< >

إعلان مونتريال

بخصوص الذكاء الاصطناعي المسؤول_

< />

# إعلان مونتريال بخصوص التطوير المسؤول للذكاء الاصطناعي لعام ٢٠١٨

# جدول المحتويات


| | |
|---|---|
| قراءة الإعلان | ٥ |
| تمهيد | ٧ |
| **المبادئ** | |
| ١. مبدأ الرفاهية | ٨ |
| ٢. احترام مبدأ الاستقلال | ٩ |
| ٣. حماية الخصوصية والسرية | ١٠ |
| ٤. مبدأ التضامن | ١١ |
| ٥. مبدأ المشاركة الديمقراطية | ١٢ |
| ٦. مبدأ المساواة | ١٣ |
| ٧. مبدأ إدراج التنوع | ١٤ |
| ٨. مبدأ الحيطة | ١٥ |
| ٩. مبدأ المسؤولية | ١٦ |
| ١٠. مبدأ التنمية المستدامة | ١٧ |
| المسرد | ١٨ |
| اعتمادات | I |
| الشركاء | II |

# قراءة الإعلان

## ما الغرض من الإعلان؟

يتضمن إعلان مونتريال بخصوص التطوير المسؤول للذكاء الاصطناعي ثلاثة أهداف رئيسية، وهي كما يلي:

١. وضع إطار عمل أخلاقي لتطوير الذكاء الاصطناعي ونشره؛

٢. توجيه التحول الرقمي حتى يستفيد الجميع من هذه الثورة التقنية؛

٣. فتح منتدى وطني ودولي للنقاش لتحقيق تنمية متكافئة، وشاملة ومستدامة بيئيًا للذكاء الاصطناعي إجمالاً.

## عمَّ يدور الإعلان؟

### المبادئ

يتضمن الهدف الأول من الإعلان تحديد المبادئ والقيم الأخلاقية التي تعزز الاهتمامات الأساسية للأفراد والمجموعات. تبقى هذه المبادئ السارية على مجال الذكاء الرقمي والاصطناعي عامة ومجرّدة. من المهم أن تضع النقاط التالية في الاعتبار
لقراءتها بشكل صحيح:

> على الرغم من تقديمها في قائمة، فإنه لا يوجد تسلسل هرمي. إذ لا يقل المبدأ الأخير أهمية عن المبدأ الأول. ومع ذلك، من الممكن إيلاء الأهمية لأحد المبادئ عن الآخر، أو اعتبار أحد المبادئ أكثر ملاءمة عن غيره، وفقًا للظروف المحيطة.

> على الرغم من تنوع هذه المبادئ، فإنه يجب تفسيرها باستمرار لمنع حدوث أي تعارض من شأنه أن يمنع تطبيقها. تُحدَّد حدود تطبيق أحد المبادئ، كقاعدة عامة، من خلال مجال تطبيق مبدأ آخر.

> **على الرغم من أنها تعكس الثقافة الأخلاقية والسياسية للمجتمع الذي نشأت به، فإنها توفر الأساس للحوار الثقافي والدولي.**

> **على الرغم من احتمالية تفسيرها بطرق مختلفة، فإنه لا يمكن تفسيرها بأي طريقة. من الضروري أن يكون التفسير متسقًا.**

> **على الرغم من كونها مبادئ أخلاقية، فإنه يمكن ترجمتها إلى لغة سياسية وتفسيرها بأسلوب قانوني.**

بدءًا من تلك المبادئ، وضعت بعض التوصيات التي تهدف إلى اقتراح إرشادات توجيهية لإكمال التحول الرقمي ضمن إطار العمل الأخلاقي للإعلان. تهدف إلى تغطية بعض الموضوعات الرئيسية الشاملة لعدة قطاعات للتفكير في الانتقال نحو مجتمع حيث يساعد الذكاء الاصطناعي في تعزيز المصلحة العامة: الإدارة الخوارزمية، ومحو الأمية الرقمية، والإدراج الرقمي للتنوع والاستدامة البيئية.

## لمن يوّجه الإعلان؟

يوّجه إعلان مونتريال إلى أي شخص، ومؤسسة، وشركة ترغب في المشاركة في التطوير المسؤول للذكاء الاصطناعي، سواء أكان للمساهمة علميًا أو تقنيًا، أو لتطوير المشاريع الاجتماعية، أو لوضع القواعد (اللوائح والقوانين) التي تنطبق عليها، لتتمكن من خوض مناهج سيئة أو غير حكيمة، أو لتتمكن من تنبيه الرأي العام عند اللزوم.

كما أنه موجه إلى الممثلين السياسيين، سواء أُجري انتخابهم أو تنصيبهم، الذين يتوقع منهم مواطنوهم
تقييم تطورات التغييرات الاجتماعية، ووضع إطار عمل سريعًا حتى يسمح بالانتقال الرقمي الذي يخدم المصلحة العامة،
وتوقع المخاطر التي يُمثلها تطوير الذكاء الاصطناعي.

## إعلان وفقًا لأي طريقة؟

نشأ الإعلان من عملية تداول شاملة تبدأ الحوار بين المواطنين، والخبراء، والمسؤولين الحكوميين، والمساهمين في الصناعة، والمؤسسات المدنية والجمعيات المهنية. وتُعد مزايا هذا النهج ثلاثية:

١. التوسط الجماعي للخلافات الاجتماعية والأخلاقية لنظام الذكاء الاصطناعي؛

٢. تحسين جودة التفكير في الذكاء الاصطناعي المسؤول؛

٣. تعزيز شرعية مقترحات الذكاء الاصطناعي المسؤول.

