

مايكروسوفت تسهل توليد الفيديو بالذكاء الاصطناعي

الخميس 11 يناير 2024 06:07 م

تقود مايكروسوفت الموجة التالية من توليد الفيديو بالذكاء الاصطناعي، إذ طرحت الشركة نموذجًا جديدًا لتوليد الفيديو بالذكاء الاصطناعي يسمى [DragNUWA](#). ويهدف هذا النموذج إلى توفير تحكم دقيق في توليد الفيديو من خلال استخدام النص والصور والمسار بصفته ثلاثية عوامل تحكم أساسية لتسهيل توليد مقطع فيديو يمكن التحكم فيه بشكل كبير من الجوانب الدلالية والمكانية والزمانية. وتتسابق شركات الذكاء الاصطناعي لإتقان توليد الفيديو بالذكاء الاصطناعي، وأصدر العديد من اللاعبين في هذا المجال على مدار الأشهر القليلة الماضية نماذج قادرة على توليد مقاطع فيديو مختلفة بالاعتماد على المطالبة المستندة إلى النص والمطالبة المستندة إلى الصورة. ويتيح نموذج DragNUWA للمستخدمين بمعالجة الكائنات أو الكائنات داخل الصور مباشرة، ويترجم هذه الإجراءات بسلاسة إلى حركات الكاميرا أو حركات الكائنات، مما يؤدي إلى توليد الفيديو المقابل. ويضيف النموذج التوليد المستند إلى المسار بصفته أسلوبًا جديدًا إلى جانب الأساليب المعروفة التي تشمل المطالبة المستندة إلى النص والمطالبة المستندة إلى الصورة. ويتيح ذلك للمستخدمين معالجة الكائنات أو إطارات الفيديو بأكملها عبر مسارات محددة. ويوفر هذا طريقة سهلة لتوليد فيديو يمكن التحكم فيه بشكل كبير من الجوانب الدلالية والمكانية والزمانية، مع ضمان إخراج عالي الجودة في الوقت نفسه. وقدمت مايكروسوفت [المعايير القابلة للتعليم](#) للنموذج بشكل مفتوح المصدر، ووفرت [عرضًا توضيحيًا](#) للمشروع، مما يسمح للمجتمع بتجربته. ويتمحور توليد الفيديو بالذكاء الاصطناعي حول النص أو الصورة أو المدخلات المستندة إلى المسار، وواجه كل نهج صعوبات في توفير تحكم دقيق في المخرجات المطلوبة. ويفشل الجمع بين النص والصور بمفرده في نقل تفاصيل الحركة المعقدة الموجودة في الفيديو، وقد لا تمثل الصور والمسارات الأشياء المستقبلية بشكل مناسب، ويمكن أن تؤدي النصوص والمسارات إلى الغموض عند التعبير عن المفاهيم المجردة. واقترح فريق الذكاء الاصطناعي في مايكروسوفت في شهر أغسطس 2023 نموذج DragNUWA للتغلب على هذه المشكلة، إذ إنه نموذج قائم على الانتشار مفتوح المدى يجمع بين العوامل الثلاثة. ويتيح ذلك للمستخدمين التحديد الدقيق للنص والصورة والمسار المطلوب في الإدخال للتحكم في جوانب، مثل حركات الكاميرا، ويشمل ذلك تأثيرات التكبير أو التصغير، أو حركة الكائن في الفيديو الناتج. ويوفر المسار تفاصيل الحركة، وتعطي النصوص تفاصيل عن الأشياء المستقبلية، وتضيف الصور التمييز بين الكائنات. وزعمت مايكروسوفت في اختباراتها أن النموذج كان قادرًا على تحقيق حركات دقيقة للكاميرا والكائنات بمسارات سحب مختلفة.