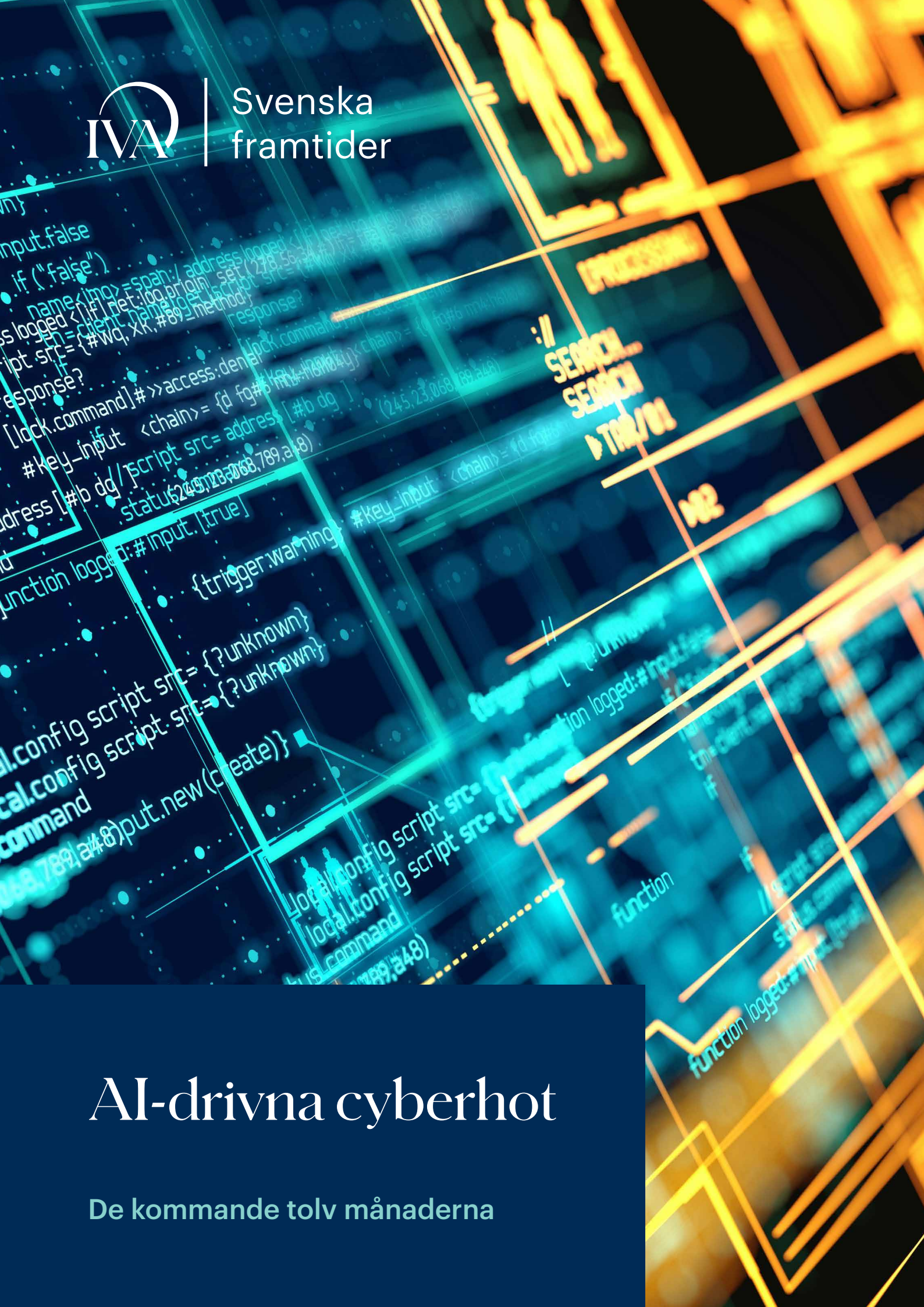


Svenska framtider

# AI-drivna cyberhot

## De kommande tolv månaderna

# Innehåll

# Förord

IVA driver projektet Svenska framtider med syftet att formulera en vision för Sverige som ledande inom teknik och innovation år 2035. Svenska framtider samlar aktörer från akademi, näringsliv och offentlig sektor för att identifiera möjligheter, utmaningar och strategiska riktningar för framtida konkurrenskraft, hållbar utveckling och säkerhet.

Projektet omfattar bland annat arbetsgrupper som analyserar utmaningar och möjligheter inom olika teknikområden. Grupperna tar fram rapporter som ger en översikt av nuläget och framtidsutsikterna inom det aktuella området, samt presenterar konkreta handlingsförslag. De fungerar också som underlag för arbetet med att forma en övergripande vision för Sverige år 2035.

Artificiell intelligens förändrar snabbt förutsättningarna för cybersäkerhet. Förmågor som tidigare krävt specialistkompetens är nu på väg att bli billigare, snabbare och mer tillgängliga för såväl angripare som försvarare. Den övergripande fråga som arbetsgruppen i denna rapport haft i uppdrag att besvara är: Vad bör vi i Sverige göra under de kommande tolv månaderna, för att minska de mest sannolika och allvarliga cyberhoten, givet den accelererande användningen av AI som vi ser idag?

Denna rapport är avsiktligt kortfattad och översiktlig. Den presenterar bland annat ett begränsat antal prioriterade inriktningar som bedöms vara relevanta för samhället i stort. Avsikten är att arbetet ska tas vidare med en mer fördjupad rapport. Rapporten har tagits fram i nära dialog med Nationellt cybersäkerhetscenter (NCSC) och AI Sweden.

Som i alla IVA-projekt har deltagarna medverkat i sin personliga kapacitet och inte som företrädare för de organisationer där de är verksamma. Rapportens analyser och rekommendationer bygger på den erfarenhet och kunskap som deltagarna bidragit med, samt på de diskussioner som dessa bidrag givit upphov till. Arbetsgruppen står bakom rapporten som helhet, men det innebär inte att samtliga medlemmar nödvändigtvis står bakom varje enskild formulering.

Arbetsgruppen för AI-drivna cyberhot har arbetat under perioden maj–juni 2026.