يُعد وضع المبادئ والتوصيات عملاً مشتركًا يتضمن مجموعة متنوعة من المشاركين في الأماكن العامة، أو في مجالس إدارة المؤسسات المهنية، أو في اجتماعات الخبراء الدولية، أو في مكاتب الأبحاث، أو الصفوف الدراسية أو عبر الإنترنت، وبنفس الحزم دائمًا.

## ماذا بعد الإعلان؟

نظرًا لأن الإعلان يهتم بالتقنية التي كانت تتقدم بثبات منذ الخمسينيات، والتي تزداد وتيرة ابتكاراتها الرئيسية بأسلوب أسي، فإنه من الضروري النظر إلى الإعلان بأنه مستند إرشادي مفتوح، يجب مراجعته ومواءمته وفقًا لتطور المعرفة والتقنيات، فضلاً عن تعليق المستخدم على استخدام الذكاء الاصطناعي في المجتمع. في نهاية عملية وضع الإعلان، وصلنا إلى نقطة البداية لإجراء محادثة مفتوحة وشاملة حول مستقبل البشرية الذي تخدمه تقنيات الذكاء الاصطناعي.

# تمهيد

لأول مرة في تاريخ البشرية، يمكن إنشاء أنظمة ذاتية يمكنها تنفيذ المهام المعقدة التي كان يُعتقد أنه يمكن للذكاء الطبيعي وحده تنفيذها، مثل: معالجة كميات كبيرة من المعلومات، والحساب والتنبؤ، والتعلّم، والتكيف مع الظروف المتغيرة، والتعرّف على الأشياء وتصنيفها. نظرًا للطبيعة غير المادية لهذه المهام، وقياسًا على الذكاء البشري، نُجري تعيين هذه الأنظمة واسعة النطاق ضمن الاسم العام للذكاء الاصطناعي. يُشكِّل الذكاء الاصطناعي نموذجًا رئيسيًا للتقدم العلمي والتقني، والذي يمكن أن ينتج عنه فوائد اجتماعية هائلة من خلال تحسين ظروف المعيشة والصحة، وتيسير العدالة، وتكوين الثروات، ودعم السلامة العامة، وتخفيف تأثير الأنشطة البشرية على البيئة والمناخ. لا تنحصر الآلات الذكية على أداء حسابات أفضل من البشر؛ لكن يمكنها أيضًا التفاعل مع الكائنات الحية، والحفاظ على جماعتهم ورعايتهم.

ومع ذلك، يُمثِّل تطور الذكاء الاصطناعي تحديات أخلاقية ومخاطر اجتماعية مهمة. في الواقع، يمكن أن تقيد الآلات الذكية اختيارات الأفراد والمجموعات، وتخفض معايير المعيشة، وتعطل تنظيم العمل وسوق العمل، وتؤثر في السياسة، وتتعارض مع الحقوق الأساسية، وتُفاقم عدم المساواة الاجتماعية والاقتصادية، وتؤثر في الأنظمة البيئية والمناخ والبيئة. على الرغم من أن التقدم العلمي، والحياة المجتمعية دائمًا ما يحملان الخطر، فإن ذلك يرجع إلى المواطنين لتحديد الغايات الأخلاقية والسياسية التي تمنح المخاطر التي نواجهها في عالم يعوزه اليقين أهميةً.

كلما قلت مخاطر نشر الذكاء الاصطناعي، ازدادت فوائده. يتمثل الخطر الأول لتطوير الذكاء الاصطناعي في الإيهام بأنه يمكننا السيطرة على المستقبل من خلال الحسابات. إن تقليص المجتمع إلى سلسلة من الأرقام وحكمه من خلال الإجراءات الحسابية وهم لا يزال يقود طموحات الإنسان. لكن عندما يتعلق الأمر بالشؤون الإنسانية، نادرًا ما يتشابه الغد مع اليوم، ولا يمكن أن تحدد الأرقام القيمة الأخلاقية، ولا ما هو مرغوب اجتماعيًا.

تُعد مبادئ الإعلان الحالي مثل النقاط على البوصلة الأخلاقية التي ستساعد في توجيه تطوير الذكاء الاصطناعي تجاه الغايات الأخلاقية والاجتماعية المرغوبة. كما أنها توفر أيضًا إطار عمل أخلاقيًا يعزز حقوق الإنسان المعترف بها دوليًا في المجالات المتأثرة بتطبيق الذكاء الاصطناعي. بشكل عام، تضع المبادئ المذكورة الأساس لتنمية الثقة الاجتماعية تجاه أنظمة الذكاء الاصطناعي.

هذا وتستند مبادئ الإعلان الحالي إلى الاعتقاد الشائع بأن البشر يسعون إلى النمو بصفتهم كائنات اجتماعية تتمتع بالأحاسيس، والأفكار، والمشاعر، ويسعون لتحقيق إمكاناتهم من خلال ممارسة قدراتهم العاطفية، والأخلاقية، والفكرية بحرية. يتعين على مختلف المساهمين من القطاعين العام والخاص وواضعي السياسات على المستوى المحلي، والقومي، والدولي، ضمان توافق تطوير الذكاء الاصطناعي وانتشاره مع حماية القدرات والأهداف البشرية الأساسية، والمساهمة في تحقيقها بشكل كامل. وسعيًا إلى بلوغ هذا الهدف، يجب على الفرد تفسير المبادئ المقترحة بطريقة منطقية، في ظل مراعاة بعض السياقات الاجتماعية، والثقافية، والسياسية، والقانونية لتطبيقها.

# ١

# مبدأ الرفاهية

**يجب أن يسمح تطوير أنظمة الذكاء الاصطناعي (AIS) واستخدامها بتنمية رفاهية جميع الكائنات الحية.**

١. يجب أن يساعد نظام AIS الأفراد على تحسين ظروفهم المعيشية، وصحتهم وظروف عملهم.

