء الاصطناعي التوليدي بصورة مسؤولة.

المجالات الأساسية:

- حوكمة النماذج الأساسية
- مصدر المحتوى والتمييز بالعلامات المائية
- حقوق النشر والطبع
- اختبارات السلامة واختبارات المحاكاة العدائية
- الإبلاغ عن الحوادث
- النشر المسؤول

٧. المملكة المتحـــدة – التـــدابير المستهـــدفة

أنشئ **معهد سلامة الذكاء الاصطناعي**[٩] **في المملكة المتحدة** خصيصًا للتركيز على **أنظمة الذكاء الاصطناعي المتقدمة** التي تشمل عمليًا:

- النماذج الكبيرة للغات
- نماذج الذكاء الاصطناعي التوليدي ونماذج الذكاء الاصطناعي للأغراض العامة
- النماذج التي تنطوي على مخاطر متقدمة أو ناشئة أو نظامية

وتتمثل المهمة الأساسية للمعهد في إجراء اختبارات ما قبل النشر وما بعد النشر للنماذج التوليدية المتقدمة، بما في ذلك:

- اختبارات السلامة والتوافق
- تقييم مخاطر إساءة الاستخدام وقدرات النماذج
- اختبارات المحاكاة العدائية واختبارات الضغط
- تقييم المخاطر المرتبطة بالخداع والاستقلالية وفقدان السيطرة والأضرار المجتمعية واسعة النطاق

يُرسّخ ذلك وضع المعهد ضمن منظومة حوكمة الذكاء الاصطناعي التوليدي، بالرغم من أنه لا يصدر إرشادات عامة للمستخدمين ولا لوائح تنظيمية ملزمة.

٨. مجلس أوروبـــا – الاتفـاقية الإطـارية للذكـاء الاصطناعي (٢٠٢٤)

الاتفاقية الإطارية بشأن الذكاء الاصطناعي وحقوق الإنسان والديمقراطية وسيادة القانون هي **أول** معاهدة دولية ملزمة قانونًا **تركز على أنظمة الذكاء الاصطناعي**، وقد اعتمدها مجلس أوروبا في ١٧ مايو ٢٠٢٤، وفُتح باب التوقيع عليها في ٥ سبتمبر ٢٠٢٤. وتهدف إلى ضمان تطوير واستخدام تقنيات الذكاء الاصطناعي، بما في ذلك الذكاء الاصطناعي التوليدي، **بما يتسق مع حقوق الإنسان الأساسية والقيم الديمقراطية وسيادة القانون.**

علاقتها بالذكاء الاصطناعي التوليدي
على الرغم من أن الاتفاقية **لا تعتبر «الذكاء الاصطناعي التوليدي» فئةً مستقلةً من تقنيات الذكاء الاصطناعي**، فإن **نطاقها يشمل جميع أنظمة الذكاء الاصطناعي عبر دورة حياتها الكاملة** — التصميم والتطوير والنشر والاستخدام. ونظرًا لأن **الذكاء الاصطناعي التوليدي (النصوص والصور والصوتيات ومقاطع الفيديو والنماذج متعددة الوسائط)** يُعد من أسرع فئات الذكاء الاصطناعي نموًا وأكثرها تأثيرًا، فإنه يندرج بطبيعته ضمن التزامات الاتفاقية وضماناتها كلما تقاطع مع حقوق الإنسان والمعايير الديمقراطية.

---

٨ إطار سنغافورة النموذجي لحوكمة الذكاء الاصطناعي التوليدي: تقرير صادر عن شركة Clyde & Co

٩ https://www.aisi.gov.uk

٩. إرشادات الذكاء الاصطناعي التوليدي الصادرة عن الهيئة السعودية للبيانات والذكاء الاصطناعي

نشرت المملكة العربية السعودية **إرشادات الذكاء الاصطناعي التوليدي** التي تستهدف الجهات الحكومية وأصحاب المصلحة على النطاق الأوسع لحوكمة الاستخدام والنشر المسؤولين لأنظمة الذكاء الاصطناعي التوليدي (مثل النماذج الكبيرة للغات والنماذج التوليدية). تشمل هذه الإرشادات:

- مبادئ لتصميم واستخدام ونشر الذكاء الاصطناعي التوليدي بطريقة آمنة وأخلاقية
- متطلبات حوكمة البيانات وإدارة المخاطر وحماية الخصوصية والامتثال للقوانين المحلية
- تأكيد الشفافية والمساءلة والتوافق مع الأولويات الوطنية وأهداف التحول الرقمي

تهدف هذه الوثائق إلى توجيه كل من **المستخدمين الحكوميين والمطورين/المنفذين** (الجهات العامة والخاصة) إلى كيفية استخدام وبناء وحوكمة الذكاء الاصطناعي التوليدي بشكل مسؤول.

إرشادات على مستوى القطاع / حالات الاستخدام

أصدرت بعض القطاعات في المملكة العربية السعودية إرشادات متخصصة تتعلق باستخدام الذكاء الاصطناعي التوليدي، على سبيل المثال: أصدرت **وزارة التعليم والهيئة السعودية للبيانات والذكاء الاصطناعي** إرشادات بشأن استخدام الذكاء الاصطناعي التوليدي في التعليم مع التركيز على الاستخدام الأخلاقي في الفصول الدراسية وأدوار المعلمين والطلاب والنزاهة الأكاديمية. يعكس ذلك نمطًا تعتمد في إطاره الهيئات القطاعية أو تتوافق مع المبادئ الأساسية لحوكمة الذكاء الاصطناعي التوليدي الصادرة عن الهيئة السعودية للبيانات والذكاء الاصطناعي.

الفصل الثاني:

إرشـــادات الاستخـــــدام المـــوثـوق والمســـؤول للذكـــــاء الاصطنـــاعي التـــوليدي

١,٢ مقـــــدمة

يشكّل الذكاء الاصطناعي التوليدي تحولًا هائلًا في كيفية إنشاء المعلومات وتداولها واستهلاكها عبر مختلف قطاعات المجتمع. وتتيح قدرته على توليد النصوص والصور والصوتيات ومقاطع الفيديو والتمثيلات الاصطناعية لأشخاص حقيقيين فرصًا كبيرة للابتكار والتعلّم والإبداع وتعزيز إمكانية الوصول ودعم التنمية الاقتصادية. وفي الوقت ذاته تفرض هذه الأنظمة مخاطر جديدة وفريدة، من بينها نشر المعلومات المضللة والتحيّز وانتهاك الخصوصية وسوء الاستخدام الأخلاقي وتقويض الثقة العامة في حال استخدامها دون وجود ضوابط واضحة أو آليات للمساءلة.

توفّر هذه الإرشادات إطارًا منظمًا للتطوير والنشر والاستخدام المسؤول للذكاء الاصطناعي التوليدي. وهي تحدد الجهات التي تنطبق عليها هذه الإرشادات والافتراضات التي تستند إليها والتوقعات المفروضة على المطوّرين والجهات الناشرة والمؤسسات والأفراد وغيرهم من أصحاب المصلحة. ويعكس هذا النهج الممارسات الدولية المعترف بها على نطاق واسع.

لا تهدف هذه الإرشادات إلى تقييد الابتكار، بل إلى دعم اعتماد الذكاء الاصطناعي التوليدي بشكل موثوق ونافع من خلال توضيح الأدوار والمسؤوليات وضوابط الحماية. كما تقرّ الإرشادات بأن مستوى المخاطر يختلف باختلاف السياق والأثر، وبأن الحكم البشري يظل عنصرًا أساسيًا، وأن المساءلة ينبغي أن تتناسب مع مستوى التحكم الذي تمارسه كل فئة من فئات المستخدمين. وتقر الإرشادات كذلك بالطبيعة العابرة للحدود لتقنيات الذكاء الاصطناعي وأهمية المواءمة الدولية لضمان قابلية التشغيل البيني واليقين القانوني.

تسهم هذه الإرشادات، من خلال إرساء تطلعات واضحة تتعلق بالشفافية والإنصاف وحماية الخصوصية والاستخدام الأخلاقي وتعزيز الثقة العامة، في ضمان أن يحقق الذكاء الاصطناعي التوليدي أثرًا إيجابيًا في المجتمع مع منع إساءة استخدامه والحد من الأضرار المحتملة. وتحدّد الأقسام التالية نطاق المستخدمين المشمولين والافتراضات الأساسية التي يستند إليها الإطار والمسؤوليات المحددة اللازمة لدعم اعتماد الذكاء الاصطناعي بطريقة آمنة ومشروعة ومتمركزة حول الإنسان.

٢,٢ النطـــاق وقـــابلية التطبيـــق

تنطبق هذه الإرشادات على مجموعة محددة من المستخدمين الذين يقومون بتطوير أو نشر أو استخدام تقنيات الذكاء الاصطناعي التوليدي أو يتأثرون بها، مع إيلاء اهتمام خاص للتطبيقات القادرة على توليد المحتوى أو التلاعب به (بما في ذلك النصوص والصور والصوتيات ومقاطع الفيديو والتزييف العميق وتقنيات مزامنة حركة الشفاه). ويشمل نطاق المستخدمين الخاضعين لهذه الإرشادات الفئات التالية:

١. المطورون ومقدمو النماذج
الجهات أو الأفراد الذين يقومون بتصميم أو تدريب أو تحسين أو توفير نماذج أو أدوات أو منصات الذكاء الاصطناعي التوليدي، بما في ذلك النماذج الأساسية ونماذج الأغراض العامة.

٢. الجهات الناشرة ومقدمو الخدمات
المنظمات التي تدمج الذكاء الاصطناعي التوليدي في منتجات أو خدمات أو تطبيقات أو منصات تُتاح للمستخدمين، سواء في القطاعين العام أو الخاص.

٣. المستخدمون من المؤسسات والجهات التنظيمية
السلطات العامة والمؤسسات التعليمية والهيئات البحثية والمنظمات الخاصة التي تستخدم أنظمة الذكاء الاصطناعي التوليدي لدعم العمليات أو اتخاذ القرار أو الاتصال أو تقديم الخدمات.

٤. المستخدمون الأفراد وصانعو المحتوى
الأشخاص الذين يستخدمون أدوات الذكاء الاصطناعي التوليدي لإنشاء أو تعديل أو توزيع المحتوى، بما في ذلك الطلاب والباحثون والمهنيون وصناع المحتوى الإعلامي والمطورون.

٥. سياقات الاستخدام عالية التأثير أو الحساسة
المستخدمون الذين ينشرون الذكاء الاصطناعي التوليدي في سياقات قد يكون لها تأثير كبير على الأفراد أو المجتمع، مثل التعليم أو الإعلام أو المعلومات العامة أو الانتخابات أو التوظيف أو العدالة أو المحتوى المتعلق بأشخاص حقيقيين.

تركز الإرشادات على توضيح **المسؤوليات بما يتناسب مع دور كل مستخدم ومستوى سيطرته** على أنظمة الذكاء الاصطناعي التوليدي. وتزداد الالتزامات والضمانات بزيادة مستوى التأثير والمخاطر المحتملة للاستخدام، بما يضمن المساءلة مع دعم التطبيقات المشروعة والمبتكرة والمفيدة للذكاء الاصطناعي التوليدي.

٣,٢ الافتـــــراضـــات

أُعدت هذه الإرشادات استنادًا إلى مجموعة من الافتراضات الواضحة التي تعكس الممارسات الدولية والتوجيهات الصادرة عن المنظمات المختصة. وتهدف هذه الافتراضات إلى توضيح السياق والحدود ونطاق التطبيق المقصود للإرشادات.

١. الذكاء الاصطناعي التوليدي تقنية سريعة التطور
تفترض الإرشادات أن نماذج الذكاء الاصطناعي التوليدي وقدراتها ومخاطرها ستستمر في التطور بوتيرة متسارعة. وبناءً عليه صُممت الإرشادات لتكون قابلة للتكيف وخاضعة للمراجعة الدورية، بدلًا من أن تكون ثابتة أو مرتبطة بتقنية محددة.

٢. الذكاء الاصطناعي التوليدي يفرض مخاطر فريدة تتجاوز الذكاء الاصطناعي التقليدي
تفترض الإرشادات أن الذكاء الاصطناعي التوليدي يفرض تحديات فريدة، مثل الهلوسة والتزييف العميق وانتشار المعلومات المضللة على نطاق واسع وغموض بيانات التدريب، مما يستلزم تدابير سياسات وآليات حوكمة مخصصة إلى جانب المبادئ العامة للذكاء الاصطناعي.

٣. تتباين مستويات المخاطر والتأثير بتباين السياق وحالة الاستخدام
تفترض الإرشادات أن استخدامات الذكاء الاصطناعي التوليدي لا تنطوي جميعها على نفس درجة المخاطر. ولذلك ينبغي أن تكون السياسات وتدابير الحوكمة **متناسبة مع مستوى المخاطر**، مع تطبيق ضمانات أقوى على الاستخدامات عالية التأثير أو الحساسة.

