الهيئة الاتحادية للموارد البشرية الحكومية
Federal Authority For Government Human Resources

United Arab Emirates

الإرشادات الخاصة باستخدام وتقييم الذكاء الاصطناعي في تقييم المواهب

SHL.

www.fahr.gov.ae

الإرشادات الخاصة باستخدام وتقييم
الذكاء الاصطناعي
في تقييم المواهب

حقوق الطبع © 2023 الهيئة الاتحادية للموارد البشرية الحكومية، وشركة SHL

جميع حقوق الطبع محفوظة.

لا يجوز دون الحصول على إذن خطي من الهيئة الاتحادية للموارد البشرية الحكومية أو شركة SHL، استخدام أي من المواد التي يتضمنها هذا الدليل، أو استنساخها أو نقلها، كلياً أو جزئياً، في أي شكل وبأي وسيلة، سواءً بطريقة إلكترونية أو آلية، بما في ذلك الاستنساخ الفوتوغرافي أو التسجيل أو استخدام أي نظام من نظم تخزين المعلومات واسترجاعها، إلا كما هو منصوص عليه في شروط وأحكام استخدام الأدلة الخاصة بالهيئة.

ص.ب 2350، أبو ظبي، الإمـــارات العربيـــة المتحـــدة هاتف 4036000 2 971+ فاكس 6266767 2 971+

ص.ب 5002، دبـــــي، الإمـــارات العربيـــة المتحـــدة هاتف 2319000 4 971+ فاكس 2959888 4 971+

المحتويات

نبذة عن SHL	6
شكر وعرفان	7
ملخص تنفيذي	8
استراتيجية الموارد البشرية للذكاء الاصطناعي	15
ما هو تقييم المواهب؟	16
ما هو الذكاء الاصطناعي؟	17
كيف يمكن أن يستفيد تقييم المواهب من الذكاء الاصطناعي؟	20
ما هي المخاطر التي ينطوي عليها استخدام الذكاء الاصطناعي في تقييم المواهب؟	22
المخاطر القانونية	22
مخاطر العلاقات العامة	27
مخاطر الفعالية	27
السؤال الأول: إلى أي مدى تعتبر بيانات التدريب مُلائمة؟	34
نوعية البيانات	34
كمية البيانات	35
تمثيل للبيانات	36
خصوصية البيانات وحمايتها	36
تطبيق السؤال الأول على دراسة الحالة الافتراضية	36
ورقة تقييم لتقييم الذكاء الاصطناعي: ملاءمة بيانات التدريب	37
مراجعة نوعية البيانات	39
مراجعة كمية البيانات	40
مراجعة تمثيل البيانات	40
مراجعة خصوصية البيانات وحمايتها	41

المحتويات

السؤال الثاني: كيف تقوم الخوارزمية باتخاذ القرارات؟	42
تطبيق السؤال الثاني على دراسة الحالة الافتراضية	43
ورقة تقييم لتقييم الذكاء الاصطناعي: الشفافية	45
السؤال الثالث: ما مـــدى انحيــاز القرارات؟	46
تطبيق السؤال الثالث على حالة دراسة افتراضية	48
ورقة تقييم لتقييم الذكاء الاصطناعي: التحيز	49
السؤال الرابع: ما مدى صحــة القـــرارات؟	50
التحقق من صحة المحتوى	51
التحقق من صحة البنية	52
التحقق من الصحة ذات الصلة بالمعيار	52
تطبيق السؤال 4 على دراسة الحالة الافتراضية	56
ورقة تقييم لتقييم الذكاء الاصطناعي: التحقق من الصحة	57
السؤال الخامس: إلى أي مدى تكون القرارات نهائيـة؟	58
تطبيق السؤال الخامس على دراسة الحالة الافتراضية	60
ورقة تقييم الذكاء الاصطناعي: الإشراف	60
السؤال السادس: كيف يتم إبلاغ المرشحين؟	61
تطبيق السؤال السادس على دراسة الحالة الافتراضية	62
ورقة تقييم الذكاء الاصطناعي: إبلاغ المرشح	63
بطاقة قياس استخدام الجهات الحكومية للذكاء الاصطناعي لتقييم المواهب	64
نظرة مستقبلية مستقبل الذكاء الاصطناعي في مجــال الموارد البشريـة	69
المراجــع	70
مسرد المصطلحات	71

نبذة عن SHL

SHL شركة عالمية متخصصة في تقنيات الموارد البشرية وعلوم القياس النفسي، حيث تعمل على تطوير الشركات من خلال توظيف طاقات الأفراد والعلوم وأحدث التقنيات.

وتقدم SHL مجموعة من الحلول للمؤسسات، من بينها بيانات القوى العاملة لتوفير القوى البشرية اللازمة للمؤسسات للمساهمة في الاستفادة القصوى من طاقات موظفيها وزيادة الإنتاجية في العمل. كما تزوّد مسؤولي التوظيف وقادة المؤسسات برؤى حول الأفراد على مستوى المؤسسة ككل، وفرق العمل، وعلى المستوى الفردي أيضاً.

مع 45 عاماً من الخبرة في مجال إدارة المواهب، تعد الشركة شريك التقنية الموثوق فيه لأكثر من 10,000 شركة على مستوى العالم، بما ذلك 50% من شركات قائمة Fortune Global 500، و80% من شركات مؤشر FTSE 100.

لمزيد من المعلومات حول شركة SHL ...يمكنكم زيارة موقعها الإلكتروني:
www.shl.com

SHL.

شكر وعرفان

محمـــد فريـــد المدير العام، الشرق الأوسط وأفريقيا

ديفيـــد إدواردز رئيس حلول المواهب، الشرق الأوسط

ديـــالا جـــرار مدير تطويـــر الأعمال، الشرق الأوسط

ســـارة جوتيريز رئيـــس قســـم العلـــوم، SHL

جيمس ميـــدن باحـــث أول سابـــق، SHL

جيف جونسـون باحـــث رئيســـي، SHL

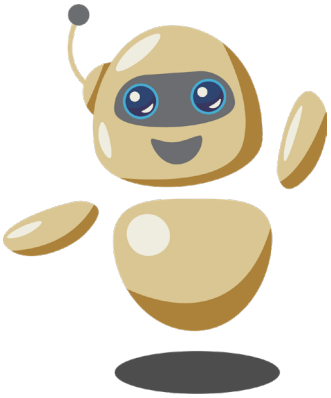
الهيئة الاتحادية للموارد البشرية الحكومية
Federal Authority For Government Human Resources

ملخص تنفيذي

يهدف هذا الدليل بشكل رئيسي إلى تقديم أفضل الممارسات والتوجيهات العامة، وإطار العمل الذي يمكن أن تستخدمه الوزارات والجهات الاتحادية داخل دولة الإمارات العربية المتحدة في تقييم الأنظمة والتقنيات التي تدمج وتستخدم الذكاء الاصطناعي (AI) لتقييم الأفراد؛ من أجل اتخاذ القرارات المتعلقة بالتوظيف. وفي الحقيقة، فإن الذكاء الاصطناعي هو مجال دراسي يهدف بشكل عام إلى تطوير البرامج والآلات الرقمية التي يمكنها عرض بعض الخصائص التي تشبه الذكاء، أو الحكم على مستوى الإنسان. ويتضمن التعلم الآلي (ML)، وهو حقل فرعي من الذكاء الاصطناعي، استخدام الخوارزميات الرياضية التي يتم نشرها وتكييفها لتحقيق أقصى قدر من التنبؤ بأنماط العلاقات داخل مجموعات البيانات، ويعمل التعلم الآلي على تمكين خوارزميات الكمبيوتر من التعلم من مجموعات البيانات، دون أن تكون مبرمجة بشكل خاص.

تقييم المواهب هو مصطلح واسع لعملية أو نشاط تقييم المهارات الأساسية والسمات والخبرات والكفاءات الخاصة بالأفراد (أي المتقدمين للوظائف أو الموظفين)،

واستخدام هذه المعلومات في اتخاذ قرارات مستنيرة تتعلق بالتوظيف، ويمكن إجراء عملية تقييم المواهب إما عندما يتقدم الفرد لوظيفة، وهو ما يعرف باسم تقييمات ما قبل التوظيف، أو بعد تعيين الفرد وذلك لأغراض التطوير، وهو ما يُعرف باسم تقييمات ما بعد التوظيف. وتجدر الإشارة إلى وجود العديد من أساليب وأشكال التقييم المختلفة، بعضها يتطلب قدراً قليلاً من التكنولوجيا، مثل المقابلة الشخصية، والبعض الآخر متطور مثل محاكاة الواقع الافتراضي الواقعي للغاية للأداء الوظيفي.



وعلى غرار العديد من الصناعات الأخرى، تشهد عملية تقييم المواهب فترة من النمو والابتكار من خلال تطبيق تقنيات الذكاء الاصطناعي؛ والتي تساهم في تحسين درجات التقييمات التقليدية، مع وعدٍ بجعل التقييمات أكثر دقة، كما أنه يمكّن من تطوير تقييمات جديدة تعمل على الحصول على معلومات من المرشحين لم تكن متوفرة من قبل، مثل البيانات المرئية والصوتية، ولا شك أن القدرة على تضمين اللغة المنطوقة والمعلومات المرئية في التقييم تفتح إمكانية تطوير محاكاة واقعية للغاية لسيناريوهات العمل من أجل استخدامها في التقييم، كما يمكن أن تساعد أساليب الذكاء الاصطناعي في تبسيط عملية التقييم، وتمكين المرشح من الحصول على تجربة أفضل.

وعلى الرغم من أن تقنية الذكاء الاصطناعي توفر مجموعة متنوعة من الفوائد المحتملة، فإن أي نوع من تقييم المواهب ينطوي على بعض المخاطر، وهذا هو الحال أيضاً بالنسبة للتقييمات التي تستخدم أساليب الذكاء الاصطناعي، فبعض المخاطر المرتبطة بالتقييمات التي تستخدم أساليب الذكاء الاصطناعي تتضمن المخاطر القانونية ومخاطر العلاقات العامة ومخاطر الفعالية.

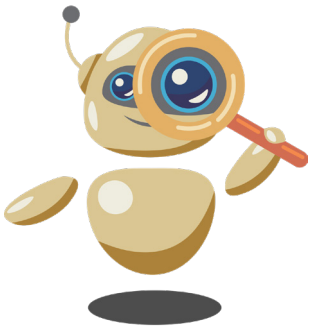
وتندرج المخاطر القانونية المرتبطة باستخدام التقييمات التي تستخدم الذكاء الاصطناعي ضمن فئتين:

(أ) حماية البيانات

(ب) التحيز

وضعت العديد من الدول لوائح قانونية تحدد الاستخدام المسموح به لتطبيق أساليب الذكاء الاصطناعي الخاصة بتحليل البيانات المتعلقة بالأفراد؛ وكذلك اصدرت دولة الإمارات قانون حماية البيانات الشخصية الذي يتسم بمرونة عالية لتمكين أنظمة الذكاء الاصطناعي في الدولة ودعم تبنيها، حيث تتعلق هذه اللوائح بالحماية القانونية التي

تكتنف مسألة جمع واستخدام البيانات الشخصية. ويحدث التحيز في التقييم عندما يمُيز التقييم بشكل غير عادل ضد أحد الأفراد بناءً على واحدة أو أكثر من خصائص الفرد أو خلفيته (على سبيل المثال، العِرق، الإثنية، الدين، البلد الأصلي، الجنس، العمر، حالة الإعاقة). وتتمثل إحدى فوائد أساليب الذكاء الاصطناعي في تقييم المواهب في أنها يمكن أن تؤدي إلى تقليل هذا التحيز عن طريق الحد من تأثير الحكم البشري الذاتي، كما يمكن أن يؤدي دمج الذكاء الاصطناعي في التقييم إلى زيادة التحيز وتوطيده، إذا لم يتم بشكل صحيح وتحت إشراف الخبراء.



ينطوي استخدام أساليب الذكاء الاصطناعي في تقييم المواهب أيضاً على مخاطر تتعلق بتصور الجمهور للمؤسسة إذا لم يتم تصميم طرق الذكاء الاصطناعي وتطويرها بطريقة حذرة ومدروسة. على سبيل المثال، يمكن للمرشحين الذين يتم تقييمهم باستخدام التقييم القائم على الذكاء الاصطناعي، والذين يقررون أنه غير عادل أو تدخلي، مشاركة ردود الفعل السلبية هذه مع شبكتهم وعلى وسائل التواصل الاجتماعي. وقد تضر هذه التصورات والآراء السلبية بسمعة المؤسسة، ويمكن أن يكون لها تأثير سلبي على كمية ونوعية المتقدمين إلى المؤسسة.

مخاطر الفعالية هي مخاطر استخدام تقييم قائم على الذكاء الاصطناعي لا يعمل كما هو متوقع. إن تعقيد مصادر البيانات وتقنيات النمذجة التي تم تمكينها بواسطة أساليب الذكاء الاصطناعي يعني أنه حتى الخبراء الذين يصممون تقييماً للذكاء الاصطناعي قد لا يعرفون بالضبط كيف يعمل وكيف يصل إلى توقعاته و/أو قراراته، ويكمن الخطر هنا في أن التقييم قد يبدو أنه يعمل بشكل جيد في الدراسات التجريبية، لكن الخبراء لا يعرفون كيف يقوم باتخاذ القرارات. وفي هذه الحالة، لا يمكنهم التنبؤ بدقة أو توقع مدى جودة أدائه عند استخدامه في سيناريوهات جديدة، مثل اتخاذ قرارات بشأن بيانات الحياة الواقعية في ظروف جديدة لم ترها الخوارزميات من قبل.

وللمساعدة في تجنب المخاطر وفهم ما يجعل تقييم الذكاء الاصطناعي مُصمماً بشكل جيد، فإن هذه الإرشادات تعمل على توجيه صانعي القرار خلال عملية تقييم التقييم القائم على الذكاء الاصطناعي؛ حيث تتضمن هذه الإرشادات سلسلة من ستة أسئلة عامة ومجموعة من المقاييس المرتبطة لتقييم هذا التقييم. وتقدم الإرشادات أيضاً الأسئلة الرئيسية التالية التي يجب طرحها للتوضيح في حالة عدم توفر المعلومات المطلوبة بسهولة أو عدم توفيرها بواسطة مزود التقييم:

السؤال الأول

إلى أي مدى تعتبر بيانات التدريب ملائمة؟

من الأهمية بمكان استخدام النوعية والكمية الصحيحة من البيانات عند تطوير وتقييم الذكاء الاصطناعي؛ حيث يجب أن تكون البيانات دقيقة وخالية نسبياً من الأخطاء، وذات صلة بالتقييم الجاري تطويره. ويجب أن تكون كمية البيانات كافية لتحليلات البيانات المطلوبة لتطوير الخوارزمية والتحقق من صحتها، كما يتعين أن تكون البيانات ممثلة لمجموعة المتقدمين المعنيين، وجميع المجموعات ذات الصلة (على سبيل المثال، العمر والجنس وحالة الإعاقة)، على أن يتم تخزين أي بيانات تُستخدم لتطوير أو استخدام تقييم الذكاء الاصطناعي بطريقة توفر أكبر قدر ممكن من الخصوصية والحماية.

السؤال الثاني

كيف تقوم الخوارزمية باتخاذ القرارات؟

من الأهمية بمكان أن يعرف أي مستخدم لتقييم يستخدم أساليب الذكاء الاصطناعي كيف يعمل التقييم والذكاء الاصطناعي الأساسي، وهذا المفهوم يعرف باسم الشفافية.

وينبغي الإشارة إلى أن التقييمات التي تعتبر شفافة هي تلك التقييمات التي يمكن من خلالها لمطوري التقييم والمستخدمين شرح كيفية عمل خوارزمية الذكاء الاصطناعي، وكيف وصلت إلى قرار معين (على سبيل المثال، توظيف مرشح أو عدم توظيفه)، ومن المؤكد أن استخدام تقييم شديد الشفافية سوف يؤدي إلى تقليل مخاطر المشكلات القانونية، وتلك المتعلقة بصورة المؤسسة ومخاطر الفعالية.

السؤال الثالث

ما مدى انحياز القرارات؟

في سياق الذكاء الاصطناعي، يُستخدم مصطلح «التحيز» للإشارة إلى خوارزمية الذكاء الاصطناعي التي تؤدي إلى التمييز ضد مجموعة معينة من الأشخاص (على سبيل المثال، عِرق أو فئة عمرية أو جنس معين)، بغض النظر عما إذا كان هذا التمييز عادلاً أو غير عادل. ولتجنب الدمج غير المقصود للتحيز في تقييم الذكاء الاصطناعي، يجب إجراء دراسة شاملة التحيز في كل خطوة من خطوات عملية تطوير التقييم.

السؤال الرابع

ما مدى صحة القرارات؟

الصحة هي الدرجة التي يدعم بها الدليل والنظرية تفسيرات درجات الاختبار. إن التحقق من صحة التقييم هو العملية التي يتم من خلالها إثبات صحة التقييم، كما أن التحقق الشامل من صحة التقييم هو أفضل ممارسة في كل من تقييم المواهب والذكاء الاصطناعي. وسوف نصفُ الطرق الثلاثة الأكثر شيوعًا لجمع أدلة التحقق من الصحة لدعم استخدام التقييم (دليل متعلق بالمحتوى والبناء والمعيار) وكيفية تقييم دراسات التحقق من الصحة.

السؤال الخامس

إلى أي مدى تكون القرارات نهائية؟

يجب تصميم تقييمات الذكاء الاصطناعي لتوفير المعلومات التي يتم استخدامها، إلى جانب المعلومات المتوفرة من مصادر أخرى (عند الضرورة)، من الجانب البشري لاتخاذ قرارات بشأن الموظفين الحاليين أو المحتملين في المؤسسة. وبناءً عليه، لا ينبغي تصميم تقييمات الذكاء الاصطناعي لاتخاذ هذه القرارات دون إشراف بشري.

السؤال السادس

كيف يتم إبلاغ المرشحين؟

يجب إبلاغ المرشحين بموعد استخدام الذكاء الاصطناعي لتسجيل ردودهم على التقييم، مع تقديم شرح كافٍ لكيفية عمل الذكاء الاصطناعي.

لكي يفي الذكاء الاصطناعي بالمأمول منه في تقديم تقييمات أكثر دقة وجاذبية للمواهب، يجب تطوير هذه التقييمات واستخدامها وفقاً لمبادئ وممارسات إرشادية قوية. ومع استمرار تطور اللوائح القانونية في جميع أنحاء العالم، يمكن أن يؤدي الاستخدام غير الملائم للذكاء الاصطناعي في التقييمات إلى انتهاكات قانونية وأخلاقية، مما قد يعيق بشكل كبير تطوير تقييمات الذكاء الاصطناعي. ويمكن استخدام الإرشادات الواردة في هذا المستند للمساعدة في معالجة المشهد سريع التطور والمعقد للذكاء الاصطناعي في تقييم المواهب.