٢. يجب أن يسمح نظام AIS للأفراد بمتابعة تفضيلاتهم، طالما أنهم لا يسببون الضرر لغيرهم من الكائنات الحية.

٣. يجب أن يسمح نظام AIS للأشخاص بممارسة إمكاناتهم العقلية والجسدية.

٤. يجب ألا يصبح نظام AIS مصدرًا للشقاء، ما لم يسمح لنا بتحقيق رفاهية أعلى من التي يمكن أن يحققها الأشخاص بطريقة أخرى.

٥. يجب ألا يساهم استخدام نظام AIS في زيادة التوتر، أو القلق، أو الشعور بالمضايقة من خلال البيئة الرقمية لأحد الأشخاص.

# احترام مبدأ الاستقلال

٢

**يجب تطوير نظام AIS واستخدامه في ظل مراعاة استقلال الأفراد، وباعتزام زيادة تحكم الأفراد في حياتهم ومحيطهم.**

١. يجب أن يسمح نظام AIS للأفراد بتحقيق أهدافهم الأخلاقية ومفهومهم لحياة تستحق العيش.

٢. يجب ألا يطوَّر نظام AIS أو يُستخدم لفرض نمط حياة معين على الأفراد، سواء كان ذلك بطريقة مباشرة أو غير مباشرة،
من خلال تنفيذ طرق قمعية للمراقبة والتقييم
أو آليات تحفيزية.

٣. يجب ألا تستخدم المؤسسات العامة نظام AIS لتعزيز مفهوم معين للحياة الجيدة أو تشويهها.

٤. من الضروري تمكين المواطنين فيما يتعلق بالتقنيات الرقمية من خلال ضمان الوصول إلى نماذج ذات صلة
من المعرفة، لتعزيز تعلُّم المهارات الأساسية (محو الأمية الرقمية والإعلامية)، وتعزيز تنمية التفكير النقدي.

٥. يجب ألا يطوَّر نظام AIS لنشر معلومات، أو أكاذيب أو دعاوى غير جديرة بالثقة، ويجب تصميمه بهدف احتواء نشرها.

٦. يجب أن يتجنب تطوير نظام AIS إنشاء التبعية عبر تقنيات جاذبة للانتباه أو محاكاة
الصفات البشرية (المظهر، والصوت، وغير ذلك) بطرق من شأنها أن تسبب التباسًا بين نظام AIS والبشر.

٣

# مبدأ
# حماية الخصوصية والسرية

**يجب حماية الخصوصية والسرية من تدخل نظام AIS وأنظمة الاستحواذ على البيانات أرشفتها (DAAS).**

١. يجب حماية المساحات الشخصية التي لا يخضع فيها الأشخاص للمراقبة أو التقييم الرقمي من
تدخل نظام AIS وأنظمة الاستحواذ على البيانات وأرشفتها (DAAS).

٢. يجب حماية سرية الأفكار والانفعالات بشكل صارم من استخدامات نظام AIS وأنظمة DAAS القادرة على إلحاق الضرر، ولا سيما الاستخدامات التي تفرض أحكامًا أخلاقية على خيارات الأشخاص
وأنماط حياتهم.

٣. يجب أن يتمتع الأشخاص دائمًا بحق قطع الاتصال الرقمي
في حياتهم الخاصة، ويجب أن يوفر نظام الذكاء الاصطناعي صراحة الخيار لقطع الاتصال في فترات زمنية منتظمة، دون تشجيع الأشخاص
على البقاء متصلين.

٤. يجب أن يتمتع الأفراد بسيطرة واسعة على المعلومات المتعلقة بتفضيلاتهم. يجب ألا ينشئ نظام AIS ملفات تعريف تفضيلية للتأثير على سلوك الأفراد دون موافقتهم الحرة والمطلعة.

٥. يجب أن تضمن أنظمة DAAS سرية البيانات وإخفاء
هوية الملف التعريفي الشخصي.

٦. يجب أن يكون كل شخص قادرًا على ممارسة السيطرة الواسعة على بياناته الشخصية، ولا سيما عندما يتعلق الأمر بجمعها، واستخدامها، ونشرها. يجب ألا يكون وصول الأفراد إلى نظام AIS والخدمات الرقمية مشروطًا بالتخلي عن التحكم ببياناتهم الشخصية أو ملكيتها.

٧. يجب أن يتمتع الأفراد بحرية التبرع ببياناتهم الشخصية للمؤسسات البحثية للمساهمة في تحسين المعرفة.

٨. يجب ضمان سلامة الهوية الشخصية. يجب ألا يُستخدم نظام الذكاء الاصطناعي لمحاكاة مظهر أحد الأشخاص، أو صوته أو أي صفات فردية أخرى أو تغييرها للإضرار بسمعته أو التلاعب بأشخاص آخرين.

٤

# مبدأ
# التضامن

**يجب أن يتوافق تطوير نظام AIS مع الحفاظ على أواصر التضامن بين الأشخاص والأجيال.**

١. يجب ألا يهدد نظام AIS الحفاظ على تحقيق العلاقات الإنسانية المعنوية والنفسية، ويجب تطويره بهدف تعزيز هذه العلاقات والحد من تعرض الأشخاص للضعف والعزلة.

٢. ينبغي تطوير نظام AIS بهدف التعاون مع البشر في المهام المعقدة، وتعزيز العمل التعاوني بين البشر.

٣. ينبغي ألا يُنفَّذ نظام AIS لاستبدال الأشخاص في المهام التي تتطلب علاقات إنسانية جيدة، بل ينبغي تطويره لتيسير هذه العلاقات.

٤. ينبغي أن تضع أنظمة الرعاية الصحية التي تستخدم نظام AIS في اعتبارها أهمية علاقات المريض مع الأسرة وموظفي الرعاية الصحية.