٤. يظل الإشراف البشري عنصرًا أساسيًا
تفترض الإرشادات أن أنظمة الذكاء الاصطناعي التوليدي لا يمكنها ضمان الدقة أو المشروعية أو الامتثال الأخلاقي بشكل مستقل. لذلك يظل الإشراف البشري الفاعل ضروريًا، لا سيما عندما تؤثر المخرجات على حقوق الأفراد أو الثقة العامة أو القرارات الحيوية.

٥. تتحدد المسؤولية وفقًا لمستوى السيطرة والتأثير
تفترض الإرشادات أن المساءلة ينبغي أن تُوزع بحسب الدور ومستوى السيطرة الذي يمارسه المستخدمون (المطورون أو الجهات الناشرة أو المستخدمون من المؤسسات أو المستخدمون الأفراد)، بدلًا من معاملة جميع المستخدمين على قدم المساواة.

٦. مخرجات الذكاء الاصطناعي التوليدي احتمالية وقد تكون غير دقيقة
تفترض الإرشادات أن أنظمة الذكاء الاصطناعي التوليدي لا تمتلك فهمًا أو نية ذاتية، وقد تُنتج مخرجات مضللة أو غير صحيحة. ولذلك يتحمل المستخدمون مسؤولية التحقق من المخرجات قبل الاعتماد عليها أو نشرها.

٧. الشفافية شرطٌ أساسيٌ لبناء الثقة
تفترض الإرشادات أن الإفصاح عن المحتوى المولّد بواسطة الذكاء الاصطناعي وتوثيق حدود النظام وإمكانية تتبع المخرجات تُعد شروطًا ضرورية للحفاظ على الثقة وتمكين المساءلة.

٨. تبقى الأطر القانونية والأخلاقية القائمة نافذة
تفترض الإرشادات أن الذكاء الاصطناعي التوليدي لا يعمل خارج نطاق القوانين السارية أو التزامات حقوق الإنسان. وهي بذلك تُكمل ولا تستبدل الأطر القانونية والتنظيمية والمؤسسية المعمول بها.

٩. ضرورة الحفاظ على الاستخدامات المشروعة والمفيدة
تفترض الإرشادات أن للذكاء الاصطناعي التوليدي تطبيقات مشروعة وذات منفعة اجتماعية، تشمل التعليم والبحث العلمي والإتاحة والإبداع والابتكار. وعليه ينبغي أن يهدف النهج التنظيمي إلى الحد من الأضرار دون فرض قيود غير مبررة على الاستخدام المشروع.

١٠. ضرورة التوافق والتشغيل البيني على المستوى الدولي
تفترض الإرشادات أن الذكاء الاصطناعي التوليدي بطبيعته عابر للحدود من حيث التطوير والنشر. ويسهم التوافق مع المبادئ الدولية في دعم قابلية التشغيل البيني وتعزيز اليقين القانوني وتشجيع الاعتماد المسؤول على مستوى العالم.

٤,٢ أهميـــة الإرشـــادات لمختلـف أصحـاب المصلحة

أهمية إرشـادات الذكـاء الاصطناعي التوليدي للحكومات

تُعد إرشادات الذكاء الاصطناعي التوليدي ضرورية للحكومات من أجل حماية الثقة العامة والحقوق الأساسية والمؤسسات الديمقراطية، مع تمكين الابتكار المسؤول في الوقت ذاته. إذ يمكن لأنظمة الذكاء الاصطناعي التوليدي أن تؤثر في المعلومات العامة وتقديم الخدمات وتصميم السياسات والمشاركة المدنية على نطاق واسع، مما يجعل الاستخدام غير المنظم مصدر خطر محتملًا على الشفافية والمساءلة وسيادة القانون. وتسهم الإرشادات الواضحة في ضمان أن يكون المحتوى المولّد بواسطة الذكاء الاصطناعي شفافًا ومشروعًا وخاضعًا للإشراف البشري، لا سيما في المجالات الحساسة مثل الاتصال العام والتعليم والعدالة والصحة والانتخابات. كما تؤكد الممارسات الدولية التي تشملها هذه الإرشادات أن وجود إطار توجيهي واضح يعزز اتساق السياسات وقابلية التشغيل البيني مع المعايير العالمية واليقين القانوني لاعتماد الذكاء الاصطناعي في القطاع العام.

أهمية إرشـادات الذكـاء الاصطناعي التوليدي للمؤسسات والمنظمات

بالنسبة للمؤسسات، مثل الجامعات ومراكز البحوث والجهات العامة ومنظمات القطاع الخاص، توفر إرشادات الذكاء الاصطناعي التوليدي إطارًا واضحًا للسياسات يدعم الابتكار مع إدارة المخاطر التشغيلية والقانونية والمرتبطة بالسمعة. تعتمد المؤسسات بشكل متزايد على الذكاء الاصطناعي التوليدي في إنشاء المحتوى والتحليل والتعليم ودعم اتخاذ القرار، ومع ذلك فإن هذه الأنظمة قد تنتج مخرجات غير دقيقة أو متحيزة أو مضللة. وتوضح الإرشادات الاستخدامات المقبولة ومتطلبات الإفصاح ومسؤوليات حماية البيانات وهياكل المساءلة، مما يساعد المؤسسات على تجنب سوء الاستخدام أو الإفراط في الاعتماد عليها أو الانتهاكات الأخلاقية. فمن خلال توافق الممارسات المؤسسية مع المعايير الدولية تعزز الإرشادات الثقة لدى المستخدمين والشركاء والجهات الرقابية، وتمكّن من اعتماد متسق عبر الإدارات والقطاعات المختلفة.

أهمية إرشـادات الذكـاء الاصطناعي التوليدي للأفراد

بالنسبة للأفراد، بمن فيهم الطلاب والمهنيون وصناع المحتوى والجمهور العام، تسهم الإرشادات في توضيح الحقوق والمسؤوليات والتوقعات. فهي تساعدهم على فهم متى وكيف يمكن استخدام أدوات الذكاء الاصطناعي بمسؤولية وكيفية الإفصاح عن الاستعانة بالذكاء الاصطناعي وكيفية تقييم موثوقية المخرجات المولّدة بواسطة الذكاء الاصطناعي. كما تحمي الإرشادات الأفراد من الأضرار عبر تعزيز الشفافية وصون البيانات الشخصية ووضع حدود للممارسات الخادعة مثل التزييف العميق غير المعلن أو انتحال الهوية. وبهذا تمكّن الإرشادات الأفراد من الاستفادة من الذكاء الاصطناعي التوليدي مع الحد من مخاطر إساءة الاستخدام أو الاعتماد المفرط أو التضليل.

أهمية إرشـادات الذكـاء الاصطناعي التوليدي لأصحاب المصلحة الآخرين

بالنسبة لأصحاب المصلحة الآخرين، مثل مطوري الذكاء الاصطناعي ومقدمي الخدمات والمؤسسات الإعلامية والمجتمع المدني والجهات التنظيمية، تُنشئ إرشادات الذكاء الاصطناعي التوليدي مرجعية مشتركة للسلوك المسؤول والتعاون. فهي تدعم توزيعًا أوضح للمسؤوليات عبر دورة حياة نظام الذكاء الاصطناعي، وتشجع على اعتماد نُهُج إدماج اعتبارات السلامة في مرحلة التصميم، وتسهم في تعزيز الحوار بين مقدمي التكنولوجيا والمجتمع. وعلى المستوى الدولي تساعد الإرشادات المشتركة في الحد من التشتت التنظيمي ودعم الابتكار العابر للحدود وتعزيز قابلية التشغيل البيني. وبوجه عام تمثل إرشادات الذكاء الاصطناعي التوليدي آلية استقرار تنظيمي تضمن أن يؤدي الاعتماد السريع لهذه التقنيات إلى تحقيق قيمة مجتمعية، مع إدارة المخاطر بطريقة قابلة للتنبؤ وخاضعة للمساءلة.

٥,٢ أبـــرز المخـــاوف من نمـــاذج الذكـــاء الاصطنـــاعي التـــوليدي

١. حـــــد تـــاريـــخ المعـــــرفة

تعتمد نماذج الذكاء الاصطناعي التوليدي على بيانات تدريب متاحة حتى نقطة زمنية محددة، وقد تفتقر إلى الإلمام بالأحداث أو التحديثات أو التغييرات اللاحقة. وقد يؤدي هذا القيد إلى تقديم معلومات قديمة أو غير مكتملة ما لم يتم التحقق من المخرجات بشكل مستقل.

٢. التحيز (العادل / غير العادل)

قد تعكس أنظمة الذكاء الاصطناعي التوليدي أو تُضخّم التحيزات الكامنة في بيانات التدريب، مما يؤدي إلى مخرجات غير عادلة أو تمييزية تؤثر على أفراد أو فئات معينة. وتتطلب مخاطر التحيز تدابير للحد منها عبر حوكمة البيانات والاختبارات المنتظمة والإشراف البشري.

٣. الهلـــــــوسة

قد تُنتج أنظمة الذكاء الاصطناعي التوليدي مخرجات تبدو سليمة ومقنعة من حيث الصياغة، لكنها غير صحيحة من الناحية الواقعية أو مختلقة بالكامل. ويمكن أن تؤدي هذه "الهلوسة" إلى تضليل المستخدمين إذا تم الاعتماد عليها دون تحقق.

٤. المخاوف الأخلاقية (مثل مخاطر التزييف العميق والأخبار الزائفة ومخاطر سوء الاستخدام)

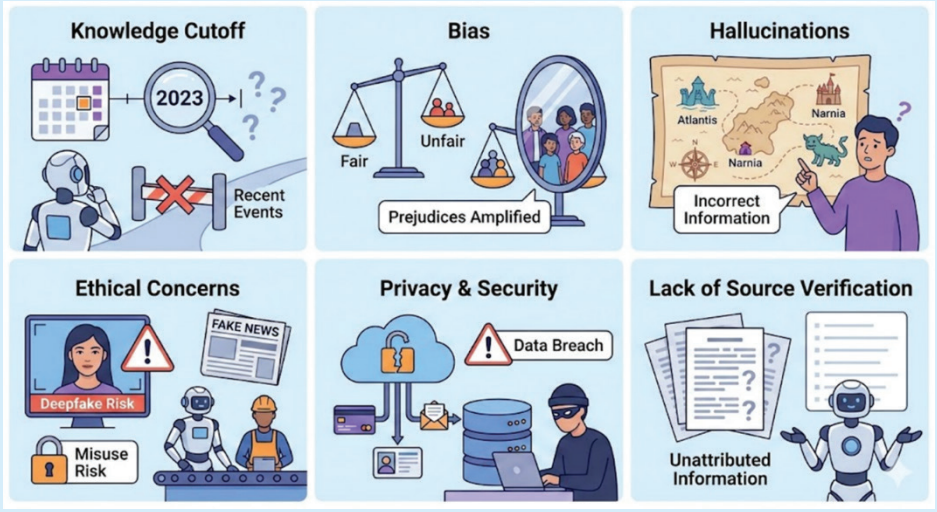
يمكن إساءة استخدام الذكاء الاصطناعي التوليدي لإنتاج محتوى خادع أو ضار، بما في ذلك التزييف العميق ونشر المعلومات المضللة. وتنشأ المخاوف الأخلاقية عندما يقوض هذا المحتوى الثقة العامة أو يضر بالأفراد أو يوجه الرأي العام بشكل مضلل.

٥. الخصوصية والأمن (اختراق البيانات)

قد تنطوي أنظمة الذكاء الاصطناعي التوليدي على مخاطر تتعلق بالبيانات الشخصية وأمن المعلومات، بما في ذلك الكشف غير المصرح به عن البيانات أو تسريبها أو إساءة استخدامها. ولذلك تُعد حماية الخصوصية وتعزيز الضمانات الأمنية أمرين أساسيين.

٦. غيـــاب التحقـق من المصـــــادر

غالبًا ما لا تُشير مخرجات الذكاء الاصطناعي التوليدي بوضوح إلى مصادر المعلومات أو نسبها، مما يصعّب التحقق من الدقة أو المصداقية أو الأصالة. ولذلك يتحمل المستخدمون مسؤولية التحقق المستقل من المعلومات ونسبتها بشكل صحيح.