استراتيجية الموارد البشرية للذكـــــاء الاصطنـــــاعي

الرؤيـــة:

أن تكـون الحكومـة الاتحاديـة لدولـة الإمـارات العربيـة المتحـدة واحـدة مـن أكـثر الحكومـات التـي يتـم إدارتهـا بـذكاء عـلى المسـتويين الإقليمـي والعالمـي مـن خـلال الاسـتفادة مـن الـذكاء الاصطناعـي في تبسـيط مهـام المـوارد البشريـة.

الرسـالة:

تضمـين الـذكاء الاصطناعـي في المـوارد البشريـة لخلـق قيمـة مؤسسـية تعتمـد عـلى البيانـات.

الأهـداف:

- إنشاء منتجات وخدمات ذكية وتصميم عمليات موارد بشرية ذكية.
- زيادة إنتاجية الموظفين ورضا العملاء وتحقيق قدر أكبر من الكفاءة.
- توحيد وأتمتة البيانات للمساعدة في عملية اتخاذ القرار الاستراتيجي.
- خلق التآزر وتحقيق الاستفادة القصوى من التكاليف وخفضها وتحسين تخصيص الموارد.

ويمكـن دعـم رؤيـة الهيئـة في بنـاء قـوة عاملـة حكوميـة سـعيدة ومبتكـرة، ورسـالتها في القيـام بذلـك باسـتخدام حلـول مبتكـرة وفعالـة، باسـتخدام الـذكاء الاصطناعـي في تقييـم المواهـب.

ما هو تقييم المواهب؟

تقييم المواهب هو مصطلح واسع لعملية أو نشاط تقييم المهارات الأساسية والسمات والخبرات والكفاءات الخاصة بالأفراد (أي المتقدمين للوظائف أو الموظفين)، واستخدام هذه المعلومات في اتخاذ قرارات مستنيرة تتعلق بالتوظيف، ويمكن إجراء عملية تقييم المواهب إما عندما يتقدم الفرد لوظيفة، وهو ما يعرف باسم تقييمات ما قبل التوظيف، أو بعد تعيين الفرد وذلك لأغراض التطوير، وهو ما يُعرف باسم تقييمات ما بعد التوظيف. وتجدر الإشارة إلى وجود العديد من أساليب وأشكال التقييم المختلفة، بعضها يتطلب قدراً قليلاً من التكنولوجيا، مثل المقابلة الشخصية، والبعض الآخر متطور مثل محاكاة الواقع الافتراضي الواقعي للغاية للأداء الوظيفي.

وفي الحقيقة، يوجد بحث أكاديمي وافٍ حول فعالية تقييم المواهب يعود تاريخه إلى بداية القرن الماضي، ويجمع الباحثون في هذه الأبحاث على أن تقييمات المواهب، عند تطويرها واستخدامها بشكل صحيح، تعد طريقة فاعلة لتعزيز القرارات المتعلقة بالتوظيف، فيما يتعلق بالمرشحين للوظائف الفردية و/أو الموظفين الحاليين.

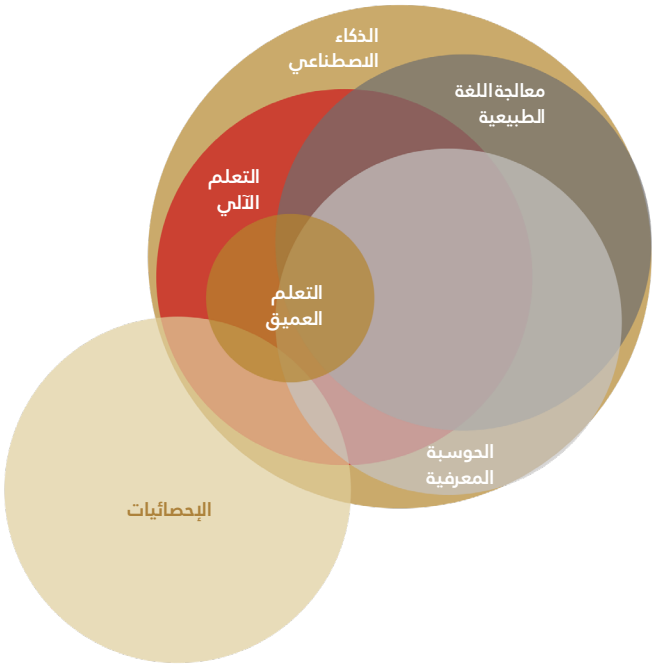
ويعتبر علم القياس النفسي جزءًا لا يتجزأ من العديد من أشكال تقييم المواهب، ويعرف القياس النفسي بأنه الممارسة العلمية لقياس الحالات النفسية والسمات والسلوك.

وتشتمل التطورات الأخيرة في تقييم المواهب على تركيز جديد على تجربة المرشحين الذين يخضعون للتقييمات؛ حيث كان هذا الاتجاه مدفوعاً بمعدلات البطالة المنخفضة منذ الأزمة الاقتصادية التي عصفت بالعالم في العام 2008، كما أنه تزامن مع النمو في مجال تجربة المستخدم (UX)، وزيادة استخدام منصات التكنولوجيا الجديدة (مثل الأجهزة المحمولة) لتنفيذ التقييمات.

وعلى غرار العديد من الصناعات الأخرى، تشهد عملية تقييم المواهب فترة من النمو والابتكار من خلال تطبيق تقنيات الذكاء الاصطناعي؛ والتي تساهم في تحسين درجات التقييمات التقليدية، مع وعدٍ بجعل التقييمات أكثر دقة، كما أنه يمكّن من تطوير تقييمات جديدة تعمل على الحصول على معلومات من المرشحين لم تكن متوفرة من قبل، مثل البيانات المرئية والصوتية.

ما هو الذكاء الاصطناعي؟

الذكاء الاصطناعي (AI) هو مصطلح واسع تطور التعريف الخاص به بمرور الوقت، ويميل المعنى الدقيق إلى الاختلاف عبر مختلف المجالات. وبشكل عام، فإن الذكاء الاصطناعي هو مجال دراسي يهدف بشكل عام إلى تطوير البرامج والآلات الرقمية التي يمكنها عرض بعض الخصائص التي تشبه الذكاء أو الحكم على مستوى الإنسان (Poole & Mackworth ، 1998) ، وهو حقل يشمل العديد من الحقول الفرعية الأخرى (انظر الشكل 1) لتحقيق هذا الغرض العام.

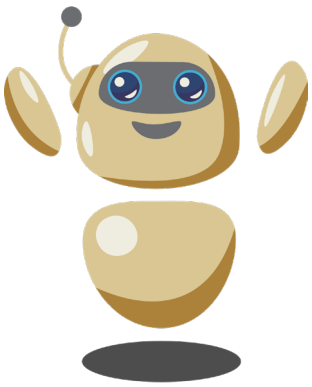


الشكل (1) العديد من الحقول الفرعية التي تقع ضمن الذكاء الاصطناعي.

أدت التطورات المبكرة التي قام بها باحثو الذكاء الاصطناعي إلى خوارزميات قائمة على القواعد صُمّمت لتتبع عمليات محددة بوضوح (مثل لعب لعبة الطاولة)، وقد تم ترميز هذه الخوارزميات يدوياً لاتباع مجموعة طويلة من القواعد، والتي تضمنت بالضرورة مدخلات الخبراء المتخصصين. وكانت نقطة التحول الرئيسية في تاريخ الذكاء الاصطناعي هي تطوير التعلم الآلي واستخدامه. يتضمن التعلم الآلي (ML)، وهو حقل فرعي من الذكاء الاصطناعي، استخدام الخوارزميات الرياضية التي يتم نشرها وتكييفها لتحقيق أقصى قدر من التنبؤ بأنماط العلاقات داخل مجموعات البيانات. ويعمل التعلم الآلي، كما يوحي اسمه، على تمكين خوارزميات الكمبيوتر من التعلم من مجموعات البيانات دون أن تتم برمجتها بشكل خاص. وتكمن أهمية التعلم الآلي في أنه يمكّن برامج الكمبيوتر من التعلم بما يتجاوز مستوى الذكاء البشري. على سبيل المثال، فإن تطبيق أو نظام الذكاء الاصطناعي الذي تم تطويره من خلال التعلم الآلي سوف يتفوق عادةً على تطبيق الذكاء الاصطناعي الذي تم تطويره من خلال الترميز اليدوي، حيث أن خوارزمية التعلم الآلي قادرة على التعلم بشكل متكرر من التعرض الإضافي للبيانات. هذا هو الشكل الأخير لتطوير نظام الذكاء الاصطناعي، والذي يتضمن التعلم الآلي، وهذا هو محور هذه الورقة، ويشار إليه عادةً عند استخدام مصطلح الذكاء الاصطناعي اليوم.

وينبغي الإشارة إلى أن محور فكرة التعلم الآلي هو مفهوم التدريب، أو التدريس، أو برنامج الكمبيوتر. ويتم تدريس البرنامج في البداية على مجموعة بيانات تدريبية، ثم يحاول البرنامج إظهار تعلمه على مجموعة بيانات غير مرئية (بواسطة برنامج الكمبيوتر)، والمعروفة باسم مجموعة الاختبار. ويتم تقييم أداء برنامج الكمبيوتر في مجموعة الاختبار من قبل البشر ويعتبر مقبولاً أو غير مقبول. وإذا كان الأداء مقبولاً، فقد يستمر البرنامج في إجراء مزيد من الاختبارات أو قد يتم نشره (حسب الموقف المحدد). وإذا لم يعمل البرنامج بشكل كافٍ في مجموعة بيانات الاختبار، فسوف يعود لتلقي المزيد من التدريب. وقد أدى استخدام التعلم الآلي، ولا سيما مجموعة فرعية من التعلم الآلي المعروفة باسم التعلم العميق، إلى ظهور برامج الذكاء الاصطناعي التي يمكن أن تتفوق على الخبراء البشريين في مهام محددة بدقة (مثل الشطرنج).

تعد معالجة اللغة الطبيعية (NLP) حقلاً فرعياً آخر من فروع الذكاء الاصطناعي، وتتضمن معالجة الاتصال المنطوق والمكتوب باستخدام لغة بشرية (أي «طبيعية»). وتمكّن معالجة اللغة الطبيعية من تحويل المعنى الذي تنقله الكلمات إلى بيانات يمكن استخدامها بواسطة خوارزميات التعلم الآلي. وتستخدم العديد من أنظمة الذكاء الاصطناعي اليوم هذا المزيج من معالجة اللغة الطبيعية والتعلم الآلي، لتمكين معالجة وفهم وتوليد الكلام بواسطة أنظمة الكمبيوتر، مثل Alexa من Amazon وSiri من Apple.



كيف يمكن أن يستفيد تقييم المواهب من الذكاء الاصطناعي؟

شهد مجال تقييم المواهب فترات متعددة من الابتكار في القرن الماضي، وكان من أبرزها الانتقال من الاختبار الورقي والقلم الرصاص إلى الكمبيوتر، ثم الإنترنت، ثم التقييمات القائمة على الهاتف المحمول. وبالتوازي مع هذه التطورات التكنولوجية، كانت هناك تطورات في مجال الرياضيات وراء القياسات النفسية والذكاء الاصطناعي، مما أتاح نماذج جديدة وأكثر دقة للقياس والتنبؤ بالمتغيرات الرئيسية المتعلقة بالموهبة. يمر تقييم المواهب حالياً بلحظة محورية أخرى من الابتكار مع دمج الأساليب والتكنولوجيا المأخوذة من مجال الذكاء الاصطناعي؛ حيث يؤدي تطبيق هذه التقنيات والأساليب إلى زيادة قدرات التقييمات السيكومترية، وتطوير طرق جديدة قد تكون قادرة على تحسين دقة التنبؤ بما يتجاوز ما كان يُعتقد سابقاً أنه ممكن.

يتضمن الجدول 1 أدناه بعض الفوائد المحتملة التي يمكن تحقيقها من خلال التقييمات التي تستخدم تقنيات وأساليب الذكاء الاصطناعي.

الفائدة	الوصف
التنبؤ بشكل أفضل	يمكن أن تؤدي أساليب الذكاء الاصطناعي، مثل التعلم العميق، إلى دقة أعلى في التنبؤ بنتائج المواهب الرئيسية (مثل الأداء، معدل الدوران).
أقل تحيزاً	عند استخدامها بشكل صحيح، يمكن لأساليب الذكاء الاصطناعي تحديد مقدار التحيز في التقييم وتقليله بسهولة أكبر.
أساليب جديدة	يحتوي مجال الذكاء الاصطناعي على مجموعة واسعة من التقنيات القادرة على نمذجة العلاقات بين المعلومات الصوتية والمرئية.
الرد المنطوق	يعمل الاستخدام المشترك لطرق معالجة اللغة الطبيعية والتعلم الآلي على تمكين المرشحين من الرد بإجاباتهم على التقييم، مما يتيح طريقة طبيعية أكثر بكثير للمرشح للمشاركة في التقييم.
أكثر واقعية	تفتح القدرة على تضمين اللغة المنطوقة والمعلومات المرئية في التقييم إمكانية تطوير محاكاة واقعية للغاية لسيناريوهات العمل لاستخدامها في التقييم.
تجربة أفضل	يمكن أن يساعد استخدام أساليب التعلم الآلي في تبسيط عملية التقييم، وتمكين المرشح من الحصول على تجربة أفضل.

الجدول (1) فوائد الذكاء الاصطناعي في تقييم المواهب.

ما هي المخاطر التي ينطوي عليها استخدام الذكاء الاصطناعي في تقييم المواهب؟

ينطوي تقييم المواهب من أي نوع على مخاطر مرتبطة به، وهذا هو الحال أيضاً بالنسبة للتقييمات التي تستخدم أساليب الذكاء الاصطناعي. لذلك، تنطبق نفس الإرشادات واللوائح العامة التي تسترشد باستخدام التقييمات التقليدية أيضاً على تقييمات الذكاء الاصطناعي (على سبيل المثال، منع التحيز، وضمان الارتباط الوظيفي). ومع ذلك، هناك بعض الاعتبارات الإضافية المتعلقة باستخدام التقييمات التي تستخدم أساليب الذكاء الاصطناعي، والتي يتم مناقشتها أدناه.

المخاطر القانونية

تندرج المخاطر القانونية المرتبطة باستخدام التقييمات التي تستخدم الذكاء الاصطناعي ضمن فئتين:

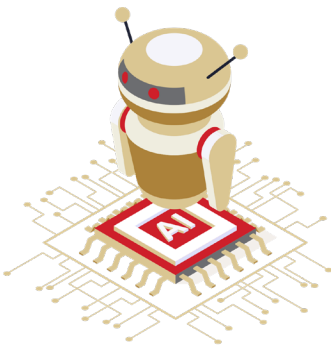
(أ) حماية البيانات.

(ب) التحيز.

تصف الأقسام التالية هذه الفئات من المخاطر القانونية.

حماية البيانات

تتعلق الفئة الأولى من المخاطر القانونية بجمع البيانات واستخدامها. وقد وضعت العديد من الدول لوائح قانونية تحدد الاستخدام المسموح به لتطبيق أساليب الذكاء الاصطناعي الخاصة بتحليل البيانات المتعلقة بالأفراد؛ حيث تتعلق معظم هذه اللوائح باستخدام الذكاء الاصطناعي داخل دولة الإمارات العربية المتحدة والاتحاد الأوروبي والولايات المتحدة، وترتبط بالحماية القانونية التي تكتنف مسألة جمع واستخدام البيانات الشخصية. يتم عرض أمثلة على هذه اللوائح في الجدول 2.



المنطقة	التشريع/ التوجيه	الوصف
دولة الإمارات العربية المتحدة	قانون اتحادي رقم 45 بشأن حماية البيانات الشخصية	ينص قانون اتحادي رقم 45 بشأن حماية البيانات الشخصية على حق المعالجة والمعالجة المؤتمتة، حيث يحق لصاحب البيانات الاعتراض على القرارات التي تصدر عن المعالجة المؤتمتة وتكون لها تبعات قانونية أو تؤثر بشكل جسيم على صاحب البيانات، بما في التنميط. وأن يتعين على المتحكم بالبيانات إدخال العنصر البشري في مراجعة قرارات المعالجة المؤتمتة بناءً على طلب صاحب البيانات.
الاتحاد الأوروبي	اللائحة العامة لحماية البيانات (GDPR)	تتطلب اللائحة العامة لحماية البيانات معالجة البيانات الشخصية لتكون عادلة وقانونية وشفافة. ويتعين على الشركات الكشف عن استخدام الذكاء الاصطناعي للمتقدمين وتقديم معلومات كافية حول كيفية استخدام بياناتهم حتى يتمكن المتقدمون من اتخاذ قرار مستنير بالرفض أو الموافقة (Liem et al., 2018). وتتضمن اللائحة العام لحماية البيانات (GDPR) أيضاً الحق في عدم التعرض لاتخاذ قرارات آلية فقط، مما يعني أن المتقدمين لهم الحق (في ظروف معينة) في الحصول على تدخل بشري، والتعبير عن وجهة نظرهم بشأن القرار، والحصول على حق الاستئناف ضد القرار.
	إطار عمل تدقيق الذكاء الاصطناعي في مفوضية المعلومات (ICO)	في أوروبا، لا توجد لائحة محددة بشأن الذكاء الاصطناعي سارية المفعول في الوقت الحالي، ولكن تم إصدار إرشادات من قبل الجهات التنظيمية المسؤولة عن خصوصية البيانات (وخاصةً مفوضية المعلومات في المملكة المتحدة (ICO) والتي تحتوي على توصيات محددة.

المنطقة	التشريع/ التوجيه	الوصف
الولايات المتحدة	قانون «إجراء المقابلات باستخدام الفيديو بالذكاء الاصطناعي في إلينوي	يتطلب قانون «إجراء المقابلات باستخدام الفيديو بالذكاء الاصطناعي في إلينوي» من أصحاب العمل الحصول على موافقة المتقدمين لاستخدام الذكاء الاصطناعي في عملية التوظيف (Bologna، 2019). بالإضافة إلى ذلك، يتطلب القانون من أصحاب العمل شرح العملية وإتلاف البيانات عند الطلب.
	قانون الإبلاغ عن الائتمان العادل (FCRA)	ينظم قانون الإبلاغ عن الائتمان العادل (FCRA) جمع المعلومات الائتمانية للمستهلك والوصول إلى تقارير الائتمان، كما أنه وثيق الصلة بتقييم المواهب؛ لأنه ينص على أنه لا ينبغي لأي مؤسسة الاحتفاظ بقاعدة بيانات سرية تُستخدم لاتخاذ قرارات بشأن حياة الأشخاص، وأن الأفراد يجب أن يكون لهم الحق في رؤية المعلومات الموجودة في قواعد البيانات هذه والطعن فيها، وأنه يجب أن تنتهي صلاحية المعلومات الموجودة في قاعدة البيانات هذه بعد فترة زمنية معقولة.

الجدول (2) اللوائح المتعلقة باستخدام الذكاء الاصطناعي في دولة الإمارات العربية المتحدة والاتحاد الأوروبي والولايات المتحدة.