٥. ينبغي ألا يُشجع تطوير نظام AIS على السلوك القاسي تجاه الروبوتات المُصممة لتُماثل مظهر البشر أو المخلوقات غير البشرية أو سلوكهم.

٦. ينبغي أن يساعد نظام AIS في تحسين إدارة المخاطر، وتهيئة الظروف لبناء مجتمع يتمتع بتوزيع أكثر إنصافًا وتبادلاً للمخاطر الفردية والجماعية.

٥

# مبدأ
# المشاركة الديمقراطية

**ينبغي أن يفي نظام AIS بمعايير المفهومية والقدرة على التبرير وإمكانية الوصول، وأن يخضع للتدقيق الديمقراطي والنقاش والمراقبة.**

١. يجب أن تكون عمليات AIS المعنية باتخاذ القرارات المؤثرة على حياة الشخص، أو نوعية حياته، أو سمعته مفهومة لمبتكريها.

٢. يجب أن تكون القرارات التي يتخذها نظام AIS والتي تؤثر في حياة الشخص أو نوعيتها، أو سمعته مبررة دائمًا بلغة يفهمها الأشخاص ممن يستخدمونها أو ممن يتعرضون لعواقب استخدامها. يتمثل التبرير في جعل الشفافية أهم العوامل والمعايير التي تُشكِّل القرار، ويجب أن يتخذ نفس شكل التبرير
الذي قد نتوقعه من شخص يتخذ نفس القرار.

٣. يجب أن يكون استخدام رمز الخوارزميات، سواءً كان عامًا أو خاصًا في متناول السلطات العامة المختصة وأصحاب المصلحة
لأغراض التحقق والمراقبة.

٤. يجب الإبلاغ عن اكتشاف أخطاء تشغيل نظام AIS، والتأثيرات غير المتوقعة أو غير المرغوب فيها، وانتهاكات الأمان، وتسريبات البيانات إلى الجهات العامة المختصة، وأصحاب المصلحة، والمتأثرين بالوضع.

٥. يجب أن يكون رمز خوارزميات اتخاذ القرارات التي تستخدمها السلطات العامة في متناول الجميع، باستثناء
الخوارزميات التي تُمثِّل خطرًا كبيرًا لخطر جسيم في حالة إساءة استخدامها، وذلك وفقًا لمتطلبات الشفافية في القرارات العامة.

٦. بالنسبة إلى أنظمة AIS العامة التي لها تأثير كبير على حياة المواطنين، فإنه يجب أن تتوفر للمواطنين الفرصة والمهارات لمناقشة المعايير الاجتماعية لنظم AIS وأهدافها والحدود من استخدامها.

٧. يجب أن نكون قادرين في جميع الأوقات على التحقق من إجراء أنظمة AIS لما بُرمجت عليه وتحقيق الغرض من استخدامها.

٨. يجب أن يكون أي شخص يستخدم هذه الخدمة على دراية بما إذا كان نظام AIS قد اتخذ قرارًا متعلقًا به أو قد يمسه.

٩. يجب أن يتمكن أي مستخدم في خدمة تستخدم روبوتات الدردشة على تحديد ما إذا كان يتفاعل مع نظام AIS أو مع شخص حقيقي.

١٠. يجب أن تظل بحوث الذكاء الاصطناعي مفتوحة وفي متناول الجميع.

٦

# مبدأ
# المساواة

**يجب أن يساهم تطوير نظام AIS واستخدامه في إنشاء مجتمع عادل ومنصف.**

١. يجب وضع أنظمة AIS وتدريبها على ألا تخلق التمييز أو تعززه أو تمارسه على أساس، ضمن أمور أخرى، الاختلافات الاجتماعية، أو الجنسية، أو العرقية، أو الثقافية، أو الدينية.

٢. يجب أن يساهم تطوير نظام AIS في القضاء على علاقات الهيمنة بين المجموعات والأشخاص استنادًا إلى الاختلافات في السلطة أو الثروة أو المعرفة.

٣. يجب أن يتمخض عن تطوير نظام AIS فوائد اجتماعية واقتصادية للجميع، من خلال الحد من أوجه عدم المساواة والضعف الاجتماعي.

٤. يجب أن يتوافق تطوير نظام AIS الصناعي مع ظروف العمل المقبولة في كل خطوة من دورة حياته، بدءًا من استخراج الموارد الطبيعية إلى إعادة التدوير، بما في ذلك معالجة البيانات.

٥. يجب اعتبار النشاط الرقمي لمستخدمي نظام AIS والخدمات الرقمية عملاً يسهم في وظيفة الخوارزميات و ينتج عنه قيمة.

٦. يجب ضمان قدرة الجميع على الوصول إلى الموارد الأساسية، والأدوات الرقمية والمعرفة.

٧. يجب أن ندعم تطوير الخوارزميات العامة، والبيانات المفتوحة اللازمة لتدريبها، وتوسيع استخدامها، كهدف يرمي للإنصاف الاجتماعي.

# مبدأ إدراج التنوع

٧

**يجب أن يتوافق تطوير نظام AIS واستخدامه مع وجوب الحفاظ على التنوع الاجتماعي والثقافي، وألا يقيد نطاق خيارات نمط الحياة أو التجارب الشخصية.**

١. يجب ألا يؤدي تطوير AIS واستخدامه إلى تجانس المجتمع من خلال توحيد السلوكيات والآراء.

