Explicabilité de l'IA Générative

L'objectif de cette formation est de donner aux participants les compétences pour analyser, documenter, justifier et auditer les décisions prises par des systèmes d'IA générative, en combinant outils techniques, principes éthiques, sémantique métier et cadres réglementaires.

Pré-requis : *Connaissances générales sur l'IA ou les LLMs, capacité à lire ou construire un prompt / prompt chain.*

Modalité :
- Distanciel en classe virtuelle
- E-learning : à venir
- Présentiel

Communauté: community.reconvert.net

Durée totale : 21 H (3 jours)

PLAN DETAILLE

Pourquoi l'explicabilité est essentielle

Définitions : interprétabilité vs explicabilité
Spécificités des modèles génératifs et des LLMs
Cas d'usage critiques : juridique, médical, RH, finance
Risques liés à l'absence d'explication : confiance, adoption, conformité

Explicabilité des LLMs – limites et leviers

Comment fonctionnent les LLMs : boîtes noires ou systèmes réductibles ?
Prompts, mémoire, outputs : où sont les biais ?
Limitations structurelles :
+ instabilité
+ hallucination
+ manque de traçabilité
Reproductibilité : challenge ou mirage ?

Méthodes d'explicabilité dans le contexte GenAI

Prompt engineering orienté interprétabilité
Approches « chain of thought », step-by-step reasoning
Justification générée vs preuve calculée
Éléments observables dans LangChain : logs, agents, tools

Atelier 1 : Réponse LLM et justification pas-à-pas

Analyse comparative d'une génération correcte et incorrecte
Reconstruction de la chaîne de raisonnement
Visualisation du contexte et du prompt complet

Tracer, comprendre et expliquer via LangChain

Composants traçables dans LangChain : agents, tools, chain logs
Logging, callback handlers
Prompt templates explicites
Introduction à TruLens, PromptLayer
Helicone, LangSmith
Création d'un pipeline traçable

Utiliser les ontologies et graphes pour expliquer

Structurer le savoir pour mieux l'expliquer
Graphes de connaissance + LLM = contexte interprétable
Ontologies métiers : support d'explication intelligible pour l'utilisateur
Dialogue entre agent LLM et graphe structuré

Création d'un assistant explicable avec support métier

Conception d'un agent qui justifie ses réponses à partir d'un graphe/ ontologie
Rédaction de prompts auto-explicatifs
Journalisation complète de la requête à la réponse

Explicabilité de l'IA Générative

L'objectif de cette formation est de donner aux participants les compétences pour analyser, documenter, justifier et auditer les décisions prises par des systèmes d'IA générative, en combinant outils techniques, principes éthiques, sémantique métier et cadres réglementaires.

Pré-requis : *Connaissances générales sur l'IA ou les LLMs, capacité à lire ou construire un prompt / prompt chain.*

Modalité :
- Distanciel en classe virtuelle
- E-learning : à venir
- Présentiel

Communauté: community.reconvert.net

Durée totale : 21 H (3 jours)

PLAN DETAILLE

Explicabilité et cadre réglementaire

Ce que demandent le RGPD et l'IA Act (droit à l'explication, transparence).
Obligation de documentation, logs, reproductibilité
Interfaces explicables : comment afficher une justification intelligible. Le rôle de l'explicabilité dans les DPIA et les évaluations de risque

Méthodes d'audit des systèmes GenAI

Créer un log explicatif : prompt + contexte + sources + raisonnement
Contrôle qualité des réponses générées (hallucination, cohérence, biais)
Inclusion d'agents « critique » ou de score d'explication
Évaluation humaine des raisonnements

Atelier final : Construire un système de réponse auditable

Cas : assistant juridique ou RH
Mise en place d'un flux complet avec justification, score de confiance, log d'audit
Démo d'une interface explicative (textuelle + graphique)