الدورة التدريبية: تقنيات زيادة البيانات لنماذج الذكاء الاصطناعي القوية

#AI4596

الدورة التدريبية: تقنيات زيادة البيانات لنماذج الذكاء الاصطناعي القوية

مقدمة الدورة التدريبية / لمحة عامة:

يقدم BIG BEN Training Center هذه الدورة التدريبية المتخصصة حول تقنيات زيادة البيانات لنماذج الذكاء الاصطناعي القوية، وهي مصممة لعلماء البيانات، ومهندسي التعلم الآلي، والباحثين، والمطورين الذين يعملون في مجال الذكاء الاصطناعي (AI) والتعلم العميق (Deep Learning)، ويواجهون تحديات نقص البيانات (Data Scarcity)، أو توازن البيانات (Data Imbalance)، أو الحاجة إلى تحسين أداء النماذج. تُعد زيادة البيانات (Data Augmentation) تقنية حيوية لـتوسيع مجموعات البيانات التدريبية بشكل فعال، مما يساعد على تحسين تعميم النماذج (Model Generalization)، وتقليل الانحياز (Bias)، وزيادة قوة نماذج الذكاء الاصطناعي، خاصةً في مهام مثل تصنيف الصور (Image Classification)، ومعالجة اللغات الطبيعية (Natural Language Processing – NLP)، والرؤية الحاسوبية (Computer Vision). تهدف الدورة إلى تزويد المشاركين بـالمعرفة والمهارات اللازمة لـتطبيق مجموعة واسعة من تقنيات زيادة البيانات**، بدءاً من التحويلات البسيطة (Simple Transformations) وصولاً إلى الأساليب المتقدمة المدعومة بالذكاء الاصطناعي مثل الشبكات التوليدية التنافسية (GANs) لـتوليد بيانات اصطناعية (Synthetic Data). ستغطي الدورة مفاهيم أهمية البيانات في الذكاء الاصطناعي، وأنواع زيادة البيانات** (مثل التدوير، والقلب، والقص في الصور؛ والمرادفات، وإعادة الصياغة في النصوص)، وتقنيات زيادة البيانات المتخصصة لكل من البيانات المرئية والنصية والسمعية. سيتعلم المشاركون كيفية اختيار تقنيات زيادة البيانات المناسبة لـمشاكل تعلم آلي معينة، وتطبيقها باستخدام أطر عمل شائعة مثل TensorFlow وPyTorch، وتقييم تأثير زيادة البيانات على أداء النموذج. تهدف الدورة إلى تمكين المختصين من التغلب على قيود البيانات، وبناء نماذج ذكاء اصطناعي أكثر قوة ومرونة، وتحقيق نتائج أفضل في مشاريع الذكاء الاصطناعي. نستلهم في هذه الدورة من أعمال البروفيسور يان ليكون (Yann LeCun)، Yann LeCun، أحد رواد التعلم العميق، الذي أكد على أهمية البيانات الكافية والمتنوعة لـتدريب نماذج فعالة.

الفئات المستهدفة / هذه الدورة التدريبية مناسبة لـ:

- علماء البيانات.
- مهندسي التعلم الآلي.
- مهندسي التعلم العميق.
- الباحثين في مجال الذكاء الاصطناعي.
- مطورين الذكاء الاصطناعي.
- طلاب الدراسات العليا في علوم الحاسوب والذكاء الاصطناعي.
- المهتمين بتحسين أداء نماذج التعلم العميق.
- المتخصصين في الرؤية الحاسوبية.
- المتخصصين في معالجة اللغات الطبيعية.
- المختصين في تحليل البيانات الكبيرة.

القطاعات والصناعات المستهدفة:

- تكنولوجيا المعلومات والبرمجيات.
- البحث والتطوير (R&D).
- الرعاية الصحية (توليد بيانات طبية).
- السيارات ذاتية القيادة (بيانات القيادة).
- الأمن والمراقبة.
- الزراعة الذكية.
- الخدمات المالية (بيانات المعاملات).
- التجارة الإلكترونية (بيانات المنتجات والصور).
- الألعاب (توليد الأصول).
- التعليم والبحث الأكاديمي.