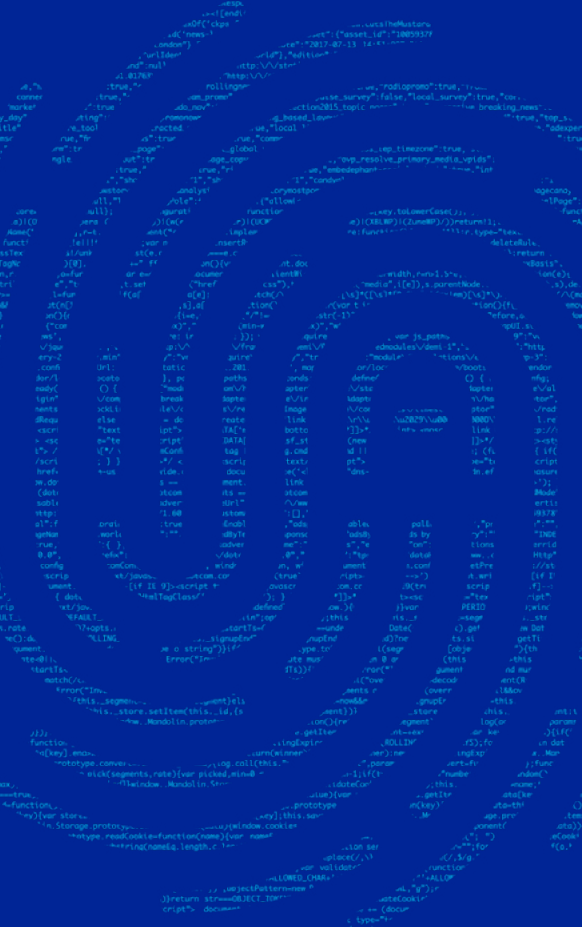
٢. من اللحظة التي يتم فيها تصميم الخوارزميات، يجب أن يضع تطوير AIS في اعتباره العديد من تعبيرات التنوع الاجتماعي والثقافي الموجودة في المجتمع.

٣. يجب أن تتسم بيئات تطوير الذكاء الاصطناعي بالشمولية وتعكس تنوّع الأفراد والجماعات في المجتمع، وهذا سواءً كان في مجال البحث أو الصناعة.

٤. يجب على AIS تجنب استخدام البيانات المكتسبة لإبقاء الأفراد في ملف تعريف المستخدم أو لإصلاح هويتهم الشخصية أو لحصرهم في فقاعة التصفية، والتي من شأنها تقييد إمكاناتهم لتحقيق التنمية الشخصية وحصرها، خاصة في مجالات مثل التعليم، أو العدالة، أو الأعمال التجارية.

٥. يجب ألا يُطوَّر AIS أو يُستخدم بهدف الحد من حرية التعبير عن الأفكار، أو الفرصة لسماع آراء متنوعة، والتي تُعد شروطًا أساسية لإقامة مجتمع ديمقراطي.

٦. يجب تنويع عروض AIS، لكل فئة من الخدمات، لمنع الاحتكارات الفعلية من تشكيل الحريات الفردية وتقويضها.



# مبدأ الحيطة

**٨**

**يجب أن يتوخى كل شخص يشارك في تطوير الذكاء الاصطناعي الحذر من خلال التوقع، بقدر الإمكان، الآثار السلبية لاستخدام نظام AIS واتخاذ التدابير المناسبة لتحاشيها.**

١. من الضروري تطوير آليات تراعي إمكانات الاستخدام المزدوج، المفيدة والضارة، من أبحاث الذكاء الاصطناعي وتطوير AIS (سواء كان عامًا أو خاصًا) للحد من الاستخدامات الضارة.

٢. عندما تُعرِّض إساءة استخدام نظام AIS الصحة العامة أو السلامة للخطر وتتسم باحتمالية كبيرة لحدوثها، فإنه من الحكمة تقييد الوصول المفتوح والنشر العام للخوارزمية.

٣. يجب أن يفي نظام AIS بمتطلبات الموثوقية والأمان، والنزاهة الصارمة، وأن يخضع لاختبارات لا تعرض حياة الأشخاص للخطر أو تضر بنوعية حياتهم أو تؤثر سلبًا في سمعتهم أو سلامتهم النفسية وذلك قبل طرحها في السوق سواء كانت معروضة مقابل رسوم أو مجانًا. يجب أن تكون هذه الاختبارات مفتوحة أمام السلطات العامة المختصة وأصحاب المصلحة.

٤. يجب أن يتفادى تطوير نظام AIS مخاطر إساءة استخدام بيانات المستخدم وحماية سلامة البيانات الشخصية وسريتها.

٥. يجب أن تُشَارك الأخطاء والعيوب المكتشفة في AIS وSAAD علنًا، وعلى نطاق عالمي، من قبل المؤسسات العامة والشركات في القطاعات التي تُشكِّل خطرًا كبيرًا على السلامة الشخصية والتنظيم الاجتماعي.

# ٩

# مبدأ المسؤولية

**يجب ألا يسهم تطوير نظام AIS واستخدامه في تقليل مسؤولية البشر عند ضرورة اتخاذ القرارات.**

١. يمكن أن يتحمل البشر وحدهم المسؤولية عن القرارات الناشئة عن التوصيات التي يقدمها AIS، والإجراءات الناتجة عن ذلك.

٢. يجب أن يتخذ انسان القرار النهائي، في جميع المجالات التي يجب فيها اتخاذ قرار من شأنه التأثير في حياة الشخص أو نوعيتها، أو سمعته، وحين يسمح الوقت والظرف بهذا، وينبغي أن يكون هذا القرار حرًا ومبنيًا على علمٍ تام.

٣. يجب أن يتخذ البشر القرار بقتل شخص ما دائمًا، وألا تنقل مسؤولية اتخاذ هذا القرار إلى أحد أنظمة AIS.

٤. يتحمل الأشخاص ممن يأذنون لنظام AIS بارتكاب جريمة أو جنحة أو السماح لنظام AIS بارتكابها بسبب إهمالهم، مسؤولية هذه الجريمة أو الجنحة.

