بايدو تطور نظاما لتحويل "النص إلى كلام" قادرا على تقليد الصوت البشري بدقة ولحظيا

الأحد 28 مايو 2017 08:05 م

أعلنت شركة بايدو الصينية عن إطلاق الجيل الثاني من نظام تحويل النص إلى كلام التابع لها "ديب فويس" Deep Voice، وذلك بعد ثلاثة أشهر فقط من إطلاق الجيل الأول، ليأتي "ديب فويس 2" مع تحسينات كبيرة تبشر بمساعدات رقمية تتفاعل مع المستخدمين كما لو كانت أناس حقيقيين.

وكانت عملاقة البحث الصينية قد أطلقت في شهر شباط/فبراير الماضي "ديب فويس 1"، وهو نظام لتوليد الأصوات البشرية الاصطناعية تمامًا باستخدام الشبكات العصبية العميقة.

وقالت بايدو إنه وعلى عكس أنظمة تحويل النص إلى كلام العصبية البديلة، عمل "ديب فويس 1" في الوقت الحقيقي، إذ يُجمِّع الصوت بأسرع ما يمكن لتشغيله، مما يجعله قابلًا للاستخدام في التطبيقات التفاعلية مثل الإعلام وواجهات المحادثة، مثل المساعدات الرقمية.

وأضافت الشركة أنها، ومن خلال تدريب الشبكات العصبية العميقة القادرة على التعلم من كميات كبيرة من البيانات والميزات البسيطة، أنشأت نظامًا مرنًا على نحو لا يصدق لتوليف صوت عالي الجودة في الوقت الحقيقي.

ومع أن نظام "ديب فويس 1" كان قادرًا على إنتاج كلام لا يمكن تمييزه تقريبًا من صوت الإنسان الفعلي في أول استماع، إلا أن قدرات النظام كانت محدودة بالتعلم من صوت واحد فقط في المرة الواحدة، ويتطلب ساعات طويلة من الصوت لبناء عينة.

أما "ديب فويس 2" الجديد، فقد قالت بايدو إنها، وفي غضون ثلاثة أشهر فقط، تمكنت من توسيع نطاق النظام من 20 ساعة من الكلام وصوت واحد إلى مئات الساعات مع المئات من الأصوات مع القدرة على تقليدها تمامًا. هذا بالإضافة إلى قدرة النظارة على التعلم من مئات الأصوات الفريدة في أقل من نصف ساعة من البيانات لكل متكلم مع تحقيق جودة صوت عالية.

وأوضحت الشركة أن "ديب فويس 2" قادر على التعلم لتوليد خطاب من خلال إيجاد الصفات المشتركة بين الأصوات المختلفة. وعلى عكس جميع أنظمة تحويل النص إلى كلام السابقة، يتعلم "ديب فويس 2" هذه الصفات من الصفر، دون أي توجيه بشأن ما يجعل الأصوات قابلة للتمييز.

ونشرت بايدو على قسم الأبحاث على موقعها الإلكتروني مجموعة من العينات من نظام "ديب فويس 2"، الذي دُرِّب بالاستماع على ما يقرب من 100 متكلم. وكان لكل متكلم إيقاع كلام، ولهجة، ونبرة، وعادات النطق خاصة به، وقد تمكن النظام تقليد كل ذلك بالضبط تقريبًا.

وتعتقد بايدو أن هذه التقنية ستكون مفيدة لخدمات المساعدة الرقمية التي يُتحكَّم بها من خلال الأوامر الصوتية وتتفاعل من خلال التحدث مع مستخدميها. كما ترى لها إمكانات في تطبيقات تحويل النص إلى كلام مثل الكتب الإلكترونية.