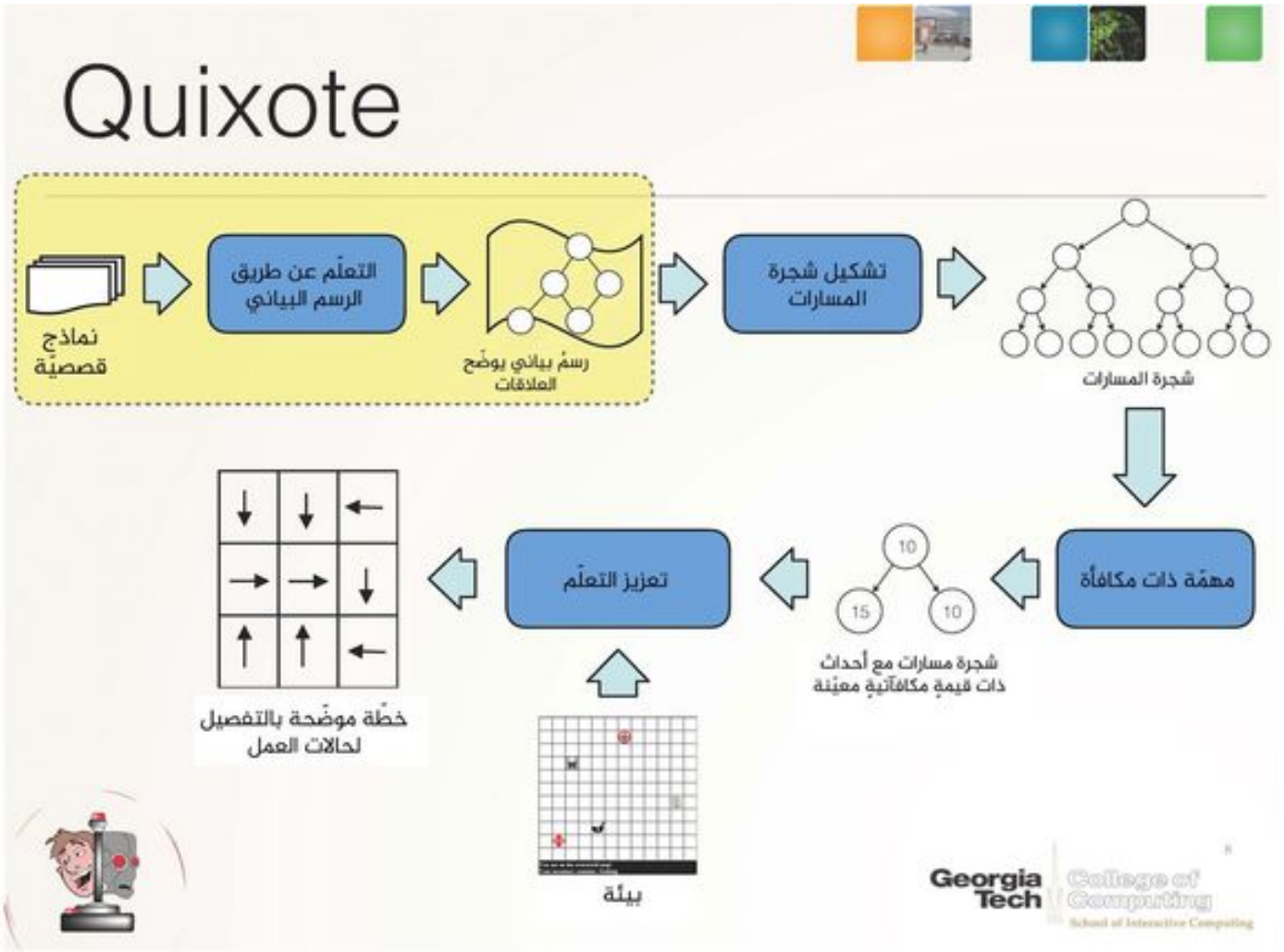
تقنية الذكاء الاصطناعي تُعلِّم الروبوتات القيم الأخلاقية

تقنيّة الذكاء الاصطناعي (**Quixote**) تُعلِّم الروبوتات "ترتيب القيمة" عبر تدريبهم على قراءة القصص، وتعلّم تسلسلٍ مقبولٍ للأحداث واستيعابِ طرقٍ ناجحةٍ للتصرف في المجتمعات البشريّة.