التحيز

تتعلق الفئة الثانية من المخاطر القانونية بالتحيز. يحدث التحيز في التقييم عندما يُميز التقييم بشكل غير عادل ضد أحد الأفراد بناءً على واحدة أو أكثر من خصائص الفرد أو خلفيته (على سبيل المثال، العرق، الإثنية، الدين، البلد الأصلي، الجنس، العمر، حالة الإعاقة). وتتمثل إحدى فوائد أساليب الذكاء الاصطناعي في تقييم

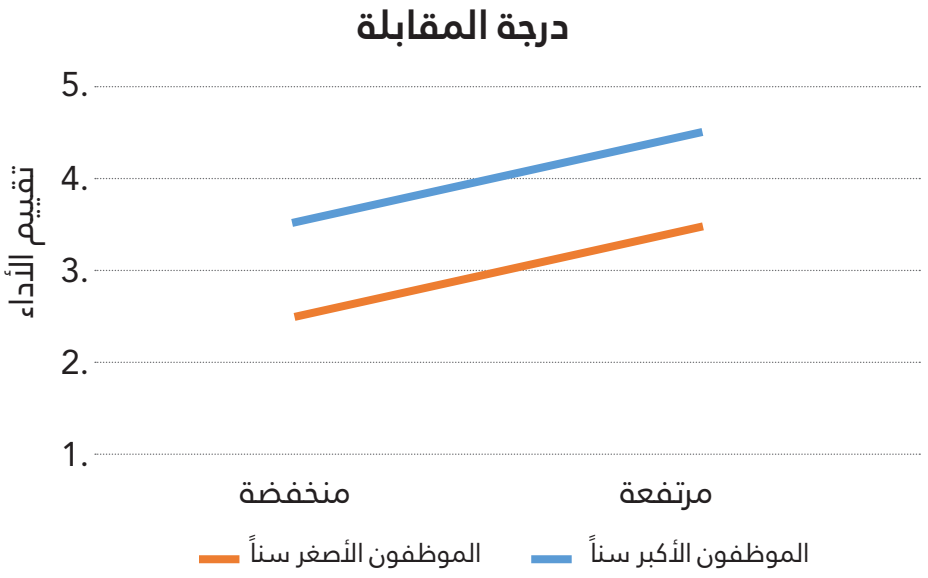
المواهب في أنها يمكن أن تؤدي إلى تقليل هذا التحيز عن طريق الحد من تأثير الحكم البشري الذاتي، كما يمكن أن يؤدي دمج الذكاء الاصطناعي في التقييم إلى زيادة التحيز وتوطيده، إذا لم يتم بشكل صحيح وتحت إشراف الخبراء.

على سبيل المثال، قد يتم إنشاء تقييم مقابلة عبر الفيديو للتنبؤ بمن سيكون صاحب الأداء الجيد من خلال ردود المتقدمين، وقد تتمثل العملية العامة لربط الاستجابات بالأداء في جعل شاغلي الوظائف الحاليين يكملون المقابلة، مع جمع تقييمات الأداء الخاصة بشاغلي الوظائف هؤلاء من المشرفين عليهم. وينبغي الانتباه إلى أن تقييمات الأداء هي أحكام ذاتية يتخذها المشرفون، ومع ذلك، قد تتأثر تقييماتهم بعوامل غير ذات صلة بكيفية ارتباط استجابات المرشحين للأداء الوظيفي اللاحق. على سبيل المثال، قد يميل المشرفون إلى إعطاء درجات أعلى للموظفين الأكبر سناً لأي سبب من الأسباب، مثل (أ) غالباً ما يكون المشرفون أكبر سناً وقد يمنحون درجات أعلى للموظفين الذين يشبهونهم بشكل أكبر، (ب) يميل الموظفون الأكبر سناً إلى امتلاك المزيد من الخبرة، أو (ج) قد يكون المشرفون أكثر دراية بأداء الموظف الموجود في وظيفته لفترة أطول.

يوضح الشكل 2 الشكل الذي قد تبدو عليه خطوط التنبؤ للموظفين الأكبر سناً والأصغر سناً في الموقف أعلاه، مع درجات مقابلة أقل مرتبطة بتقييمات أداء أقل، ودرجات مقابلة أعلى مرتبطة بتقييمات أداء أعلى بنفس المعدل، ولكن تقييمات الأداء تكون أعلى باستمرار بالنسبة لكبار السن من الموظفين. وإذا تم استخدام هذه البيانات لتدريب خوارزمية الذكاء الاصطناعي، فقد تعطي الخوارزمية درجات أعلى للردود التي تحتوي على كلمات أو عبارات غالباً ما يستخدمها كبار السن، ودرجات أقل للردود التي تحتوي على كلمات أو عبارات يقل احتمال استخدامها من قبل كبار السن. وسوف يؤدي هذا إلى شكل خفي من التحيز في الخوارزمية التي تفضل كبار السن على الشباب.

ومن أجل تحقيق الفائدة المحتملة للذكاء الاصطناعي في تقليل التحيز، يجب تصميم تقييمات الذكاء الاصطناعي لتحقيق الإنصاف من البداية؛ حيث يجب أن يكون هذا الاعتبار للعدالة حاضراً في جميع مراحل عملية التطوير، بدلاً من الاعتماد على اختبار واحد للتحيز بعد تطوير التقييم. ولحسن الحظ، مثلما يمكن لتقييم الذكاء الاصطناعي أن «يتعلم» أفضل طريقة للتنبؤ بالنتائج المتعلقة بالعمل، يمكنه أيضًا تعلم كيفية تجنب التحيز، فمن خلال تصميم التقييم ليكون صالحاً (فعالاً) وعادلاً (أخلاقياً) من البداية، يتم تقليل مخاطر التحيز بشكل كبير.

وتجدر الإشارة إلى أن تقييم الذكاء الاصطناعي الذي تم تصميمه وتطويره بنجاح مع مراعاة الإنصاف منذ البداية سوف يتطلب مدخلات الخبراء المتخصصين (SMEs) وإشرافهم في عملية تقييم المواهب؛ حيث سيتمكن هؤلاء الخبراء من تصميم دراسة مناسبة لتطوير الخوارزمية، وكذلك إبلاغ المطورين التكنولوجيين بتقييم الذكاء الاصطناعي بشأن الميزات التي من المحتمل أن (أ) تكون مرتبطة بالوظيفة، و (ب) تشكل خطر التحيز. وبالإضافة إلى مدخلات الخبراء المتخصصين، يعد الاختبار الصارم والمستمر طوال عملية التطوير أمراً هاماً للغاية لمنع التحيز من التسلل إلى تقييم الذكاء الاصطناعي.



الشكل (2) مثال على التحيز بسبب الدرجات المتوقعة المختلفة على أساس عضوية المجموعة.

مخاطر العلاقات العامة

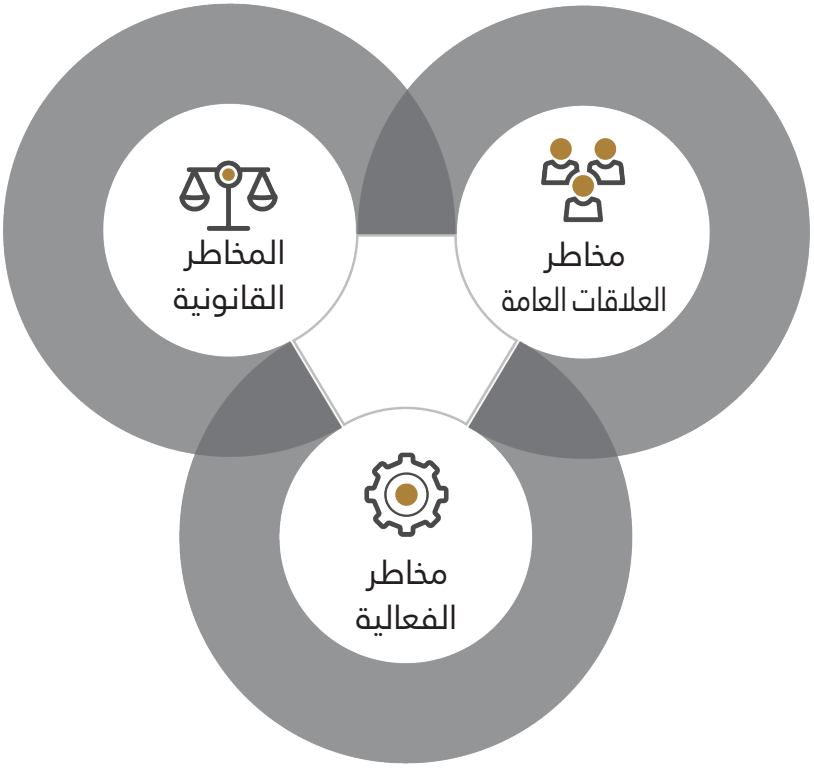
ينطوي استخدام أساليب الذكاء الاصطناعي في تقييم المواهب أيضاً على مخاطر تتعلق بتصور الجمهور للمؤسسة إذا لم يتم تصميم طرق الذكاء الاصطناعي وتطويرها بطريقة حذرة ومدروسة. على سبيل المثال، يمكن للمرشحين الذين يتم تقييمهم باستخدام التقييم القائم على الذكاء الاصطناعي، والذي يقررون أنه غير عادل أو تدخلي، مشاركة ردود الفعل السلبية هذه مع شبكتهم وعلى وسائل التواصل الاجتماعي. وقد تضر هذه التصورات والآراء السلبية بسمعة المؤسسة، ويمكن أن يكون لها تأثير سلبي على كمية ونوعية المتقدمين إلى المؤسسة. ولهذا فإن خطر إلحاق الضرر بالعلاقات العامة يؤكد بشكل أكبر على أهمية تطوير تقييمات الذكاء الاصطناعي وتقديمها بتوجيه من الخبراء المتخصصين في هذا المجال؛ حيث يمكن لخبراء تقييم المواهب التعرف بسهولة أكبر على ميزات التقييم التي يمكن أن تخلق ردود فعل سلبية، والتوصية بالتغييرات التي يمكن إجراؤها لتحسين تجارب المرشح.

مخاطر الفعالية

تتمثل الفئة الأخيرة من فئات المخاطر الثلاث في مخاطر استخدام التقييم القائم على الذكاء الاصطناعي الذي لا يفعل ببساطة ما يدعي أنه يفعله - وهو التقييم غير المجدي. إن تعقيد مصادر البيانات وتقنيات النمذجة التي تم تمكينها بواسطة أساليب الذكاء الاصطناعي يعني أنه حتى الخبراء الذين يصممون تقييماً للذكاء الاصطناعي قد لا يعرفون بالضبط كيف يعمل، وكيف يصل إلى توقعاته و/أو قراراته، ويكمن الخطر هنا في أن التقييم قد يبدو أنه يعمل بشكل جيد في الدراسات التجريبية، لكن الخبراء لا يعرفون كيف يقوم باتخاذ القرارات. وفي هذه الحالة، لا يمكنهم التنبؤ بدقة أو توقع مدى جودة أدائه عند استخدامه في سيناريوهات جديدة، مثل اتخاذ قرارات بشأن بيانات الحياة الواقعية في ظروف جديدة لم تعهدها الخوارزميات من قبل (مثل الأدوار الوظيفية الجديدة، وأنواع مختلفة من إدخال البيانات، والجنسيات المختلفة للمتقدمين، والمستجيبين بأنواع مختلفة من اللهجات).

لذلك، مرةً أخرى، من الضروري أن يشارك الخبراء في كلٍ من الذكاء الاصطناعي وتقييم المواهب في تصميم وتطوير وتنفيذ تقييم الذكاء الاصطناعي؛ حيث إن خبراء تقييم المواهب يكونون على دراية جيدة بالطرق المختلفة للتحقق من صحة الاستنتاجات التي يتم الحصول عليها من التقييم، كما يمكنهم المساعدة في التأكد من أن التقييم يقيس ما يقصد قياسه، ويقيس الخصائص ذات الصلة بالوظيفة، ويعمم قدرته على التنبؤ بالمواقف الجديدة.

ثلاثة مخاطر رئيسية
في استخدام الذكاء الاصطناعي في تقييم المواهب

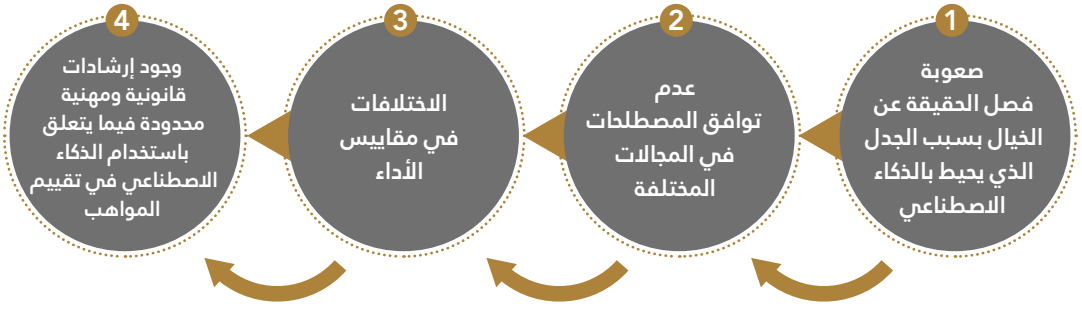


تقييمـــات الذكاء الاصطناعي

ناقشـت الأقسـام السـابقة الحاجـة إلى الخـبرة في تطويـر التقييـمات القائمـة عـلى الـذكاء الاصطناعـي، غـير أن الخـبرة وحدهـا ليسـت كافيـة لتطويـر تقييـمات الـذكاء الاصطناعـي الرائـدة في الصناعـة، فهنـاك حاجـة أيضـاً إلى وجـود طريقـة منظمـة وموحـدة لمراجعـة التقييـمات، حتـى تلـك التـي وضعهـا الخـبراء، لتحديـد مـدى فعاليتهـا وتحديـد مخاطـر التحيـز، وهـذا لـن يتيـح إجـراء مراجعـة شـاملة لتقييـم واحـد فحسـب، بـل سـيوفر أيضـاً معيـاراً يمكـن عـلى أسـاسـه مقارنـة تقييـمات الـذكاء الاصطناعـي المتعـددة أثنـاء عمليـة المراجعـة.

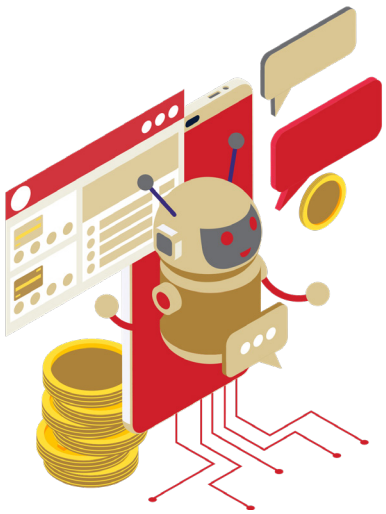
وتجـدر الإشـارة إلى وجـود العديـد مـن التحديـات التـي تواجـه ذلـك الأمـر. أولاً، هنـاك الكثـير مـن الجـدل حـول صناعـات التكنولوجيـا والأعـمال التـي تتعلـق بفوائـد وقـدرات تقنيـات الـذكاء الاصطناعـي، وقـد يكـون مـن الصعـب فصـل الحقيقـة عـن الخيـال في هـذا السـيناريو. ثانيـاً، يعتبـر كلٌ مـن الـذكاء الاصطناعـي وتقييـم المواهـب مجالـين تقنييـن، ولـكل منهـما مصطلحـات يحتمـل أن تكـون مربكـة، وغالبـاً لا تتوافـق هـذه المصطلحـات عـبر مختلـف المجـالات (عـلى سـبيل المثـال، مصطلـح «التحيـز» في الـذكاء الاصطناعـي يعنـي شـيئاً مختلفـاً عـن مصطلـح «التحيـز» في تقييـم المواهـب)، ويمكـن أن يمثـل هـذا الأمـر تحديـات بالنسـبة للمسـتخدمين غـير التقنييـن الذيـن يحاولـون فهـم كيفيـة عمـل التقييـم. ثالثـاً، غالبـاً مـا تُسـتخدم مقاييـس الأداء المختلفـة بنـاءً عـلى طريقـة التقييـم وكيـف تـم تطويرهـا (عـلى سـبيل المثـال، مربـع R، الدقـة مقابـل الحساسـية، منحنـى ROC)، مـما يـؤدي إلى وجـود تحديـات عنـد محاولـة مقارنـة الدقـة التنبؤيـة لتقييمـين أو أكـثر. رابعـاً، الإرشـادات المتعلقـة بأفضـل الممارسـات عنـد اسـتخدام الـذكاء الاصطناعـي لتقييـم المواهـب محـدودة لأن هـذا يعـد تطـوراً جديـداً في هـذا المجـال. وهنـاك إرشـادات قانونيـة ومهنيـة محـدودة فيـما يتعلـق بكيفيـة اسـتخدام الـذكاء الاصطناعـي وعـدم وجـوب اسـتخدامه عنـد تقييـم الأفـراد.

أربعة تحديات في مراجعة تقييمات الذكاء الاصطناعي



ونظراً للتعقيد الذي يكتنف هذا الموقف، تم تطوير هذه المجموعة من الإرشادات للمساعدة في توجيه صانعي القرار خلال عملية تقييم التقييم القائم على الذكاء الاصطناعي؛ حيث تتضمن هذه الإرشادات سلسلة من ستة أسئلة عامة، ومجموعة من المقاييس المرتبطة لتقييم هذا التقييم. وتقدم الإرشادات أيضاً الأسئلة الرئيسية التالية التي يجب طرحها للتوضيح في حالة عدم توفر المعلومات المطلوبة بسهولة أو عدم توفيرها بواسطة مزود التقييم.

تم تطوير أفضل الممارسات والتوصيات لمراجعة التقييمات القائمة على الذكاء الاصطناعي (SHL، 2020). انظر الجدول 3 للحصول على ملخص بشأن هذا الأمر. وقد تم إعداد الإرشادات الواردة في هذا المستند لتتماشى مع أفضل هذه الممارسات. يحتوي الجدول 4 على مخطط بأفضل الممارسات بشأن أسئلة إطار مراجعة تقييم الذكاء الاصطناعي التي نقدمها في بقية هذه الوثيقة.



أفضل الممارسات	الوصف
1. تحديد متطلبات البيانات	مراعاة تقليل البيانات وجودتها وتنوعها وأمنها.
2. إعطاء الأولوية للشفافية	تطوير ذكاء اصطناعي يتسم بالشفافية ويوضح قابلية تفسير النتائج (أي عدم وجود خوارزميات «الصندوق الأسود») وتكون قرارات ومنهجيات أنظمة الذكاء الاصطناعي قابلة للتفسير، إلى الحد الذي تسمح به التكنولوجيا المتاحة.
3. التصميم بطريقة تحقق العدالة	وضع أسس العدل في التقييم من البداية.
4. التحقق بدقة	وضع تقييمات الذكاء الاصطناعي في مستوى عالٍ فيما يتعلق بأدلة الصلاحية.
5. دمج الإشراف البشري	يجب ألا يتخذ أي تقييم للذكاء الاصطناعي قرارات دون إشراف بشري.
6. الكشف عن النية	تقديم إشعار وشرح وطلب الموافقة (عند الضرورة) من المرشحين الذين سيتم تقييمهم بواسطة الذكاء الاصطناعي.

