الدورة التدريبية: الذكاء الاصطناعي القابل للتفسير (XAI) - فهم سلوك نموذج الذكاء الاصطناعي



14 - 18 ديسمبر 2026
طوكيو
5800 € (للشخص الواحد)
Ref: #AI9816_322413

مقدمة الدورة التدريبية / لمحة عامة:

يقدم BIG BEN Training Center هذه الدورة التدريبية المتخصصة حول الذكاء الاصطناعي القابل للتفسير (XAI) - فهم سلوك نموذج الذكاء الاصطناعي، وهي مصممة لعلماء البيانات، ومهندسي التعلم الآلي، والمطورين، والباحثين، ومديري المنتجات، وصناع القرار الذين يسعون لبناء ثقة أكبر في أنظمة الذكاء الاصطناعي (AI)، وفهم كيفية اتخاذ نماذج التعلم الآلي (Machine Learning) لقراراتها. في عالم تزداد فيه تطبيقات الذكاء الاصطناعي تعقيداً وانتشاراً في قطاعات حساسة مثل الرعاية الصحية، والتمويل، والعدالة الجنائية، أصبح فهم الشفافية (Transparency)، والقابلية للتفسير (Interpretability)، وإمكانية التتبع (Traceability) أمراً بالغ الأهمية. تهدف الدورة إلى تجاوز مفهوم "الصندوق الأسود" (Black Box) لنماذج الذكاء الاصطناعي، واستكشاف الأدوات والتقنيات التي تمكننا من فهم وتفسير التنبؤات والقرارات الصادرة عنها. ستغطي الدورة مفاهيم الذكاء الاصطناعي القابل للتفسير، وأهميته التنظيمية والأخلاقية، والتقنيات المحلية والعالمية للتفسير (Local and Global Interpretability)، مثل LIME (Local Interpretable Model-agnostic Explanations)، SHAP (Shapley Additiveg explanations)، وInterpretML، وتفسير الشبكات العصبية (Neural Network Interpretability)، وتحديد أهمية الميزات (Feature Importance). سيتعلم المشاركون كيفية تطبيق هذه التقنيات على نماذج تعلم آلي مختلفة، وتقييم مدى قابلية التفسير، وتوصيل الرؤى المكتسبة إلى أصحاب المصلحة غير التقنيين. تهدف الدورة إلى تمكين المختصين من بناء نماذج ذكاء اصطناعي أكثر شفافية ومسؤولية، وتعزيز الثقة في أنظمة الذكاء الاصطناعي، والامتثال للمتطلبات التنظيمية المتعلقة بالذكاء الاصطناعي المسؤول. نستلهم في هذه الدورة من أعمال البروفيسور سينثيا رودين (Cynthia Rudin)، التي تدافع عن بناء نماذج ذكاء اصطناعي قابلة للتفسير بطبيعتها بدلاً من تفسير نماذج الصندوق الأسود.

الفئات المستهدفة / هذه الدورة التدريبية مناسبة لـ:

- علماء البيانات.
- مهندسي التعلم الآلي.
- مهندسي الذكاء الاصطناعي.
- محللي البيانات.
- مديري المنتجات.
- المطورين في مجال الذكاء الاصطناعي.
- المهندسين المعماريين للذكاء الاصطناعي.
- مسؤولي الامتثال (Compliance Officers).
- مديري المخاطر.
- الباحثين في الذكاء الاصطناعي.

القطاعات والصناعات المستهدفة:

- الخدمات المالية (تقييم الائتمان، كشف الاحتيال).
- الرعاية الصحية (التشخيص، خطط العلاج).
- القانون والعدالة (تطبيقات العدالة الجنائية).
- التأمين (تقييم المخاطر، تسعير البوليصات).
- القطاع الحكومي (صنع القرار، أنظمة الدعم).
- التصنيع (الصيانة التنبؤية، مراقبة الجودة).
- الخدمات الاستشارية.
- التكنولوجيا والبرمجيات.
- البحث والتطوير.
- الخدمات اللوجستية.