## Arbetsgrupp

**Pontus Johnson** (ordförande), KTH, IVA-ledamot
**Anne-Marie Eklund Löwinder**, Amelsec, IVA-ledamot

**Erik Ekudden**, Ericsson, IVA-ledamot
**Patrik Fältström**, Netnod, IVA-ledamot
**Daniel Gillblad**, Recorded Future, IVA-ledamot
**Fredrik Heintz**, Linköpings universitet, IVA-ledamot
**Simin Nadjm-Tehrani**, Linköpings universitet, IVA-ledamot
**Anders Sandberg**, Institutet för framtidsstudier, IVA-ledamot
**Staffan Truvé**, Recorded Future, IVA-ledamot

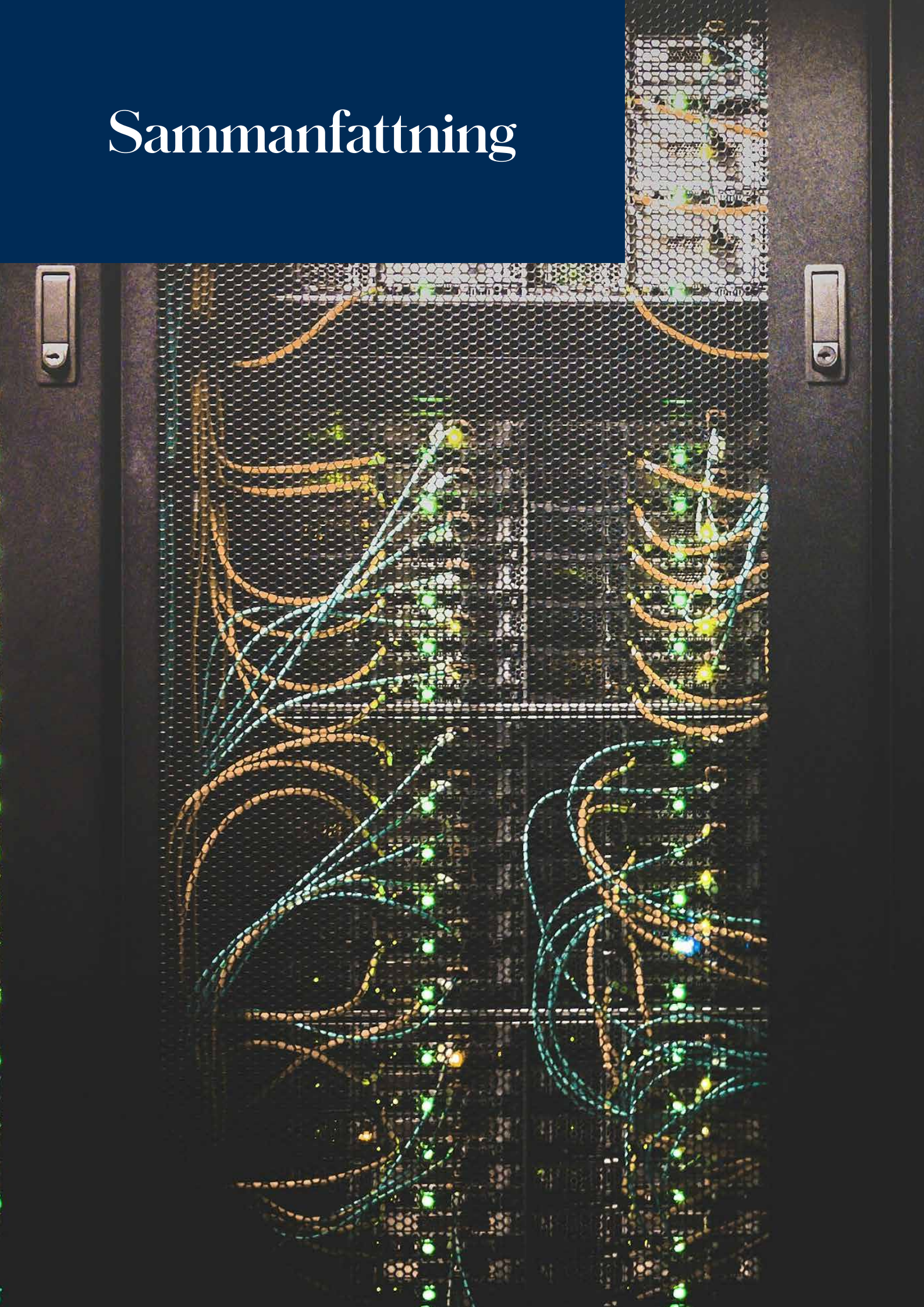
## Stöd till arbetsgruppen

**Per Hjertén**, IVA, projektledare
**Musard Balliu**, KTH, författare

## Rapporten har tagits fram i nära dialog med

**John Billow**, Nationellt cybersäkerhetscenter (NCSC)
**Otto Elmgart**, Nationellt cybersäkerhetscenter (NCSC)
**Robert Valsjö**, Nationellt cybersäkerhetscenter (NCSC)
**Robert A. Bridges**, AI Sweden
**Mauricio Muñoz**, AI Sweden
**Mats Nordlund**, AI Sweden
**Tommy Schönberg**, AI Sweden

# Sammanfattning



Artificiell intelligens (AI) förändrar tempot i cybersäkerhetsarbetet. Den centrala utmaningen är att förkorta tiden mellan att risken uppstår och hanteras. Slutsatsen i den här rapporten är att AI ökar hastigheten och skalan i hur många sårbarheter kan hittas och utnyttjas. Det handlar om att AI ökar räckvidden, tempot och genomslaget av cyberhot, samtidigt som organisationer måste säkra de AI-system som de tar i bruk. Avancerade AI-modeller kan vara ett stöd vid programmering, felsökning, analys och tekniskt resonemang. Och när de ges tillgång till verktyg och data blir de praktiska hjälpmedel i försvaret mot långa kedjor av skadlig cyberaktivitet.

Detta förändrar riskbilden för oss i Sverige. Många organisationer saknar en tydlig bild av sin digitala miljö och har inte tillräckligt robusta rutiner för att hantera behörigheter, hålla system uppdaterade, minska onödig exponering och upptäcka incidenter i tid. AI ökar risken kopplad till dessa svagheter, eftersom AI hjälper angripare att hitta och utnyttja dem i högre takt än tidigare. Trösklarna i kostnad och kunskap sjunker för angriparna, samtidigt som kraftfulla modeller och återanvändbara verktyg sprids till en bredare krets av aktörer.