الجدول (3) أفضل ممارسات شركة SHL لاستخدام الذكاء الاصطناعي في تقييم المواهب.

أفضل ممارسات تقييم الذكاء الاصطناعي	أسئلة إطار مراجعة تقييم الذكاء الاصطناعي التابع للهيئة الاتحادية للموارد البشرية الحكومية
1. تحديد متطلبات البيانات	1. إلى أي مدى تعتبر بيانات التدريب مُلائمة؟
2. إعطاء الأولوية للشفافية	2. كيف يقوم الذكاء الاصطناعي باتخاذ القرارات؟
3. التصميم بطريقة تحقق العدالة	3. ما مدى انحياز قرارات الذكاء الاصطناعي؟
4. التحقق بدقة	4. ما مدى دقة أو صحة قرارات الذكاء الاصطناعي؟
5. دمج الإشراف البشري	5. إلى أي مدى تكون قرارات الذكاء الاصطناعي نهائية؟
6. الكشف عن النية	6. ما مقدار المعلومات التي يحصل عليها المرشحون؟

الجدول (4) تحديد إطار مراجعة التقييم التابع لـ «الهيئة الاتحادية للموارد البشرية الحكومية» فيما يتعلق بأفضل ممارسات شركة SHL.

تقدم الأقسام التالية وصفاً للإرشادات، بالإضافة إلى دراسة حالة افتراضية لكيفية استخدام الإرشادات ومقاييس النتائج.

دراسة حالة افتراضية

نظرًا للتغيرات الأخيرة في الاقتصاد المحلي وسوق العمل، بدأت إحدى المؤسسات في تلقي زيادة بمقدار ثلاثة أضعاف في الطلبات المقدمة لشغل وظائف خاصة بالمبتدئين، ولا شك أن هذه الزيادة في الطلبات المقدمة تؤدي إلى إجهاد فريق الموارد البشرية بالمؤسسة، والذين يتعذر عليهم مراجعة كل طلب قبل اتخاذ قرارات بشأن المرشحين الذين يجب أن يدخلوا المرحلة الثانية من عملية الاختيار – وهي إجراء مقابلة عبر الهاتف. وبناءً عليه، قررت إدارة الموارد البشرية الاستثمار في التكنولوجيا التي ستمكنها من مراجعة جميع الطلبات، بهدف تعزيز عملية التوظيف ومخرجاتها.

تكشف مراجعة التقنيات المتاحة لحل هذه الحاجة عن وجود تقييم قائم على الذكاء الاصطناعي يستخدم المعلومات الموجودة في السير الذاتية للتنبؤ بالأداء الوظيفي. وتزعم المواد التسويقية الواردة من المؤسسة التي طورت هذا التقييم أن استخدام التقييم يؤدي إلى زيادة بنسبة 20٪ في تقييمات المدير لأداء الموظف، مع تقليل تكاليف التوظيف.

وبناء عليه، فأنت مسؤول عن تحديد تقنية التقييم لتحقيق الهدف الذي حددته قيادة الموارد البشرية. كيف ستشرع في تحديد ما إذا كان يجب عليك استخدام تقييم السيرة الذاتية بواسطة الذكاء الاصطناعي؟

استخدم إطار العمل الوارد في الأقسام التالية للوصول إلى إجابة.

السؤال الأول:

إلى أي مدى تعتبر بيانات التدريب مُلائمة؟

الاعتبار الأول فيما يتعلق بتطوير التقييم يتعلق بالبيانات التي تم استخدامها أو التي سيتم استخدامها لتطويره، ولذا فإنه من الأهمية بمكان أن يتم استخدام نوعية وكمية البيانات الصحيحة، وهذا سيختلف لكل تقييم. وستكون هناك حاجة إلى بعض الخبرة في هذا المجال عند مراجعة البيانات المستخدمة لتطوير التقييم.

نوعية البيانات

تعد نوعية البيانات من أهم الاعتبارات المتعلقة بتطوير التقييم؛ حيث يمكن اعتبار البيانات التي يستند عليها تقييم الذكاء الاصطناعي هي الأساس في أداء التقييم، وذلك لأن التقييم سيتم تدريبه على هذه البيانات، وبالتالي سيتعلم منها. وفي حالة ما إذا كانت البيانات تفتقر إلى الجودة الكافية، فإن الجودة (أي الدقة التنبؤية) الخاصة بالتقييم الناتج ستكون محدودة.

كيف يمكنك التمييز بين البيانات عالية الجودة والبيانات الرديئة؟ البيانات عالية الجودة دقيقة وخالية نسبياً من الأخطاء وذات صلة بالتقييم الجاري تطويره. على سبيل المثال، إذا تم استخدام معالجة اللغة الطبيعية لتحليل الاستجابات المنطوقة (مثل مقابلة عن طريق الفيديو أو تمرين تقمص الأدوار)، فمن الضروري أن تمثل ترجمة الصوت إلى نصٍ يمثل الردود الأصلية بدقة.

أما إذا نتج عن برنامج الترجمة نسبة عالية نسبياً من الأخطاء، فستعتمد خوارزمية الذكاء الاصطناعي على معلومات خاطئة وقد لا يتم تقييم المستجيبين بشكل عادل.

السؤال الأول:

إلى أي مدى تعتبر بيانات التدريب مُلائمة؟

وسوف تختلف تفاصيل كيفية تقييم نوعية البيانات حسب كل حالة وحسب نوع البيانات، على سبيل المثال، يتم تقييم جودة محتوى السيرة الذاتية بشكل مختلف عن جودة الردود المنطوقة. ويجب على مطوري التقييم تقديم أكبر قدر ممكن من المعلومات حول نوعية البيانات التي تم استخدامها لتطوير تقييم الذكاء الاصطناعي. وإذا لم تكن هذه المعلومات متاحة (كأن تكون متوفرة في دليل فني)، فيجب إرسال استفسار عن نوعية البيانات المستخدمة لتدريب الخوارزمية إلى مطور التقييم.

كمية البيانات

الاعتبار الثاني فيما يتعلق بالبيانات التي يتم تطوير تقييم الذكاء الاصطناعي استناداً عليها هو كمية البيانات؛ حيث تتطلب التقييمات التي تستخدم الذكاء الاصطناعي في كثيرٍ من الأحيان، ولكن ليس دائماً، كميات أكبر بكثير من البيانات المطلوبة للتطوير مقارنةً بالتقييمات التقليدية، وهذا صحيح بشكل خاص عندما يتم تطوير شكل جديد من التقييم لا يحظى بدعم مسبق في المؤلفات العلمية، أو بوجود سابقة تم استخدامها على نطاق واسع في الممارسات الخاصة بصناعة تقييم المواهب. ومن الأمثلة على هذا التقييم هو تطبيق خوارزميات الذكاء الاصطناعي لاستخراج تعابير الوجه أثناء إجراء مقابلة عن طريق الفيديو، واستخدام هذه المعلومات للتنبؤ بالأداء الوظيفي. في هذا السيناريو، ونظراً لوجود بحث مسبق محدود يدعم استخدام تعبيرات الوجه كأساس لقرارات التوظيف، فسوف يحتاج مطور التقييم إلى تقديم دليلٍ مقنع على أن هذه ممارسة توظيف سليمة. ويمكن تحقيق هذا الدليل المقنع من خلال كمية كبيرة كافية من البيانات المستخدمة لتدريب الخوارزمية (على الرغم من أن العوامل الأخرى المذكورة في هذا الدليل تحتاج أيضاً إلى النظر فيها بدقة في هذا السيناريو).

ما هي كمية البيانات المطلوبة؟

سيختلف العدد الدقيق لنقاط البيانات المطلوبة وفقاً لمجموعة متعددة من المعايير (على سبيل المثال، الغرض من التقييم ونوع البيانات وحداثة التقييم)، وهنا يجب تقديم أساس منطقي واضح لعدد نقاط البيانات المستخدمة لتدريب خوارزمية الذكاء الاصطناعي في دليل فني. وإذا لم تكن هذه المعلومات متاحة بسهولة، فيجب إرسال استفسار عن الأساس المنطقي لكمية البيانات المستخدمة لتدريب الخوارزمية إلى مطور التقييم.

السؤال الأول:

إلى أي مدى تعتبر بيانات التدريب مُلائمة؟

تمثيل للبيانات

يجب أن تكون البيانات المستخدمة لتطوير تقييم الذكاء الاصطناعي ممثلة لمجموعة المتقدمين المعنية وجميع المجموعات ذات الصلة (على سبيل المثال، العمر والجنس وحالة الإعاقة). ولتحقيق ذلك، يجب تضمين البيانات المأخوذة من الأفراد من جميع المجموعات ذات الصلة في التدريب على تقييم الذكاء الاصطناعي، وقد يتطلب ذلك الإفراط في أخذ عينات استراتيجية؛ حيث يتم تضمين بيانات إضافية لمجموعة من الأشخاص لا يمثلون نسبة كبيرة من المتقدمين - على سبيل المثال الأفراد الذين يعانون من إعاقة – وذلك لضمان إدراج نقاط بيانات كافية في عملية التدريب. ويجب على مطوري التقييم تقديم معلومات تتعلق بالنهج المستخدم لضمان إضفاء الطابع التمثيلي للبيانات أثناء تطوير تقييم الذكاء الاصطناعي.

خصوصية البيانات وحمايتها

من الأهمية بمكان أن يتم تخزين أي بيانات تُستخدم لتطوير أو استخدام تقييم الذكاء الاصطناعي بطريقة توفر أكبر قدر ممكن من الخصوصية والحماية. إن العديد من التقييمات الجديدة التي تستخدم أساليب الذكاء الاصطناعي، مثل المقابلات عبر الفيديو التي يتم تحديد درجتها بواسطة الذكاء الاصطناعي، تلتقط وتخزن بيانات الفيديو التي تكشف عن هوية الفرد وقد تحتوي على معلومات حساسة. وتحدد اللوائح الحديثة، مثل اللائحة العامة لحماية البيانات للاتحاد الأوروبي (GDPR)، متطلبات حماية هذه المعلومات وتخزينها. ولهذا، يجب على مطوري التقييم تقديم معلومات كافية فيما يتعلق بالطرق المستخدمة لتأمين وحماية بيانات التقييم، وأي معايير يستوفونها (على سبيل المثال، اللائحة العامة لحماية البيانات).

تطبيق السؤال الأول على دراسة الحالة الافتراضية

يسلط هذا القسم الضوء على كيفية تطبيق كل من اعتبارات بيانات التدريب (نوعية وكمية البيانات وتمثيلها وخصوصيتها/وحمايتها) على دراسة الحالة الافتراضية -مراجعة تقييم الذكاء الاصطناعي الذي يستخدم المعلومات الموجودة في السير الذاتية للتنبؤ بنتائج الموظف.

1

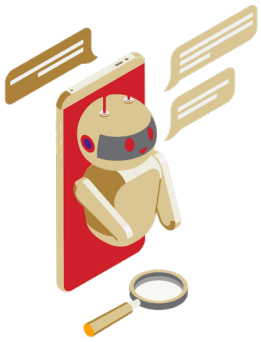
السؤال الأول:

إلى أي مدى تعتبر بيانات التدريب مُلائمة؟

السؤال الأول الذي يجب طرحه في هذا السيناريو هو ما إذا كان التقييم سيكون مخصصاً أم منتجاً عاماً جاهزًا. وفي الحقيقة، يتطلب تطوير نسخة مخصصة من تقييم السير الذاتية بواسطة الذكاء الاصطناعي أن يقوم المتعامل بتزويد مُورّد التقييم بنوعية وكمية كافية من البيانات التي سيتم استخدامها لتدريب تقييم الذكاء الاصطناعي. وبعد تطوير التقييم، يجب إجراء دراسة التحقق من قبل المُورّد بالتعاون مع المتعامل. وينبغي الإشارة إلى أنه كان من الممكن مسبقاً التحقق من صحة إصدار جاهز للتقييم باستخدام مجموعة مختلفة من البيانات غير الخاصة بالعميل. ويمكن التأكيد على وجود مفاضلات بين كل من هذه الطرق، لكن القرار الأفضل لمؤسسة ما سوف يختلف بناءً على عوامل خاصة بموقفها وطبيعة عملها.

ورقة تقييم لتقييم الذكاء الاصطناعي: ملاءمة بيانات التدريب

يقدم الجدول 5 ورقة تقييم يمكن استخدامها لتقييم ملاءمة بيانات التدريب التي تم جمعها لتقييم الذكاء الاصطناعي. يتم إنشاء درجتين لكل مقياس من مقاييس ملاءمة البيانات، الدرجة الأولى هي إما 1 أو 0 بناءً على ما إذا كان الوصف التفصيلي للمقياس متوفراً في دليل فني أو مستند آخر يتم توفيره للمتعامل. أما الدرجة الثانية فتعتمد على تقييم شخصي لمدى استيفاء المقياس لمعيار ما، وهي مصنفة على مقياس من 0 إلى 3. ويجب أن يصدر هذا الحكم خبير متخصص (SME) لديه خبرة في تقييم المواهب. يحتوي العمود الأخير على درجات المثال بناءً على دراسة الحالة الافتراضية الخاصة بنا، على أن يتم تقديم الأساس المنطقي لهذه الدرجات في الأقسام التالية.



1

السؤال الأول:

إلى أي مدى تعتبر بيانات التدريب مُلائمة؟

المقياس	الوصف	الدرجة
نوعية البيانات	أ. يتوفر وصف تفصيلي لأسباب اختيار جميع أنواع البيانات وإدراجها (نعم = 1، لا = 0). ب. مدى ملاءمة نوعية البيانات لحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: 1 ب: 2
كمية البيانات	أ. يتوفر وصف تفصيلي لأسباب حجم العينة المستخدمة لتدريب تقييم الذكاء الاصطناعي (نعم = 1، لا = 0). ب. مدى ملاءمة كمية البيانات لحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: 1 ب: 3
تمثيل البيانات	أ. يتوفر وصف تفصيلي لتمثيل العينة المستخدمة لتدريب تقييم الذكاء الاصطناعي (نعم = 1، لا = 0). ب. مدى تمثيل البيانات لحالة الاستخدام (غير ممثل = 0، تمثيل منخفض = 1، تمثيل متوسط = 2، تمثيل عالٍ = 3).	أ: 1 ب: 1
أمن البيانات	أ. يتوفر وصف تفصيلي لكيفية تخزين البيانات التي تم جمعها من خلال التقييم وحمايتها (نعم = 1، لا = 0). ب. مدى ملاءمة تخزين وحماية البيانات التي تم جمعها لحالة الاستخدام (غير آمنة /غير محمية = 0، أمن/حماية منخفضة = 1، أمن/حماية متوسطة = 2، أمن/حماية عالية = 3)	أ: 1 ب: 3
النتيجة الإجمالية لمدى ملاءمة بيانات التدريب (اجمع جميع القيم في عمود «النتيجة»)		16/13

الجدول (5) ورقة تقييم لتقييم ملاءمة بيانات التدريب.

السؤال الأول:

إلى أي مدى تعتبر بيانات التدريب مُلائمة؟

مراجعة نوعية البيانات

في دراسة حالة فحص السيرة الذاتية بواسطة الذكاء الاصطناعي، يمكن التحقق من جودة البيانات المستخدمة لتطوير التقييم من خلال الاستفسار عن دقة التحول من السير الذاتية في شكلها الأصلي (على سبيل المثال، PDF ، Word) إلى الإصدار المخزن في قاعدة البيانات (مثل متجه النص). على سبيل المثال، هل كل كلمة في السيرة الذاتية الأصلية موجودة في متجه الكلمات (word vector) وبنفس الترتيب بالضبط؟ بالإضافة إلى ذلك، يجب فحص نوعية بيانات النتائج. وإذا تم استخدام تقييمات الأداء السنوية الخاصة بالمشرف، فقد يمثل ذلك تحدياً لأن هذه الأنواع من التقييمات لم يتم تصميمها لهذا النوع من المهام التنبؤية. أما من الناحية المثالية، فسيتم تطوير مقياس أداء وظيفي مخصص يوفر تقييماً أكثر شمولاً لأداء الموظف مقابل الكفاءات المطلوبة لهذه الوظيفة. والسبب في ذلك هو أن مراجعات الأداء السنوية تفتقر عادةً إلى التباين - على سبيل المثال، ربما يحصل 70٪ من الموظفين في مؤسسة ما على درجة أداء سنوية تبلغ «3»، بينما يتلقى 10٪ «2»، ويحصل 10٪ على «4»، ويحصل 5٪ على «1»، بينما يحصل 5٪ الآخرون على «5.» هذا النقص في التمييز الذي يشمل معظم القوى العاملة سيجعل من الصعب على خوارزمية الذكاء الاصطناعي تعلم العلاقات بين العناصر الموجودة في السيرة الذاتية والأداء الوظيفي. لذلك، وعلى الرغم من أنه يمكن أن تكون هناك بيانات عالية الجودة من جانب مؤشرات التنبؤ (بيانات السيرة الذاتية)، فقد يظل قياس النتيجة منخفض الجودة. أما من الناحية المثالية في هذه الحالة، فسيتم تطوير مقياس تقييم أداء مخصص لجمع بيانات النتائج التي يمكن الاستناد عليها في التحقق من صحة التقييم، على الرغم من أن هذا قد يكون مُكلّفاً للقيام به على نطاق واسع.

لتقديم مثال لتقييم نوعية البيانات الخاصة بورقة التقييم الموضحة في الجدول 5، نفترض أن فحص السير الذاتية بواسطة الذكاء الاصطناعي في دراسة الحالة لها معدل تحويل ودقة نسخ عالي جداً تبلغ نسبتها 99.7٪، وهذا يعني أن بيانات السيرة الذاتية المستخدمة لتطوير التقييم قريبة جداً من شكلها الأصلي، وبالتالي فإننا نعتبر بيانات السيرة الذاتية عالية الجودة، كما نفترض أنه سيتم التحقق من صحة التقييم مقابل درجات مراجعة أداء الموظف السنوية. ووفقاً لما تمت مناقشته أعلاه، فإن هذا النوع من بيانات الأداء له العديد من القيود، لذلك فإننا نعتبر هذا الجانب من البيانات منخفضة الجودة. ومن خلال دمج الجودة العالية للبيانات الموجودة في جانب مؤشرات التنبؤ، مع الجودة المنخفضة للبيانات الموجودة في جانب المعايير، نكون قد وصلنا إلى تصنيف «الملاءمة المتوسطة» (2 من 3) لدرجة نوعية البيانات الواردة في الجدول 5.

السؤال الأول:

إلى أي مدى تعتبر بيانات التدريب مُلائمة؟

مراجعة كمية البيانات

بالنظر إلى أن بيانات السيرة الذاتية وفيرة داخل المؤسسات، لا سيما بالنسبة لوظائف المبتدئين التي يكثر عليها الطلب، فقد يكون هناك قدراً كبيراً جداً من السير الذاتية المتاحة بسهولة لتدريب الخوارزمية عليها. وفي الواقع، يمكن بسهولة أن يكون هناك عشرات إن لم يكن مئات الآلاف من السير الذاتية التي جمعتها المؤسسة (المتعامل) بمرور الوقت، ولهذا يجب توفير عدد الحالات المستخدمة لتدريب الخوارزمية والأساس المنطقي لاختيار هذا الرقم من قبل مطور التقييم.