٥. عندما يُحدث نظام AIS ضررًا أو تلفًا، وثبت أنه يمكن الاعتماد على النظام وأنه قد استُخدم على النحو المنشود، فإنه من غير المعقول إلقاء اللوم على الأشخاص المشاركين في تطويره أو استخدامه.

# مبدأ التنمية المستدامة

١٠

**يجب تنفيذ تطوير نظام AIS واستخدامه لضمان استدامة بيئية قوية لكوكب الأرض.**

١. يجب أن تهدف أجهزة AIS، وبنيتها التحتية الرقمية، والكائنات ذات الصلة التي تعتمد عليها، مثل مراكز البيانات، إلى تحقيق أعلى كفاءة في استخدام الطاقة والحد من انبعاثات الغازات الدفيئة على مدار دورة حياتها بالكامل.

٢. يجب أن تهدف أجهزة AIS، وبنيتها التحتية الرقمية، والكائنات ذات الصلة التي تعتمد عليها، إلى توليد أقل قدر ممكن
من النفايات الكهربائية والإلكترونية، وأن تتكفل بإجراءات الصيانة، والإصلاح، وإعادة التدوير وفقًا
لمبادئ الاقتصاد الدائري.

٣. يجب أن تقلل أجهزة AIS، وبنيتها التحتية الرقمية والكائنات ذات الصلة التي تعتمد عليها، من تأثيرنا في النظم الإيكولوجية والتنوع البيولوجي في كل مرحلة من مراحل دورة حياتها، خاصةً فيما يتعلق باستخراج الموارد، والتخلص النهائي من المعدات عندما تبلغ نهاية مدة صلاحيتها.

٤. يجب أن تدعم كل من الجهات الفاعلة العامة والخاصة التطوير المسؤول بيئيًا لنظام AIS من أجل مكافحة إهدار الموارد الطبيعية والسلع المنتجة، وبناء سلاسل الإمداد والتجارة المستدامة، والحد من التلوث العالمي.

# المسرد

## الخوارزميات
إن الخوارزمية هي طريقة لحل المشكلات من خلال سلسلة محدودة وغير مبهمة من العمليات. وبشكل أكثر تحديدًا في سياق الذكاء الاصطناعي، هي سلسلة العمليات المطبقة لإدخال البيانات
وتحقيق النتيجة المرجوة.

## الذكاء الاصطناعي (AI)
يشير الذكاء الاصطناعي (AI) إلى سلسلة من التقنيات التي تسمح للآلة بمحاكاة تعلم سلوك الإنسان، أي التعلم، والتنبؤ، واتخاذ القرارات، وإدراك البيئة المحيطة بها. يُطبَّق الذكاء الاصطناعي على البيانات الرقمية وذلك في حالة نظام الحوسبة.

## نظام الذكاء الاصطناعي(AIS)
إن نظام الذكاء الاصطناعي هو أي نظام حسابي يستخدم خوارزميات الذكاء الاصطناعي، سواءً كان برنامجًا أو كائنًا متصلاً أو روبوتًا.

## روبوت الدردشة Chatbot
chatbot هو نظام يستخدم الذكاء الاصطناعي ويمكنه التحدث مع مستخدميه بلغة طبيعية.

## نظام الاستحواذ على البيانات وأرشفتها (DAAS)
يُشير DAAS إلى أي نظام حسابي يمكنه جمع البيانات وتسجيلها. تُستخدم هذه البيانات في النهاية لتدريب أنظمة الذكاء الاصطناعي، أو كمعايير لاتخاذ القرار.

## مبررات القرار
يمكن تبرير قرار اتخذه نظام AIS عندما تكون هناك أسباب غير بديهية تحفزه، أو إذا كان يمكن تقديم هذه الأسباب
بلغة طبيعية.

## التعلم العميق
التعلم العميق هو أحد فروع التعلم الآلي الذي يستخدم شبكات عصبية اصطناعية على عدة مستويات. وهو التقنية الكامنة وراء أحدث طفرات الذكاء الاصطناعي.

## المشاعات الرقمية
إن المشاعات الرقمية هي التطبيقات أو البيانات التي ينتجها أحد المجتمعات. وهي قابلة للمشاركة بسهولة ولا تتلف عند استخدامها بخلاف السلع المادية. لذلك، وبخلاف البرامج الاحتكارية، تعتبر البرامج مفتوحة المصدر، والتي غالبًا ما تكون نتيجة للتعاون بين المبرمجين، من المشاعات الرقمية نظرًا لأن شفرة برمجيتها مفتوحة ومتاحة للجميع.

## انقطاع الاتصال الرقمي
يُشير انقطاع الاتصال الرقمي إلى التوقف المؤقت أو الدائم للفرد عن ممارسة النشاط عبر الإنترنت

## المعرفة الرقمية
تُشير المعرفة الرقمية للفرد إلى قدرته على الوصول إلى المعلومات، وإدارتها، وفهمها، ودمجها، وإبلاغها، وتقييمها، وإنشائها بأمان وبشكل مناسب من خلال الأدوات الرقمية والتقنيات الشبكية للمشاركة في الحياة الاقتصادية والاجتماعية.

## فقاعة المرشح
يشير تعبير فقاعة المرشح (أو فقاعة التصفية) إلى المعلومات "المرشحة" والتي تصل
للفرد على شبكة الإنترنت. توفر الخدمات المختلفة مثل الشبكات الاجتماعية أو محركات البحث نتائج مخصصة لمستخدميها. يمكن أن يكون لهذا نفس تأثير عزل الأفراد (داخل "الفقاعات") إذ لم يعد بإمكانهم الوصول إلى المعلومات المشتركة.