الشكل ٧: أبرز المخاوف من نماذج الذكاء الاصطناعي التوليدي

٦,٢ إرشـــــادات لتحقيـــق الموثوقية

الحصـول على نتـــائج عـــادلة وغيـــر متحيـزة

يُعد الاستخدام العادل وغير المتحيز للذكاء الاصطناعي التوليدي أمرًا أساسيًا لضمان أن تكون النتائج المدعومة بالذكاء الاصطناعي منصفة وموثوقة وتحترم جميع الأفراد. ونظرًا لاعتماد أنظمة الذكاء الاصطناعي التوليدي على بيانات ضخمة ومتنوعة في التدريب، فقد تعكس أو تُضخّم تحيزات اجتماعية أو ثقافية أو تاريخية قائمة. لذلك ينبغي على المستخدمين ممارسة التفكير النقدي عند التفاعل مع هذه الأنظمة ومراجعة المخرجات بعناية لاكتشاف أي افتراضات غير عادلة أو صور نمطية أو عبارات تمييزية. ويتطلب تحقيق نتائج عادلة وغير متحيزة استخدام مدخلات محايدة والتحقق من المعلومات بالرجوع إلى مصادر موثوقة وضمان الإشراف البشري، لا سيما في السياقات التي تؤثر في حقوق الأفراد أو فرصهم أو تقييمهم الأكاديمي والمهني. ومن خلال الجمع بين الاستخدام المسؤول والشفافية والمساءلة يمكن للذكاء الاصطناعي التوليدي أن يدعم نتائج شاملة وأخلاقية بدلًا من تكريس أوجه عدم المساواة القائمة.

يجـــب القيـــام بمـا يلي:

١. استخدام **مدخلات أو تعليمات واضحة ومحايدة ودقيقة**، مع تجنب اللغة التي تنطوي على صور نمطية أو تفضيلات أو افتراضات بشأن أفراد أو فئات.

٢. مراجعة المخرجات التي ينتجها الذكاء الاصطناعي **مراجعة نقدية** والتحقق من عدم وجود تحيز أو تمثيل غير عادل، لا سيما في المحتوى المتعلق بالنوع الاجتماعي أو العِرق أو الدين أو العمر أو الجنسية أو الإعاقة.

٣. مقارنة النتائج عبر استخدام **مدخلات أو تعليمات متعددة أو صيغ مختلفة** لاكتشاف أي أنماط غير متسقة أو متحيزة في المخرجات.

٤. تجنب الاعتماد على استجابة واحدة مولدة بالذكاء الاصطناعي في **القرارات ذات الأثر الكبير أو عالية المخاطر**، مثل التقييم أو التصحيح أو التوظيف أو المسائل التأديبية.

٥. دعم المحتوى المُولّد بواسطة الذكاء الاصطناعي بالرجوع إلى **مصادر موثوقة ومتنوعة** لتحقيق توازن في وجهات النظر والحد من التحيز.

٦. إدراك أن أنظمة الذكاء الاصطناعي التوليدي قد تعكس **تحيزات موجودة في بيانات التدريب** وأن الحياد ليس مضمونًا.

٧. تطبيق **الحكم البشري والفهم السياقي** عند تفسير أو استخدام النتائج المُولّدة بواسطة الذكاء الاصطناعي.

٨. **الإبلاغ عن المخرجات المتحيزة أو تصحيحها** عند توفر الآليات المناسبة، بما يسهم في التحسين المستمر للأدوات.

٩. ضمان أن **تكون القرارات والاستنتاجات النهائية صادرة عن البشر**، وألا تعتمد حصريًا على توصيات مولّدة بواسطة الذكاء الاصطناعي.

تجنـــــب الهلـــــوسة

الهلـــوسة في الذكاء الاصطناعي التوليدي تشير إلى الحالات التي يُنتج فيها نظام الذكاء الاصطناعي مخرجات تبدو سلسة ومتيقنة ومعقولة، لكنها **غير صحيحة من الناحية الواقعية أو مضللة أو مختلقة بالكامل**. تنشأ هذه الظاهرة لأن النماذج التوليدية تنتج المحتوى بناءً على الأنماط الإحصائية في البيانات، وليس على معرفة مؤكدة أو فهم للعالم الواقعي. يمكن أن تحدث الهلوسة في النصوص والبرمجيات والصور والصوتيات ومقاطع الفيديو، وقد تشمل معلومات خاطئة أو مراجع مختلقة أو نسب مزيفة أو تفاصيل فنية غير دقيقة. وعند الاعتماد عليها دون تحقق قد تؤدي الهلوسة إلى أخطاء ومعلومات مضللة وقرارات غير مناسبة، خصوصًا في السياقات التعليمية أو المهنية أو القطاع العام. لهذا السبب تؤكد الممارسات الدولية على ضرورة الإشراف البشري والتحقق من المصادر الموثوقة والاستخدام المسؤول لمخرجات الذكاء الاصطناعي التوليدي بدلاً من اعتبارها مصادر موثوقة أو نهائية.

يجـــب القيـــام بمـا يلي:

١. اعتبار جميع مخرجات الذكاء الاصطناعي التوليدي اقتراحات احتمالية، وليست حقائق مؤكدة أو مصادر موثوقة.

٢. التحقـــق من المعلـــومات الحيـــوية باستخـــدام مصـــادر مـــوثوقة ومستقلة قبـــل الاعتمـــاد على المحتـــوى المـــولَّد أو مشـــاركته.

٣. تجنـــب استخـــدام الذكـــاء الاصطناعي التوليدي باعتباره مصـــدرًا وحيـــدًا للمعلـــومات الواقعية أو القانونية أو الطبية أو التقنية أو الأكاديمية.

٤. طلـــب الاستشهـــادات أو المصـــادر عنـــد الاقتضـــاء والتـــأكد بشكـــل مستقل من دقتها وملاءمتها.

٥. الحـــذر عنـــد تقديم الذكـــاء الاصطناعي لمخـــرجات متيقنـــة أو مفصلـــة للغـــاية دون أدلـــة واضحة.

٦. استخـــدام مـــدخـلات أو تعليمـــات واضحـــة ومحـــددة وتقـــديم سيـــاق ذي صلـة لتقليل الغموض الـــذي قد يـــزيد من خطـــر الهلـــوسة.

٧. تقسيـــم المهـــام المعقـــدة إلى خطـــوات أصغـــر وتحقـــق من المخـــرجات تـــدريجيًا بـــدلاً من الاعتـــماد على استجـــابة واحـــدة.

٨. اختبـــــار ومـــراجعة الأكـــواد أو الحســـابات أو المخـــرجات الفنية المـــولدة قبـل الاستخدام أو النشر.

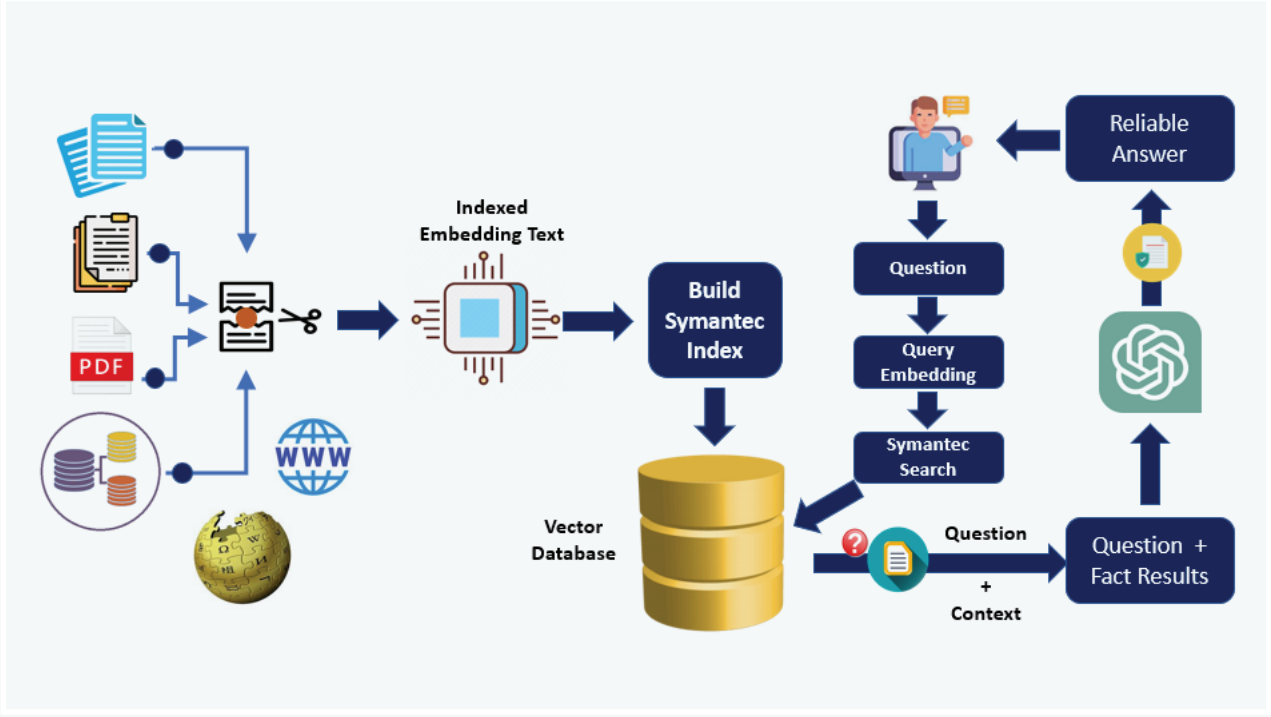
٩. تطبيق الحكـــم البشـــري والخبـــرة المتخصصـة لتقييم مـــدى معقـــولية واتساق النتـــائج المولدة.

١٠. تجنب الاعتمـــاد المفـــرط على الذكاء الاصطناعي التوليدي في السياقات ذات الأثر الكبير أو المخاطر العـــالية دون مـــراجعة ومـــوافقة بشـــرية.

١١. عنـد استخدام الذكاء الاصطناعي التوليدي مع **البيانات الخاصة أو الملكية أو المؤسسية** يجب استخدام **تقنيات الاسترجاع القائمة على التمثيلات المضمنة** (مثل **التوليد المعزز بالاسترجاع**) لتأصيل مخرجات النموذج في مصادر داخلية موثوقة وتقليل خطر الهلوسة.

١٢. تشجيـع النمـوذج على تـوليد إجـابات بـديلة أو طـرق حـل مختلفة، ومقـارنتها من خـلال تحـديد المـزايا والقيـود والافتـراضات الأساسية لكـل منهـا.

١٣. دعـــم المطـــالبات الرئيسية بمـــراجع موثوقة وموثقة، وإعطـــاء الأولـــوية للمصادر الأولية أو الراسخة، مع الـــتأكد من إمكـــانية تتبع الاستشهادات والتحقـــق منها.



الشكل ٨: استخدم تقنيات الاسترجاع القائمة على التمثيلات المضمنة

المـــوثـوقية والســـلامة

لتحقيق مخرجات موثوقة وآمنة عند استخدام الذكاء الاصطناعي التوليدي يجب على المستخدمين إدراك أن المخرجات التي يولدها النظام احتمالية بطبيعتها وقد تحتوي على أخطاء أو محتوى غير مقصودة. تتطلب الموثوقية مراجعة دقيقة والتحقق من صحة ومطابقة المخرجات قبل الاعتماد عليها أو مشاركتها. وتُعزّز السلامة عند استخدام الذكاء الاصطناعي التوليدي ضمن مستويات محددة وتجنب التطبيقات عالية المخاطر أو الحساسة دون إشراف مناسب. ويبقى الحكم البشري والمساءلة في صلب تفسير وتطبيق النتائج المولّدة بالذكاء الاصطناعي. تساعد الحدود الواضحة للاستخدام والشفافية والوعي الأخلاقي في تقليل سوء الاستخدام والأضرار غير المقصودة. تضمن هذه الممارسات مجتمعة أن يدعم الذكاء الاصطناعي التوليدي اتخاذ القرارات المستنيرة مع الحفاظ على السلامة والثقة.

يجـــب القيـــام بمـا يلي:

١. تحديد دور النظام وأسلوب التواصل (الشخصية الافتراضية) لضمان أن يكون الأسلوب والعمق ومستوى الخبرة مناسبًا للمهمة والسياق.

٢. تحديد شخصية الجمهور المستهدف وتكييف الشروحات والمصطلحات ومستوى التفاصيل بما يتوافق مع خلفيتهم واحتياجاتهم ومدى إلمامهم بالموضوع.

٣. تحديد **غرض وسياق** المهمة بوضوح قبل استخدام الذكاء الاصطناعي التوليدي، بما في ذلك نوع المخرجات المطلوبة وكيفية استخدامها.

٤. توضيح الاستفسارات الغامضة أو غير الواضحة من خلال طرح أسئلة متابعة دقيقة لفهم أهداف المستخدم والقيود والنتائج المتوقعة. إعادة صياغة الأسئلة المعقدة إلى عناصر أوضح وأكثر تركيزًا لتحسين دقة الرد وتقليل سوء الفهم.

٥. استخدام **مدخلات أو تعليمات واضحة ومحددة وخالية من الغموض** لتقليل الأخطاء أو المخرجات غير الدقيقة أو غير الملائمة.