بالنسبة لدراسة الحالة، نفترض أنه سيتم التحقق من صحة التقييم على 10,000 سيرة ذاتية ودرجات مراجعة الأداء المرتبطة بها بناءً على ثلاث سنوات من البيانات. لذلك، فإن درجات التقييم ستكون عالية (3 من 3) بالنسبة لكمية البيانات وفقاً لما هو واردٌ في الجدول 5.

مراجعة تمثيل البيانات

في حين أنه قد يكون هناك عدد كبير من السير الذاتية التي يمكن تدريب الخوارزمية عليها، فإن معظم السير الذاتية لن تتضمن المعلومات التي ستكون مطلوبة لتقييم تمثيل مجموعة المتقدمين (مثل العرق أو حالة الإعاقة). لهذا السبب، قد لا تكون هناك بيانات مصنفة كافية فيما يتعلق بهذه الأنواع من المتغيرات الديموغرافية لتضمينها في عملية تدريب الخوارزمية. ومن الجوانب الأخرى الهامة التي يجب مراعاتها هو ما إذا كانت مجموعة المرشحين سوف تتغير بمرور الوقت، على سبيل المثال، قد تؤدي التحولات في سوق العمل إلى اختلافات في التعليم أو الخبرة بين المرشحين بمرور الوقت. وقد تكون النتيجة أن مجموعة المرشحين التي تم تطوير الخوارزمية استناداً عليها لا تمثل مجموعة المرشحين التي يتم تطبيق الخوارزمية عليها.

بالنسبة لدراسة الحالة، نفترض أنه من المتوقع أن تظل مجموعة المرشحين مستقرة نسبياً بمرور الوقت. ومع ذلك، سوف نفترض أيضاً أن البيانات تفتقر إلى المعلومات الديموغرافية الكافية فيما يتعلق بعضوية المجموعة. لذلك، لن يكون من الممكن إجراء تحليلات لتقييم التحيز أو الإنصاف المحتمل لمجموعات ديموغرافية معينة. ولهذا السبب، حصل التقييم على درجات منخفضة (1 من 3) في تمثيل البيانات وفقاً لما هو واردٌ في الجدول 5.

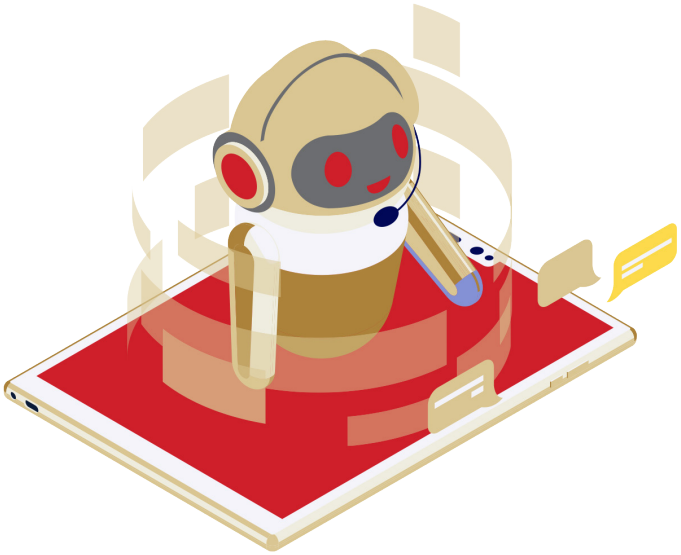
السؤال الأول:

إلى أي مدى تعتبر بيانات التدريب مُلائمة؟

مراجعة خصوصية البيانات وحمايتها

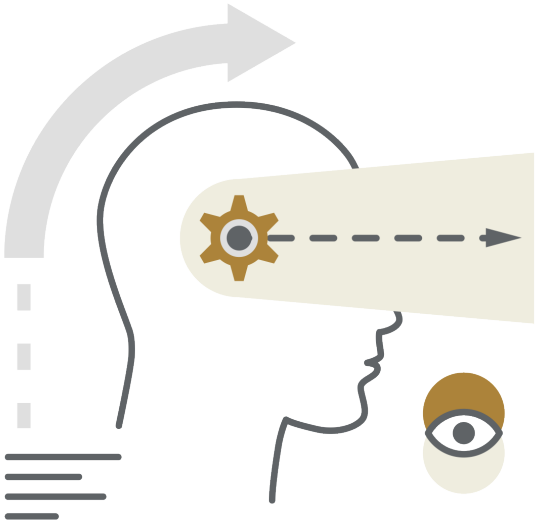
يجب أن يكون لدى مزوّد التقييم بيان بشأن كيفية تخزين وحماية أي بيانات يتم جمعها كجزء من عملية التقييم، على أن يصف هذا البيان مدى قوة عمليات إخفاء هوية البيانات والأمان الخاصة بالمزوّد وكيفية مقارنتها بأعلى المعايير الدولية (مثل اللائحة العامة لحماية البيانات).

وبالنسبة لدراسة الحالة، نفترض أن المؤسسة التي توفر التقييم تعمل وفقًا لمعايير اللائحة العامة لحماية البيانات (GDPR) فيما يتعلق بكيفية الحصول على البيانات الشخصية وتخزينها. لذلك، سوف يسجل التقييم درجات عالية (3 من 3) في خصوصية البيانات وحمايتها وفقاً لما هو واردٌ في الجدول 5.



السؤال الثاني:

كيف تقوم الخوارزمية باتخاذ القرارات؟



من الأهمية بمكان أن يعرف أي مستخدم لتقييم يستخدم أساليب الذكاء الاصطناعي كيف يعمل التقييم والذكاء الاصطناعي الأساسي، وهذا المفهوم يعرف باسم الشفافية، أي كيف تقوم خوارزميات الذكاء الاصطناعي باتخاذ القرارات؟ وينبغي الإشارة إلى أن التقييمات التي تعتبر شفافة هي تلك التقييمات التي يمكن من خلالها لمطوري التقييم والمستخدمين شرح كيفية عمل خوارزمية الذكاء الاصطناعي وكيف وصلت إلى قرار معين (على سبيل المثال، توظيف مرشح مقابل عدم توظيفه)، ومن المؤكد أن استخدام تقييم شديد الشفافية سوف يؤدي إلى تقليل مخاطر المشكلات القانونية وتلك المتعلقة بصورة المؤسسة ومخاطر الفعالية، وتمكين المستخدم من التخطيط للمستقبل فيما يتعلق باستمرار ملاءمة التقييم للأدوار المعينة التي يتم استخدامه من أجلها.

وفي الحقيقة، نجد أن المعايير الدولية تدعم استخدام التقييمات الشفافة، على سبيل المثال تنص مبادئ جمعية علم النفس الصناعي والتنظيمي (SIOP) للتحقق من صحة واستخدام إجراءات اختيار الموظفين (2018) على أن «المتغيرات المختارة كمؤشرات للتنبؤ يجب أن يكون لها أساس نظري أو منطقي أو تجريبي، ويجب تحديد الأساس المنطقي لاختيار مؤشر التنبؤ (مؤشرات التنبؤ). ومن المرجح أن يقدم مؤشر التنبؤ دليلاً على الصحة إذا كان هناك سبب وجيه أو نظرية لافتراض وجود علاقة بينه وبين السلوك المصمم للتنبؤ به».

السؤال الثاني:
كيف تقوم الخوارزمية باتخاذ القرارات؟

وهناك مثالان آخران هما إرشادات الاتحاد الأوروبي الأخلاقية للذكاء الاصطناعي الجدير بالثقة (المفوضية الأوروبية، 2019) والإرشادات العالمية للذكاء الاصطناعي الصادرة عن (The Public Voice 2018). بالإضافة إلى ذلك، تتضمن اللائحة العامة لحماية البيانات الخاصة بالاتحاد الأوروبي (GDPR)، والإرشادات التنظيمية المرتبطة بها حق الأفراد في تلقي تفسير، بل والطعن، في قرار مؤتمت فقط من خلال خوارزمية ما. لذلك، يجب أن تكون الأعمال الداخلية - الطريقة التي يتخذ بها التقييم القرارات - معروفة لمطور التقييم والمستخدم، مع ضرورة توثيقها بشكل كافٍ (على سبيل المثال، في دليل فني).

وينبغي الالتفات إلى أنه يمكن جعل تقييم الذكاء الاصطناعي أكثر شفافية من خلال العديد من الأساليب المختلفة؛ حيث تتمثل إحدى هذه الطرق في السماح فقط بإدراج المتغيرات التي لها ارتباط مفاهيمي بالوظيفة المستهدفة في عملية التدريب النموذجية، وهناك طريقة أخرى تتمثل في استخدام خوارزميات بسيطة وسهلة التفسير فقط (مثل الانحدار اللوجستي) في التقييم، أما الخيار الثالث فهو استخدام طرق جديدة ومتطورة نسبياً، تُعرف باسم الذكاء الاصطناعي القابل للتفسير (XAI)، والذي يهدف إلى تقديم تفسير حتى لأكثر الخوارزميات تعقيداً. وفي الممارسة العملية، قد تكون هناك حاجة لمجموعة من الخيارات المذكورة أعلاه لتطوير تقييم شفاف بما فيه الكفاية.

تطبيق السؤال الثاني على دراسة الحالة الافتراضية

تعد الشفافية الخاصة بتقييم الذكاء الاصطناعي مفهوماً مهماً يعكس قدرة المستخدم على فهم كيفية اتخاذ القرارات (على سبيل المثال، توظيف فرد مقابل عدم توظيفه). وتعتبر تفاصيل الخوارزمية التي يستخدمها التقييم لاتخاذ القرارات سرية، وبالتالي فإنه من غير المحتمل أن يعرفها مستخدم التقييم. ومع ذلك، يجب على مطور التقييم توفير درجة معينة من الرؤية المناسبة في الأعمال الداخلية للتقييم، علماً بأن درجة الشفافية المطلوبة قد تختلف حسب حالة الاستخدام (على سبيل المثال، قرارات الاختيار مقابل تدريب الموظفين)، ولكن يجب على مطور تقييم الذكاء الاصطناعي تقديم وصف شامل للمستخدمين المحتملين بشأن كيفية اتخاذ التقييم، مع ضرورة أن تتضمن هذه الشفافية معلومات حول جوانب الفرد التي تستخدمها الخوارزمية لاتخاذ القرارات، وكيف يتم ترجيح هذه المتغيرات ودمجها.

السؤال الثاني:

كيف تقوم الخوارزمية باتخاذ القرارات؟

في السيناريو الافتراضي الخاص بسيرتنا الذاتية لتسجيل الدرجات، يمكن أن يبدأ التحقيق في شفافية التقييم بمراجعة أي مواد داعمة للتقييم، مثل الدليل الفني. ويجب أن تركز المراجعة أولاً على تحديد المعلومات المأخوذة من السيرة الذاتية للمرشح المضمنة في خوارزمية الذكاء الاصطناعي، فمثلاً ربما تبحث الخوارزمية فقط عن الكلمات الرئيسية أو العبارات التي تطابق الوظيفة الحالية التي يتم استخدام التقييم من أجلها، أو ربما تبحث فقط عن أنواع معينة من المعلومات مثل التعليم. بالإضافة إلى ذلك، ينبغي تقديم معلومات بشأن العملية التي تم من خلالها تحديد هذه المعلومات على أنها مهمة، وهل كانت العملية مدفوعة بالكامل بالبيانات؟ أو هل كان هناك مجال ومدخلات من الخبراء المتخصصين في أنواع المتغيرات التي يجب تحديدها واستخدامها في بناء الخوارزمية؟ بالإضافة إلى ذلك، يجب وصف بعض المعلومات المتعلقة بالطريقة التي يتم بها ترجيح هذه المتغيرات ودمجها، على سبيل المثال ربما يتم استخدام نموذج انحدار المربعات الصغرى الاعتيادية (OLS) لوزن متغيرات السيرة الذاتية ودمجها.

وبالنسبة لدراسة حالة تقييم السيرة الذاتية الخاصة بنا، نفترض أنه تم الاستعانة بالخبراء المتخصصين في تقييم المواهب لتحديد المتغيرات التي كان من المتوقع أن يكون لها علاقة بالأداء الوظيفي في الوظائف الموضوعة قيد الدراسة. لقد استطاع الخبراء وضع تصنيف لعناوين الوظائف التي يمكن استخدامها لتحديد الخبرة في الوظائف السابقة التي كانت مشابهة للوظيفة الحالية المعنية. على سبيل المثال، بالنسبة لوظيفة خدمة العملاء، كان يُنظر إلى الخبرة السابقة في مركز الاتصال على أنها تحتوي على الأقل على بعض الكفاءات المطلوبة للدور الجديد. بالإضافة إلى ذلك، يمكن أيضاً تحديد المسؤوليات المدرجة أسفل كل مسمى وظيفي في السيرة الذاتية الخاصة بهذا التصنيف. وقد اختار الخبراء المتخصصون أيضاً تضمين الوقت في كل مقدار من الوظائف في الخوارزمية، والنتيجة هي تقدير مقدار الوقت الذي أمضاه كل مرشح في كل فئة من فئات الوظائف المدرجة في التصنيف، ثم تم بعد ذلك استخدام هذه المعلومات للتنبؤ بتقييمات الأداء الوظيفي اللاحقة لذلك الفرد.

نفترض أيضاً أن ترجيح هذه المعلومات ودمجها قد تم تطويرها من خلال اختبار عدة نماذج مختلفة، وكان النموذج الأفضل أداءً هو انحدار المربعات الصغرى الاعتيادية (OLS)، مع تقديم المقاييس المتعلقة بأداء النموذج (أي نموذج R-squared).

بالنظر إلى هذه الافتراضات، فإن التقييم في دراسة الحالة هذه يسجل درجات

السؤال الثاني:

كيف تقوم الخوارزمية باتخاذ القرارات؟

عالية في الشفافية؛ حيث يتم توفير المعلومات المتعلقة بالمعلومات الموجودة في السيرة الذاتية التي تستخدمها الخوارزمية لإنشاء درجة متوقعة، كما أن هناك وصفاً لكيفية ترجيح الخوارزمية ودمج هذه المعلومات (أي نموذج انحدار المربعات الصغرى الاعتيادية (OLS)). بالإضافة إلى ذلك، شارك الخبراء المتخصصون في تقييم المواهب في تحديد المعلومات الخاصة بالسير الذاتية التي يجب أن يستخدمها التقييم، وقد أدى ذلك إلى تصنيف الوظائف بشكلٍ يساعد على جعل التقييم أكثر شفافية. لذلك، فإن ورقة التقييم الخاصة بنا، الواردة في الجدول 6، تعطي درجة 3 لعناصر المعلومات المستخدمة ووزن المعلومات ودمجها.

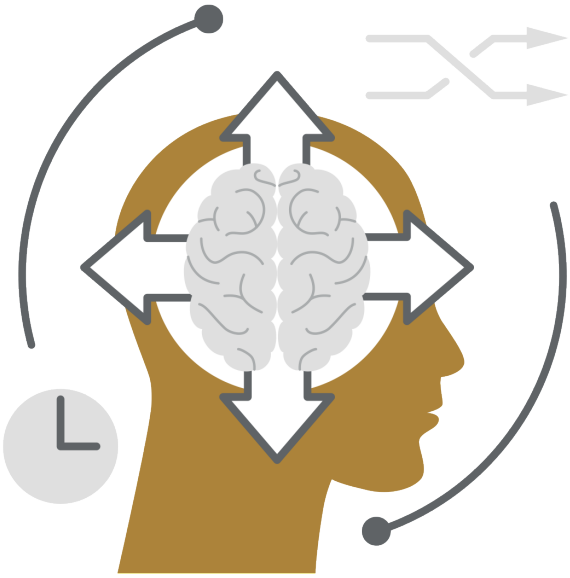
ورقة تقييم لتقييم الذكاء الاصطناعي: **الشفافية**

المقياس	الوصف	الدرجة
المعلومات المستخدمة	أ. يتوفر وصف للمعلومات التي تستخدمها خوارزمية الذكاء الاصطناعي للوصول إلى التنبؤ (نعم = 1، لا = 0).	أ: 1
	ب. المعلومات ملائمة لحالة الاستخدام (غير ملائمة = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	ب: 3
ترجيح المعلومات ودمجها	أ. يتوفر وصف للطريقة التي يتم بها ترجيح المعلومات المذكورة أعلاه ودمجها بواسطة خوارزمية الذكاء الاصطناعي للوصول إلى التنبؤ (نعم = 1، لا = 0).	أ: 1
	ب. المنهجية ملائمة لحالة الاستخدام (غير ملائمة = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	ب: 3
النتيجة الإجمالية للشفافية		8/8

الجدول (6) ورقة تقييم لتقييم الشفافية.

السؤال الثالث:

ما مـــدى انحيــاز القرارات؟



من المهم أن يتم تطوير تقييمات الذكاء الاصطناعي بطريقة لا تؤدي إلى اتخاذ قرارات متحيزة. ففي سياق الذكاء الاصطناعي، يُستخدم مصطلح «التحيز» للإشارة إلى خوارزمية الذكاء الاصطناعي التي تؤدي إلى التمييز ضد مجموعة معينة من الأشخاص (على سبيل المثال، عِرق أو فئة عمرية أو جنس معين)، بغض النظر عما إذا كان هذا التمييز عادلاً أو غير عادل، وقد تعكس هذه التحيزات تحيزات مجتمعية أكبر (مثل العنصرية والتفرقة العمرية والتمييز على أساس الجنس).

السؤال الثالث:

ما مـــدى انحيــاز القرارات؟

ولتجنب الدمج غير المقصود للتحيز في تقييم الذكاء الاصطناعي، يجب إجراء دراسة شاملة للتخلص من التحيز في كل خطوة من خطوات عملية تطوير التقييم. على سبيل المثال، يمكن اتخاذ الخطوات التالية أثناء تطوير وتقييم التقييم القائم على الذكاء الاصطناعي للمساعدة في تقليل التحيز المحتمل:

1. يمكن أن ينشأ التحيز من خلال اللغة أو السياقات التي قد يتعذر الوصول إليها نسبياً لمجموعة معينة، أو ربما يوجد عنصر ما قد يسيء إلى أفراد من مجموعات معينة أو يجعلهم غير مرتاحين (على سبيل المثال، من خلال الترويج لصورة نمطية معينة)، ولهذا يجب على الخبراء متعددي الثقافات مراجعة جميع الأسئلة، أو تسجيل درجات التقييم، أو أي نص آخر بحثاً عن أي تحيز أو عدوانية أو تأثيرات ثقافية محتملة.