Arbetsgruppen bedömer att de kommande tolv månaderna bör förstås som en kapplöpning mellan offensiv och defensiv användning av AI. Angripare kan använda AI för att arbeta snabbare, skala upp sina försök och anpassa sina metoder. Försvarare kan använda samma teknik för kodgranskning, prioritering av sårbarheter, övervakning, analys, incidenthantering och utbildning. Den avgörande frågan är om vi i Sverige kan få AI att höja försvararnas produktivitet snabbare än den höjer angriparnas förmåga. För styrelser, ledningsgrupper och beslutsfattare i såväl privat som offentlig sektor pekar detta mot tre omedelbara prioriteringar: att veta vilka tjänster och beroenden som måste fungera, att korta tiden från insikt om en risk till handling och att ansvarsfullt införa AI i sitt cybersäkerhetsarbete.

Åtgärderna behöver hänga ihop. AI ökar hastigheten i angreppscykeln, vilket tvingar fram en motsvarande ökning av försvarets hastighet. Snabbhet är viktigt, men förmågan att agera snabbt behöver vila på god styrning, tydliga prioriteringar och kapacitet att förutse risker innan de blir incidenter. AI förändrar tempot i cybersäkerheten, vidgar angreppsytan och ställer nya krav på hur AI-system styrs och säkras. Det grundläggande cybersäkerhetsarbete som varje organisation behöver ha på plats måste därför bedrivas med betydligt ökad

uthållighet, disciplin och flexibilitet. Styrningen måste bli mer riskbaserad och kunna arbeta i ett tempo som vi inte sett tidigare, för att stödja beslut i rätt tid i en föränderlig miljö. Det kräver ett tydligt ansvar för kritiska tjänster, för AI-stödda beslut och för beroenden till leverantörer, liksom mandat att agera under en incident. Samverkan och informationsdelning måste ske mycket snabbare och bli mer operativ och mer förtroendefull. Slutligen räcker det inte att göra samma arbete snabbare. AI måste integreras i cybersäkerhetsarbetet. Det bör ske på ett kontrollerat och ansvarsfullt sätt med tester och styrning utan att skapa nya ohanterade risker.

# En ny verklighet – därför är frågan brådskande

Digitaliseringen har redan gjort samhällskritiska tjänster beroende av mjukvara, data, nätverk och komplexa leveranskedjor. I Sverige syns detta beroende i den digitala infrastruktur vi använder dagligen, som BankID och betaltjänster, i kommuner, hälso- och sjukvård och uppkopplade industrimiljöer, och i beroendet av internationella leverantörer av moln, mjukvara och AI. IVAs tidigare rapport *Cybersäkerhet för ökad konkurrenskraft* framhöll att Sveriges höga digitaliseringsgrad skapar både möjligheter och sårbarhet, och vidare att cybersäkerhet hänger samman med konkurrenskraft och motståndskraft (IVA, 2022). Det gäller fortfarande framåt. AI lägger till ett nytt lager i detta beroende. Tekniken gör att befintliga cyberrisker lättare övergår i operativa hot, genom att öka hastigheten, skalan och räckvidden i hur sårbarheter kan hittas och utnyttjas. Två kompletterande behov bör därför hanteras tillsammans från början, att skydda sig mot AI-drivna cyberhot och att säkra de AI-system som organisationer använder, oavsett om de används för cybersäkerhet eller för andra ändamål.

Utvecklingen drivs av flera olika faktorer men det är den samlade effekten av flera förändringar som förstärker varandra. AI-modellerna blir med tiden alltmer kapabla. Verktygen som använder modellerna har blivit bättre. Kunskapen om hur mo-

dellerna används sprids. Resultatet är att fler aktörer kan utföra fler uppgifter, snabbare och med mindre expertkunskap än tidigare. Det gäller både angripare och försvarare.

AI bör i första hand betraktas som en produktivitetsteknik, inte som en självständig aktör. Det innebär inte att automatiserade eller agentliknande AI-system ska underskattas. Det innebär att ansvaret ligger kvar hos de människor och organisationer som utformar, driftsätter och förlitar sig på systemen. Många av de viktigaste riskerna bottnar fortfarande i svaga säkerhetsnivåer, oklara ansvarsförhållanden och bristande förståelse för beroenden i leveranskedjan.

En organisation som inte kan identifiera sina exponerade system, som inte vet vilka leverantörer som bär upp kritiska funktioner eller som saknar fungerande rutiner för incidenthantering blir inte säker genom att fokusera på valet av AI-verktyg. Grundläggande cybersäkerhet, som segmentering, övervakning, patchning, behörighetsstyrning och ansvarsutkrävande, är fortsatt viktigare än vilket varumärke en modell har.

Storbritanniens nationella cybersäkerhetscenter (UK NCSC) bedömer att AI redan gör delar av cyberintrång mer effektiva och att tekniken fram till 2027 sannolikt ökar intrångens omfatt-

ning och verkan främst genom att förstärka befintliga metoder snarare än att skapa helt nya (UK NCSC, 2025). Det stämmer överens med arbetsgruppens bild. Sårbarheter kan hittas snabbare, många angrepp kräver mindre expertkunskap och repetitiva uppgifter som spaning och social manipulation kan skalas upp med en låg kostnad.

Samtidigt är AI inte felfritt. Tekniken ger upphov till misslyckade försök, oförsiktiga handlingar som kan aktivera intrångsdetekteringssystem och felaktiga svar. Försvarare kommer att behöva avgöra hur de ska hantera en större mängd aktivitet, både meningsfull och av låg kvalitet. Det praktiska trycket är ändå tydligt. Om upptäckt och utnyttjande av sårbarheter går snabbare blir tiden mellan exponering, upptäckt och angrepp kortare. Organisationer med god kontroll över tillgångar, identiteter, patchning, beroenden och säkerhetsprocesser förblir betydligt mer motståndskraftiga än organisationer utan sådan kontroll. Men även organisationer med starkt cybersäkerhetsarbete kommer behöva korta sina beslutscykler, pröva nya försvarsmetoder, till exempel med hjälp av AI, och betrakta de egna AI-systemen som en del av angreppsytan.