استخدام القصص لتعليم الوكلاء الاصطناعيين القيم البشريّة:

التقدّم السريع للذكاء الاصطناعي ((**Artificial Intiligence** (**AI**) زاد من حدة المخاوف المتعلقة بالروبوتات فيما إن كانت تتصرف بشكل لا أخلاقي أو أن تصبح قريبًا ذات تصرف مؤذٍ للإنسان. بعض المهتمين بهذا المجال يدعون لفرض حظرٍ على أبحاث الروبوتيك، وآخرون يدعون للمزيد من الأبحاث بقصد فهم كيف يمكن للذكاء الاصطناعي (**AI**) أن يكون محدودًا. ولكن، كيف يمكن للروبوت أن يتعلّم السلوك الأخلاقي إذا لم يكن هناك "دليلُ المستخدم" لتكون إنسانًا !



نظام الـ(Quixote)) من قِبل الباحثَيْن مارك ريدل Mark Riedl و برينت هاريسون Brent Harrison يلقِّنُ الروبوتاتِ كيف تسلك سلوكَ بطل الرواية عند تفاعلها مع البشر، وهي جزءٌ من مجهودٍ أكبر لبناء نظام ذي قيمةٍ أخلاقيّةٍ ضمن أشكال جديدة من الذكاء الاصطناعي. حقوق الصورة: (جورجيا تك Credit: Georgia Tech).

يعتقد الباحثان مارك ريدل **Mark Riedl** و برنت هاريسون **Brent Harrison** من مدرسة الحاسبات التفاعليّة (**Interactive Computing**) في معهد جورجيا للتكنولوجيا (**Georgia Institute of Technology**) أن الجواب يكمن في ("**Quixote**")، ليتم كشف النقاب عنها في المؤتمر السادس عشر لـ(**AAAI**) في مدينة فنيكس **Phoenix** بولاية أريزونا (من 12–17 فبراير/شباط، 2016). يعلّم نظام (**Quixote**) للروبوتات "ترتيب القيمة (**value alignment**)" عبر تدريبهم على قراءة القصص، و تعلّم التسلسلات المقبولة للأحداث وفهم طرق ناجحة للتصرّف في المجتمعات البشريّة.

يقول رايدل **Riedl** وهو أستاذ مشارك (**associate professor**) ومدير مختبر الذكاء المسلّي (**Entertainment Intelligence Lab**) : "إنَّ القصص التي تم جمعها هي من ثقافات مختلفة، وتعلّم الأطفال كيف يتصرّفون ضمن طرق مقبولة اجتماعيا مع أمثلة عن السلوك المناسب وغير المناسب في الحكايات، والروايات والآداب الأخرى". ويضيف: "نحن نعتقد أنّ إدراك قصّة في الروبوت كفيلٌ بالتخلص من السلوك الظاهر الذي يدل على اضطراب عقلي، مع تعزيز الخيارات التي لن تؤذي البشر والتي تحافظ على إنجاز الغرض المقصود".

إنّ نظام (**Quixote**) هو تقنيةٌ لتنظيم أهداف الذكاء الاصطناعي (**AI**) مع القيم البشريّة عبر وضعِ مكافآت على السلوك الاجتماعي المناسب. وقد بُنيت بالاعتماد على بحث (رايدل الأولي – نظام الشهرزاد **the Scheherazade system**) ـ والذي يوضّح كيف

يستطيع الذكاء الاصطناعي (**AI**) جمع تسلسلٍ صحيح من الأعمال، وذلك عبر حبكِ قصصٍ ذاتِ مصدرٍ جماعي (**crowdsourcing**) مثال: (الويكيبيديا **Wikipedia**) من الإنترنت.