٦. **التحقق من جميع المخرجات التي يولدها الذكاء الاصطناعي**، خصوصًا الحقائق والبيانات والبرمجيات والنصائح التقنية باستخدام مصادر موثوقة وموثقة.

٧. التعامل مع المحتوى المولد بواسطة الذكاء الاصطناعي على أنه **احتمالي وقد يكون غير دقيق**، وليس باعتباره معلومة مؤكدة أو رسمية.

٨. تطبيق **الحكم البشري والتفكير النقدي** لتقييم الدقة والملاءمة والتحيز والكمال قبل الاعتماد على المخرجات أو مشاركتها.

٩. تجنب استخدام الذكاء الاصطناعي التوليدي باعتباره **المصدر الوحيد للمعلومات** في القرارات التي قد تؤثر على السلامة أو الحقوق أو التقييم الأكاديمي أو الخدمات أو السمعة.

١٠. اختبار ومراجعة الأكواد أو الحسابات أو التوصيات التي يولدها النظام قبل نشرها أو اعتمادها.

١١. الانتباه إلى **الهلوسة** أو المعلومات القديمة أو لغة الثقة المفرطة التي قد تخفي عدم اليقين أو الأخطاء.

١٢. استخدام أدوات الذكاء الاصطناعي التي توفر **الشفافية بشأن القيود واستخدام البيانات والأغراض المقصودة** حيثما توفرت.

١٣. الرجوع إلى **خبير أو مشرف بشري** عند غموض المخرجات أو وجود مخاطر عالية أو احتمالية وجود عواقب كبيرة.

دقـــة النتـــائج

للحصول على نتائج دقيقة من أنظمة الذكاء الاصطناعي التوليدي يجب على المستخدمين تقديم مدخلات واضحة ومحددة ومنظمة تعكس الهدف المرجو أو السؤال المطروح. ويجب مراجعة المخرجات التي يولدها الذكاء الاصطناعي بشكل نقدي والتحقق منها، لا سيما عند استخدامها لأغراض أكاديمية أو تقنية أو مهنية. يجب على المستخدمين التحقق من المعلومات المهمة بمقارنة المخرجات مع مصادر موثوقة ومعتمدة وعدم الاعتماد على إجابة واحدة فقط من النظام. ونظرًا لأن أنظمة الذكاء الاصطناعي التوليدي قد تنتج محتوى ناقصًا أو غير صحيح، يظل الحكم البشري ضروريًا لتقييم الدقة والملاءمة. ويمكن لتحسين صياغة الأسئلة وطرح استفسارات متابعة أن يعزز جودة المخرجات، لكنه لا يغني عن عملية التحقق. وفي النهاية تقع المسؤولية على المستخدم بشأن صحة النتائج وطريقة استخدامها.

يجـــب القيـــام بمـا يلي:

١. تحديد السؤال أو المهمة بوضوح وتقديم مدخلات محددة ومنظمة بدلًا من التعليمات الغامضة.

٢. تـــوفير سياق كـــافٍ وحـــدود ومعلـومات خلفية لتوجيه الذكـــاء الاصطناعي التـــوليدي نحو مخرجات دقيقة وذات صلة.

٣. تقسيم المهام المعقدة إلى خطوات أصغر يمكن إدارتها بدلًا من طلب كل شيء في مدخل واحد.

٤. التحقق من المعلومات الناتجة عن الذكاء الاصطناعي بمقارنة النتائج مع مصادر موثوقة ومعتمدة، خصوصًا فيما يتعلق بالمحتوى الواقعي أو التقني أو الأكاديمي.

٥. تطبيق التفكير النقدي والحكم البشري لمراجعة وتحرير وتصحيح المخرجات قبل الاعتماد عليها أو تقديمها.

٦. اختبار والتحقق من صحة الأكواد أو الحسابات أو التوصيات التقنية التي يولدها الذكاء الاصطناعي عمليًا بدلًا من افتراض صحتها.

٧. تجنب الاعتماد على الذكاء الاصطناعي للحصول على معلومات تتطلب دقة زمنية أو بيانات حديثة ما لم يتم التحقق من ذلك خارجيًا.

٨. الانتباه إلى أن الذكاء الاصطناعي التوليدي قد يولد إجابات واثقة لكنها خاطئة، ويجب اعتباره أداة مساعدة وليست مصدرًا موثوقًا بشكل مطلق.

حمـــــاية الخصـــوصية

لحماية الخصوصية عند استخدام الذكاء الاصطناعي التوليدي يجب على المستخدمين توخي الحذر بشأن المعلومات التي يشاركونها مع أدوات الذكاء الاصطناعي. يجب عدم مشاركة البيانات الشخصية أو الحساسة أو السرية ما لم يكن استخدامها مصرحًا به بوضوح وضروريًا. يجب أن يكون المستخدمون على علم بأن المدخلات والمخرجات قد يتم تخزينها أو تسجيلها من قبل مقدمي الخدمة، مما قد يشكل مخاطر على الخصوصية في حال الإفراط في مشاركة المعلومات. ينبغي مراجعة المحتوى الذي يولده الذكاء الاصطناعي لضمان عدم كشفه عن تفاصيل خاصة بالأفراد أو المؤسسات بشكل غير مقصود. حيثما أمكن يجب تفعيل إعدادات الخصوصية وخيارات تقليل البيانات. ويظل الحفاظ على الخصوصية متوقفًا على الاستخدام المسؤول والمستنير والمراقبة البشرية المستمرة.

يجـــب القيـــام بمـا يلي:

١. تجنب إدخال أو مشاركة المعلومات الشخصية أو الحساسة (مثل أرقام الهوية الوطنية أو العناوين أو أرقام الهواتف أو كلمات المرور أو الصور أو الملفات الصوتية أو مقاطع الفيديو أو البيانات الصحية أو البيانات المالية) مع أدوات الذكاء الاصطناعي التوليدي إلا إذا كانت مصرحًا بها بوضوح وضرورية.

٢. حماية خصوصية الآخرين بعدم رفع أو توليد محتوى يتضمن بيانات شخصية عن زملاء الدراسة أو الزملاء في العمل أو أطراف ثالثة دون موافقتهم.

٣. افتراض أن المدخلات والمخرجات قد يتم تخزينها أو تسجيلها من قبل مزودي الخدمة واستخدام الذكاء الاصطناعي التوليدي وفق ذلك.

٤. استخدام بيانات مجهولة الهوية أو تخيلية عند ممارسة الاختبارات أو التعلم باستخدام أدوات الذكاء الاصطناعي التوليدي.

٥. الالتزام بالقوانين والسياسات المؤسسية لحماية البيانات، بما في ذلك السياسات المتعلقة بالبيانات الشخصية والسرية والاستخدام المقبول.

٦. تجنب مشاركة المعلومات السرية أو المقيدة، بما في ذلك المستندات الداخلية أو مواد الامتحانات أو الأكواد المحمية بحقوق ملكية أو السجلات غير العامة.

٧. مراجعة إعدادات الخصوصية وشروط الاستخدام لأدوات الذكاء الاصطناعي قبل استخدامها، خاصة عند إنشاء حسابات.

٨. الإبلاغ عن أي مخاوف تتعلق بالخصوصية أو استخدام البيانات بشكل غير مصرح به للجهة أو المؤسسة المختصة.

٩. مراجعة اتفاقية مستوى الخدمة وسياسة الخصوصية دائمًا المقدمة من مقدم خدمة نموذج الذكاء الاصطناعي قبل استخدام الخدمة.[١٠]

الشفـــافية وقـــابلية التفسير

الشفافية وقابلية التفسير تعدان أساسيتين للاستخدام المسؤول للذكاء الاصطناعي التوليدي، إذ تمكّنان المستخدمين وأصحاب المصلحة من فهم متى وكيف تؤثر أنظمة الذكاء الاصطناعي على النتائج. يجب أن تشير أدوات الذكاء الاصطناعي التوليدي بوضوح إلى متى يكون المحتوى مُولّدًا بواسطة الذكاء الاصطناعي أو بمساعدته، وأن توفر معلومات يسهل الوصول إليها حول هدفها وحدودها والسلوك المتوقع لها. النتائج القابلة للتفسير تدعم اتخاذ القرار المستنير من خلال تمكين المستخدمين من تقييم موثوقية ومناسبة مخرجات الذكاء الاصطناعي، بدلًا من اعتبارها معلومات موثوقة أو نهائية. عندما يُستخدم الذكاء الاصطناعي التوليدي في سياقات عالية التأثير أو مواجهة للجمهور، يجب توفير وضوح إضافي لتفسير كيفية إنتاج ومراجعة المخرجات. كما تدعم الشفافية المساءلة من خلال إمكانية تتبع القرارات والاعتراض على النتائج وتصحيح الأخطاء. تساعد الشفافية وقابلية التفسير مجتمعتين في بناء الثقة وتقليل مخاطر سوء الاستخدام أو الاعتماد المفرط أو الضرر غير المقصود.

يجـــب القيـــام بمـا يلي:

١. **إبلاغ المستخدمين والجمهور** بوضوح عندما يكون المحتوى أو المخرجات أو التوصيات مولدة بواسطة الذكاء الاصطناعي أو ساعد فيها الذكاء الاصطناعي بشكل كبير.

٢. فهم وتوصيل **الهدف المقصود والقيود والاستخدام المناسب** لنظام الذكاء الاصطناعي التوليدي المستخدم.

١٠. نموذج لسياسة الخصوصية: https://privacy.openai.com/policies

٣. تجنب تقديم المخرجات المولدة بواسطة الذكاء الاصطناعي على أنها موضوعية تمامًا أو موثوقة أو من إنتاج بشري دون الإفصاح.

٤. مراجعة وتقييم النتائج المولدة بواسطة الذكاء الاصطناعي باستخدام **الحكم البشري**، خصوصًا عند تأثيرها على القرارات أو نتائج التعلم أو الفهم العام.

٥. القدرة على **التفسير، بلغة بسيطة**، لكيفية إسهام الذكاء الاصطناعي التوليدي في المخرجات النهائية، بما في ذلك دوره في التوليد أو التحرير أو اقتراح المحتوى.

٦. الحفاظ على **وثائق أو سجلات** بشأن كيفية استخدام أدوات الذكاء الاصطناعي التوليدي في المهام أو المشاريع أو العمليات المؤسسية عند الاقتضاء.

٧. التحقق، وعند الضرورة **مراجعة المعلومات** التي ينتجها الذكاء الاصطناعي مقارنة بمصادر موثوقة ومعتمدة.

٨. استخدام أنظمة الذكاء الاصطناعي التوليدي التي **توفر معلومات واضحة عن قدرات النموذج والقيود المعروفة**، بما في ذلك المخاطر مثل الهلوسة أو التحيز.

٩. تجنب الاعتماد على مخرجات الذكاء الاصطناعي التوليدي في سياقات عالية التأثير إلا إذا كان يمكن تقديم **تفسير ومبرر منطقيين**.

١٠. دعم إجراءات الشفافية مثل **الوسم بالعلامات التعريفية أو ملاحظات الإفصاح أو آليات تتبع مصدر المحتوى** عند مشاركة المخرجات المولدة بالذكاء الاصطناعي.

النتـــــــائج المحـــــدثة

للحصول على أحدث النتائج عند استخدام الذكاء الاصطناعي التوليدي يجب أن يكون المستخدمون على علم بأن العديد من النماذج تعتمد على بيانات تدريب لها حد معرفي ثابت، وقد لا تعكس أحدث التطورات أو الأحداث أو التحديثات. لذلك يجب عدم افتراض أن مخرجات الذكاء الاصطناعي التوليدي حديثة أو مكتملة تلقائيًا. يُشجع المستخدمون على التحقق من المعلومات الناتجة عن الذكاء الاصطناعي باستخدام مصادر موثوقة ومحدثة، لا سيما في المواضيع الحساسة من الناحية الزمنية مثل التكنولوجيا أو السياسات أو الصحة أو المعلومات العامة. وعندما تكون الدقة والحداثة أمرًا حاسمًا جب استخدام الذكاء الاصطناعي التوليدي كأداة مساعدة، وليس باعتباره المصدر الوحيد للمعلومات. ويساعد تطبيق الحكم البشري والتحقق على تقليل مخاطر الاعتماد على نتائج قديمة أو غير دقيقة.

يجـــب القيـــام بمـا يلي:

١. العلم بأن العديد من أدوات الذكاء الاصطناعي التوليدي لها **حد تاريخ معرفي** وقد لا تعكس أحدث الأحداث أو القوانين أو البيانات أو التطورات التكنولوجية.

٢. **التحقق من المعلومات الحرجة أو الحساسة من الناحية الزمنية** باستخدام مصادر موثوقة ومحدثة، مثل المواقع الرسمية أو المنشورات الحكومية أو المجلات الأكاديمية أو وسائل الإعلام الموثوقة.