# De kommande tolv månaderna

Det kommande året väntas innebära en fortsatt ökning av AI-driven cyberförmåga. Starkare modeller, bättre verktyg och mer kapabla AI-agenter gör det enklare att analysera system, lära sig okända verktyg, felsöka problem och automatisera repetitivt arbete. UK NCSC och Storbritanniens institut för AI-säkerhet (UK AISI) menar att försvarare bör utgå från att åtminstone vissa angripare redan har tillgång till kapabla AI-verktyg, samtidigt som de betonar att försvarare kan använda samma förmågor för att skaffa sig ett övertag (UK NCSC och UK AISI, 2026). De mest avancerade förmågorna kan inledningsvis vara koncentrerade till stater och andra resursstarka aktörer, men utvecklingen är också betydelsefull för alla organisationer i takt med att verktygen blir lättare att använda och anpassa.

Sårbarheter och svaga säkerhetsrutiner får därmed större konsekvenser. Svaga lösenord, felkonfigurerade system, gammal mjukvara och bristfällig loggning är inte några nya problem. Det som förändras är hastigheten och kostnaden för att hitta och utnyttja dem. AI kan göra sambandet av vardagliga svagheter mer synlig och mer användbar för angripare. AI-stödd mjukvaruutveckling kan också tillföra risk, om funktionsduglig kod tas fram och driftsätts snabbare än vad organisationer hinner granska, testa och underhålla på ett säkert sätt

(AI Sweden, 2026). Den kan dessutom göra mer skräddarsydda angrepp ekonomiskt möjliga, särskilt när målet är värdefullt. Ekonomiskt motiverade aktörer kommer ofta att fortsätta välja enklare mål, men aktörer som söker störning eller strategisk effekt kan också rikta AI-stödda insatser mot mer härdade miljöer.

AI-system och AI-stödda arbetsflöden blir i sig mer attraktiva mål. Frågan handlar inte bara om huruvida en modell kan ge skadliga tekniska råd. Det handlar också om vad ett AI-system tillåts göra inom en organisation, vilka data den har tillgång till och vilka åtgärder det kan utlösa. Det är ofta inte svårt att lura AI-system. En angripare kan försöka få en AI-assistent att röja information, följa dolda instruktioner, fatta ett olämpligt beslut eller missbruka en affärsprocess. Vi bör utgå från att organisationer kommer att möta olika former av manipulation som är svåra att förutse. NIST:s arbete med antagonistisk maskininlärning understryker att säkerheten i AI-system måste hanteras genom hela systemens livscykel och inte läggas till i efterhand (NIST, 2025).

Samtidigt som den allmänna tillgången till AI fortsätter att breddas är en ytterligare fråga under de kommande tolv

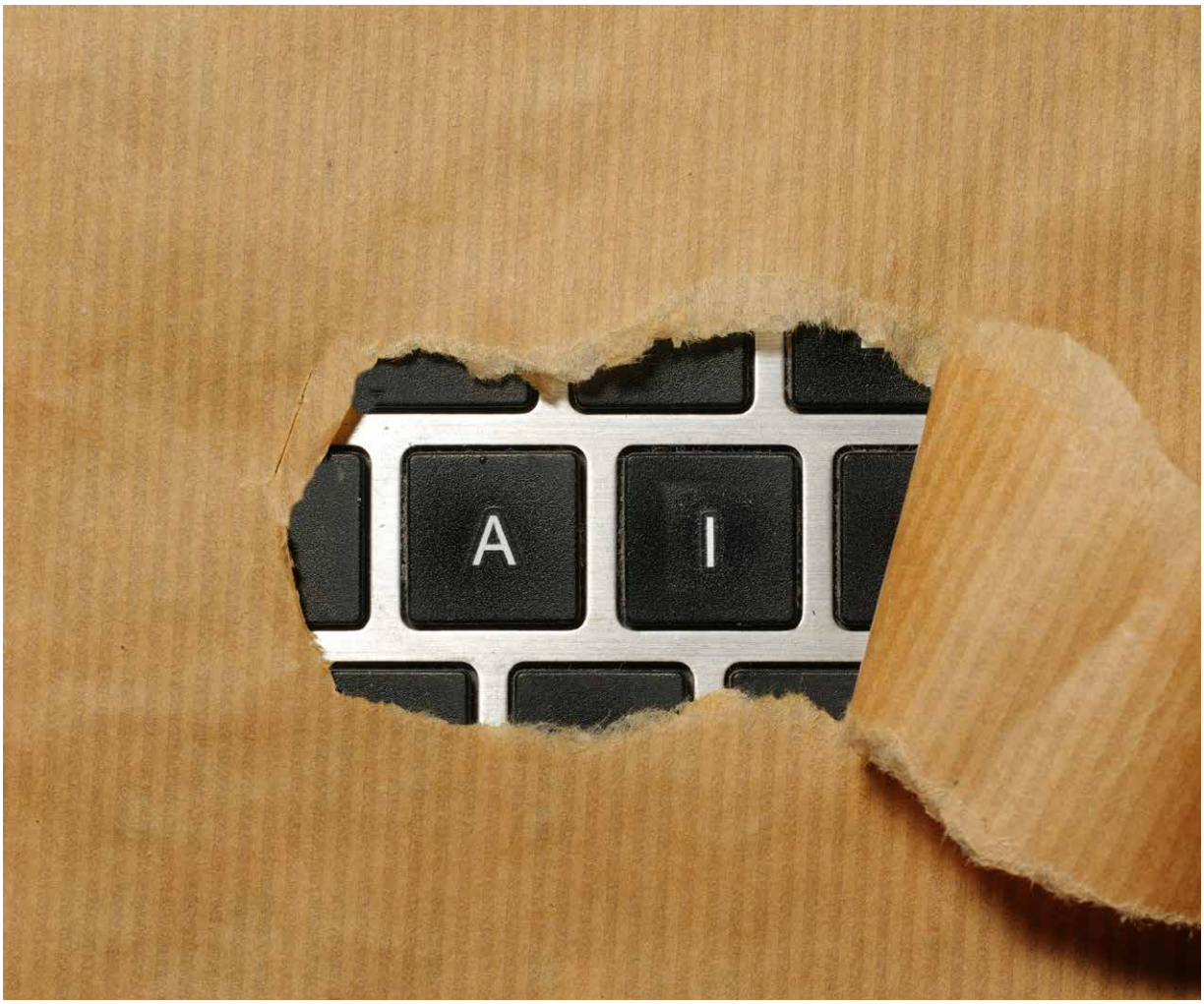
månaderna den tillförlitliga tillgången till de specifika modeller, tjänster och verktyg som organisationer kan komma att bli beroende av. När AI blir mer användbart för cybersäkerhetsarbete blir tillgången till kapabla AI-system i sig en del av frågan om motståndskraft. Det är en praktisk och operativ fråga. Organisationer bör undvika att bygga kritiska försvarsförmågor kring tjänster som de inte kan kontrollera, ersätta eller klara sig utan. Detsamma gäller oavsett om tjänsten eller leverantören är svensk, europeisk eller global. Den avgörande frågan är om organisationen förstår sina beroenden och har realistiska alternativ om tillgången till dessa förmågor förändras.

