

Opleiding : Generatieve AI, begrijpen, rechtvaardigen en controleren van de verklaarbaarheid van LLM-modellen

LLM-beslissingen in uw AI-projecten uitleggen, documenteren en controleren

Praktijkcursus - 3d - 21u00 - Ref. LLO

Prijs : 2010 € V.B.

NEW

Deze cursus biedt deelnemers de methoden en tools om LLM-beslissingen te analyseren, te verklaren en te controleren. U leert hoe u generatieve AI kunt documenteren, rechtvaardigen en controleren in overeenstemming met ethische en regelgevende principes.

Pedagogische doelstellingen

Aan het einde van de training is de deelnemer in staat om:

- ✓ Inzicht in de strategische kwesties van verklaarbaarheid voor de adoptie van generatieve AI.
- ✓ Beoordeel de bedrijfsrisico's in verband met vertekening, fouten en ondoorzichtigheid van LLM's.
- ✓ Intern beleid opstellen om transparantie en verantwoording bij het gebruik van AI te garanderen.
- ✓ Uitlegmethoden en -hulpmiddelen inzetten om het vertrouwen en de controle te versterken.
- ✓ AI-governance afstemmen op regelgevingskaders en compliancevereisten.
- ✓ De audit en het doorlopend toezicht op GenAI-projecten binnen het bedrijf beheren.

Doelgroep

Architectes IA, data scientists, PO, juristes, fonctionnels, auditeurs internes et tout professionnel impliqué dans la responsabilité des systèmes IA.

DEELNEMERS

Architectes IA, data scientists, PO, juristes, fonctionnels, auditeurs internes et tout professionnel impliqué dans la responsabilité des systèmes IA.

VOORAFGAANDE VEREISTEN

Basiskennis van AI of LLM. Ervaring met het lezen en schrijven van prompts.

VAARDIGHEDEN VAN DE CURSUSLEIDER

De deskundigen die de cursus leiden zijn specialisten op het betreffende vakgebied. Zij werden geselecteerd door onze pedagogische teams zowel om hun vakkennis als hun pedagogische vaardigheden voor elke cursus die zij geven. Zij hebben minstens vijf tot tien jaar ervaring in hun vakgebied en oefenen of oefenden verantwoordelijke bedrijfsfuncties uit.

BEOORDELINGSMODALITEITEN

De cursusleider beoordeelt de pedagogische vooruitgang van de deelnemer gedurende de gehele cursus aan de hand van meerkeuzevragen, praktijksituaties, praktische opdrachten, ...
De deelnemer legt ook van tevoren en naderhand een test af ter bevestiging van de verworven kennis.

Voorafgaande vereisten

Basiskennis van AI of LLM. Ervaring met het lezen en schrijven van prompts.

Opleidingsprogramma

1 Waarom verklaarbaarheid essentieel is

- Definities: interpreteerbaarheid vs verklaarbaarheid.
- Specifieke kenmerken van generatieve modellen en LLM's.
- Kritische gebruikssituaties: juridisch, medisch, HR, financiën.
- Risico's verbonden aan het ontbreken van uitleg: vertrouwen, adoptie, naleving.

2 Expliceerbaarheid van LLM's, grenzen en hefbomen

- Hoe werken LLM's: zwarte dozen of herleidbare systemen?
- Prompts, geheugen, output: waar zitten de biases?
- Structurele beperkingen: instabiliteit, hallucinatie, gebrek aan traceerbaarheid.
- Reproduceerbaarheid: uitdaging of luchtspiegeling?

3 Verklaarbaarheidsmethoden in de GenAI-context

- Interpretatiegerichte prompt engineering.
- Ketenbenaderingen", stapsgewijze redeneringen.
- Gegenereerde rechtvaardiging vs. berekend bewijs.
- Waarneembare elementen in LangChain: logs, agenten, hulpmiddelen.

Praktisch werk

Analyse van een LLM-antwoord en stapsgewijze rechtvaardiging. Vergelijkende analyse van juist/onjuist genereren. Reconstructie van de redeneerketen. Visualisatie van de context en de volledige prompt

4 Traceren, begrijpen en uitleggen via LangChain

- Traceerbare componenten in LangChain: agenten, hulpmiddelen, kettinglogboeken.
- Loggen, callback-handlers, expliciete promptsjablonen.
- Inleiding tot TruLens, PromptLayer, Helicone, LangSmith.
- Creëren van een traceerbare pijplijn.

5 Ontologieën en grafieken gebruiken om uit te leggen

- Kennis structureren om het beter uit te leggen.
- Kennisgrafieken + LLM = interpreteerbare context.
- Zakelijke ontologieën: begrijpelijke verklaringen voor gebruikers.
- Dialoog tussen LLM-agent en gestructureerde grafiek.

Praktisch werk

Creatie van een verklaarbare assistent. Agent die zijn antwoorden rechtvaardigt met behulp van een grafiek/ontologie. Aanmaken van zelfverklarende prompts. Volledige logging van verzoek tot antwoord

PEDAGOGISCHE EN TECHNISCHE MIDDELEN

- De gebruikte pedagogische middelen en cursusmethoden zijn voornamelijk: audiovisuele hulpmiddelen, documentatie en cursusmateriaal, praktische oefeningen en correcties van de oefeningen voor praktijkstages, casestudies of reële voorbeelden voor de seminars.
- Na afloop van de stages of seminars verstrekt ORSYS de deelnemers een evaluatievragenlijst over de cursus die vervolgens door onze pedagogische teams wordt geanalyseerd.
- Na afloop van de cursus wordt een presentielijst per halve dag verstrekt, evenals een verklaring van de afronding van de cursus indien de stagiair alle sessies heeft bijgewoond.

TOEGANGSMODALITEITEN EN TERMIJNEN

De inschrijving dient 24 uur voor aanvang van de cursus plaatsgevonden te hebben.

TOEGANKELIJKHEID VOOR MINDERVALIDEN

Is voor u speciale toegankelijkheid vereist? Neem contact op met mevr. FOSSE, contactpersoon voor mindervaliden, via het adres psh-accueil@ORSYS.fr om uw verzoek en de haalbaarheid daarvan zo goed mogelijk te bestuderen.

6 Expliceerbaarheid en regelgevend kader

- Wat de RGPD en de IA Act vereisen (recht op uitleg, transparantie).
- Verplichting tot documentatie, logboeken, reproduceerbaarheid.
- Uitlegbare interfaces: hoe toon je een begrijpelijke rechtvaardiging.
- De rol van verklaarbaarheid in DPIA's en risicobeoordelingen.

7 Auditmethoden voor GenAI-systemen

- Maak een verklarend logboek: vraag + context + bronnen + redenering.
- Kwaliteitscontrole van gegenereerde reacties (hallucinatie, consistentie, vertekening).
- Opname van "kritieke" agenten of verklaringsscores.
- Menselijke evaluatie van redeneringen.

Praktisch werk

Een controleerbaar reactiesysteem opzetten. Casestudy: juridisch medewerker of HR-assistent. Opzetten van een volledige flow met verantwoording, auditlogboek. Demonstratie van een verklarende interface (tekst + afbeeldingen).

Data en plaats

KLAS OP AFSTAND

2026 : 23 maa., 29 juni, 28 sep., 7 dec.

PARIS LA DÉFENSE

2026 : 16 maa., 22 juni, 21 sep., 30 nov.