٣. استخدام الذكاء الاصطناعي التوليدي **لصياغة ملخصات أو استكشاف الأفكار**، ثم تحديث المخرجات يدويًا بالمعلومات الحديثة الموثوقة.

٤. عند توفرها، استخدام أدوات الذكاء الاصطناعي التوليدي **المتصلة بمصادر بيانات مباشرة أو البحث عبر الإنترنت**، مع التمييز بوضوح بين المحتوى المولد بالذكاء الاصطناعي والمعلومات الموثقة من مصادر خارجية.

٥. مراجعة التواريخ والأرقام والمراجع القانونية وبيانات السياسات قبل الاعتماد على مخرجات الذكاء الاصطناعي أو مشاركتها.

٦. تجنب استخدام الذكاء الاصطناعي التوليدي باعتباره **المصدر الوحيد** للقرارات أو المنشورات أو الإجراءات التي تعتمد على معلومات حديثة أو تتغير بسرعة.

٧. الإشارة بوضوح إلى **أي افتراضات أو عدم يقين** عند استناد المحتوى المولد بواسطة الذكاء الاصطناعي إلى معرفة قديمة.

٨. استشارة خبراء المجال أو الإرشادات الرسمية عندما تكون الدقة والحداثة أمرًا بالغ الأهمية.

الحفاظ على الاستخدام الأخلاقي وتجنب سوء الاستخدام

يعد الحفاظ على الاستخدام الأخلاقي للذكاء الاصطناعي التوليدي وتجنب سوء الاستخدام أمرًا أساسيًا لضمان الثقة والسلامة والنزاهة. يجب استخدام الذكاء الاصطناعي التوليدي لأغراض قانونية ومسؤولة فقط تحترم حقوق الإنسان وكرامته والقيم المجتمعية. ويجب على المستخدمين عدم توظيف هذه الأدوات لخداع الآخرين أو التلاعب بهم أو انتحال شخصياتهم أو نشر معلومات زائفة أو ضارة. كما يجب مراجعة جميع المخرجات الناتجة عن الذكاء الاصطناعي والتحقق منها بحكم بشري قبل الاعتماد عليها أو مشاركتها. ويجب حماية البيانات الشخصية والمعلومات السرية وحقوق الملكية الفكرية وعدم إساءة استخدامها عبر أنظمة الذكاء الاصطناعي. وأخيرًا يجب أن تبقى المسؤولية والمساءلة البشرية محور كل استخدام للذكاء الاصطناعي التوليدي.

يجـــب القيـــام بمـا يلي:

١. استخدام الذكاء الاصطناعي التوليدي فقط لأغراض قانونية ومشروعة وبناءة تتماشى مع القيم الأخلاقية وسياسات المؤسسات.

٢. تجنب استخدام الذكاء الاصطناعي لخداع الآخرين أو التلاعب بهم أو انتحال شخصياتهم أو تضليل الأفراد أو الجمهور، بما في ذلك عبر تقنيات التزييف العميق أو المعلومات الزائفة.

٣. التأكد من مراجعة المخرجات الناتجة عن الذكاء الاصطناعي والتحقق من صحتها قبل استخدامها أو نشرها.

٤. الإفصاح بوضوح عن استخدام الذكاء الاصطناعي عند الحاجة إلى الشفافية، خاصة في السياقات الأكاديمية أو المهنية أو الموجهة للجمهور.

٥. احترام حقوق الإنسان وكرامته والقيم الثقافية وتجنب إنتاج محتوى تمييزي أو يحمل كراهية أو ضار.

٦. حماية الخصوصية والبيانات الشخصية بعدم مشاركة المعلومات الحساسة أو السرية أو الشخصية إلا إذا كانت ذلك مصرحًا بها وضرورية.

٧. الامتناع عن استخدام الذكاء الاصطناعي لتجاوز القواعد أو الضوابط أو التقييمات أو آليات المساءلة.

٨. تجنب الاعتماد المفرط على الذكاء الاصطناعي والتأكد من أن الحكم والمسؤولية البشرية تظل محور عملية صنع القرار.

٩. احترام حقوق الملكية الفكرية وتجنب إنتاج أو نشر محتوى ينتهك حقوق الطبع والنشر أو حقوق الملكية.

١٠. الإبلاغ عن أي استخدام مشبوه أو مخرجات ضارة أو مخاوف أخلاقية متعلقة بالذكاء الاصطناعي للجهة أو المؤسسة المختصة.

٧,٢ إرشـــادات الذكـــاء الاصطناعي الـــوكيل

المبـــدأ الأســـاسي المـــوجـز

عند استخدام **الذكاء الاصطناعي التوليدي** ضمن أنظمة **الذكاء الاصطناعي الوكيل** أو عند استخدام **وكيل الذكاء الاصطناعي** ضمن أدوات الذكاء الاصطناعي التوليدي، يجب أن تكون استقلالية النظام محددة بوضوح دائمًا وقابلة للتدقيق وشفافة، ويجب أن تظل خاضعة للإشراف البشري والمساءلة والمسؤولية الأخلاقية. ويجب أن يبقى الحكم البشري والسيطرة البشرية في صميم جميع القرارات والإجراءات ذات الأثر العالي التي تتخذها هذه الأنظمة.

يجـــب على المستخدمين القيـام بمـا يلي:

- تحديد **أهداف ونطاق وحدود واضحة** لأنظمة الذكاء الاصطناعي الوكيل قبل النشر أو التشغيل.
- ضمان أن تعمل مكونات الذكاء الاصطناعي التوليدي **فقط ضمن مهام وصلاحيات محددة سابقًا**.
- حظر الأهداف المفتوحة أو ذاتية التوسّع دون تفويض بشري صريح.

الإشـــراف البشري وسلطة اتخـــاذ القـــرار

يجـــب على المستخدمين القيـام بمـا يلي:

- الحفاظ على **إشراف فعّال للبشر ضمن الحلقة أو سلطة الإنسان المباشرة**، خصوصًا عند اتخاذ إجراءات ذات أثر كبير.
- طلب **موافقة بشرية** على الإجراءات غير القابلة للاسترجاع أو عالية المخاطر أو المرئية للجهات الخارجية (مثل التواصل العام وتغييرات النظام والوصول إلى البيانات).
- ضمان بقاء **سلطة اتخاذ القرار النهائي** في يد البشر، وليس الوكيل الذكي.

المســـــاءلة والمســـؤولية

يجـــب على المستخدمين القيـام بمـا يلي:

- تحديد المسؤولية بوضوح عن أفعال الوكيل بناءً على من يقوم **بتصميم النظام ونشره وتكوينه وتشغيله**.
- التعامل مع المخرجات والإجراءات الناتجة عن سير العمل الوكيل على أنها **نتائج قيد الإشراف البشري**، وليست قرارات مستقلة للنظام.
- ضمان ألا تتحول المساءلة إلى نظام الذكاء الاصطناعي نفسه.

الشفـــافية وقـــابلية التتبـــع

يجـــب على المستخدمين القيـام بمـا يلي:

- تسجيل قرارات الوكيل والمدخلات أو التعليمات والإجراءات واستدعاءات الأدوات والمخرجات بطريقة قابلة للتتبع والتدقيق.
- توثيق كيفية استخدام الذكاء الاصطناعي التوليدي ضمن الذكاء الاصطناعي الوكيل (مثل التخطيط والاستدلال وتوليد المحتوى).
- توضيح حدود النظام ومستويات استقلاليته للمستخدمين وأصحاب المصلحة.

تقييـــــم المخـــاطر والضـوابط التنـــــاسبية

يجـــب على المستخدمين القيـام بمـا يلي:

- **إجراء** تقييمات للمخاطر والأثر **قبل نشر أنظمة الذكاء الاصطناعي الوكيل**.
- **تطبيق ضوابط أقوى في الحالات التي يكون فيها الوكلاء:**
  - يتفاعلون مع أنظمة العالم الحقيقي،
  - يؤثرون على حقوق الأفراد،
  - يولّدون أو يوزعون محتوى موجهًا للجمهور،
  - يعملون على مدى زمني طويل.
- إعادة تقييم المخاطر بشكل منتظم مع تغيّر القدرات أو السياقات.

السـلامة والأمـن ومنـع سـوء الاستخـدام

يجـــب على المستخدمين القيـام بمـا يلي:

- تنفيذ ضوابط لحماية النظام من **السلوك الخارج عن السيطرة** وهجمات حقن المدخلات أو التعليمات وإساءة استخدام الأدوات والإجراءات غير المقصودة.
- تقييد وصول الوكيل إلى الأنظمة أو البيانات أو واجهات برمجة التطبيقات الحساسة وفق **مبدأ الحد الأدنى من الصلاحيات.**
- تضمين آليات للطوارئ مثل **الإيقاف المؤقت والتجاوز وإيقاف التشغيل الكامل**.

الـــدقة والمـوثوقيـة والتحقـــق

يجـــب على المستخدمين القيـام بمـا يلي:

- الاعتراف بأن مخرجات الذكاء الاصطناعي التوليدي ضمن الوكلاء احتمالية وقد تكون غير صحيحة.
- طلب تنفيذ فحوصات التحقق للخطط أو الأكواد أو القرارات التي يولّدها الذكاء الاصطناعي الوكيل قبل تنفيذها.
- تجنّب نشر أنظمة الذكاء الاصطناعي الوكيل التي تعتمد فقط على الاستدلال الناتج عن الذكاء الاصطناعي وغير المتحقق منه.

حمـــاية البيـــــانات والخصـــوصية

يجـــب على المستخدمين القيـام بمـا يلي:

- تقييد البيانات المتاحة لأنظمة الذكاء الاصطناعي الوكيل إلى **الحد الضروري فقط**.
- منع الوكلاء من الاحتفاظ بالبيانات الشخصية أو الحساسة أو إعادة استخدامها أو مشاركتها دون تفويض.
- ضمان الامتثال لمتطلبات حماية البيانات والسرية المعمول بها.
- **التأكد من أن المدخلات أو التعليمات لا تكشف عن أسرار المؤسسة أو تفاصيل تشغيلية حساسة عن غير قصد**، مع الأخذ في الاعتبار أن المدخلات والمخرجات قد تُسجَّل أو تُخزن أو تُراجع من مقدمي الخدمة.

مثـــال:

X مُدخل غير مناسب (لا تستخدمه):

«قم بتحليل تقرير الحوادث الداخلية المرفق من نظام الأمن السيبراني للوزارة واقترح كيفية إصلاح الثغرات المدرجة، بما في ذلك عناوين بروتوكول الإنترنت للخوادم وأدوار المستخدمين وبيانات الوصول».

- **لماذا يشكل هذا خطرًا:**

هذا المدخل يكشف عن مستندات داخلية سرية وتفاصيل أمنية ومعلومات تشغيلية حساسة قد تُسجَّل أو تُخزن أو تُنكشف من خلال خدمة الذكاء الاصطناعي.

✓ مُدخل مناسب (بديل آمن):

«قدّم أفضل الممارسات العامة لتحسين عمليات الاستجابة لحوادث الأمن السيبراني في المؤسسات العامة دون استخدام أو الإشارة إلى أي أنظمة أو بيانات داخلية حقيقية».

الاستخـدام الأخلاقي والتصميم الـذي يركز على الإنسـان

يجـــب على المستخدمين القيـام بمـا يلي:

- تصميم أنظمة الذكاء الاصطناعي الوكيل **لدعم وتعزيز القدرات البشرية**، وليس لاستبدال الحكم البشري أو المسؤولية البشرية.
- تجنّب سلوكيات الوكيل التي تُضلل أو تخدع أو تؤثر بشكل غير مبرر على المستخدمين.
- أخذ الآثار المجتمعية والأخلاقية في الاعتبار، لا سيما في السياقات العامة أو التعليمية أو الحساسة.

المـراقبة والمـراجعة والتحسين المستمر

يجـــب على المستخدمين القيـام بمـا يلي:

- مراقبة سلوك الوكيل ونتائجه باستمرار أثناء التشغيل في العالم الحقيقي.
- وضع إجراءات للإبلاغ عن الحوادث وتصحيحها والاستفادة منها للتعلم.
- مراجعة تصميم الوكيل وصلاحياته وحالات استخدامه بشكل دوري بما يتوافق مع التوجيهات الدولية المتطورة.

٨,٢ إرشادات التزييف العميق

المبدأ التوجيهي الأساسي الموجز

يجوز استخدام تقنيات التزييف العميق ومزامنة حركة الشفاه لأغراض مشروعة فقط، شريطة ضمان الشفافية والموافقة والمساءلة وتطبيق ضوابط تمنع الخداع والضرر.