Aktuella, offentligt kända exempel pekar i samma riktning. Anthropic har rapporterat ett fall där agentisk AI-förmåga användes för att stödja en cyberspionageoperation mot mål runt om i världen (Anthropic, 2025). Ett annat angrepp, mot ett AI-stött kundtjänstflöde hos Meta, visar att risken inte bara handlar om avancerat utnyttjande av kod, utan också om manipulation av AI-drivna processer som har mandat att agera (MIT Technology Review, 2026). När Anthropic i maj 2026 stängde av tillgången till Fable 5 och Mythos 5 visade det att tillgången till avancerad AI kan förändras plötsligt av juridiska, geopolitiska eller säkerhetsmässiga skäl, vilket gör modelltillgång till en fråga om mot-

ståndskraft för AI-stött försvar (Anthropic, 2026). Dessa exempel bör inte ses som exakta prognoser för Sverige, men de visar varför säkerhet i AI, grundläggande cybersäkerhet och kontroll över kritiska beroenden alltmer hör ihop.

Myndigheter i flera andra länder drar liknande slutsatser. Nya Zeelands cybersäkerhetscenter (NCSC New Zealand) beskriver frontier AI som en utveckling med dubbla användningsområden, som kan göra kända sårbarheter, äldre system och bristande cyberhygien farligare (NCSC New Zealand, 2026). Kanadas nationella hotbildsanalys på cyberområdet och den australiska signalspaningsmyndighetens (Australian Signals Directorate) årliga hotrapport betonar båda att cyberhoten alltmer hänger samman med geopolitik, kritisk infrastruktur och förmågan hos offentliga och privata aktörer att samverka (Canadian Centre for Cyber Security, 2024; Australian Signals Directorate, 2024).

Den möjliga effekten behöver inte bli ensidig. Även försvarare får bättre verktyg. AI kan vara ett stöd vid kodgranskning, sårbarhetsanalys, övervakning, incidenthantering, triagering och utbildning. CISA:s färdplan för AI beskriver AI både som en teknik som ska säkras och som en teknik som kan stärka cyber-




säkerhetsarbetet när den införs på ett ansvarsfullt sätt (CISA, 2023). Ett realistiskt mål för oss i Sverige bör vara att göra den defensiva användningen snabbare, säkrare och mer systematisk än angriparnas offensiva användning.

# Prioriterade åtgärder

AI förändrar tempot i cybersäkerheten, men förändrar också hur mjukvara tas fram, hur angrepp kan automatiseras och hur AI-drivna processer kan missbrukas. Detta tvingar fram att försvaret ökar sin hastighet på motsvarande sätt. Den viktigaste utmaningen är inte en enskild produkt, plattform eller hotaktör. Utmaningen är att många organisationer fortfarande arbetar med processer och beslutsstrukturer som är utformade för en långsammare hantering, utan tillräcklig fokus på angripare som avsiktligt försöker slå ut deras system. Det behövs därför ett försvar som består av fem sammanhängande byggstenar: snabbhet, grundläggande cybersäkerhetsarbete, styrning, informationsdelning och samverkan samt att möta AI med AI.

## Snabbhet

Snabbhet måste bli en strategisk prioritering. Det innebär ökad hastighet med disciplin och förståelse för både förväntningar och beroenden, inte snabbhet för sakens egen skull. Organisationer behöver proaktivt korta tiden mellan att veta och agera. Det gäller sårbarhetshantering, patchning, incidenthantering, upphandling, beslutsfattande och kommunikation. Målet är en bättre balans mellan försiktighet och fördröjning. I en miljö där

angripare snabbt kan testa, anpassa och skala upp sina angrepp kan avvaktan på perfekt information, eller på den perfekta lösningen, i sig bli en risk.

Detta innebär att organisationer behöver ompröva hur de väger risk, snabbhet och handling mot varandra. Ledare och beslutsfattare i styrelserummen bör fråga sig vilka beslut som kan fattas tidigare, vilka skyddsåtgärder som kan införas tillfälligt (innan patchar finns tillgängliga) och vilka processer som kan förenklas utan att ansvarsutkrävandet går förlorat. De bör också fråga sig om ansvaret är placerat där åtgärder kan vidtas i rätt tid, och om vissa beslut behöver delegeras, godkännas i förväg eller tydliggöras innan en incident inträffar. Samma rutiner behövs när organisationer inför nya AI-verktyg. Att gå snabbt fram utan tillräckliga skyddsåtgärder kan skapa nya sårbarheter. De organisationer som är bäst förberedda är de som snabbt kan agera, lära av verkliga händelser och löpande anpassa sig.

## Grundläggande cybersäkerhetsarbete

Snabbare föränderliga hot gör det grundläggande cybersäkerhetsarbetet viktigare. Flerfaktorsautentisering, god behörighetsstyrning, identifiering av tillgångar, säker konfiguration,

loggning, övervakning, säkerhetskopiering, rutiner för incidenthantering och patchhantering är fortsatt grunden för motståndskraft (NCSC, u.å.). Detta är inte nya åtgärder, men AI gör konsekvenserna av att inte genomföra dem betydligt allvarligare.

