Risk Assessment – SentinelX / SentinelStack AB

# 1. Identifierade risker

Denna riskanalys identifierar potentiella risker relaterade till SentinelX-plattformen, i synnerhet med avseende på AI-användning inom skydd av barn, äldre och personliga enheter. Systemet hanterar känslig användardata och aktiveras ofta i hemmiljöer.

# 2. Risktyper

- Falska positiva AI-varningar (t.ex. felaktig groomingdetektion)  
- Obehörig åtkomst till fjärrstyrningsfunktioner  
- Otillräcklig dataskyddshantering vid export/loggning  
- Missbruk av övervakningsfunktioner i icke-etisk kontext  
- Systemfel som leder till förlorade bevis/loggar

# 3. Sannolikhet & påverkan

- Falsklarm: Medelhög sannolikhet, låg till medelhög påverkan  
- Fjärrstyrningsintrång: Låg sannolikhet, hög påverkan  
- Exportfel: Låg till medelhög sannolikhet, medelhög påverkan  
- Missbruk av systemet: Låg sannolikhet, hög påverkan  
- Systemkrasch: Medelhög sannolikhet, hög påverkan

# 4. Riskminskande åtgärder

- Telegram-kommandon endast till godkända ID:n (whitelist)  
- Loggning sker endast lokalt utan molnöverföring  
- AI-detektion valideras mot fler kriterier (kombinationer)  
- PDF:er signeras och krypteras där möjligt  
- Rollbaserad användning (föräldra-/adminläge)

# 5. Remaining residual risks

- AI-modellen kan i enstaka fall feltolka ton/mening  
- Förlust av enhet kan leda till att lokal logg förloras  
- I felaktiga händer kan systemet användas mot personer

# 6. Rekommenderad kontrollplan

- Regelbunden AI-modelluppdatering  
- Manualbaserad onboarding med etikavsnitt  
- Möjlighet till återkallning och avinstallation via kommandon  
- Användarfeedback via rapporteringsfunktion i PDF  
- Penetrationstest av Telegram-/CLI-gränssnitt