2. إذا تم إنشاء درجات التقييم من قبل المسؤولين عن التقييم، فيجب على جميع المسؤولين عن التقييم إكمال تدريب مكثف حول كيفية استخدام نماذج تسجيل الدرجات بشكل صحيح لتوفير تقييمات للردود تتسم بالموضوعية قدر الإمكان، مع ضرورة تقييم الردود من قبل اثنين على الأقل من المسؤولين عن التقييم. ومن المهم أن نشير إلى أن توافر مسؤولين متعديين عن التقييم سوف يساعد على التخلص من التحيزات الفردية التي قد تكون موجودة. واعتماداً على التقييم، قد يكون من الحكمة توفير التسجيلات الصوتية للردود فقط للمسؤولين عن التقييم للتأكد من أن ملامح الوجه أو المظهر لا يؤدي إلى أي تحيز.

3. عند تطوير خوارزميات لتسجيل الردود تلقائياً، قم بفحص الفروق بين درجات المجموعة بحثاً عن دليل على التأثير المعاكس ضد المجموعات المحمية. وإذا اتضح أن خوارزمية الذكاء الاصطناعي تُظهر تأثيراً سلبياً، فإننا ننظر إلى الميزات المكونة في الخوارزمية لتحديد وإزالة تلك الميزات التي يبدو أنها تنتج اختلافات غير مقصودة في الدرجات.

وينبغي الإشارة إلى ضرورة توثيق هذا النهج والنتائج في الوثائق الداعمة. على سبيل المثال، يجب تقديم البيانات التي تظهر متوسط الدرجات في التقييم من قبل المجموعات ذات الصلة من الأشخاص (مثل العرق والعمر والجنس)، مع ضرورة ملاحظة أي اختلافات كبيرة والتحقيق فيها بشكل أكبر.

السؤال الثالث:

ما مـــدى انحيــاز القرارات؟

تطبيق السؤال الثالث على حالة دراسة افتراضية

في مثال تقييم السيرة الذاتية، يجب أن يبدأ التحقيق في التحيز الموجود في التقييم بالدليل الفني؛ حيث من الضروري أن يحتوي هذا الدليل الفني على معلومات تتعلق بالتحقيقات التي تم إجراؤها لاختبار التحيز الناتج عن استخدام التقييم، إضافةً إلى المقاييس الرئيسية والتفسير الداعم. وفي الواقع، هناك نوعان من التحيز يجب التحقيق فيهما، وهما الاختلافات في درجات التقييم لمجموعات مختلفة من السكان (على سبيل المثال، العمر والعرق والجنس، مع الوضع في الاعتبار أن مجموعات التركيز قد تختلف حسب المنطقة)، والاختلافات في الدقة التنبؤية للتقييم الخاصة بمجموعات مختلفة. على سبيل المثال، سيكون من المهم التأكيد على أن الدرجات التي تم الحصول عليها من التقييم ليست أعلى أو أقل باستمرار للرجال مقارنة بالنساء، وفي حالة حدوث ذلك، فيجب وصف تفسير سبب حدوثه، إضافةً إلى ضرورة القيام بمحاولات لتقليل الفجوة في الدرجات. أما النوع الثاني من التحيز الذي يجب التحقيق بشأنه في مثال تقييم السيرة الذاتية هو ما إذا كانت الدرجات التي تم الحصول عليها من التقييم تنبؤية بشكل متساوٍ للأداء الوظيفي لكلٍ من الرجال والنساء. على سبيل المثال، إذا كانت الدرجات تُنبئُ بالأداء الوظيفي للرجال أكثر من النساء، على الرغم من حصول الرجال والنساء على درجات متساوية (أي عدم وجود فرق في متوسط الدرجات للرجال والنساء)، فإن تقييم السيرة الذاتية سيظل به شكل من أشكال التحيز.

بالنسبة لدراسة حالة تقييم السيرة الذاتية، نفترض أنه يتم توفير وصف للمتوسط والانحراف المعياري للدرجات لجميع المجموعات ذات الصلة. تكشف هذه المقاييس عن اختلاف بسيط ولكنه مهم بين متوسط الدرجات؛ بحيث يميل المرشحون الأكبر سناً إلى الحصول على درجات أعلى من المرشحين الأصغر سناً. يوضح الدليل الفني أن هذا يحدث لأن المرشحين الأكبر سناً يكون لديهم دائماً خبرة وظيفية أكبر، وبالتالي يحصلون على درجات أعلى لأن الخبرة الوظيفية هي متغير تنبؤي رئيسي تستخدمه الخوارزمية. ونظراً لأن الاختلاف في درجات التقييم يمكن تفسيره، ويتعلق بالمتغيرات التي حددها الخبراء المتخصصون باعتبارها مرتبطة بالأداء الوظيفي وبالتالي ذات صلة بالتقييم، فإن مخاطر التحيز موثقة ولكنها لا تعتبر مانعة. لذلك، يحصل التقييم على درجة متوسطة لمقياس التحيز الأول (3/2) في ورقة التقييم المعروضة في الجدول 7.

نفترض أيضًا أن الدليل الفني لا يحتوي على أي إشارة إلى أي تحقيقات حول الاختلافات في الدقة التنبؤية عبر المجموعات المختلفة. لذلك، يحصل التقييم على أدنى درجة وهي 0 لمقياس التحيز الثاني الوارد في الجدول 7.

ورقة تقييم لتقييم الذكاء الاصطناعي: **التحيز**

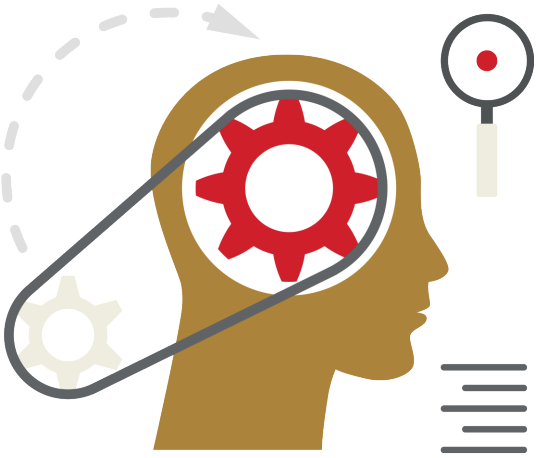
المقياس	الوصف	الدرجة
فروق جماعية كبيرة في درجات التقييم	أ. يتوفر وصف للمتوسط والانحراف المعياري للدرجات حسب المجموعات المختلفة (على سبيل المثال، العرق والجنس والعمر) (نعم = 1، لا = 0). ب.أي اختلافات في الدرجات بين المجموعات مقبولة لحالة الاستخدام (غير مقبول = 0، قبول منخفض = 1، قبول متوسط = 2، قبول عال = 3).	أ: 1 ب: 2
فروق جماعية كبيرة في دقة التقييم	أ. يتوفر وصف لدقة تنبؤات التقييم حسب المجموعات المختلفة (على سبيل المثال، العرق والجنس والعمر) (نعم = 1، لا = 0). ب.أي اختلافات في التنبؤ بين المجموعات مقبولة لحالة الاستخدام (غير مقبول = 0، قبول منخفض = 1، قبول متوسط = 2، قبول عالٍ = 3).	أ:1 ب: 3
النتيجة الإجمالية للإنصاف (عدم التحيز)		8/3

الجدول (7) ورقة تقييم لتقييم التحيز.

4

السؤال الرابع:

ما مدى صحــة القــرارات؟



الصحة هي مصطلح فني في تقييم المواهب يشير إلى «الدرجة التي يدعم بها الدليل والنظرية تفسيرات درجات الاختبار للاستخدامات المقترحة للاختبار» (AERA وآخرون، 2014؛ ص. 11). إن التحقق من صحة التقييم هو العملية التي يتم من خلالها إثبات صحة التقييم، كما أن التحقق الشامل من صحة التقييم هو أفضل ممارسة في كل من تقييم المواهب (على سبيل المثال، مبادئ جمعية علم النفس الصناعي والتنظيمي (SIOP)، 2018) والذكاء الاصطناعي (على سبيل المثال، المبادئ التوجيهية لأخلاقيات الاتحاد الأوروبي للذكاء الاصطناعي الجدير بالثقة).

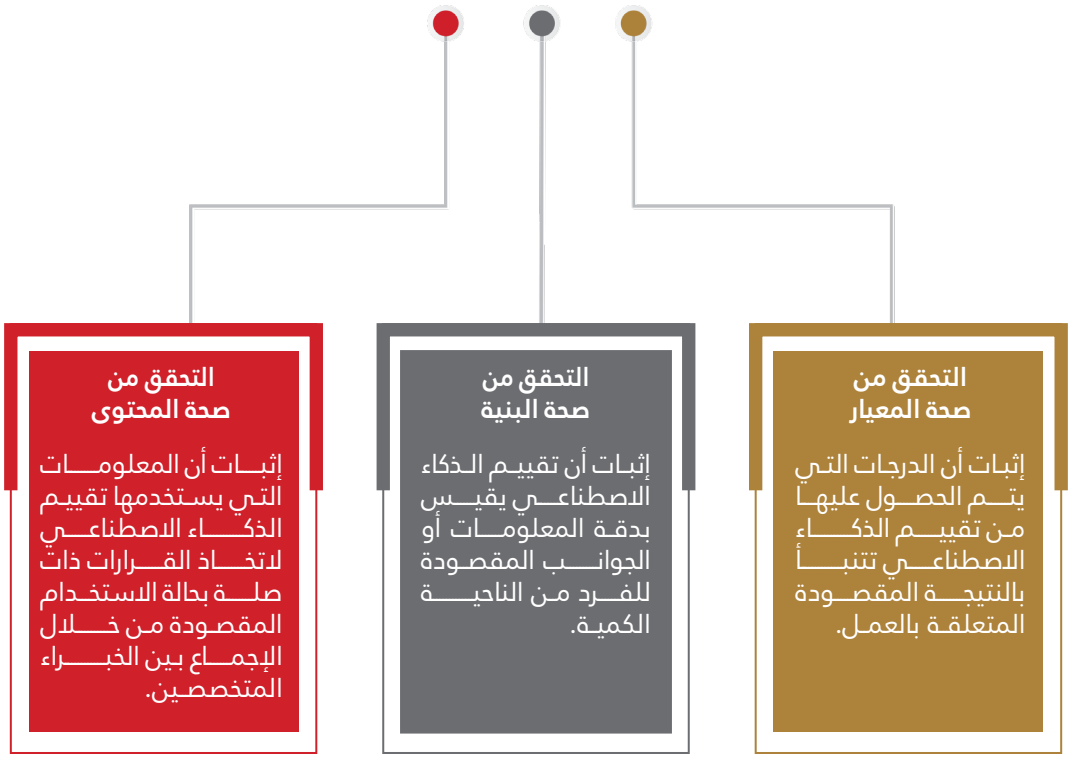
يعتبر كلا الدليلين الكميين أن التقييم يتنبأ بشكل كاف بنتائجه المقصودة، وأن النظرية التي تدعم السبب في ضرورة تنبؤه تعد أمراً مهماً في عملية التحقق من الصحة. وقد تم وصف هذا المفهوم في معايير الاختبار النفسي، والتي تنص على أن التحقق من الصحة يبدأ بـ «بيان صريح للتفسير المقترح لدرجات الاختبار، إلى جانب الأساس المنطقي لارتباط التفسير بالاستخدام المقترح، على أن يتضمن التفسير المقترح تحديد البنية التي يهدف الاختبار إلى قياسها»(AERA وآخرون، 2014؛ ص 11).

وسوف نناقش فيما يلي الطرق الأكثر شيوعاً لجمع أدلة التحقق من الحصة التي تدعم استخدام التقييم، بما في ذلك التقييمات التي تستخدم الذكاء الاصطناعي.

4

السؤال الرابع:
ما مدى صحة القرارات؟

الطرق الثلاث الأكثر شيوعاً للتحقق من صحة التقييم



التحقق من صحة المحتوى

يركز التحقق من صحة المحتوى على إثبات أن المعلومات المستخدمة في تقييم الذكاء الاصطناعي لاتخاذ القرارات ذات صلة بحالة الاستخدام المقصودة من خلال الإجماع بين الخبراء المتخصصين. على سبيل المثال، إذا اتفقت مجموعة من الخبراء المتخصصين على أن تقييم محاكاة الذكاء الاصطناعي يقوم بشكل كاف بتقييم الكفاءات الأساسية المطلوبة لمحلل البيانات، كما هو موضح من خلال تقييماتهم الشخصية، فيمكن عندئذٍ القول بأن التقييم يحتوي على دليل صحة يتعلق بالمحتوى ويدعم استخدامه في اختيار المرشحين لوظيفة محلل البيانات.

4

السؤال الرابع:
ما مدى صحة القرارات؟

التحقق من صحة البنية

يركز التحقق من صحة البنية على إثبات أن تقييم الذكاء الاصطناعي يقيس بدقة المعلومات أو الجوانب المقصودة للفرد من الناحية الكمية - والمعروفة باسم التركيبات النفسية. على سبيل المثال، إذا تم تصميم تقييم الذكاء الاصطناعي لقياس مهارات الاتصال لدى الفرد، فيمكن مقارنة نتائج التقييم بالنتائج التي يتم الحصول عليها من تقييم مختلف مصمم أيضاً لقياس مهارات الاتصال. وسوف يظهر الارتباط الكبير بين مجموعتي الدرجات الناتجة عن هذين التقييمين المختلفين ما يُعرف باسم الصلاحية المتقاربة – أي أن التقييمين المرتبطين من الناحية المفاهيمية أو النظرية (من حيث أنهما مصممان لقياس مهارات الاتصال) مرتبطان أيضاً من الناحية الكمية من حيث كون درجاتهما مترابطة للغاية، وهذه الصلاحية المتقاربة هي نوع واحد من الأدلة لدعم صحة البنية. وهناك شكل آخر من أدلة صحة البنية يُعرف باسم الصلاحية التمييزية، والتي تُستخدم لإظهار أن التقييم لا يتعلق بالدرجات التي يتم الحصول عليها من تقييم يقيس بنية غير ذات صلة. على سبيل المثال، إذا كان تقييم الذكاء الاصطناعي يقيس بدقة مهارات الاتصال، ومهارات الاتصال فقط، فيجب أن يكون غير مرتبط، أو أن يكون ارتباطه ضعيفاً للغاية، بدرجات مقياس الضمير. وينبغي أن نشير إلى أن الإثبات المشترك للصلاحية المتقاربة والتمييزية بهذه الطريقة يعد دليلاً على صحة البنية.

التحقق من الصحة ذات الصلة بالمعيار

يركز التحقق من الصحة ذات الصلة بالمعيار على إثبات أن الدرجات التي يتم الحصول عليها من تقييم الذكاء الاصطناعي تتنبأ بالنتيجة المقصودة المتعلقة بالعمل (على سبيل المثال، تقييمات المشرف للأداء الوظيفي)؛ حيث يتم التعبير عن هذا التوقع من خلال مقياس إحصائي (غالباً ما يكون معامل ارتباط)، وتوضح قيمة المقياس إلى جانب أهميته الإحصائية قوة الدليل الذي يدعم صحة المعيار المتعلق بتقييم الذكاء الاصطناعي، وغالباً ما يعتبر هذا النوع هو أقوى نوع من الأدلة لإثبات الصحة.

هناك خمسة أسئلة هامة يجب مراعاتها عند مراجعة دراسة التحقق من الصحة ذات الصلة بالمعيار الخاصة بتقييم الذكاء الاصطناعي (انظر الجدول 8).

4

السؤال الرابع:
ما مدى صحة القرارات؟

عنصر الدراسة	السؤال
1. التحليل الوظيفي	هل تم إجراء تحليل وظيفي شامل؟
2. المعيار	هل تم استخدام مقياس معياري مناسب؟
3. حجم العينة	هل كان حجم العينة كبيراً بما يكفي لتوفير قوة إحصائية كافية والسماح للعينات العشوائية المحفوظة بالتحقق المتقاطع من الصحة؟
4. مدخلات الخبراء المتخصصين	هل تم تضمين مدخلات الخبراء المتخصصين في التحليل الوظيفي، واختيار أو تطوير مقياس المعيار، واختيار الميزة التنبؤية؟
5. التحقق المتقاطع من الصحة	هل تم اختبار أداء تقييم الذكاء الاصطناعي على عينة واحدة أو أكثر من العينات العشوائية المحفوظة؟

الجدول (8) العناصر المستخدمة لتقييم دراسة التحقق من الصحة ذات الصلة بالمعيار.

توفر الأوصاف التالية لكل عنصر من العناصر الواردة في الجدول 8 إرشادات حول كيفية تصميم دراسة تحقق من الصحة قوية ومرتبطة بالمعايير الخاصة بتقييم الذكاء الاصطناعي.

1. التحليل الوظيفي

يوفر التحليل الوظيفي لمطوري التقييم معلومات أساسية حول مجال أداء الوظيفة، وعادة ما يتضمن هذا تحديد المهام الأساسية، وسلوكيات العمل الرئيسية، والكفاءات، وخصائص العامل المهمة لأداء الوظيفة بنجاح؛ حيث يساعد هذا التحليل الوظيفي في تحديد اختيار المعايير والمعلومات المتعلقة بالمرشح التي يجب تضمينها في تقييم الذكاء الاصطناعي.

4

السؤال الرابع:
ما مدى صحة القرارات؟

وعند مراجعة المعلومات الخاصة بتقييم الذكاء الاصطناعي، يتعين عليك البحث عن دليل على إجراء تحليل وظيفي مناسب، علماً بأن هناك طرق مختلفة لإجراء التحليل الوظيفي، والتي تتضمن عادةً مجموعات التركيز، أو استطلاعات الرأي، أو مزيج منهما. وقد تختلف الطريقة الدقيقة المستخدمة حسب نوع التقييم ونوع الوظيفة، ولكن الشيء الهام هو وجود وتوافر وصف تفصيلي للطرق الكامنة وراء التحليل الوظيفي ونتائجها (على سبيل المثال، دليل فني).

2. المعيار

يشير المعيار إلى المقياس المستخدم لمتغير النتيجة والذي تم تصميم تقييم الذكاء الاصطناعي للتنبؤ به، مثل الأداء الوظيفي أو مخاطر دوران الموظفين، وتعتبر جودة المعيار في دراسة التحقق من الصحة المتعلقة بالمعيار مهمة للغاية. فغالباً ما تكون مقاييس المعايير عالية الجودة هي تلك المقاييس التي استندت على التحليل الوظيفي، وتمثل جوانب مهمة من مجال أداء الوظيفة، إضافةً إلى تطويرها خصيصًا من أجل قياس مجال أداء الوظيفة المستهدفة.

عند مراجعة تقييم الذكاء الاصطناعي، من المهم جداً إجراء تحقيق في مدى ملاءمة وصلة متغير المعيار وطريقة قياسه؛ فغالباً ما تعني الكمية الكبيرة من البيانات التي تُبنى عليها نماذج الذكاء الاصطناعي أن متغير المعيار متاح بسهولة وغير مصمم لهذا الغرض المحدد (على سبيل المثال، تقييمات الأداء الوظيفي السنوية).