Uppgiften är att göra det grundläggande cybersäkerhetsarbetet mer konsekvent och flexibelt integrerat i det dagliga arbetet. Organisationer behöver förstå sin egen digitala miljö, det vill säga vilka tjänster som är viktigast, vilka system och leverantörer som tillhandahåller dem, vem som har åtkomst, hur sårbarheter hanteras och vad som måste fortsätta att fungera under en störning eller patchning, oavsett vad som händer. Cyberhygien måste behandlas som en operativ förmåga, inte som en engångsinsats för att uppfylla regelkrav. Organisationer bör också utgå från ett förhållningssätt att intrång sker, att planera utifrån att angrepp kommer att ske och att fokusera på hur de ska upptäckas, begränsas och få så liten påverkan som möjligt. När dessa grunder är på plats bör organisationer förbättra hur de upprätthålls i takt med att AI blir en del av den dagliga driften och säkerhetsarbetet.

Ett annat behov är grundläggande utbildning för användare

och beslutsfattare. De behöver inte förstå varje AI-verktyg i detalj, men de behöver veta att AI-agenter kan ha åtkomst till filer, inloggningsuppgifter, kunddata, system och andra känsliga resurser. Den förståelsen bör bli en del av den allmänna säkerhetsmedvetenheten i takt med att AI-verktyg blir en del av det dagliga arbetet.

## Styrning

Styrningen måste anpassas till en hotmiljö med högre tempo. Styrelser, ledningsgrupper och beslutsfattare i såväl privat som offentlig sektor bör behandla AI-relaterad cyberrisk som en del av den ordinarie verksamhetsrisken, inte som en avgränsad teknisk fråga. Användningen av AI tar inte bort ansvaret. Den organisation som tillhandahåller en tjänst är fortsatt ansvarig för de risker, beroenden och konsekvenser som hör samman med tjänsten.

Styrningen behöver också utgå från en tydlig bild av vad organisationen förväntas leverera. Många organisationer försöker skydda allt lika mycket, vilket gör uppgiften närmast omöjlig. Ett mer realistiskt angreppssätt är att först fastställa organisationens uppdrag och därefter vilka tjänster, informationstillgångar, system och leverantörer som är verkligt kritiska för att uppdraget

ska kunna utföras. Ledningen bör veta vilka beroenden som kan förhindra leveransen, vilka risker som ligger utanför organisationens normala hantering och var en enskild leverantör eller tjänst har blivit en enskild felpunkt (single point of failure). Detta gäller oavsett om det rör interna eller externa leverantörer eller tjänster. Samma logik gäller vid införandet av AI. Ledningen bör säkerställa att nya verktyg inte kringgår etablerade ansvarsförhållanden, säkerhetskontroller eller upphandlingskrav.

Ledningen i en organisation behöver ta ställning till ett antal frågor. Vilka funktioner är kritiska? Vilka leverantörer, system och data är de beroende av? Hur snabbt kan organisationen fatta beslut under en pågående incident? Vem har mandat att agera? Hur kan AI bidra, hur används AI och hur säkras AI-drivna processer? Styrningen bör hjälpa organisationer att ställa rätt frågor tillräckligt tidigt för att svaren ska göra skillnad. Frågorna måste ställas proaktivt och rutinmässigt, inte i efterhand som en reaktion på en inträffad incident.

## Informationsdelning och samverkan

Snabbare verksamhet kräver också snabbare samverkan. Information om hot, lärdomar och praktiska insikter behöver röra sig

snabbare mellan myndigheter, företag, forskare, aktörer som driver kritisk infrastruktur och betrodda internationella partner. Vi behöver väl fungerande och väletablerade former för samverkan som kan användas i den normala verksamheten, inte bara under stora kriser.

Informationsdelning är inte bara en teknisk fråga. Den vilar på förtroende, tydliga roller och relationer som etableras innan incidenter inträffar. Enkla och resurssnåla rutiner för krishantering bör bli en del av den ordinarie beredskapen. Målet bör vara att relevanta aktörer snabbt kan hitta varandra, förstå vilken information som kan delas och hur de i praktiken kan agera på varningssignaler.

## Att möta AI med AI

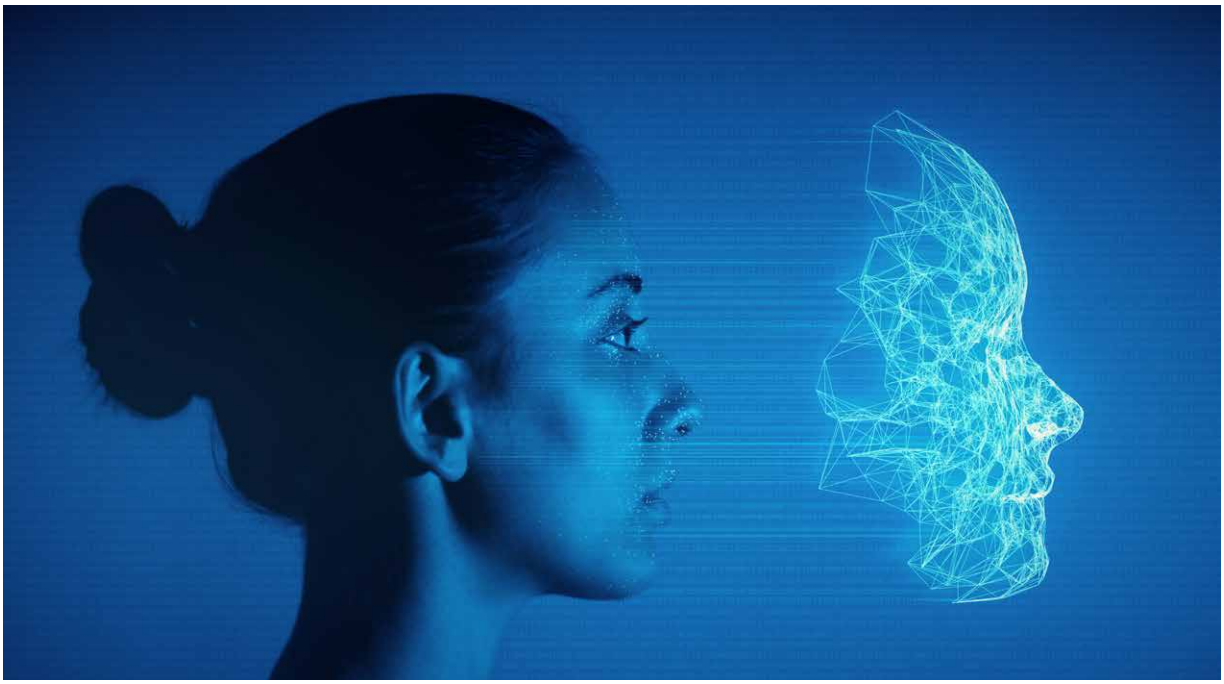
Slutligen räcker det inte att göra samma arbete snabbare. Sverige måste möta AI med AI och använda AI där det ger försvararna ett tydligt och kontrollerat övertag. Organisationer bör använda AI för att stärka kodgranskning, sårbarhetshantering, övervakning, hotanalys, incidenttriagering, beslutsstöd och utbildning. Målet bör vara att AI ska öka den defensiva produktiviteten mer än den ökar den offensiva förmågan. Detta bör ske

med en uttalad insikt om att varje ny AI-driven process också kan skapa nya beroenden, dataflöden och angreppsytor.