الشفافية والإفصاح

- **يجب** على المستخدمين ومقدّمي الخدمات **الإفصاح بوضوح** عند توليد أو تعديل الصوتيات أو مقاطع الفيديو أو الصور أو النصوص باستخدام الذكاء الاصطناعي التوليدي، بما في ذلك تقنيات مزامنة حركة الشفاه.
- يجب أن يتم الإفصاح من خلال **علامات تعريفية مرئية** أو إشعارات أو آليات توثيق المصادر التقنية، مثل البيانات الوصفية أو العلامات المائية.
- يجب حظر إزالة هذا الإفصاح أو إخفائه أو تزويره.

منع الخداع والتضليل

- **يجب عدم استخدام تقنيات التزييف العميق ومزامنة حركة الشفاه لتضليل الجمهور** بحيث يعتقد أن شخصًا حقيقيًا قال أو فعل شيئًا لم يقله أو يفعله.
- يجب إيلاء اهتمام خاص عندما يبدو المحتوى واقعيًا أو ذا طابع رسمي لأن الواقعية تزيد من مخاطر الخداع.

حماية الأفراد والموافقة

- يُحظر أو يُقيَّد بشدة استخدام صورة الشخص أو صوته أو حركات وجهه **دون موافقته**، بما في ذلك مزامنة حركة الشفاه الواقعية.
- يجب توفير حماية خاصة **للقاصرين** والأشخاص المستضعفين.

القيود في السياقات عالية المخاطر

- يجب **تقييد** استخدام تقنيات التزييف العميق ومزامنة حركة الشفاه أو **إخضاعها لضوابط معززة** في السياقات الحساسة، بما في ذلك:
  - الاتصال السياسي والانتخابات
  - المعلومات العامة والرسائل الحكومية
  - الصحافة ووسائل الإعلام الإخبارية
  - البيئات التعليمية والتدريبية
- يجب تطبيق آليات إضافية للشفافية والمراجعة والموافقة في هذه السياقات.

المســـــاءلة والمســـؤولية

- تقع مسؤولية محتوى التزييف العميق ومزامنة حركة الشفاه على من يقوم **بإنشائه أو نشره أو توزيعه**، وليس على نظام الذكاء الاصطناعي نفسه.
- يجب على المؤسسات الحفاظ على **سلاسل مساءلة واضحة** وإجراءات موافقة داخلية للاستخدامات عالية المخاطر.

الاستخدامات المشروعة والمفيدة

- يجوز السماح بالاستخدامات المشروعة لتقنيات التزييف العميق ومزامنة الشفاه، مثل **إنتاج الأفلام والدبلجة وإتاحة المحتوى لذوي الاحتياجات الخاصة والتعليم والفن الساخر والتعبير الفني**، بشرط تحقيق ما يلي:
  - الحفاظ على الشفافية،
  - الحصول على الموافقة عند الاقتضاء،
  - وألا يسبب المحتوى ضررًا أو خداعًا.

تقييم المخـــاطر والضوابط التناسبية

- قبل نشر تقنيات التزييف العميق أو مزامنة حركة الشفاه يجب إجراء **تقييمات للمخاطر والأثر** مع التركيز على الضرر المحتمل والخداع والتأثير المجتمعي.
- يجب أن تكون الضوابط **متناسبة مع مستوى واقعية** المحتوى **وحجمه ومدى وصوله إلى الجمهور.**

الـــرصـد والإبـــلاغ والمعـــالجة

- يجب على مقدّمي الخدمات والمنصّات تنفيذ آليات من أجل:
  - رصد سوء الاستخدام،
  - تمكين الإبلاغ السريع عن المحتوى الضار أو المضلل،
  - ودعم **التصحيح أو وضع العلامات التعريفية أو الإزالة في الوقت المناسب** عند حدوث ضرر.

الاستخدام الأخلاقي والذي يركز على الإنسان

- يجب تصميم واستخدام تقنيات التزييف العميق ومزامنة حركة الشفاه **لدعم التعبير المشروع والإبداع**، وليس للتلاعب أو الإكراه أو تقويض الثقة.
- يجب أن تظل كرامة الإنسان واستقلاليته وثقة المجتمع اعتباراتٍ أساسية.

٩,٢ الذكاء الاصطناعي التوليدي في التعليم والبحث العلمي

يمكن استخدام الذكاء الاصطناعي التوليدي في التعليم والبحث العلمي كأداة داعمة لتعزيز أنشطة التعلّم والبحث شريطة الحفاظ على الحكم البشري والأصالة والمسؤولية الأكاديمية. ويمكن أن يساعد الذكاء الاصطناعي التوليدي في مراجعة الأدبيات وتحليل البيانات وتحسين اللغة واستكشاف الأفكار، لكن يجب أن يظل استخدامه شفافًا ومُفصحًا عنه بصورة مناسبة. ويجب مراجعة جميع المخرجات المولدة بواسطة الذكاء الاصطناعي مراجعة نقدية والتحقق منها واعتمادها من الطلاب أو الباحثين قبل استخدامها أو نشرها.

يجب ألا يحلّ الذكاء الاصطناعي التوليدي محلّ التفكير المستقل أو يسيء تمثيل التأليف أو يختلق أو يزوّر البيانات أو يتلاعب بالنتائج أو يتجاوز عمليات التقييم أو التحكيم العلمي أو الموافقات الأخلاقية. كما لا يجوز إنتاج محتوى علمي مضلل أو منتحل أو غير مُتحقَّق منه. وتعكس الممارسات الدولية هذه التوقعات، إذ اعتمدت دور النشر والمجلات الرائدة سياسات صريحة بشأن استخدام الذكاء الاصطناعي التوليدي والإفصاح عنه بما يتماشى مع **إرشادات اليونسكو**.

الاستخـــدام المصـــرح بـــه

- **تحسين اللغة:** يجوز للمؤلفين استخدام أدوات الذكاء الاصطناعي لتحسين وضوح اللغة وقابليتها للقراءة في أوراقهم البحثية. ومع ذلك يجب أن يتم هذا تحت إشراف بشري كامل مع تحمّل المؤلفين المسؤولية الكاملة عن المحتوى.
- **تصميم الأبحاث والمنهجيات:** يمكن استخدام أدوات الذكاء الاصطناعي باعتبارها جانبًا من تصميم الأبحاث أو منهجيتها (مثل التصوير المدعوم بالذكاء الاصطناعي). ويجب وصف هذا الاستخدام بطريقة قابلة لإعادة الإنتاج في قسم المنهجية، بما في ذلك تفاصيل أداة الذكاء الاصطناعي المستخدمة.
- **متطلبات الإفصاح:** يجب الإفصاح بشفافية في الورقة البحثية عن أي استخدام للذكاء الاصطناعي التوليدي أو التقنيات المدعومة به في عملية الكتابة، على أن يظهر بيان بذلك في العمل المنشور لإبلاغ القرّاء.

الاستخـــدام غيـــر المصـــرح بـــه

- **إسناد التأليف:** لا يجوز إدراج أدوات الذكاء الاصطناعي بوصفها مؤلفين أو مؤلفين مشاركين لأن التأليف ينطوي على مسؤوليات لا يمكن إسنادها إلا إلى البشر.
- **توليد المحتوى:** يُحظر استخدام الذكاء الاصطناعي لتوليد رؤى علمية أو تعليمية أو طبية أو استخلاص استنتاجات علمية أو تقديم توصيات سريرية.
- **إنشاء الصور أو تعديلها:** لا يجوز استخدام الذكاء الاصطناعي التوليدي أو الأدوات المدعومة به لإنشاء الصور أو تعديلها في الأوراق البحثية المقدّمة إلا إذا كان ذلك جزءًا من تصميم الأبحاث أو منهجيتها كما ذُكر سابقًا.
- **عدم الإفصاح عن استخدام الذكاء الاصطناعي:** يُعدّ عدم الإفصاح عن استخدام أدوات الذكاء الاصطناعي في الورقة البحثية مخالفة لسياسات النشر، وقد يُعتبر خرقًا لأخلاقيات النشر.

١٠,٢ الإفصـــــاح

المبـــــدأ الأســـــاسي

يسهم الإفصاح في تعزيز الثقة والنزاهة، ولكنه لا يبرر سوء الاستخدام ولا يقلل من المسؤولية البشرية.

لماذا يُعد الإفصاح أمرًا ضروريًا؟

يعدّ الإفصاح عن استخدام الذكاء الاصطناعي التوليدي أمرًا ضروريًا للحفاظ على الشفافية والنزاهة الأكاديمية والمساءلة في التعليم والبحث العلمي والعمل المهني. ولأن أنظمة الذكاء الاصطناعي التوليدي يمكن أن تنتج مخرجات سلسة لكنها غير مؤكدة أو مستخلصة، فإن الإفصاح يتيح للمعلمين والمراجعين ودور النشر وأصحاب المصلحة تقييم الأصالة والمسؤولية والموثوقية بشكل صحيح.

تتطلب الممارسات الدولية، كما هو منصوص عليه في سياسات دور النشر الأكاديمية الكبرى والهيئات مثل **لجنة أخلاقيات النشر وإرشادات اليونسكو**، الإفصاح الواضح مع التأكيد أن المسؤولية تظل واقعة على عاتق المؤلف البشري.

متى يكون الإفصاح أمرًا ضروريًا؟

يجب على المستخدمين الإفصاح عن استخدام الذكاء الاصطناعي التوليدي عندما يُستخدم فيما يلي:

- توليد أو تعديل النصوص أو البرمجيات أو الصور أو تحليل بيانات أو الرسوم البيانية بشكل جوهري.
- المساعدة في صياغة المحتوى أو تلخيصه أو ترجمته أو إعادة هيكلته.
- دعم الأنشطة البحثية بما يتجاوز التدقيق اللغوي أو تصحيح الإملاء.
- الإسهام في المخرجات المقدمة للتقييم أو النشر أو الاستخدام العام.

لا يجوز إدراج الذكاء الاصطناعي التوليدي بوصفه مؤلفًا أو مؤلفًا مشاركًا، ولا ينقل الإفصاح المسؤولية من المستخدم البشري.

كيـــف يتـــم الإفصـــاح؟

يجب أن يكون الإفصاح واضحًا وموجزًا ومحددًا، ويشير إلى ما يلي:

- اسم أداة الذكاء الاصطناعي التوليدي المستخدمة.
- الغرض من استخدامها.
- مدى إسهامها.
- تأكيد مراجعة المستخدم والتحقق من المخرجات.

يمكن أن يظهر الإفصاح في قسم الشكر الواجب أو قسم المنهجية أو الحاشية أو الملحق أو بيان مخصص لاستخدام الذكاء الاصطناعي، حسب متطلبات المؤسسة أو الناشر.

نمـــــاذج الإفصــاح عن استخــدام الذكــاء الاصطناعي التوليدي

نموذج – بيان إفصاح موجز عن استخدام الذكاء الاصطناعي التوليدي

الإفصاح عن استخدام الذكاء الاصطناعي التوليدي: تم استخدام [اسم أداة/نظام الذكاء الاصطناعي التوليدي] لغرض [على سبيل المثال تحرير اللغة أو المساعدة في التشفير أو توليد الأفكار]. تمت مراجعة جميع المخرجات والتحقق منها وتحريرها من المؤلف أو المؤلفين الذين يظلون مسؤولين مسؤولية تامة عن المحتوى ودقته ونزاهته.

نموذج – بيان إفصاح عن استخدام الذكاء الاصطناعي التوليدي في المجال الأكاديمي / البحثي (موسع)

استخدام الذكاء الاصطناعي التوليدي: تم استخدام أدوات الذكاء الاصطناعي التوليدي في إعداد هذا العمل. وعلى وجه التحديد تم استخدام [اسم الأداة وإصدارها] للمساعدة في [وصف المهمة، على سبيل المثال تلخيص الأدبيات السابقة أو تحسين وضوح اللغة أو توليد مسودة شيفرة]. لم يقم نظام الذكاء الاصطناعي بتوليد نتائج بحثية أصلية أو تقديم إسهامات فكرية جوهرية. تمت مراجعة جميع المحتويات المدعومة بالذكاء الاصطناعي بشكل نقدي والتحقق منها وتنقيحها بواسطة المؤلف أو المؤلفين الذين يتحملون المسؤولية الكاملة عن المحتوى النهائي.

نموذج – بيان إفصاح عن استخدام الذكاء الاصطناعي التوليدي أداء مهمة طالب

إقرار بأداء مهمة بمساعدة الذكاء الاصطناعي: استخدمتُ [اسم أداة الذكاء الاصطناعي التوليدي] لدعم [المهمة المحددة]. أؤكد أن هذه المهمة المقدمة تعكس فهمي وعملّي الشخصيين، وأن المخرجات المولدة بواسطة الذكاء الاصطناعي قد تمت مراجعتها والتحقق منها، وأظلُ مسؤولًا مسؤولية تامة عن المحتوى المقدم.