الأقسام المؤسسية المستهدفة:

- قسم علوم البيانات.
- قسم التعلم الآلي والذكاء الاصطناعي.
- قسم الامتثال القانوني.
- قسم إدارة المخاطر.
- قسم تطوير المنتجات (الخاصة بالذكاء الاصطناعي).
- قسم البحث والتطوير (R&D).
- قسم التدقيق الداخلي.
- قسم الحوكمة والسياسات.
- قسم هندسة البرمجيات.
- قسم التحليلات.

أهداف الدورة التدريبية:

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

- فهم مفهوم الذكاء الاصطناعي القابل للتفسير وأهميته.
- التمييز بين الشفافية، والقابلية للتفسير، وإمكانية التتبع.
- تطبيق تقنيات التفسير المحلية (LIME, SHAP) على نماذج التعلم الآلي.
- استخدام تقنيات التفسير العالمية (Global Interpretability methods).
- تفسير سلوك الشبكات العصبية العميقة.
- تحديد أهمية الميزات (Feature Importance) في نماذج الذكاء الاصطناعي.
- تقييم مدى عدالة (Fairness) وموثوقية (Reliability) نماذج الذكاء الاصطناعي.
- توصيل الرؤى المكتسبة من تفسير النماذج لأصحاب المصلحة.
- التعامل مع التحديات الأخلاقية والقانونية لنشر الذكاء الاصطناعي.
- بناء نماذج ذكاء اصطناعي أكثر شفافية ومسؤولية.

منهجية الدورة التدريبية:

يعتمد BIG BEN Training Center في هذه الدورة على منهجية تدريبية تجمع بين المعرفة النظرية المتعمقة والتطبيق العملي المكثف، بهدف تمكين المشاركين من فهم وتطبيق تقنيات الذكاء الاصطناعي القابل للتفسير. تشمل المنهجية محاضرات تفاعلية تستعرض المفاهيم الأساسية لـ XAI، وأهميتها التنظيمية والأخلاقية، وأنواع التقنيات المختلفة للتفسير. تتبع هذه المحاضرات ورش عمل تطبيقية مكثفة حيث سيقوم المشاركون باستخدام أدوات ومكتبات XAI الرائدة (مثل LIME, SHAP, InterpretML) على نماذج تعلم آلي مختلفة (تصنيف، انحدار، رؤية حاسوبية). سيتم التركيز على دراسات حالة واقعية تبرز أهمية قابلية التفسير في قطاعات حساسة مثل الرعاية الصحية والخدمات المالية، وكيف يمكن لـ XAI أن يساعد في بناء الثقة والامتثال. تتضمن الدورة جلسات عمل جماعي لتحليل وتفسير مخرجات نماذج الذكاء الاصطناعي وتقديم رؤى قابلة للتنفيذ. يتلقى المشاركون تغذية راجعة مفصلة ومنتظمة من المدربين الخبراء لضمان تطوير مهاراتهم في هذا المجال الحيوي والمتنامي.

خريطة المحتوى التدريبي (محاور الدورة التدريبية):

الوحدة الأولى: مقدمة إلى الذكاء الاصطناعي القابل للتفسير (XAI).

- مفهوم الذكاء الاصطناعي القابل للتفسير وأهميته.
- تحدي "الصندوق الأسود" في نماذج الذكاء الاصطناعي.
- الشفافية، والقابلية للتفسير، وإمكانية التتبع: تعريفات وفروقات.
- أهمية XAI في القطاعات الحساسة (الصحة، التمويل، القانون).
- المتطلبات التنظيمية والأخلاقية للذكاء الاصطناعي المسؤول.
- أنواع قابلية التفسير: المحلية والعالمية.
- تطور مجال XAI وأبرز الباحثين فيه.

الوحدة الثانية: تقنيات التفسير المحلية (Local Interpretability).