AI bör användas som ett produktivitetsverktyg under tydlig mänsklig och organisatorisk kontroll, där ansvaret ligger kvar hos de människor och organisationer som beslutar om hur tekniken driftsätts och används. AI-drivet försvar bör testas, styras och integreras i det befintliga säkerhetsarbetet. De organisationer som lär sig att använda AI säkert och effektivt står bättre rustade att hantera den snabbare och mer komplexa cybermiljö som nu växer fram.

Sammantaget bidrar de beskrivna åtgärderna till en stärkt cybersäkerhet. Snabbare angrepp kräver snabbare försvar. Ett snabbare försvar kräver ett starkt grundläggande cybersäkerhetsarbete, en styrning som möjliggör handling, en effektiv samverkan och en systematisk användning av AI i försvaret. Det är detta vi bör prioritera under de kommande tolv månaderna i Sverige. Vi bör inte vänta på perfekta förutsättningar eller på att andra ska agera först. Var och en av oss bör göra det vi kan, var för sig och tillsammans med andra, vid varje given tidpunkt och utifrån de förutsättningar vi har.

# Slutsats



AI-drivna cyberhot bör hanteras som en utmaning för hela samhället. Sverige står inte ensamt inför dem och ingen enskild aktör kan lösa dem. Frågan om vad vi i Sverige bör göra är därför en fråga om delat ansvar mellan de organisationer som utformar, driver, reglerar, levererar och är beroende av kritisk digital infrastruktur.

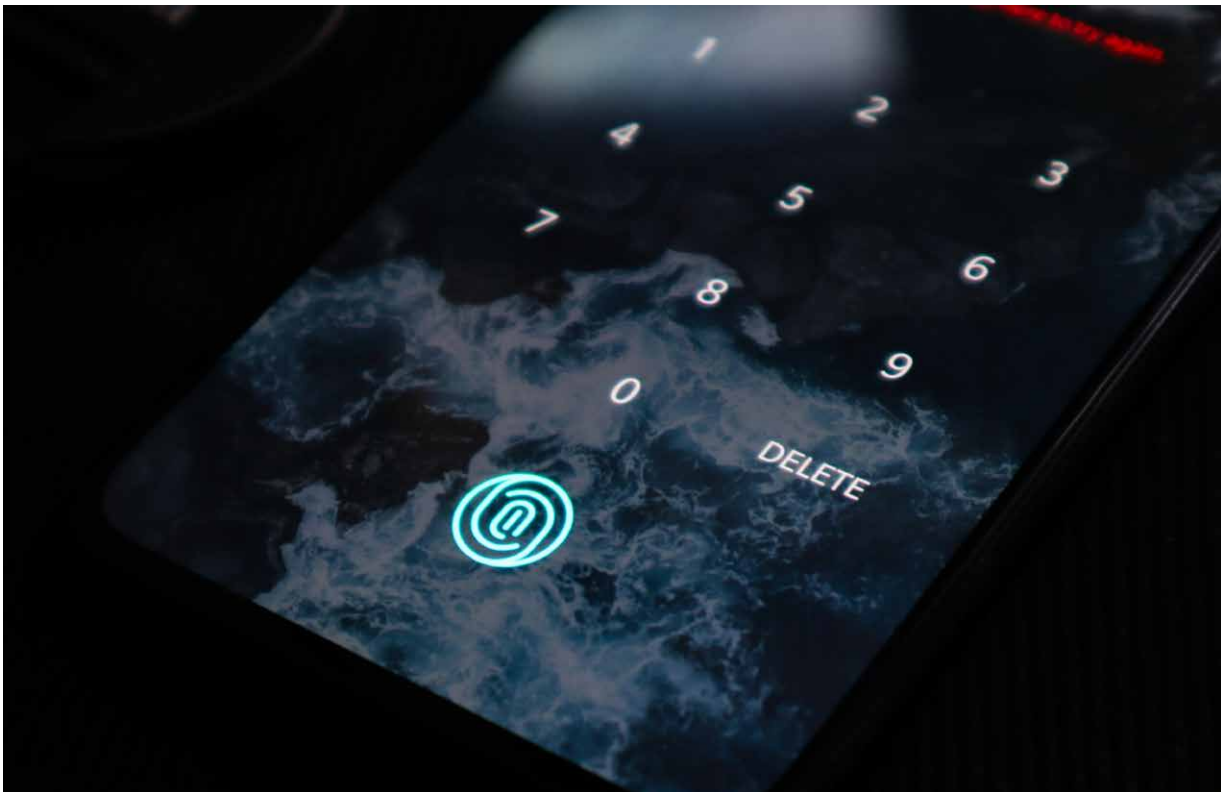
De nödvändiga avvägningarna bör hanteras öppet. Snabbare handling minskar exponeringen, men den måste stödjas av ett tydligt omdöme och tydligt ansvarsutkrävande. En bredare användning av AI kan stärka försvaret, samtidigt som den skapar nya beroenden, nya informationsflöden och nya gränser för

mandat, som i sin tur kan angripas av en motståndare. Ökad informationsdelning kan stärka den gemensamma motståndskraften, förutsatt att känslig information skyddas och att förtroendet upprätthålls.

Utmaningen det kommande året är för organisationer att förkorta tiden mellan att veta och agera. Vi bör fokusera på åtgärder som kan genomföras snabbt, prövas i verklig drift och förbättras över tid. Det innebär att prioritera de tjänster och beroenden som är viktigast, att acceptera att allt inte kan bli perfekt på en gång och att lära tillräckligt snabbt för att ligga steget före hotet. Vi kan inte vänta.