الأقسام المؤسسية المستهدفة:

- قسم علوم البيانات.
- قسم التعلم الآلي والذكاء الاصطناعي.
- قسم البحث والتطوير.
- قسم هندسة البرمجيات (فرق الذكاء الاصطناعي).
- قسم الابتكار.
- قسم تحليل البيانات.
- فريق تطوير المنتجات القائمة على الذكاء الاصطناعي.
- الفرق المتخصصة في الرؤية الحاسوبية.
- الفرق المتخصصة في معالجة اللغات الطبيعية.
- وحدات المحاكاة.

أهداف الدورة التدريبية:

بنهاية هذه الدورة التدريبية، سيكون المتدرب قد أتقن المهارات التالية:

- فهم أهمية زيادة البيانات في تحسين أداء نماذج الذكاء الاصطناعي.
- التعرف على أنواع مختلفة من تقنيات زيادة البيانات للصور.
- تطبيق تحويلات هندسية ولونية لـزيادة مجموعات بيانات الصور.
- استخدام تقنيات زيادة البيانات للنصوص (مثل تبديل الكلمات والمرادفات).
- تطبيق الشبكات التوليدية التنافسية (GANs) لـتوليد بيانات اصطناعية.
- التعامل مع مشكلة نقص البيانات وعدم توازن الفئات.
- اختيار أفضل استراتيجية لـزيادة البيانات لمشكلة محددة.
- تنفيذ تقنيات زيادة البيانات باستخدام مكتبات Python شائعة.
- تقييم تأثير زيادة البيانات على أداء النموذج.
- بناء نماذج تعلم عميق أكثر قوة ومرونة.

منهجية الدورة التدريبية:

يعتمد BIG BEN Training Center في هذه الدورة على منهجية تدريبية تجمع بين الفهم النظري لـتقنيات زيادة البيانات والتطبيق العملي المكثف، بهدف تمكين المشاركين من بناء نماذج ذكاء اصطناعي قوية حتى مع مجموعات البيانات المحدودة. تشمل المنهجية محاضرات تفاعلية تستعرض المفاهيم الأساسية لزيادة البيانات، وتحديات نقص البيانات في التعلم العميق، والأساليب التقليدية والمتقدمة لزيادة البيانات. تتبع هذه المحاضرات ورش عمل تطبيقية مكثفة حيث سيقوم المشاركون بـتطبيق تقنيات زيادة البيانات على بيانات الصور والنصوص باستخدام أطر عمل مثل TensorFlow و PyTorch، ومكتبات مثل OpenCV وNLTK. سيتم التركيز على دراسات حالة واقعية تبرز كيفية استخدام زيادة البيانات لحل مشكلات مثل تصنيف الصور الطبية ببيانات قليلة، أو تحسين نماذج معالجة اللغات الطبيعية. تتضمن الدورة جلسات عمل جماعي لـتطوير استراتيجيات زيادة بيانات مخصصة لمشاريع الذكاء الاصطناعي المختلفة. يتلقى المشاركون تغذية راجعة مفصلة ومنتظمة من المدربين الخبراء لضمان تطوير مهاراتهم في تحسين جودة البيانات وأداء النماذج.

خريطة المحتوى التدريبي (محاور الدورة التدريبية):

الوحدة الأولى: مقدمة إلى زيادة البيانات وأهميتها.

- مشكلة نقص البيانات في التعلم العميق.
- لماذا نحتاج إلى زيادة البيانات؟
- مفاهيم أساسية: التحيز، التعميم الزائد، التعميم الناقص.
- أنواع البيانات: صور، نصوص، صوت.
- الهدف من زيادة البيانات: تحسين الأداء والمرونة.
- زيادة البيانات مقابل جمع المزيد من البيانات.
- أهمية جودة البيانات في بناء نماذج قوية.

الوحدة الثانية: زيادة البيانات للصور (التحويلات الهندسية واللونية).

- التحويلات الهندسية: التدوير، الانعكاس، القص، الترجمة.
- التحويلات اللونية: تغيير السطوع، التباين، التشبع، الضوضاء.
- تطبيق مجموعات من التحويلات لتوليد بيانات متنوعة.
- تقنيات المزيج (Mix-up) وCut Mix.
- استخدام مكتبات Python لزيادة البيانات المرئية (Augmentations, Imgaug).
- دراسة حالة: زيادة البيانات لتصنيف الصور الطبية.
- أفضل الممارسات في زيادة بيانات الصور.