- أهمية التفسير المحلي لفهم التنبؤات الفردية.
- مقدمة إلى LIME (Local Interpretable Model-agnostic Explanations).
- تطبيق LIME على نماذج التصنيف والانحدار.
- مقدمة إلى SHAP (Shapley Additive explanations).
- تطبيق SHAP لفهم مساهمة الميزات في التنبؤات.
- مقارنة بين LIME وSHAP.
- أمثلة عملية لتفسير القرارات الفردية لنموذج الذكاء الاصطناعي.

الوحدة الثالثة: تقنيات التفسير العالمية (Global Interpretability) وأهمية الميزات.

- فهم السلوك العام لنموذج الذكاء الاصطناعي.
- طرق التفسير العالمية المستقلة عن النموذج (Model-agnostic).
- مخططات أهمية الميزات (Feature Importance Plots).
- تفسير نماذج الشجرة (Tree-based Models) بطبيعتها.
- تحليل اعتماد الميزات الجزئي (Partial Dependence Plots - PDPs).
- تأثيرات الميزات المعزولة (Individual Conditional Expectation - ICE plots).
- بناء نماذج قابلة للتفسير بطبيعتها (Interpretable-by-design models).

الوحدة الرابعة: تفسير الشبكات العصبية والذكاء الاصطناعي المرئي.

- تفسير الشبكات العصبية العميقة (Deep Neural Networks).
- خرائط التنشيط الفئوية (Class Activation Maps - CAMs).
- التدرجات الموجهة (Guided Backpropagation).
- أهمية الميزات في الرؤية الحاسوبية.
- تفسير قرارات نماذج الرؤية الحاسوبية (التعرف على الصور).
- تطبيقات XAI في تحليل الصور الطبية.
- تحديات تفسير النماذج المعقدة (Transformer models).

الوحدة الخامسة: تقييم XAI، الأخلاقيات، والآفاق المستقبلية.

- مقاييس لتقييم جودة التفسيرات.
- العدالة (Fairness) والتحيز (Bias) في نماذج الذكاء الاصطناعي وكيفية كشفها بـ XAI.
- XAI وخصوصية البيانات.
- كيفية توصيل تفسيرات الذكاء الاصطناعي لأصحاب المصلحة.
- التحديات القانونية والأخلاقية المتقدمة لنشر XAI.
- مستقبل الذكاء الاصطناعي القابل للتفسير والذكاء الاصطناعي المسؤول.
- بناء الثقة في أنظمة الذكاء الاصطناعي عبر الشفافية.

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:

في ظل سعينا المستمر لجعل نماذج الذكاء الاصطناعي أكثر قابلية للتفسير، هل يمكن أن يؤدي هذا التركيز الشديد على الشفافية إلى التضحية ببعض من كفاءة ودقة هذه النماذج في مهام معينة، وهل هناك نقطة توازن أمثل بين التفسيرية والأداء؟

ما الذي يميز هذه الدورة عن غيرها من الدورات؟

تتميز هذه الدورة بتقديمها نهجاً فريداً وشاملاً في مجال الذكاء الاصطناعي القابل للتفسير (XAI)، مع التركيز على فهم آليات اتخاذ القرار في نماذج الذكاء الاصطناعي المعقدة. ما يميزنا هو دمج الأسس النظرية المتينة لقابلية التفسير مع التطبيق العملي لأدوات وتقنيات XAI الرائدة، مما يتيح للمشاركين تحليل وتفسير مخرجات النماذج بفاعلية. نغطي مجموعة واسعة من تقنيات التفسير، من التقنيات المحلية التي تشرح التنبؤات الفردية إلى التقنيات العالمية التي تكشف عن السلوك العام للنموذج. الدورة تركز على تزويد المشاركين بالمهارات اللازمة لبناء ثقة أكبر في أنظمة الذكاء الاصطناعي، والامتثال للمتطلبات التنظيمية، والمساهمة في تطوير ذكاء اصطناعي مسؤول وأخلاقي، مما يجعلها ضرورية لأي محترف يتعامل مع نماذج الذكاء الاصطناعي في بيئات حساسة تتطلب الشفافية.