## GAN
اختصار لـ Generative Adversarial Network (شبكات التنافس الإنتاجي). في شبكات التنافس الإنتاجي، تتنافس شبكتان متعارضتان لاستحداث صورة. على سبيل المثال، يمكن استخدامها لإنشاء صورة، أو تسجيل، أو فيديو ليبدو حقيقيًا للإنسان.

## المفهومية
يكون AIS واضحًا عندما يملك الإنسان المعرفة اللازمة لفهم عملياته، وهذا يعني نموذجه الحسابي والعمليات التي يحددها.

## التعلم الآلي

التعلم الآلي هو أحد فروع الذكاء الاصطناعي الذي يهتم بتصميم خوارزميات تسمح للنظام بالتعلم من تلقاء نفسه.

يمكن تصنيف التقنيات المختلفة إلى ثلاثة أنواع رئيسية من للتعلم الآلي:

> يتعلم نظام الذكاء الاصطناعي (AIS) التنبؤ بقيمة من البيانات المدخلة، خلال التعلم الخاضع للإشراف. ويتطلب ذلك شرح قيمة الأزواج المُدخلة في أثناء التدريب. على سبيل المثال، يمكن للنظام أن يتعلم التعرف على كائن مميز يظهر في
صورة.

> في التعلم غير الخاضع للإشراف، يتعلم نظام AIS كيفية العثور على أوجه التشابه بين البيانات غير المشروحة، على سبيل المثال من أجل تقسيمها إلى عدة أقسام متجانسة. ومن ثمَّ يمكن للنظام التعرف على مجتمعات مستخدمي وسائل التواصل الاجتماعي.

> يمكن لنظام AIS، من خلال تعزيز التعلم، العمل على بيئته من أجل زيادة المكافأة التي يحصل عليها في أثناء التدريب. تُعد هذه التقنية التي تمكن من خلالها نظام AIS التغلب على البشر في لعبة Go أو لعبة الفيديو Dota٢.

## النشاط عبر الإنترنت

يُشير النشاط عبر الإنترنت إلى جميع الأنشطة التي يؤديها فرد ما في بيئة رقمية، سواءً كانت تلك الأنشطة تمارس على جهاز كمبيوتر أو هاتف أو غيرها من الكائنات المتصلة.

## البيانات المفتوحة

إن البيانات المفتوحة هي بيانات رقمية يمكن للمستخدمين الوصول إليها بحرية. على سبيل المثال، تُعد هذه حالة معظم نتائج أبحاث الذكاء الاصطناعي المنشورة.

## تبعية المسار

هي آلية اجتماعية تستمر من خلالها القرارات التكنولوجية أو التنظيمية أو المؤسسية، التي كانت تعتبر معقولة من قبل ولكنها تتدنى الآن، ولا تزال مستمرة في التأثير على صنع القرارات. ويُجرى الحفاظ على هذه الآلية بسبب التحيز المعرفي أو بسبب التغيير الذي قد يتطلب قدرًا كبيرًا من المال أو الجهد. وينطبق ذلك على البنية التحتية للطرق الحضرية عندما يؤدي إلى إنشاء برامج تحسين حركة المرور، بدلاً من التفكير في تغيير تنظيم النقل بانبعاثات منخفضة للغاية من الكربون. يجب أن تكون هذه الآلية معروفة عند استخدام الذكاء الاصطناعي في المشروعات الخاصة، إذ قد تعزز بيانات التدريب على التعلم الخاضع للإشراف أحيانًا استخدام النماذج التنظيمية القديمة محل النقاش في الوقت الحالي.

## البيانات الشخصية

إن البيانات الشخصية هي تلك التي تساعد في التعرف على الشخص بطريقة مباشرة أو غير مباشرة.

## التأثير الارتدادي

التأثير الارتدادي هي الآلية التي تؤدي من خلالها زيادة كفاءة استخدام الطاقة، أو تحسين الأداء البيئي للمنتجات والمعدات والخدمات إلى زيادة الاستخدام بشكل أكثر تناسبًا. على سبيل المثال، يزداد حجم الشاشة،
ويرتفع عدد الأجهزة الإلكترونية في المنزل،
وتُقطع المسافات الأكبر بالسيارة أو الطائرة.
وتكون النتيجة الشاملة هي زيادة الضغط على الموارد والبيئة.

## المصداقية

يمكن الاعتماد على نظام AIS عندما يؤدي المهمة التي صمم لها، بالطريقة المرجوة منه. المصداقية هي احتمالية النجاح التي تتراوح بين نسبتي ٥١٪ و١٠٠٪، وهذا يعني إظهار التفوق حين تسنح الفرصة. كلما كان النظام أكثر مصداقية، زادت قدرة التنبؤ بسلوكه.

## الاستدامة البيئية القوية

تعود فكرة الاستدامة البيئية القوية إلى وجوب أن يكون استهلاك الموارد الطبيعية والانبعاثات الملوثة متوافقًا مع الحدود البيئية للكوكب، ووتيرة الموارد، وتجديد النظام البيئي، واستقرار المناخ، من أجل أن تكون مستدامة.

إن الاستدامة القوية لا تسمح باستبدال الموارد الطبيعية المفقودة بموارد مصطنعة، على عكس الاستدامة الضعيفة التي تتطلب جهدًا أقل.

## التنمية المستدامة

تُشير التنمية المستدامة إلى تطور المجتمع البشري الذي يتوافق مع قدرة النظم الطبيعية على توفير الموارد والخدمات اللازمة لهذا المجتمع. وهي التنمية الاقتصادية والاجتماعية التي تلبي الاحتياجات الحالية دون تعريض وجود الأجيال المقبلة للخطر.

## التدريب

التدريب هو عملية التعلم الآلي التي من خلالها يُصمم نظام AIS نموذجًا من البيانات يعتمد أداء AIS على جودة النموذج، والذي بدوره يعتمد على كمية البيانات المستخدمة وجودتها في أثناء التدريب.