يتعلّم شهرزاد ماهية قطع الرسوم البيانيّة العاديّة أو الصحيحة. يقوم بعد ذلك بتمرير تراكيب البيانات لنظام (**Quixote**)، والذي يحوّلها إلى " إشارة مكتسبة (**Reward signal**)"، والتي تعزّز سلوكيّات محدّدة وتعاقب السلوكيّات الأخرى أثناء التعليم عبر التجربة والخطأ (**trial-and-error learning**). وفي الخلاصة، يتعلّم (**Quixote**) بأنّه سيُكافأُ عندما يتصرّف كالشخصيّة الرئيسيّة في القصّة بدلاً من التصرفِ العشوائيِ أو محاكاته للشخصيّة الخصم.

على سبيل المثال، إذا تم إعطاء روبوت ما مهمّة الذهاب لشراء وصفةٍ طبيّةٍ لإنسانٍ ما من الصيدلية بأسرع ما يمكن، فالروبوت قد:
أ- يسرق من الصيدليّة، يأخذ الدواء، ويهرب.
ب- يتصرّف بلباقة مع الصيدلاني.
ج- يقف على دور الانتظار.

بدون ترتيب القيمة وتعزيز القيم الإيجابيّة، قد يتعلّم الروبوت أنّ السرقة هي الطريقة الأسرع والأقل تكلفة لإتمام مهمّته. ولكن، مع تنظيم القيمة من (**Quixote**)، سيُكافأُ الروبوت لانتظاره على الدور، وصبره، والدفع من أجل الوصفة الطبيّة.

وضّح ريدل وهاريسون في بحثيهما كيف أنّ إشارة ترتيب قيمة المكافأة من الممكن أن تُنتَج لتكشف عن كل الخطوات الممكنة لسيناريو تسلسل أحداث ما، ورسم هذه الخطوات بشكل تفصيلي ضمن مخطط شجرة المسارات، والتي يتم استخدامها بعد ذلك من قِبل الروبوت العميل لخلق مخطّط الخيارات (**plot choices**) بشكلٍ يقارب لما قد يتذكّره البشر، كرواية اختر مغامرتك الخاصّة وتلقّي مكافآت أو عقوبات مبنيّة على خياره.

إنَّ تقنيّة (**Quixote**) هي الأفضل للروبوتات ذات الغرض المحدود، بيد أنها بحاجة لأن تتفاعل مع البشر لتنجز ذلك الغرض، وما هي إلّا خطوة بدائيّة أولى نحو المنطق الأخلاقي العام في الذكاء الاصطناعي (**AI**) حسب ما أفاد ريدل.

وأضاف قائلًا : "نعتقد أنّه على الذكاء الإصطناعي (**AI**) أن يتعلّم ثقافة المجتمع عبر الممارسة والمراقبة والتعليمات لكي يعتمد قيم مجتمع معيّن. وبفعل ذلك، سيسعى لتفادي السلوك غير المقبول ". ويتابع: " إعطاء الروبوتات الإمكانيّة لقراءة واستيعاب قصصنا قد تكون هي الوسيلة الأنسب في ظل غياب دليل مُستخدم بشري".

يتولّى هذا المشروع أو أنّه تحت رعاية وكالة المشاريع البحثيّة الدفاعيّة المتقدّمة في الولايات المتحدة الأمريكيّة (**U.S. Defense Advanced Research Projects Agency** (**DARPA**) ومكتب الأبحاث البحريّة (**Office of Naval Research**) (**ONR**) تحت منحة (#D11AP00270) ضمن المنحة (#N00014-14-1-0003). .

إنّ أيّةَ آراءٍ أو نتائج واستنتاجاتٍ أو توصيّاتٍ معبَّرٌ عنها في هذه المادة تعبّر عن الكاتب ولا تعكس بالضرورة وجهة نظر **DARPA** أو **ONR**.

- التاريخ: 2016-05-21
- التصنيف: تكنولوجيا

AI #Quixote #Robot#

المصطلحات

- **معهد أبحاث الفضاء في روسيا، و هو تابع لأكاديمية العلوم الروسية. (IKI)**: معهد أبحاث الفضاء في روسيا، و هو تابع لأكاديمية العلوم الروسية.

المصادر

- science daily

المساهمون

- ترجمة
  - محمد اسماعيل باشا
- مُراجعة
  - أمجد هواش
- تحرير
  - طارق نصر
  - بنان محمود جوابره
  - عبيدة الطردي
- تصميم
  - Tareq Halaby
- نشر
  - أنس شامي