الملحـــق (١):
التوجيه الفعّال لمدخلات الذكاء الاصطناعي التوليدي

أنمـــاط تصميـم المدخـلات

فعالية استجابات النماذج الكبيرة للغات الحوارية تتأثر بشكل كبير بوضوح وجودة مدخلات أو تعليمات المستخدم.

هندسة المدخلات: هي القدرة على صياغة المدخلات أو التعليمات واستخدامها بمهارة للتفاعل مع الأنظمة والأدوات المختلفة بشكل فعّال.

- **المدخلات:** يمكن وصف المدخلات بأنها تعليمات أو إشارات تُحفّز إجراءً أو استجابة. وفي سياق محادثات الذكاء الاصطناعي عادةً ما تكون المدخلات سطرًا من النص أو سؤالًا يوجّه استجابة نموذج الذكاء الاصطناعي.
- **أنواع المدخلات:** المدخلات النصية هي الأكثر شيوعًا، لكن المدخلات الصوتية تزداد انتشارًا مع تقدم تقنيات التعرف على الصوت. كما تُستخدم المدخلات المتعلقة بالصور في بعض التطبيقات، مثل مهام وصف الصور.

قـــــوة هنـــدسة المدخـــلات:

- **أهمية هندسة المدخلات:** يمكن للهندسة الفعّالة للمدخلات أن تعزز الكفاءة والدقة وتجربة المستخدم بشكل كبير في مختلف التطبيقات.
- **تطبيقات في مجالات مختلفة:** تُستخدم هندسة المدخلات في روبوتات المحادثة لخدمة العملاء والأعمال الفنية الرقمية والمساعدين الافتراضيين، مثل Siri وAlexa، وأدوات تحليل البيانات، وكذلك في التطبيقات المتخصصة في الرعاية الصحية والتعليم وغيرها.

نصـائح عـامة لتصميم المدخـــلات

١. **تعليمـات واضحة:** قدّم أوامر دقيقة للذكاء الاصطناعي حول المهمة، مثل «اكتب»، «صنّف»، «لخّص»، «ترجم»، «رتّب»، وغيرها.

٢. **تقديم أمثلة وخطوات:** اعرض نماذج وعمليات خطوة بخطوة لتوجيه فهم الذكاء الاصطناعي وتنفيذ المهمة.

٣. **عملية تكرارية:** اعلم أن صياغة المدخلات أو التعليمات ليست علمًا دقيقًا، وتتطلب التجربة والتحسين. جرّب مدخلات وأساليب ومقاربات مختلفة لتحقيق النتيجة المرجوة.

٤. **التفاعل مع الذكاء الاصطناعي:** انخرط في حوار مع الذكاء الاصطناعي، وقدم الملاحظات واطلب التعديلات لتحسين النتائج.

٥. **دمج الخبرة البشرية:** أدرج معرفتك وأفكارك في المدخلات لتعزيز جودة مخرجات الذكاء الاصطناعي.

٦. **الدقة:** عزّز دقة المدخلات للحصول على استجابات دقيقة وضمان تغطية شاملة للمعلومات ذات الصلة.

٧. **ماذا يجب فعله وماذا لا يجب فعله؟** تجنّب ذكر ما يجب عدم فعله، وركز على صياغة ما يجب فعله.

٨. **الاعتبارات الأخلاقية:** ناقش الجوانب الأخلاقية المتعلقة بهندسة المدخلات، مثل التحيز في نماذج اللغات وتأثير المدخلات المحتمل على سلوك المستخدمين.

٩. **فهم السياق:** صمّم المدخلات مع مراعاة سياق المهمة أو الاستفسار، بما في ذلك الكلمات المفتاحية والفروق اللغوية الدقيقة ذات الصلة.

مثال على المدخلات غير الجيدة والجيدة:

X مدخلات غير جيدة

"اكتب عن إرشادات الذكاء الاصطناعي التوليدي."

✓ مدخلات جيدة ومؤهلة تمامًا

أنت خبير في حوكمة الذكاء الاصطناعي والسياسات الرقمية في القطاع العام.

صِغ إرشادات موجزة (٦٠٠–٨٠٠ كلمة) لأعضاء هيئة التدريس والطلاب حول الاستخدام المسؤول للذكاء الاصطناعي التوليدي في التعليم والبحث العلمي.

يجب أن تشمل الإرشادات ما يلي:

- الإنصاف والتحيز والهلوسة والخصوصية وحماية البيانات والإفصاح عن استخدام الذكاء الاصطناعي والنزاهة الأكاديمية.
- التمييز بوضوح بين الاستخدامات المسموح بها (مثل تحسين اللغة أو استكشاف الأفكار) والاستخدامات غير المسموح بها (مثل تأليف غير مُفصح عنه أو بيانات مختلقة أو إساءة استخدام التزييف العميق).
- استخدام عناوين فرعية ونقاط واضحة عند الاقتضاء.
- استخدام لغة محايدة وغير تقنية مناسبة لجمهور متنوع (طلاب ومحاضرون وموظفو إدارة).
- إنهاء الإرشادات بقائمة تحقق قصيرة (٥–٧ عناصر) يمكن للطالب اتباعها قبل تقديم أي عمل مدعوم بالذكاء الاصطناعي.

أســـــاليب هنـــدسة المدخـــلات

تركز أساليب هندسة المدخلات على صياغة المدخلات وتنظيمها بطريقة توجه النماذج الكبيرة للغات لإنتاج مخرجات دقيقة وملائمة ويمكن التحكم فيها. من الممارسات الأساسية تحديد الهدف والقيود بوضوح وإضافة تفاصيل أو أمثلة لتوضيح السياق وتقليل الغموض وترتيب التعليمات بحيث يتبع النموذج خطوات منطقية في الاستنتاج.

يمكن لأساليب هندسة المدخلات مثل المدخلات قليلة الأمثلة (عرض استجابات نموذجية) ومدخلات سلسلة التفكير (تشجيع الاستدلال خطوة بخطوة) ومدخلات الدور (تعيين شخصية أو خبير للنموذج) أن تحسن جودة الاستجابات بشكل كبير. ويُعد التحسين التكراري، بما في ذلك اختبار المدخلات وتقييمها وتعديلها، ضروريًا، إذ أن جودة المدخلات تؤثر بشكل كبير على أداء المخرجات في مهام مثل الاستدلال والتلخيص والإبداع.

أصبحت هندسة المدخلات **معروفة على نطاق واسع وتُمارس** بشكل متكرر في أعمال الذكاء الاصطناعي، خصوصًا بين المطورين والمحللين والباحثين والمستخدمين المتقدمين. وقد أثبتت فائدتها في تحسين الدقة وتقليل الغموض وتوجيه سلوك النماذج الكبيرة للغات. ومع ذلك، فهي ليست محددة النتائج بالكامل: **فقد تنتج المدخلات نفسها مخرجات مختلفة قليلًا**، وتعتمد النتائج أيضًا على النموذج الأساسي وجودة البيانات وتعقيد المهمة. لذلك يُنظر إليها باعتبارها **مهارة عملية ومنهجية**، وليست حلاً مثاليًا أو شاملًا. وعادةً ما يجمع المتخصصون بين هندسة المدخلات والاختبار والتحقق والمراجعة البشرية، خاصة في المجالات الحساسة.

١. أسلوب المدخلات بدون أمثلة

التعريف

المدخلات بدون أمثلة تعني طلب أداء مهمة من الذكاء الاصطناعي دون تقديم أي مخرجات نموذجية. يعتمد النموذج فقط على وصف المهمة والسياق والقيود التي يحددها المستخدم. وتكون هذه الطريقة أكثر فاعلية عندما تكون التعليمات واضحة ومنظمة وخالية من الغموض.

أمثلة

- ترجم جملة "La vita è bella" إلى الإنجليزية.
- لخّص الفقرة التالية: إدارة البيانات أمرٌ **لا يمكن التفاوض بشأنه** لضمان ذكاء اصطناعي قانوني وآمن وعادل. يؤدي الامتثال الجزئي إلى زيادة المخاطر القانونية والأخلاقية والتشغيلية مباشرة، خاصة للأنظمة متوسطة وعالية التأثير.

٢. أسلوب المدخلات بمثال واحد

التعريف:

المدخلات بمثال واحد تعني طلب أداء مهمة من الذكاء الاصطناعي بعد تقديم مثال واحد يوضح الشكل أو النبرة أو الأسلوب المطلوب للمخرجات. يستخدم النموذج هذا المثال كمرجع لفهم شكل الاستجابة المتوقع بشكل أفضل، مع الاعتماد أساسًا على وصف المهمة والقيود المحددة من المستخدم.

أمثلة:

- السؤال: «ما هي عاصمة الولايات المتحدة؟"
- الإجابة: «عاصمة الولايات المتحدة هي واشنطن العاصمة."
- السؤال: «ما هي عاصمة فرنسا؟"
- الإجابة:

٣. أسلوب المدخلات بعدة أمثلة

التعريف:

المدخلات بعدة أمثلة تعني طلب أداء مهمة من الذكاء الاصطناعي بعد تقديم عدة مدخلات نموذجية (عدة أمثلة) توضح الشكل المطلوب والمسار التعبيري وأسلوب الاستدلال. ومن خلال التعلم من عدة أمثلة يستطيع النموذج تعميم الأنماط بشكل أفضل وإنتاج استجابات أكثر اتساقًا ودقة، خاصة في المهام المعقدة أو المتخصصة.

أمثلة:

المدخلات ١ — المُدخـــــل:

«قد ينتج الذكاء الاصطناعي التوليدي معلومات قديمة، لذا يجب على المستخدمين التحقق من الحقائق الأساسية قبل الاعتماد عليها."

المخرجات (مرجع الأسلوب):

«المعلومات المولدة بواسطة الذكاء الاصطناعي ليست دائمًا محدثة. يجب على المستخدمين التحقق من التفاصيل المهمة من مصادر موثوقة ومحدثة قبل استخدامها أو مشاركتها."

المدخلات ٢ — المُدخـــــل:
«قد تعكس أدوات الذكاء الاصطناعي أحيانًا التحيزات الموجودة في بيانات التدريب الخاصة بها."

المخرجات (مرجع الأسلوب):
«قد تعكس أنظمة الذكاء الاصطناعي التحيزات الاجتماعية أو المستندة إلى البيانات بشكل غير مقصود. من المهم وجود إشراف بشري لضمان استخدام عادل ومسؤول."

المدخلات ٣: اشرح مفهوم الهلوسة في الذكاء الاصطناعي التوليدي باستخدام نفس النبرة المحايدة والتعليمية كما في الأمثلة السابقة.

٤. أسلوب توجيه المدخلات بسلسلة التفكير

التعريف

يقصد بتوجيه المدخلات بسلسة التفكير طلب من نظام الذكاء الاصطناعي إجراء الاستدلال خطوة بخطوة قبل تقديم الإجابة النهائية. تطلب المدخلات صراحةً إظهار مراحل التفكير الوسيطة، مثل تحديد الافتراضات أو تقسيم المشكلة إلى أجزاء أو اتباع تسلسل منطقي من الخطوات. ويساعد ذلك النموذج على إنتاج مخرجات أكثر تنظيمًا ووضوحًا وموثوقية في المهام المعقدة، ولا سيما في مجالات التحليل والتخطيط وحلّ المشكلات.

مثال على المدخلات

المدخلات: حلّل المخاطر الرئيسية لاستخدام الذكاء الاصطناعي التوليدي في تقييمات طلاب الجامعات، واقترح تدابير عملية للتخفيف منها باستخدام تفكير مرحلي خطوة بخطوة.

نسق المخرجات:

- القسم ١: عرض خطوات التفكير (تحديد المخاطر ثم ربط كل خطر بإجراء ضبط مناسب)
- القسم ٢: قائمة نهائية موجزة تتضمن ٥–٧ توصيات للتخفيف في شكل نقاط

القيود:

- اجعل كل خطوة واضحة ومرتبة منطقيًا (من دون فقرات طويلة)
- تجنب المصطلحات التقنية أو القانونية إلا عند الضرورة القصوى
- لا تدرج مخاطر افتراضية أو مبالغ فيها، وركز على المخاطر الواقعية الموثقة فقط.

٥. أسلوب سلاسل التفكير

التعريف

عند تقديمك إجابة، يُرجى شرح المنطق والافتراضات التي تقف وراء اختيارك لأطر العمل البرمجية. وإذا أمكن، استخدم أمثلة محددة أو أدلة مع نماذج برمجية مرافقة لدعم إجابتك عن سبب كون إطار العمل هو الخيار الأفضل للمهمة. بالإضافة إلى ذلك يُرجى تناول أي أوجه غموض أو قيود محتملة في إجابتك من أجل تقديم استجابة أكثر اكتمالًا ودقة.