الوحدة الثالثة: زيادة البيانات المتقدمة للصور (باستخدام GANs).

- مقدمة سريعة إلى الشبكات التوليدية التنافسية (GANs).
- كيف يمكن للشبكات التوليدية التنافسية توليد بيانات اصطناعية.
- تطبيق GANs لإنشاء صور واقعية لزيادة البيانات.
- تحديات تدريب GANs لزيادة البيانات.
- أمثلة على معماريات GANs المستخدمة لزيادة البيانات (Cycle GAN).
- تقنيات توليد البيانات المشروطة (Conditional Data Generation).
- دمج البيانات الحقيقية والاصطناعية لتحسين التدريب.

الوحدة الرابعة: زيادة البيانات للنصوص والبيانات الجدولية.

- تقنيات زيادة البيانات للنصوص:
- تبديل الكلمات (Word Swapping)، إضافة المرادفات.
- إعادة صياغة الجمل (Sentence Paraphrasing).
- زيادة البيانات السياقية (Contextual Augmentation).
- استخدام نماذج اللغة الكبيرة (Large Language Models) لزيادة البيانات النصية.
- زيادة البيانات للبيانات الجدولية:
- التوليد الاصطناعي للبيانات الجدولية.
- التعامل مع البيانات العددية والفئوية.
- أمثلة على زيادة البيانات في تطبيقات معالجة اللغات الطبيعية.

الوحدة الخامسة: استراتيجيات زيادة البيانات وتقييمها.

- اختيار استراتيجية زيادة البيانات المناسبة.
- تأثير حجم مجموعة البيانات على اختيار التقنيات.
- التعامل مع عدم توازن الفئات باستخدام زيادة البيانات.
- تقييم تأثير زيادة البيانات على أداء النموذج.
- مقاييس الأداء (الدقة، الاستدعاء، F1-Score).
- الزيادة التلقائية للبيانات (Auto Augment).
- الآفاق المستقبلية في زيادة البيانات والتعلم الاصطناعي.

الأسئلة المتكررة:

ما هي المؤهلات أو المتطلبات اللازمة للمشاركين قبل التسجيل في الدورة؟

لا توجد شروط مسبقة.

كم تستغرق مدة الجلسة اليومية، وما هو العدد الإجمالي لساعات الدورة التدريبية؟

تمتد هذه الدورة التدريبية على مدار خمسة أيام، بمعدل يومي يتراوح بين 4 إلى 5 ساعات، تشمل فترات راحة وأنشطة تفاعلية، ليصل إجمالي المدة إلى 20–25 ساعة تدريبية.

سؤال للتأمل:

مع القدرة المتزايدة لتقنيات زيادة البيانات على توليد مجموعات بيانات اصطناعية واقعية بشكل متزايد، إلى أي مدى يمكننا الاعتماد على هذه البيانات في تدريب نماذج الذكاء الاصطناعي الحساسة، وهل يمكن أن يقلل هذا الاعتماد من الحاجة إلى جمع بيانات حقيقية مكلفة وعالية الجودة؟

ما الذي يميز هذه الدورة عن غيرها من الدورات؟

تتميز هذه الدورة بتقديمها نهجاً عملياً ومتقدماً في مجال زيادة البيانات (Data Augmentation)، مع التركيز على تمكين المشاركين من بناء نماذج ذكاء اصطناعي قوية حتى في ظل قيود البيانات. ما يميزنا هو دمج التقنيات التقليدية لزيادة البيانات مع الأساليب الحديثة القائمة على الذكاء الاصطناعي، مثل استخدام الشبكات التوليدية التنافسية (GANs) لـتوليد بيانات اصطناعية**. نغطي مجموعة واسعة من أنواع البيانات (صور، نصوص، جداول)، مما يوفر فهماً شاملاً لـكيفية التعامل مع تحديات البيانات المتنوعة. الدورة تركز على تزويد المشاركين بالمهارات اللازمة لـتحسين أداء نماذج التعلم العميق، والتغلب على مشكلة نقص البيانات، وتحقيق تعميم أفضل للنماذج، مما يجعلها ضرورية لأي متخصص في الذكاء الاصطناعي يسعى للارتقاء بقدراته في معالجة البيانات وبناء النماذج.