



INSIGHTS · INTELIGÊNCIA ARTIFICIAL

Privacidade não é um extra: transforma modelos em activos empresariais seguros.

Desliza →

Porquê a privacidade importa

Modelos que vazam informação sensível comprometem receitas, reputação e podem bloquear adopção interna. Garantir privacidade é garantir utilização.

Ameaças concretas

Memorização de registos, membership inference e ataques de linkage (cruzar colunas) são vetores reais que já vimos em projectos.

Técnicas aplicáveis no Fabric

Combine data masking, column-level security, dados sintéticos e DP. Use workspaces isolados, Private Link e logging para governação e auditoria.

DP em treino: um exemplo prático

Para 100k registros e 40 features: batch 256, epochs 10, clipping 1.0, noise $\approx 1.1 \rightarrow$ epsilon ≈ 2 (delta=1e-5). AUC caiu de 0,82 para $\sim 0,79-0,80$.

Custo em modelos maiores

Em NLP com 1M exemplos, aumentar noise multiplier de 1.1 para 2.0 pode subir o epsilon para ~5–8 e aumentar o impacto na utilidade (ex.: perplexidade +20 → 35%).

Mini-caso: dados sintéticos

Retalhista com 2M linhas gerou 2,2M sintéticas em ~2 horas. Linkage tests mostraram <0.5% de matches reais e correlação $\rho \approx 0,95$ nas KPIs mensais.

Operacionalizar e adoptar

Combine sintéticos para exploração (200 utilizadores) e acesso restrito a dados reais (<5 pessoas). Instrumentar DP end-to-end leva normalmente 4–10 dias de engenharia.

CONTINUA NO BLOG

Como vão proteger a privacidade nos vossos projectos de ML?

www.bconcepts.pt/insights