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

اغسطس ٢٠٢٦ ٢٨ - ٢٤

فيينا

٥٧٠٠ € (للشخص الواحد)

Ref: #AI9816_247077

مقدمة الدورة التدريبية / لمحة عامة:

سلوك نموذج الذكاء التدريبية المتخصصة حول الذكاء الاصطناعي القابل يقدم BIG BEN Training Center هذه الدورة ثقة التعلم الآلي، والمطورين، والباحثين، ومديري الاصطناعي، وهي مصممة لعلماء البيانات، ومهندسي للتفسير (- (XAI فهم الآلي(Machine Learning) أكبر في أنظمة الذكاء الاصطناعي (AI)، وفهم كيفية المنتجات، وصناع القرار الذين يسعون لبناء الاصطناعي تعقيداً وانتشاراً في قطاعات حساسة مثل لقراراتها. في عالم تزداد فيه تطبيقات الذكاء اتخاذ نماذج التعلم وإمكانية التتبع أصبح فهم الشفافية (Transparency)، والقابلية الرعاية الصحية، والتمويل، والعدالة الجنائية، إلى تجاوز مفهوم "الصندوق الأسود" ((Black Box أمراً بالغ الأهمية. تهدف الدورة (Traceability) للتفسير (Interpretability)، ستغطي الدورة مفاهيم والتقنيات التي تمكننا من فهم وتفسير التنبؤات لنماذج الذكاء الاصطناعي، واستكشاف الأدوات التنظيمية والأخلاقية، والتقنيات المحلية والعالمية الذكاء الاصطناعي القابل للتفسير، وأهميته والقرارات الصادرة عنها. SHAP، و(Explanations مثل LIME (Local Interpretable Model-agnostic للتفسير (Local and Global Interpretability)، Feature العصبية (Neural Network Interpretability)، ، وتفسير الشبكات InterpretML، و(Shapley Additive explanations) آلي مختلفة، وتقييم مدى قابلية سيتعلم المشاركون كيفية تطبيق هذه التقنيات على وتحديد أهمية الميزات (.Importance) إلى أصحاب المصلحة نماذج تعلم

اصطناعي أكثر شفافية ومسؤولية، غير التقنيين. تهدف الدورة إلى تمكين المختصين من التفسير، وتوصيل الرؤى المكتسبة الدورة من أعمال والامتثال للمتطلبات التنظيمية المتعلقة بالذكاء وتعزيز الثقة في أنظمة الذكاء الاصطناعي، بناء نماذج ذكاء عن بناء نماذج ذكاءCynthia Rudin البروفيسور سينثيا رودين(Cynthia Rudin) ، الاصطناعي المسؤول. نستلهم في هذه نماذج الصندوق الأسود. اصطناعي قابلة للتفسير بطبيعتها بدلاً من تفسير ، التي تدافع

لـ: الفئات المستهدفة / هذه الدورة التدريبية مناسبة

- 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:

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

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

منهجية الدورة التدريبية:‏

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

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.
- تعريفات وفروقات: الشفافية، والقابلية للتفسير، وإمكانية التتبع.
- 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.
- LIME (Local Interpretable Model-agnostic Explanations) مقدمة إلى.
- تطبيق LIME على نماذج التصنيف والانحدار.
- SHAP (Shapley Additive explanations) مقدمة إلى.
- تطبيق SHAP لفهم مساهمة الميزات في التنبؤات.
- مقارنة بين LIME و SHAP.
- 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).

الاصطناعي المرئي: الوحدة الرابعة: تفسير الشبكات العصبية والذكاء

- 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)

المستقبلية: الوحدة الخامسة: تقييم XAI، الأخلاقيات، والآفاق

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

الأسئلة المتكررة:

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

لا توجد شروط مسبقة.

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

المدة إلى ٢٥٢٠- بمعدل يومي يتراوح بين ٤ إلى ٥ ساعات، تشمل فترات تمتد هذه الدورة التدريبية على مدار خمسة أيام، ساعة تدريبية. راحة وأنشطة تفاعلية، ليصل إجمالي

سؤال للتأمل:

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

ما الذي يميز هذه الدورة عن غيرها من الدورات؟

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