مثال على المدخلات

١. خطـــوة التأمل

- كلما قمت بتوليد إجابة:

١. اشرح المنطق والافتراضات الرئيسية وراء الاختيار
٢. قدّم مقطعًا برمجيًا قصيرًا واحدًا على الأقل أو مثال استخدام ملموسًا
٣. حدّد أوجه الغموض المحتملة أو المفاضلات أو القيود في التوصية
٤. اذكر بإيجاز إطار عمل بديلًا أو اثنين ولماذا وقع الاختيار عليهما

أوص بإطار عمل الويب الأنسب لتطوير بوابة خدمات عامة آمنة وواسعة النطاق (مثل منصة حكومة إلكترونية)، ثم قدّم تأملًا في مبررات الاختيار.

٦. أسلوب توجيه المدخلات القائم على تحديد الشخصية والجمهور المستهدف

التعريف

يقصد بتوجيه المدخلات القائم على تحديد الشخصية إسناد دور مهني أو هوية محددة لنظام الذكاء الاصطناعي (مثل مستشار سياسات أو محلل أمن سيبراني أو محاضر جامعي) بحيث تتوافق الاستجابات مع النبرة المتوقعة ومستوى الخبرة والمنظور المطلوب.

أما توجيه المدخلات القائم على الجمهور المستهدف فيعني تحديد الفئة المستهدفة من الاستجابة بوضوح (مثل الطلاب أو القيادات العليا أو المطورون)، بحيث تكون اللغة ودرجة التفصيل والأسلوب مناسبة لتلك الفئة. ويساعد الجمع بين الأسلوبين على إنتاج مخرجات أوضح وأكثر صلة بالسياق ووعيًا به.

مثال على المدخلات

المدخلات: اشرح مخاطر إساءة استخدام تقنيات التزييف العميق بلغة مبسطة ومسؤولة وتثقيفية.
الشخصية (دور الذكاء الاصطناعي): مستشار حوكمة الذكاء الاصطناعي في القطاع العام.
الجمهور المستهدف: طلاب السنة الجامعية الأولى ممن لديهم معرفة عامة بالثقافة الرقمية.

٧. أسلوب قائمة التحقق من الحقائق

التعريف

يعني أسلوب قائمة التحقق من الحقائق طلبًا صريحًا من نظام الذكاء الاصطناعي للتحقق من التصريحات أو البيانات الرئيسية بالرجوع إلى مصادر موثوقة وتقديم نتيجة التحقق في نسق قائمة منظمة للتحقق. يُدرج كل تصريح أو بيان مع حالته (مثل دقيق / دقيق جزئيًا / إحصاءات زائفة / مضلل / غير صحيح)، يلي ذلك شرح موجز مع الإشارة إلى مصدر موثوق. يساعد هذا الأسلوب على الحد من المعلومات المضللة، خاصة في سياقات السياسات والتعليم والمحتوى الموجّه للجمهور، إذ تُعد الدقة والشفافية أمرين أساسيين.

مثال على المدخلات

- إذا طرحتُ عليك سؤالًا، فقم بإعداد قائمة بالحقائق الرئيسية، وأدرج هذه القائمة في نهاية الملخص.
- **السؤال:** راجع وتحقق من صحة البيانات المتعلقة بمخاطر الذكاء الاصطناعي التوليدي، واعرض نتيجة التحقق بوضوح.

٨. أسلوب المدقق المعرفي

التعريف

عندما يطرح المستخدم سؤالًا، يُتوقع من النظام توليد ثلاثة أسئلة إضافية تساعد المستخدم على تقديم إجابة أكثر دقة. يجب أن يفترض النظام أن معرفة المستخدم بالموضوع محدودة، وأن يوضح أية مصطلحات مستخدمة ليست من المعرفة العامة. عند إجابة المستخدم عن هذه الأسئلة الثلاثة، يقوم النظام بدمج الإجابات لإنتاج الإجابة النهائية على السؤال الأصلي.

مثال على المدخلات

- عند طرح سؤال عليك،
- قم بتوليد خمسة أسئلة إضافية لفهم سياقي. أدمج إجاباتي لتحديد نقاط ضعفي، واقترح دورة تدريبية مناسبة لي.
- عندما أسأل عن أثر تغيّر المناخ في الزراعة،
- **السؤال:** قم بتوليد ثلاثة أسئلة إضافية لتوضيح جوانب التغييرات التي أسأل عنها. بعد الإجابة عنها، أدمج المعلومات لتقديم استجابة للسؤال الأصلي.

٩. أسلوب الإجابات البديلة

التعريف

يشير **أسلوب الإجابات البديلة** (المعروف أيضًا باسم **أسلوب النُهج البديلة**) في هندسة المدخلات إلى أسلوب منظم يُستخدم للحصول على نموذج من نماذج الذكاء الاصطناعي (مثل النماذج الكبيرة للغات) لتوليد عدة إجابات أو حلول مختلفة لنفس المشكلة أو المهمة، بدلاً من إجابة واحدة فقط. يشجع هذا الأسلوب النموذج على استكشاف **طرق تفكير مختلفة** أو **استراتيجيات متنوعة** لحل السؤال نفسه، مما يساعد على مقارنة الخيارات واختيار الأفضل.

مثال على المدخلات

- لكل مهمة أعطيك إياها، ضع أفضل ٣ نُهج بديلة مع ذكر مزاياها وعيوبها.
- السؤال: كيف يمكنني تأمين ملفي على حاسوبي الشخصي؟

١٠. أسلوب توفير مراجع موثوقة ومعتمدة

التعريف

المصدر الموثوق والمعتمد هو **مصدر معلومات يُعرف على نطاق واسع بأنه موثوق وذو مصداقية ومعتمد في مجاله**، ويمكن الاعتماد عليه لدعم البيانات الواقعية أو الاستنتاجات في البحث أو الكتابة أو اتخاذ القرار. وعادةً ما يأتي هذا من **خبراء أو مؤسسات راسخة أو منشورات محكّمة أو منظمات رسمية أو وسائل إعلام موقرة**، بحيث يمكن التحقق من دقتها وموثوقيتها ويعترف بها الآخرون في نفس المجال.

مثال على المدخلات

- تصرف كخبير في منهجية **آجايل** لتطوير البرمجيات. اشرح كيفية إدارة اجتماع **سكرم** اليومي، بما في ذلك الغرض منه وهيكله والمشاركين وتحديد الوقت وأفضل الممارسات الشائعة.
- **السؤال:** لأي بيانات واقعية استشهد بمصادر موثوقة ومعتمدة، وأدرج الروابط المباشرة للمصادر في إجابتك.

١١. أسلوب تحسين صياغة السؤال

التعريف

تحسين صياغة السؤال (ويُعرف غالبًا باسم **أسلوب تحسين صياغة السؤال**) هو ممارسة تهدف إلى **تحسين أو توضيح السؤال بشكل تكراري**، بحيث يفهم نموذج الذكاء الاصطناعي التوليدي نيتك بشكل أفضل ويستطيع تقديم إجابة أكثر دقة وملاءمة وتركيزًا. وعادةً ما يتضمن ذلك فحص السؤال الأصلي وإعادة صياغته لتقليل الغموض أو إضافة السياق المفقود، ثم استخدام النسخة المحسّنة لتوجيه استجابة الذكاء الاصطناعي.

مثال على المدخلات:

- كلما طرحتُ سؤالًا عن تطوير البرمجيات، اقترح نسخة أفضل من السؤال تأخذ في الاعتبار المزايا والعيوب لكل خيار، مثل الأداء والأمان وسهولة التطوير.
- **السؤال:** أحتاج إلى تطوير آلية لتحديث ملف العميل بعد أن يضع طلبًا على الموقع الإلكتروني، أي النهجين التاليين أنسب؟ إطلاق مشغّل قاعدة بيانات أم تنفيذ عبارة تحديث أخرى لغة الاستعلامات الهيكلية لتعمل فورًا بعد أن يقدم العميل طلبه؟

١٢. أسلوب التفاعل العكسي

التعريف

أسلوب التفاعل العكسي هو أسلوب تفاعلي منظم **يقود نموذج الذكاء الاصطناعي من خلاله الحوار عن طريق طرح أسئلة على المستخدم** لجمع المعلومات اللازمة، بدلاً من النهج التقليدي الذي يطرح فيه المستخدم الأسئلة ويجيب النموذج. يساعد هذا «العكس» في الأدوار النموذج على اكتشاف السياق أو التفاصيل المفقودة التي يحتاجها لأداء المهمة بشكل أكثر فعالية.

مثال على المدخلات

- من الآن فصاعدًا أود أن تطرح عليّ أسئلة لنشر تطبيق .NET على خدمات معلومات الإنترنت في خدمات أمازون ويب. عندما تحصل على معلومات كافية لنشر التطبيق، أنشئ نصًا برمجيًا لأتمتة عملية النشر.
- اطرح الأسئلة عليّ واحدًا تلو الآخر. أفضّلُ أن تكون الأسئلة متعددة الاختيارات.
- استمر في طرح الأسئلة حتى أرتكب أقل من ثلاثة أخطاء متتالية. اطرح عليّ السؤال الأول.
- أود أن تطرح عليّ أسئلة لاختبار معرفتي بالنظام الشمسي.
- يجب أن تطرح الأسئلة حتى أجيب عليها كلها بشكل صحيح.

الملحـــق (٢):
المـــــــــــــراجـــع

١. قانون الاتحاد الأوروبي للذكاء الاصطناعي
of the European Parliament and of the Council of 13 June 2024 1689/Regulation (EU) 2024
laying down harmonised rules on artificial intelligence and amending Regulations (EC) No
and 1139/EU) 2018) ,858/EU) 2018) ,2013/EU) No 168) ,2013/EU) No 167) ,2008/300
Artificial) 1828/and (EU) 2020 797/EU, (EU) 2016/90/and Directives 2014 2144/(EU) 2019
.Intelligence Act)Text with EEA relevance

٢. تقوم منظمة التعاون الاقتصادي والتنمية بتحديث مبادئ الذكاء الاصطناعي لمواكبة التطورات التكنولوجية السريعة
OECD updates AI Principles to stay abreast of rapid technological developments | OECD

٣. OpenAI – أفضل الممارسات في هندسة المدخلات:
https://platform.openai.com/docs/guides/prompt-engineering

٤. Microsoft Learn – أساليب هندسة المدخلات:
https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/prompt-engineering

٥. Stanford HAI – فهم تصميم المدخلات للنماذج الكبيرة للغات:
https://hai.stanford.edu/news/promise-and-pitfalls-prompting-large-language-models

٦. اليونسكو – إرشادات الذكاء الاصطناعي التوليدي في التعليم والبحث العلمي
https://www.unesco.org/en/generative-ai

٧. التطور مع الابتكار: تحديث مبادئ الذكاء الاصطناعي الصادرة عن منظمة التعاون الاقتصادي والتنمية لعام ٢٠٢٤
https://oecd.ai/en/wonk/evolving-with-innovation-the-2024-oecd-ai-principles-update

٨. تحديث مبادئ الذكاء الاصطناعي لمنظمة التعاون الاقتصادي والتنمية لمواكبة المخاطر الجديدة والمتزايدة الناتجة عن الذكاء الاصطناعي للأغراض العامة والتوليدي private-ai.com

٩. إرشادات استخدام الذكاء الاصطناعي التوليدي في الإدارة العامة
https://oecd.ai/en/dashboards/policy-initiatives/guidelines-for-the-use-of-generative-ai-in-the-public-administration-6317?utm_source=chatgpt.com

١٠. نموذج من سياسات الخصوصية: https://privacy.openai.com/policies

١١. تنشر المفوضية المسودة الأولى من مدونة السلوك بشأن تمييز ووضع العلامات التعريفية للمحتوى المولد بواسطة الذكاء الاصطناعي
https://digital-strategy.ec.europa.eu/en/news/commission-publishes-first-draft-code-practice-marking-and-labelling-ai-generated-content

١٢. نظرة عامة على مسودة التدابير المتعلقة بالذكاء الاصطناعي التوليدي:
https://www.chinalawtranslate.com/en/overview-of-draft-measures-on-generative-ai

١٣. الإطار النموذجي لسنغافورة لحوكمة الذكاء الاصطناعي من أجل الذكاء الاصطناعي التوليدي
Singapore's Model AI Governance Framework for Generative AI : Clyde & Co

١٤. أبحاث دقيقة في مجال الذكاء الاصطناعي لتمكين حوكمة متقدمة للذكاء الاصطناعي
https://www.aisi.gov.uk

١٥. الوثائق (الرسمية) لعملية هيروشيما التابعة لقمة السبع – البيان الختامي للقادة.
G7 Summit – Hiroshima Process documents (official) Leaders_Communique_01_en.pdf