

Grzegorz Cebula
Systems Engineer
Fortinet

„Bezpieczeństwo suwerennych środowisk AI – jak chronić lokalne modele LLM przed atakami, nadużyciami i wyciekiem danych”

Suwerenność AI nie wynika wyłącznie z uruchomienia modelu LLM we własnej infrastrukturze. Osiąga się ją dopiero wtedy, gdy organizacja kontroluje cały przepływ danych i proces podejmowania decyzji: kto, z jaką tożsamością, do którego modelu, z jakim kontekstem, narzędziami i politykami bezpieczeństwa może się komunikować, a także jak realizowane są logowanie i audyt.

Lokalny model językowy nie wysyła danych do publicznej chmury, ale samo to nie gwarantuje bezpieczeństwa. Nadal może zostać zmanipulowany promptem, uzyskać dostęp do danych wykraczających poza uprawnienia użytkownika, wykonać nieautoryzowane działania za pośrednictwem agentów lub API, ujawnić informacje w generowanej odpowiedzi albo zostać wykorzystany do kosztownego wyczerpania zasobów GPU. Dlatego bezpieczeństwo zaczyna się nie na serwerze z akceleratorem AI, lecz na styku użytkownika, aplikacji, danych, narzędzi i modelu.

Podczas prezentacji pokażę, jak projektować i zabezpieczać suwerenne środowiska AI wykorzystujące lokalne modele językowe, od warstwy użytkownika i aplikacji, przez dane i mechanizmy RAG, aż po sam model, agentów AI oraz udostępniane im narzędzia. Omówię również, jak utrzymać pod pełną kontrolą cały krytyczny łańcuch przetwarzania: dane, inferencję, polityki bezpieczeństwa, logowanie i audyt. Dzięki temu suwerenność obejmuje całe rozwiązanie AI, a nie wyłącznie sam model.