Denna rapport är medvetet kort och översiktlig och kommer att följas upp av en mer detaljerad senare i år. De kommande tolv månaderna bör användas till att höja kvaliteten och tempot i försvaret med starkare cyberhygien, snabbare processer, mer operativ informationsdelning, en styrning som ställer rätt frågor och praktiskt införande av AI i cybersäkerhetsarbetet. Den viktigaste slutsatsen är att det avgörande övertaget kommer att tillhöra de organisationer som snabbare kan omsätta kunskap i handling, med tydligt ansvar, betrodd samverkan och ansvarsfull användning av AI.

# Bilaga




# Metod

Den här delrapporten följer en snabb men spårbar metod. Syftet är inte att utveckla en helt ny teori om AI-drivna cyberhot. I stället utgår den från redan publicerade internationella rapporter och prövar hur stor samstämmighet som finns kring deras mest relevanta slutsatser.

Arbetet inleddes med en genomgång av internationell litteratur från myndigheter, institut för AI-säkerhet, standardiseringsorgan, cybersäkerhetsorganisationer och ledande AI-laboratorier.

Ur dessa källor destillerades ett antal påståenden som grupperades i fyra områden: problembeskrivning, nuläge, prognos och rekommendationer. Syftet var att identifiera vad litteraturen i stort tycks vara enig om, var det råder fortsatt osäkerhet och vilka slutsatser som är relevanta för oss i Sverige inom en tolvmånadershorisont.

De destillerade påståendena bedömdes genom en expertenkät med svarsalternativen Instämmer, Instämmer inte och Avstår, med möjlighet att lämna fritextkommentarer för varje del. Svaren sammanställdes och klassificerades utifrån graden av samstämmighet. Slutsatser med hög samstämmighet kunde användas som gruppens egen ståndpunkt. Slutsatser med lägre samstämmighet behandlades mer försiktigt och tillskrevs, där det var relevant, den underliggande litteraturen.

Arbetsgruppen användes för att pröva om problembeskrivningen var rimlig, om viktiga aspekter saknades, om prognosen var väl avvägd och om de föreslagna inriktningarna var genomförbara inom de kommande tolv månaderna. Den färdiga rapporten presenterar en koncis syntes av de områden där gruppen fann tillräcklig samstämmighet för en delrapport.

# Referenser

AI Sweden. (2026). *AI and Cyber Resilience*. https://www.ai.se/en/ai-labs/secure-ai/ai-and-cyber-resilience/

Anthropic. (2025). *Disrupting AI-enabled espionage operations*. https://www.anthropic.com/news/disrupting-AI-espionage

Anthropic. (2026). *Statement on the US government directive to suspend access to Fable 5 and Mythos 5*. Anthropic. https://www.anthropic.com/news/fable-mythos-access

Australian Signals Directorate. (2024). *Annual Cyber Threat Report 2023–2024*. https://www.cyber.gov.au/about-us/view-all-content/reports-and-statistics/annual-cyber-threat-report-2023-2024

Canadian Centre for Cyber Security. (2024). *National Cyber Threat Assessment 2025–2026*. https://www.cyber.gc.ca/en/guidance/national-cyber-threat-assessment-2025-2026

Cybersecurity and Infrastructure Security Agency. (2023). *Roadmap for Artificial Intelligence*. https://www.cisa.gov/resources-tools/resources/roadmap-ai

IVA. (2022). *Cybersäkerhet för ökad konkurrenskraft*. Kungl. Ingenjörsvetenskapsakademien. https://www.iva.se/publicerat/rapport-cybersakerhet-for-okad-konkurrenskraft/

MIT Technology Review. (2026). *The Meta hack shows there is more to AI security than Mythos*. https://www.technologyreview.com/2026/06/05/1138437/the-meta-hack-shows-theres-more-to-ai-security-than-mythos/

National Institute of Standards and Technology. (2025). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*. NIST AI 100-2e2025. https://csrc.nist.gov/pubs/ai/100/2/e2025/final

National Cyber Security Centre Sweden (NCSC-SE). (n.d.). *10 rekommenderade säkerhetsåtgärder*. https://www.ncsc.se/sv/cybersakerhet/10-rekommenderade-sakerhetsatgarder/

National Cyber Security Centre New Zealand. (2026). *Cyber readiness in the Frontier AI era*. https://www.ncsc.govt.nz/protect-your-organisation/cyber-readiness-in-the-frontier-ai-era/

UK AI Security Institute. (2026). *Our evaluation of Claude Mythos Preview's cyber capabilities*. https://www.aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities

UK National Cyber Security Centre. (2025). *Impact of AI on cyber threat from now to 2027*. https://www.ncsc.gov.uk/report/impact-ai-cyber-threat-now-2027

UK National Cyber Security Centre and UK AI Security Institute. (2026). *Why cyber defenders need to be ready for frontier AI*. https://www.ncsc.gov.uk/blogs/why-cyber-defenders-need-to-be-ready-for-frontier-ai

Kungl. Ingenjörsvetenskapsakademien är en fristående akademi med uppgift att främja tekniska och ekonomiska vetenskaper samt näringslivets utveckling. I samarbete med näringsliv och högskola initierar och föreslår IVA åtgärder som stärker Sveriges konkurrenskraft.

**Utgivare:** Kungl. Ingenjörsvetenskapsakademien (IVA), 2026
Box 5073, SE-102 42 Stockholm
**Telefon:** 08-791 29 00

**IVA-R 541**
**ISSN:** 1100-5645
**ISBN:** 978-91-89181-87-8

**Författare:** Musard Balliu, KTH
**Projektledning:** Per Hjertén, IVA
**Grafisk form:** Pelle Isaksson, IVA

Denna rapport finns att ladda ned på www.iva.se.

IVAs visionsprojekt **Svenska framtider** ska resultera i en väl förankrad och tydlig vision för Sverige som ledande teknik- och innovationsland år 2035 – med fokus på konkurrenskraft, hållbarhet och säkerhet.

AI SWEDEN

IVA
Kungl. IngenjörsVetenskaps Akademien

NATIONELLT CYBERSÄKERHETSCENTER
En del av FRA