3. حجم العيّنة

حجم العينة هو اعتبار مهم لدراسة التحقق من الصحة ذات الصلة بالمعيار؛ حيث يجب أن يكون حجم العينة كبيراً بما يكفي لتقديم تقديرات ثابتة للترجيحات المرتبطة بخوارزمية الذكاء الاصطناعي ولاختبار أهميتها الإحصائية. تتوفر تحليلات الطاقة لتحديد الحجم الأدنى من العينة للتأكد من ملائمة اختبارات الأهمية الإحصائية (على سبيل المثال، Cohen، 1992).

قد تتراوح الإرشادات غير الرسمية المستخدمة بكثرة لتحديد أحجام العينات المناسبة لدراسة التحقق من صحة التقييم من حوالي 100 إلى 300، على الرغم من أن هذا الرقم قد يكون أعلى أو أقل اعتماداً على نوع خوارزمية

الذكاء الاصطناعي وعدد المتغيرات المضمنة وتحليلات البيانات المطلوبة. على سبيل المثال، تتطلب دراسات التحيز الإحصائي في التنبؤ في كثير من الأحيان أكثر من 400 حالة للحصول على قوة كافية (Aguinis & Stone-Romero ، 1997).

عند مراجعة حجم العينة المستخدمة لتطوير تقييم الذكاء الاصطناعي، ابحث عن دليل على الأساس المنطقي وراء قرار استخدام حجم عينة معين. هل وضع مطورو التقييم تفكيراً دقيقاً وخبرة متخصصة في تطوير خوارزمية الذكاء الاصطناعي في هذا القرار؟ أم أن اختيار حجم العينة يعتمد على ما كان مناسباً أو متاحاً بسهولة؟

4. مدخلات الخبراء المتخصصين

بينما يعتمد الاختبار النهائي للصلاحية في دراسة التحقق من الصحة ذات الصلة بالمعيار على العلاقة التجريبية بين درجات التقييم والمعيار، لا تزال مدخلات الخبراء المتخصصين مهمة للغاية أثناء تصميم التقييم وتطويره، فغالباً ما يؤدي الاستعانة بالخبراء المتخصصين أثناء تصميم وتطوير تقييم الذكاء الاصطناعي إلى إنتاج تقييمات أداء أفضل.

عند مراجعة تقييم الذكاء الاصطناعي، ابحث عن دليل على الاستعانة بالخبراء المتخصصين في تقييم المواهب، وليس في الذكاء الاصطناعي فقط، أثناء تصميم التقييم وتطويره.

5. التحقق المتقاطع من الصحة

يتضمن التحقق المتقاطع من صحة التقييم اختبار دقة الدرجات الناتجة عن التقييم في مجموعة بيانات جديدة، مع مراعاة أن هذه خطوة مهمة للغاية لأن دقة خوارزمية الذكاء الاصطناعي عادةً ما تكون أعلى بكثير ضمن مجموعة البيانات التي تم تطويرها عليها، وغالباً ما تنخفض عند اختبار أداء الخوارزمية على مجموعة بيانات جديدة. لذلك، تعد تحليلات التحقق من الصحة طريقة مهمة للتأكد من أن الدرجات التي يتم الحصول عليها من التقييم ستظل صالحة عند استخدامها عملياً لتقييم المرشحين.

جدير بالذكر أن هناك عدة طرق للتحقق من صحة خوارزمية الذكاء الاصطناعي، وقد يختلف اختيار استراتيجية التحقق المتقاطع من الصحة حسب نوع التقييم

4

السؤال الرابع:
ما مدى صحة القرارات؟

ونوع الوظيفة وحالة استخدام التقييم المقصودة. ومن الأهمية بمكان أن تعرف أنه تم إجراء دراسة تحقق من الصحة شاملة ومناسبة، وأنه تم الإبلاغ عن نتائجها وأنها متاحة للمراجعة.

باختصار، ربما يكون نهج التحقق من الصحة ذات الصلة بالمعيار هو الأكثر استخداماً للتحقق من صحة تقييمات الذكاء الاصطناعي. ويمكن تحديد دراسات التحقق من الصحة ذات الصلة بالمعيار التي تفي بأغلبية المعايير الخمسة المذكورة أعلاه للحصول على دليل أقوى على الصلاحية فيما يتعلق باستخدام الدرجات التي يتم الحصول عليها من التقييم للمساعدة في اتخاذ القرارات المتعلقة بالتوظيف.

تطبيق السؤال 4 على دراسة الحالة الافتراضية

عند مراجعة تقييم الذكاء الاصطناعي بحثًا عن دليل على الصحة، ابحث أولاً عن نوع دليل الصحة الذي استند عليه تقييم الذكاء الاصطناعي.

بالنسبة لدراسة حالة تقييم السيرة الذاتية، نفترض أنه تم استخدام نهج التحقق من الصحة ذات الصلة بالمعيار. يؤكد التشاور مع الخبراء المتخصصين في تقييم المواهب أن هذا النوع من التحقق من الصحة مناسب للغاية لهذا التقييم، ولكن إضافة دليل التحقق من صحة المحتوى أو صحة البنية كان من شأنه أن يوفر أساساً منطقياً أقوى لاستخدام التقييم. لذلك، نعطي تصنيفاً بقيمة 2 درجة للمقياس الأول في ورقة تقييم الصلاحية، وهو مدى ملاءمة استراتيجية التحقق من الصحة (كما هو موضح في الجدول 9).

الخطوة التالية هي مراجعة نتائج دراسة التحقق من الصحة. نظراً لأن تقييم الذكاء الاصطناعي للسير الذاتية استخدم دراسة تحقق من الصحة ذات صلة بالمعيار، يجب مراجعة العوامل الواردة في الجدول 8 لتحديد مدى قبول نتائج التحقق من الصحة. وسوف نضع الافتراضات التالية لتقييم الذكاء الاصطناعي للسير الذاتية مقابل العوامل الخمسة الواردة في الجدول 8:

- **التحليل الوظيفي:** يوفر الدليل الفني لتقييم الذكاء الاصطناعي للسير الذاتية وصفاً تفصيلياً للتحليل الوظيفي الذي تم إجراؤه.
- **المعيار:** استخدم تقييم الذكاء الاصطناعي للسير الذاتية درجات مراجعة الأداء السنوية للمعيار، لكن هناك مشكلات في استخدام هذا النوع من المقاييس كمعيار (كما هو موضح سابقًا في هذا المستند).

4

السؤال الرابع:
ما مدى صحة القرارات؟

- **حجم العينة:** 50000 سيرة ذاتية وما يرتبط بها من تقييمات مراجعة أداء السنة الأولى من قبل المشرفين، وهو حجم عينة كبير جداً وأكثر من كافٍ لجميع تحليلات البيانات.
- **مدخلات الخبراء المتخصصين:** يوفر الدليل الفني أوصافاً تفصيلية للمجالات التي تم فيها تقديم مدخلات الخبراء المتخصصين أثناء تصميم التقييم وتطويره.
- **التحقق المتقاطع من الصحة:** لا يذكر الدليل الفني استخدام عينات عشوائية محفوظة غير مرئية كجزء من عملية التحقق من الصحة.

يتم منح التقييم درجة 2 بعد مراجعة العوامل الواردة في الجدول 8؛ وذلك بسبب وجود مشكلات مع المعيار المستخدم، كما لم يتم إجراء تحقق متقاطع من الصحة باستخدام عينة عشوائية محفوظة غير مرئية.

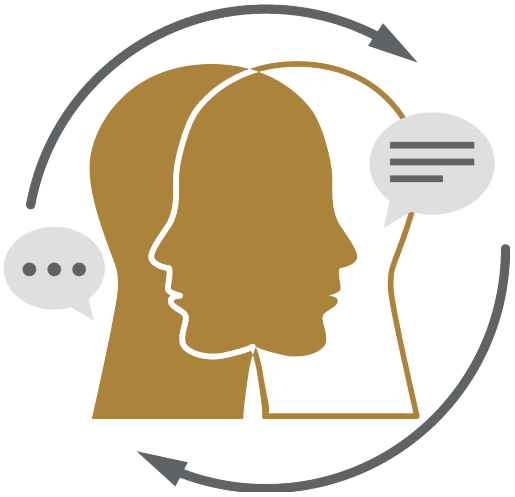
ورقة تقييم لتقييم الذكاء الاصطناعي: **التحقق من الصحة**

المقياس	الوصف	الدرجة
استراتيجية التحقق من صحة التقييم	أ. يتوفر وصف للاستراتيجية المستخدمة للتحقق من صحة التقييم (نعم = 1، لا = 0). ب. استراتيجية التحقق من الصحة ملائمة للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: 1 ب: 2
دراسة التحقق من صحة التقييم	أ. يتوفر وصف لنتائج دراسة التحقق من الصحة (نعم = 1 ، لا = 0). ب. تم تصميم وتنفيذ دراسة التحقق بشكل جيد (غير مقبول = 0، قبول منخفض = 1، قبول متوسط = 2 ، قبول عالٍ = 3).	أ: 1 ب: 2
النتيجة الإجمالية للتحقق من الصحة		8/6

الجدول (9) ورقة تقييم لتقييم التحقق من الصحة.

السؤال الخامس:

إلى أي مدى تكون القرارات نهائيـة؟



يجب أن تكون هناك قدرة على إتاحة الإشراف والتدخل البشري في أي قرارات يتخذها تقييم الذكاء الاصطناعي، مع الوضع في الاعتبار أن الإرشادات واللوائح الرائدة في هذه الصناعة في جميع أنحاء العالم تدعم هذا الاقتراح. على سبيل المثال، تنص الإرشادات المعمول بها في أوروبا على أن «لجميع الأفراد الحق في الحصول على قرار نهائي يتم اتخاذه من قبل شخص ما» (The Public Voice، 2018)، وأنه «يجب ضمان توافر آليات الإشراف المناسبة، والتي يمكن تحقيقها من خلال نهج توافر التدخل البشري في هذه الدائرة وعلى رأسها» (إرشادات المفوضية الأوروبية للذكاء الاصطناعي الجدير بالثقة، 2019).

وبموجب اللائحة العامة لحماية البيانات في الاتحاد الأوروبي، إذا تم اتخاذ القرار بالوسائل الآلية فقط (أي أنه لا توجد مساهمة بشرية ذات مغزى في القرار)، فإنه يحق للفرد طلب التدخل البشري، والتعبير عن وجهة نظره، والطعن في القرار.

هذا ويختلف مقدار الإشراف البشري المطلوب لتقييم معين للذكاء الاصطناعي؛ فعلى الطرف الأدنى من المشهد، يمكن للذكاء الاصطناعي أن يكون حراً في اتخاذ القرارات دون أي إشراف بشري - حيث يتمتع الذكاء الاصطناعي هنا باستقلالية كاملة، وفي الطرف الأعلى من المشهد، يمكن دمج الإشراف البشري في تقييم الذكاء الاصطناعي بحيث يجب على الإنسان أولاً الموافقة على أي قرار يتخذه الذكاء الاصطناعي. أما من الناحية العملية، سيكون مقدار الإشراف البشري المطلوب لمعظم تقييمات الذكاء الاصطناعي في مكان ما في منتصف هذا المشهد.

السؤال الخامس:

إلى أي مدى تكون القرارات نهائية؟

وتجدر الإشارة إلى أن الإشراف البشري على تقييم الذكاء الاصطناعي يمكن أن يحدث أثناء كلٍ من تطوير التقييم والاستخدام المستمر له. ويمكن للمطورين الذين يشرفون على إنشاء تقييم الذكاء الاصطناعي والتحقق من صحته، تحديد الدرجات الفاصلة التي تحدد نتائج المرشحين من إجراء التقييم، ووضع المعايير التي يمكن للذكاء الاصطناعي العمل بموجبها (على سبيل المثال، العمل مع استقلالية كاملة، أو مع بعض الإشراف البشري).

ويعمل مستخدم التقييم (على سبيل المثال، مسؤول التوظيف أو مدير التوظيف) على توفير الإشراف على تقييم الذكاء الاصطناعي باستخدام توصيات الذكاء الاصطناعي كمعلومات يتم دمجها مع معلومات من مصادر أخرى في اتخاذ القرار. على سبيل المثال، إذا توقع تقييم الذكاء الاصطناعي أن المرشح لديه إمكانات عالية تؤهله للأداء الجيد في وظيفة معينة، لكن مدير التوظيف لا يوافق على ذلك استناداً إلى المعلومات الأخرى المتاحة عن المرشح، فعندئذ يكون لمدير التوظيف الحرية في التدخل واختيار عدم القيام بتوظيف هذا المرشح، والعكس صحيحٌ أيضاً، أي إذا حدد تقييم الذكاء الاصطناعي أن المرشح ليس مناسباً، لكن مسؤول التوظيف يعتقد خلاف ذلك، فإنه يمكن لمسؤول التوظيف تجاوز قرار تقييم الذكاء الاصطناعي. يوضح هذان المثالان تصميم وتطوير تقييم الذكاء الاصطناعي الذي يخضع للإشراف البشري. في هذه السيناريوهات، قد يكون عدد المرات التي يختارها المستخدم البشري لتجاوز التوصية الصادرة عن تقييم الذكاء الاصطناعي نادراً جداً ولا يحدث إلا في ظل ظروف محددة (على سبيل المثال، استثناء محدد أو عملية تصعيد)، لكن النقطة المهمة هنا هي أن الإنسان لديه القدرة والفرصة للتدخل عند الحاجة.

النقطة الأساسية هنا هي أن تقييمات الذكاء الاصطناعي يجب تصميمها لتوفير المعلومات التي يتم استخدامها، إلى جانب المعلومات المتوفرة من مصادر أخرى (عند الضرورة)، من الجانب البشري لاتخاذ قرارات بشأن الموظفين الحاليين أو المحتملين في المؤسسة، أي أنه لا ينبغي تصميم تقييمات الذكاء الاصطناعي لاتخاذ هذه القرارات دون إشراف بشري.

وبغض النظر عن الإرشادات أو المتطلبات التنظيمية، فإن الإشراف البشري على تقييم الذكاء الاصطناعي يعد ممارسة جيدة للمؤسسة ولموظفيها وللمجتمع ككل، وهنا ينبغي أن نؤكد على أنه يجب تطوير تقييمات الذكاء الاصطناعي مع وجود إشراف بشري طوال العملية بأكملها - جمع البيانات وتنقيتها، واستخراج الميزات وتطويرها، وتدريب النموذج واختباره وتشغيله.

5

السؤال الخامس:

إلى أي مدى تكون القرارات نهائية؟

تطبيق السؤال الخامس على دراسة الحالة الافتراضية

عند مراجعة تقييم الذكاء الاصطناعي بحثاً عن دليل على الإشراف البشري، ابحث عن المدى الذي تكون فيه القرارات مدفوعة من جانب نواتج الخوارزمية في كل نقطة قرار في عملية التوظيف.

سوف نقوم بعمل الافتراضات التالية لمثال فحص السيرة الذاتية: يقوم المرشح باستكمال طلب الوظيفة وإرسال السيرة الذاتية. يتم تسجيل السيرة الذاتية تلقائياً من خلال تقييم الذكاء الاصطناعي، ويتم إرفاق هذه النتيجة بالتطبيق ليقوم مسؤول التوظيف بمراجعتها. توفر النتيجة رقماً على مقياس من 1-100، مع وجود عتبة - تعتبر الدرجات الأقل من 50 مخاطرة عالية للأداء الأقل من المتوسط، ثم يتم استخدام هذه النتيجة من قبل مسؤولي التوظيف كأداة مساعدة في اتخاذ القرار، إلى جانب معلومات أخرى في التطبيق، لمساعدتهم على تحديد ما إذا كانوا سيدعون المرشح لإجراء مقابلة هاتفية أم لا. نظراً لأن مستخدم التقييم في هذا السيناريو يمكنه تحديد ما إذا كان سيتبع نصيحة تقرير التقييم فيما يتعلق بدعوة المرشح لمقابلة عبر الهاتف، فإنه يحصل على درجة عالية فيما يتعلق بالإشراف البشري (3) في ورقة التقييم الموضحة في الجدول 10.

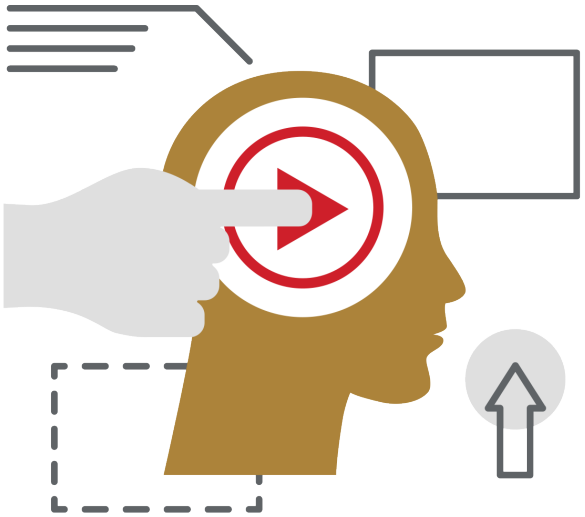
ورقة تقييم الذكاء الاصطناعي: **الإشراف**

المقياس	الوصف	الدرجة
فرصة التدخل البشري	أ. يتوفر وصف لمستوى الرقابة البشرية المضمن في التقييم (نعم = 1، لا = 0). ب. مستوى الإشراف البشري ملائم للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: 1 ب: 3
النتيجة الإجمالية للإشراف		4/4

الجدول (10) ورقة تقييم لتقييم الإشراف.

السؤال السادس:

كيف يتم إبلاغ المرشحين؟



تشير أفضل الممارسات الرائدة في الصناعة إلى أن المرشحين يتم إبلاغهم بالوقت الذي سيتم فيه استخدام الذكاء الاصطناعي لتسجيل ردودهم على التقييم، مع ضرورة توفير شرح كافٍ لكيفية عمل الذكاء الاصطناعي.

ويجب النظر فيما إذا كان يجب طلب الحصول على موافقة مستنيرة من المرشحين حتى يتمكن الذكاء الاصطناعي من تسجيل ردودهم. وفي هذه الحالة، يجب تزويد المرشحين الذين لا يقدمون موافقتهم بطريقة بديلة ومكافئة للتقييم (على سبيل المثال، إجراء مقابلة شخصية تقليدية بدلاً من مقابلة افتراضية يتم تحديد درجتها بواسطة الذكاء الاصطناعي). وعند حدوث ذلك، يجب ألا يكون قرار المرشح بعدم تقديم الموافقة ضاراً بفرصه في الاختيار.