# الاعتمادات

إن كتابة إعلان مونتريال للتنمية المسؤولة للذكاء الاصطناعي هو نتاج عمل فريق علمي جامعي مشترك متعدد التخصصات يعتمد على عملية التشاور فيما بين المواطنين والحوار مع الخبراء وأصحاب المصلحة في تطوير الذكاء الاصطناعي.

**Christophe Abrassart**، أستاذ مساعد في كلية التصميم، والمدير المشارك للمختبر Lab Ville Prospective التابع لكلية التخطيط بجامعة مونتريال، وعضو في مركز Centre de recherche en éthique ((CRÉ

**Yoshua Bengio**، أستاذ محاضر في قسم علوم الحاسوب وبحوث العمليات، UdeM، ومدير الشئون العلمية في MILA وIVADO

**Guillaume Chicoisne**، مدير البرامج العلمية في IVADO

**Nathalie de Marcellis-Warin**، أستاذة محاضرة في جامعة بوليتكنيك مونتريال ( Polytechnique Montréal)، والرئيس والمدير التنفيذي لمركز البحوث المشتركة بين الجامعات والتحليل بين المنظمات (CIRANO)

**Marc-Antoine Dilhac**، أستاذ مساعد في قسم الفلسفة، بجامعة مونتريال، ورئيس مجموعة الأخلاقيات والسياسة في مركز Centre de recherche en éthique (CRÉ)، رئيس الأبحاث الكندية في الأخلاقيات العامة والنظرية السياسية ومدير معهد Institut Philosophie Citoyenneté Jeunesse

**Sébastien Gambs**، أستاذ علوم الحاسوب بجامعة كيبيك في مونتريال، كندا، ورئيس برنامج البحث الكندي للحفاظ على الخصوصية والتحليل الأخلاقي للبيانات الضخمة.

**Vincent Gautrais**، أستاذة محاضر في كلية الحقوق بجامعة مونتريال؛ ومدير مركز de recherche en droit public (CRDP)؛ رئيس L.R. Wilson Chair في تكنولوجيا المعلومات وقانون التجارة الإلكترونية.

**Martin Gibert**، مستشار الأخلاقيات في IVADO، وباحث في مركز Centre de recherche en éthique (CRÉ)

**Lyse Langlois**، أستاذ محاضر ونائب عميد كلية العلوم الاجتماعية، ومدير معهد Institut d'éthique appliquée (IDÉA)؛ وباحث في مركز البحوث المشتركة بين الجامعات على العولمة والعمل (CRIMT)

**François Laviolette**، أستاذ محاضر في قسم علوم الكمبيوتر وهندسة البرمجيات، جامعة لافال؛ ومدير مركز Centre de recherche en données massives (CRDM)

**Pascale Lehoux**، أستاذة محاضرة في كلية الصحة العامة بجامعة مونتريال (ESPUM)؛ رئيسة قسم الابتكار المسؤول في الصحة.

**Jocelyn Maclure**، أستاذ محاضر في كلية الفلسفة بجامعة لافال، ورئيس لجنة الأخلاقيات في العلوم والتكنولوجيا في كيبيك (CEST)

**Marie Martel**، أستاذة الدراسات العليا في كلية École de bibliothéconomie et des sciences de l'information، بجامعة مونتريال

**Joëlle Pineau**، أستاذة مساعدة في كلية علوم الحاسوب بجامعة ماكجيل؛ ومديرة Facebook AI Lab في مونتريال؛ والمديرة المشاركة في مختبر Reasoning and Learning Lab

**Peter Railton**، أستاذ جامعي متميز حاصل على أستاذية Gregory S. Kavka؛ وعلى أستاذية John Stephenson Perrin، بقسم الفلسفة بجامعة ميشيجان، وزميل الأكاديمية الأمريكية للفنون والعلوم

**Catherine Régis**، أستاذة مساعدة في كلية الحقوق بجامعة مونتريال؛ ورئيسة البحث الكندي في الثقافة التعاونية في القانون والسياسة الصحية؛ وباحثة منتظمة في مركز Centre de recherche en droit public (CRDP)

**Christine Tappolet**، أستاذة محاضرة في قسم الفلسفة، UdeM، ومديرة مركز Centre de recherche en éthique ((CRÉ

**Nathalie Voarino**، حاصلة على الدكتوراه في أخلاقيات علم البيولوجيا في جامعة مونتريال

# شركاؤنا

Université de Montréal

CRE
CENTRE DE RECHERCHE EN ETHIQUE

LAB VILLE PROSPECTIVE

IVADO
INSTITUT DE VALORISATION DES DONNÉES

SAT

Mila

ICRA
Programme
IA et
société

Université de Montréal | designNsociété

polEtHics
CHAIRE DE RECHERCHE DU CANADA
ÉTHIQUE PUBLIQUE ET THÉORIE POLITIQUE

CIRANO
*Allier savoir et décision*
Centre interuniversitaire de recherche en analyse des organisations

Québec
Fonds de recherche – Nature et technologies
Fonds de recherche – Santé
Fonds de recherche – Société et culture

UNIVERSITÉ LAVAL
Institut d'éthique appliquée

MUSÉE DE LA CIVILISATION
Québec

CENTRE D'ÉTUDES ET DE RECHERCHES INTERNATIONALES
CÉRIUM
Université de Montréal

crdm_ul
CENTRE DE RECHERCHE EN DONNÉES MASSIVES DE L'UNIVERSITÉ LAVAL

Canada

Centre Culturel Canadien Paris
Canadian Cultural Centre Paris

MEC
Maison des étudiants canadiens

< >

# Montréal Declaration Responsible AI_

</ >

montrealdeclaration-responsibleai.com