السؤال السادس:

كيف يتم إبلاغ المرشحين؟

تطبيق السؤال السادس على دراسة الحالة الافتراضية

عند مراجعة تقييم الذكاء الاصطناعي لكيفية إبلاغ المرشحين، ابحث عن وصف أو لقطة شاشة لما يراه المرشح. هل هناك بيان بأنه سيتم استخدام الذكاء الاصطناعي؟ هل يوجد وصف لكيفية عمل الذكاء الاصطناعي؟ هل هناك طلب الحصول على موافقة مستنيرة من المرشح؟ هل يوجد وصف لخيار بديل للتقييم إذا لم يقدم المرشح موافقته على أن يتم تقييم درجة ردوده بواسطة الذكاء الاصطناعي؟

بالنسبة لمثال فحص السيرة الذاتية، نفترض أنه يتم توفير إشعار على الصفحة التي يقوم مقدم الطلب بتحميل سيرته الذاتية عليها؛ حيث ينص هذا الإشعار على أنه سيتم استخدام خوارزمية الذكاء الاصطناعي، التي تم تطويرها للمساعدة في تحديد الأفراد المناسبين للوظيفة بناءً على المعلومات الواردة في سيرتهم الذاتية، وذلك لإنتاج درجة مناسبة للوظيفة من محتوى سيرتهم الذاتية، كما يوضح هذا الإشعار أن هذه النتيجة ستكون مرئية لفريق الموارد البشرية في الشركة وسيتم استخدامها، إلى جانب معلومات أخرى مأخوذة من طلبهم، لتحديد ما إذا كان الفرد سيتم دعوته للمشاركة في الخطوة التالية من عملية الاختيار.

ومع ذلك، نفترض أيضاً أن هذا موقف يجب أن تكون فيه الموافقة المستنيرة مطلوبة لاستخدام تقييم الذكاء الاصطناعي. ونفترض أنه لا يتم تقديم أي معلومات للمرشحين بشأن كيفية عمل الذكاء الاصطناعي، ولا يُطلب منهم تقديم الموافقة على أن يتم تحديد الدرجات بواسطة الذكاء الاصطناعي، ولا يوجد ذكر لإجراء تقييم بديل يمكن استخدامه لتسجيل المرشحين الذين يفضلون ذلك. بناءً على هذه الافتراضات، سوف يشتمل التقييم على درجة عالية لإبلاغ المرشح باستخدام الذكاء الاصطناعي، ولكن ستكون هناك درجات منخفضة للمقاييس الأخرى الواردة في ورقة التقييم المعروضة في الجدول 11.

السؤال السادس:

كيف يتم إبلاغ المرشحين؟

ورقة تقييم الذكاء الاصطناعي: **إبلاغ المرشح**

المقياس	الوصف	الدرجة
إبلاغ المرشح باستخدام الذكاء الاصطناعي	أ. وصف لدرجة إبلاغ المرشحين باستخدام الذكاء الاصطناعي متاح (نعم = 1، لا = 0). ب. الكشف عن استخدام الذكاء الاصطناعي ملائم للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: 1 ب: 3
إبلاغ المرشحين بكيفية عمل الذكاء الاصطناعي	أ. يتوفر وصف لدرجة إبلاغ المرشحين بكيفية عمل الذكاء الاصطناعي (نعم = 1، لا = 0). ب. الكشف عن كيفية عمل الذكاء الاصطناعي ملائم للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: 0 ب: 0
ضرورة الحصول على موافقة مستنيرة من المرشح	أ. يتوفر وصف بما إذا كانت الموافقة المستنيرة مطلوبة من المرشحين، ولماذا/ولم لا (نعم = 1، لا = 0). ب. إدراج أو استبعاد الموافقة المستنيرة ملائم للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: 0 ب: 0
يوجد تقييم بديل متاح	أ. يتوفر وصف بما إذا كان يمكن استخدام نموذج بديل للتقييم لأولئك الذين يرفضون تحديد درجاتهم بواسطة الذكاء الاصطناعي، ولماذا/ولم لا (نعم = 1، لا = 0). ب. إدراج أو استبعاد شكل بديل من التقييم أمر ملائم (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: 0 ب: 0
النتيجة الإجمالية لإبلاغ المرشح		16/4

الجدول (11) ورقة تقييم لتقييم إبلاغ المرشح.

بطاقة قياس استخدام الجهات الحكومية للذكاء الاصطناعي لتقييم المواهب

في هذا القسم، نقدم أوراق تقييم فارغة لكل سؤال من الأسئلة التي يمكن استخدامها للحكم على مدى ملاءمة تقييمات الذكاء الاصطناعي.

السؤال الأول: إلى أي مدى تعتبر بيانات التدريب ملائمة؟		
المقياس	الوصف	الدرجة
نوعية البيانات	أ. يتوفر وصف تفصيلي لأسباب اختيار جميع أنواع البيانات وإدراجها (نعم = 1، لا = 0). ب. مدى ملاءمة نوعية البيانات لحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
كمية البيانات	أ. يتوفر وصف تفصيلي لأسباب حجم العينة المستخدمة لتدريب تقييم الذكاء الاصطناعي (نعم = 1، لا = 0). ب. مدى ملاءمة كمية البيانات لحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
تمثيل البيانات	أ. يتوفر وصف تفصيلي لتمثيل العينة المستخدمة لتدريب تقييم الذكاء الاصطناعي (نعم = 1، لا = 0). ب. مدى تمثيل البيانات لحالة الاستخدام (غير ممثل = 0، تمثيل منخفض = 1، تمثيل متوسط = 2، تمثيل عالٍ = 3).	أ: /1 ب: /3
أمن البيانات	أ. يتوفر وصف تفصيلي لكيفية تخزين البيانات التي تم جمعها من خلال التقييم وحمايتها (نعم = 1، لا = 0). ب. مدى ملاءمة تخزين وحماية البيانات التي تم جمعها لحالة الاستخدام (غير آمنة /غير محمية = 0، أمن/حماية منخفضة = 1، أمن/حماية متوسطة = 2، أمن/حماية عالية = 3)	أ: /1 ب: /3
النتيجة الإجمالية لمدى صلة بيانات التدريب (مجموع جميع الدرجات المعينة في عمود «الدرجة»/مجموع جميع الدرجات الممكنة)		/16

السؤال الثاني: كيف تقوم الخوارزمية باتخاذ القرارات؟		
المقياس	الوصف	الدرجة
المعلومات المستخدمة	أ. يتوفر وصف للمعلومات التي تستخدمها خوارزمية الذكاء الاصطناعي للوصول إلى التنبؤ (نعم = 1، لا = 0). ب. المعلومات ملائمة لحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
ترجيح المعلومات ودمجها	أ. يتوفر وصف للطريقة التي يتم بها ترجيح المعلومات المذكورة أعلاه ودمجها بواسطة خوارزمية الذكاء الاصطناعي للوصول إلى التنبؤ (نعم = 1، لا = 0). ب. المنهجية ملائمة لحالة الاستخدام (غير ملائمة = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
النتيجة الإجمالية للشفافية		/8

السؤال الثالث: ما مدى انحياز القرارات؟		
المقياس	الوصف	الدرجة
فروق جماعية كبيرة في درجات التقييم	أ. يتوفر وصف للمتوسط والانحراف المعياري للدرجات حسب المجموعات المختلفة (على سبيل المثال، العرق والجنس والعمر) (نعم = 1، لا = 0). ب. أي اختلافات في الدرجات بين المجموعات مقبولة لحالة الاستخدام (غير مقبول = 0، قبول منخفض = 1، قبول متوسط = 2، قبول عال = 3).	أ: /1 ب: /3
فروق جماعية كبيرة في دقة التقييم	أ. يتوفر وصف لدقة تنبؤات التقييم حسب المجموعات المختلفة (على سبيل المثال، العرق والجنس والعمر) (نعم = 1، لا = 0). ب. أي اختلافات في التنبؤ بين المجموعات مقبولة لحالة الاستخدام (غير مقبول = 0، قبول منخفض = 1، قبول متوسط = 2، قبول عالٍ = 3).	أ: /1 ب: /3
النتيجة الإجمالية للإنصاف (عدم التحيز)		/8

السؤال الرابع: ما مدى صحة القرارات؟		
المقياس	الوصف	الدرجة
استراتيجية التحقق من صحة التقييم	أ. يتوفر وصف للاستراتيجية المستخدمة للتحقق من صحة التقييم (نعم = 1، لا = 0). ب. استراتيجية التحقق من الصحة ملائمة للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
دراسة التحقق من صحة التقييم	أ. يتوفر وصف لنتائج دراسة التحقق من الصحة (نعم = 1 ، لا = 0). ب. تم تصميم وتنفيذ دراسة التحقق بشكل جيد (غير مقبول = 0، قبول منخفض = 1، قبول متوسط = 2 ، قبول عالٍ = 3).	أ: /1 ب: /3
النتيجة الإجمالية للتحقق من الصحة		/8

السؤال الخامس: إلى أي مدى تكون القرارات نهائية؟		
المقياس	الوصف	الدرجة
فرصة التدخل البشري	أ. يتوفر وصف لمستوى الرقابة البشرية المضمن في التقييم (نعم = 1، لا = 0). ب. مستوى الإشراف البشري ملائم للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
النتيجة الإجمالية للإشراف		/4

السؤال السادس: كيف يتم إبلاغ المرشحين؟		
المقياس	الوصف	الدرجة
إبلاغ المرشح باستخدام الذكاء الاصطناعي	أ. وصف لدرجة إبلاغ المرشحين باستخدام الذكاء الاصطناعي متاح (نعم = 1، لا = 0). ب.الكشف عن استخدام الذكاء الاصطناعي ملائم للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
إبلاغ المرشحين بكيفية عمل الذكاء الاصطناعي	أ. يتوفر وصف لدرجة إبلاغ المرشحين بكيفية عمل الذكاء الاصطناعي (نعم = 1، لا = 0). ب.الكشف عن كيفية عمل الذكاء الاصطناعي ملائم للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
ضرورة الحصول على موافقة مستنيرة من المرشح	أ. يتوفر وصف بما إذا كانت الموافقة المستنيرة مطلوبة من المرشحين، ولماذا/ولم لا (نعم = 1، لا = 0). ب.إدراج أو استبعاد الموافقة المستنيرة ملائم للتقييم وحالة الاستخدام (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
يوجد تقييم بديل متاح	أ. يتوفر وصف بما إذا كان يمكن استخدام نموذج بديل للتقييم لأولئك الذين يرفضون تحديد درجاتهم بواسطة الذكاء الاصطناعي، ولماذا/ ولم لا (نعم = 1، لا = 0). ب. إدراج أو استبعاد شكل بديل من التقييم أمر ملائم (غير ملائم = 0، ملاءمة منخفضة = 1، ملاءمة متوسطة = 2، ملاءمة عالية = 3).	أ: /1 ب: /3
النتيجة الإجمالية لإبلاغ المرشح		/16

نظرة مستقبلية

مستقبل الذكاء الاصطناعي في مجـال الموارد البشريـة

يَعـدُ وصـول التقييـمات القائمـة عـلى الـذكاء الاصطناعـي بإحـداث ثـورة في تقييـم المواهـب، ولكـن لا يـزال يتعين تحديـد مدى تأثير الـذكاء الاصطناعـي في هـذا المجال؛ حيـث مـن الممكـن أن يتـم تجاوز سـقف التحقـق الحالي مـن صحة التقييـمات، وتقليـل التحيـز، وأن تصبـح التقييـمات أكـثر جذبـاً وإمتاعـاً للمرشـحين. وبغـض النظـر عـن النتائـج الدقيقـة، مـن الممكـن أن يكـون تقييـم المواهـب عـلى وشـك تحقيـق نقلـة نوعيـة هائلـة مع دمج الـذكاء الاصطناعـي.

وإلى جانـب التحسـينات في تقييـم المواهـب واختيارهـا، قـد يجلـب الـذكاء الاصطناعـي المزيـد مـن الفوائـد للمـوارد البشريـة مـن خـلال زيـادة الكفـاءة، وأتمتـة المهـام، والقـدرة عـلى اتخـاذ قـرارات اسـتراتيجية مسـتنيرة مـن كميـة البيانـات المتزايـدة باسـتمرار والتـي يمكـن للمـوارد البشريـة الوصـول إليهـا. وسـوف تمُكّـنُ هـذه الفوائـد بدورهـا المـوارد البشريـة مـن الاسـتمرار في تقديـم المزيـد والمزيـد مـن القيمـة للمؤسسـات؛ علاوةً عـلى أنـه مـن المتوقـع أن يحـدث كل هـذا خـلال فـترة زمنيـة قصيـرة نسـبياً.

وعـلى المـدى الطويـل، توقـع أن تـرى التطـور المتزايـد باسـتمرار في تقييـمات الـذكاء الاصطناعـي، الأمـر الـذي سـوف يأخـذ تقييـم المواهـب في عـوالم كل مـن الواقـع المعـزز والافـتراضي. ففـي هـذه العـوالم الافـتراضية، لـن يتمكـن المرشـحون مـن إلقـاء ردودهـم بطريقـة طبيعيـة فحسـب، بـل سـيكونون أيضـاً قادريـن عـلى التحـرك والتصـرف بطريقـة طبيعيـة. ويمكـن أن تـؤدي تقييـمات الـذكاء الاصطناعـي التـي تـم تطويرهـا إلى تسـخير مزيـج مـن التكنولوجيـا الجديـدة والـذكاء الاصطناعـي وعلـوم التقييـم بنجـاح في عمليـات محاكاة غنيـة للغايـة وعاليـة الدقـة تعيـد تحديـد عتبـات المسـتويات المقبولـة مـن التحقـق مـن الصحـة وتجربـة المرشـح.

وكي تفـي تقييـمات الـذكاء الاصطناعـي بهـذه الوعـود، يجـب تطويرهـا واسـتخدامها وفقـاً لإرشـادات وممارسـات قويـة. وينبغـي الإشـارة إلى أنـه مـع اسـتمرار تطـور اللوائـح القانونيـة في جميـع أنحـاء العـالم، يمكـن أن يـؤدي الاسـتخدام غـير الملائـم للـذكاء الاصطناعـي في التقييـمات إلى انتهـاكات قانونيـة وأخلاقيـة، مـما قـد يعيـق بشـكل كبـير تطـوير تقييـمات الـذكاء الاصطناعـي. ويمكـن اسـتخدام الإرشـادات الـواردة في هـذا المسـتند للمسـاعدة في معالجـة المشـهد سريـع التطـور والمعقـد للـذكاء الاصطناعـي في تقييـم المواهـب.

المراجـــع

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. Journal of Applied Psychology, 82, 192206-.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Bologna, M.J. (2019, May 30). 'Hiring robots' restrictions passed by Illinois legislature. Bloomberglaw. https://news.bloomberglaw.com/daily-laborreport/hiring-robots-restrictions-passed-by-illinois-legislature.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155–159.

High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy artificial intelligence. Office for Official Publications of the European Communities.

Liem, C. C., Langer, M., Demetriou, A., Hiemstra, A. M., Wicaksana, A. S., Born, M. P., & König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In H. J. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlütürk, U. Güçlü, & M. van Gerven, (Eds.), Explainable and interpretable models in computer vision and machine learning (pp. 197253-). Cham, Switzerland.

Poole, D. L., & Mackworth, A. K. (2010). Artificial intelligence: Foundations of computational agents. New York: Cambridge University Press.

SHL (2020). The ethical and effective use of artificial intelligence to assess talent. Arlington, VA: Author.

Society for Industrial Organizational Psychology, Inc (2018). Principles for the validation and use of personnel selection procedures (5th ed.) Bowling Green, OH: Author.

The Public Voice (2018). Universal Guidelines for Artificial Intelligence. https://thepublicvoice.org/ai-universal-guidelines/

مسرد المصطلحات

تقييم الذكاء الاصطناعي: في سياق تقييم المواهب، يشير هذا المصطلح إلى التقييمات التي تستخدم الذكاء الاصطناعي. وبشكل أكثر تحديداً، يشير "تقييم الذكاء الاصطناعي" إلى أي تحليل غير بشري لاستجابات المشاركين يقوم باستخدام التعلم الآلي، أو معالجة اللغة الطبيعية (NLP)، أو أساليب وتقنيات النمذجة الأخرى ذات الصلة (مثل التعلم العميق والتحليل الدلالي الكامن) التي تقوم بتحديد الدرجات بناءً على سمات الأشخاص (على سبيل المثال، المعرفة والمهارات والكفاءات)، أو نتائج العمل المتوقعة للأفراد (على سبيل المثال: احتمالية دوران الموظفين).

الخوارزمية: عملية أو سلسلة من الخطوات يتبعها جهاز الحاسوب لإكمال مهمة ما.

الذكاء الاصطناعي (AI): فرع من علوم الكمبيوتر يتعامل مع محاكاة السلوك الذكي في أجهزة الحاسوب.

التحيز: في سياق تقييم المواهب، يشير التحيز إلى صفات التقييم الذي يعاقب بشكل غير عادل مجموعة من المرشحين بسبب الجنس، أو العرق، أو الأصل الإثني، أو العمر، أو حالة الإعاقة، أو غيرها من الخصائص المحمية قانوناً.

الصندوق الأسود: أي نظام ذكاء اصطناعي تكون العمليات الحسابية أو الخوارزميات الأساسية له غير معروفة.

المعيار: في سياق تقييم المواهب، يشير إلى متغير النتيجة الذي يتم على أساسه التحقق من صحة التقييم (عند استخدام دراسة التحقق ذات الصلة بالمعيار).

التعلم العميق: شكل معقد من أشكال التعلم الآلي، يشار إليه أحياناً باسم "الشبكة العصبية الاصطناعية"، وهو مستوحى من بنية وعمل الخلايا العصبية البيولوجية.

الإنصاف: في سياق تقييم المواهب، يعد مصطلح الإنصاف مصطلحاً واسعاً يشمل معاملة جميع المرشحين على نحو متساوٍ، مع المساواة في الوصول إلى الهياكل التي يتم قياسها من خلال التقييم، وممارسات التوظيف غير التمييزية أو نتائج ممارسات التوظيف.

الميـــــــــزات: مصطلح يُستخدم في علوم الحاسوب لتمثيل المتغيرات ا لمستقلة .

التحليـل الوظيفـي: الدراسة المنهجية والتوثيق الخاص بمهام ومسؤوليات الوظيفة، بالإضافة إلى المعرفة والمهارات والقدرات والخصائص الأخرى (KSAO) المطلوبة لأداء الوظيفة.

التعلـــم الآلي: طريقة آلية لتحليل البيانات، والتعرف على الأنماط، وبناء النماذج التي يمكنها التعلم من البيانات واتخاذ القرارات بأقل قدر ممكن من التدخل البشري.

اللغة الطبيعية: أي لغة تطورت بشكل طبيعي من خلال الاستخدام، على عكس لغة الحاسوب.

معالجة اللغة الطبيعية (NLP): حقل فرعي من اللغويات وعلوم الكمبيوتر والذكاء الاصطناعي الذي يدرس معالجة وتحليل بيانات اللغة الطبيعية.

(مجموعة) بيانات الاختبار/العينة العشوائية المحفوظة: البيانات المستخدمة لاختبار (أو التحقق من صحة) نموذج.

(مجموعة) بيانات التدريب: البيانات المستخدمة في تدريب نموذج.

مُستخدم (التقييم): في هذا المستند، يشير مصطلح "مستخدم"، عند الإشارة إلى تقييم ما، إلى فرد داخل مؤسسة يحتاج إلى معلومات التقييم (على سبيل المثال، مسؤول التوظيف أو مدير التوظيف)، ما لم ينص على خلاف ذلك.

التحقق من الصحة: الدرجة التي يدعم بها الدليل والنظرية تفسيرات درجات الاختبار للاستخدامات المقترحة للاختبارات.

www.fahr.gov.ae