

Remissyttrande över Europeiska kommissionens förslag till förordning om utveckling av moln och AI, KOM(2026)502 (Fi2026/01676)

Yttrande från privatperson

Alexander Karlsson

Datum 2026-09-30

SAMMANFATTNING

Jag ser positivt på CADA:s övergripande mål att stärka Europas egen kompetens, infrastruktur och beräkningskapacitet inom moln och AI. Jag avstyrker dock förslaget i dess nuvarande utformning, eftersom den stora utbyggnaden av beräkningskapacitet enligt min bedömning inte kombineras med tillräckliga krav och satsningar på säkerheten kring avancerad AI. Mer beräkningskapacitet gör det möjligt att utveckla och använda allt mer avancerade AI-system. Om EU bygger upp kapaciteten för sådan utveckling behöver därför även förmågan att förstå, testa och kontrollera systemen byggas ut. Under 2026 har flera av världens ledande AI-företag haft allvarliga incidenter där AI-agenter fått obehörig åtkomst till verkliga system. För mig visar det att vår förmåga att kontrollera allt mer kapabla och autonoma system fortfarande är begränsad.

Jag anser därför att Sverige i den fortsatta behandlingen av CADA bör verka för att utbyggnaden av europeisk AI- och beräkningskapacitet tydligare kopplas till AI-säkerhet, oberoende granskning och rapportering av allvarliga incidenter. EU bör också arbeta för större internationellt samarbete kring säkerheten för den mest avancerade AI:n och för överenskommelser som går att kontrollera i praktiken, inklusive tydliga gränser för när utvecklingen behöver bromsas om riskerna blir för stora. Min oro är att konkurrensen mellan länder och företag annars gör att utvecklingen går snabbare än säkerhetsarbetet.

ÖVERGRIPANDE SYNPUNKTER PÅ UTVECKLINGEN AV AVANCERAD AI

Utvecklingen av artificiell intelligens har under de senaste åren gått mycket snabbt. AI-system får allt större förmåga att lösa komplexa uppgifter, använda digitala verktyg och arbeta självständigt. Utvecklingen innebär stora möjligheter, men enligt min bedömning också risker som behöver ges större politisk uppmärksamhet.

Ett av de mest centrala problemen enligt mig är AI-anpassning (eng. *alignment*), det vill säga svårigheten att säkerställa att avancerade AI-systems mål och beteenden förblir förenliga med mänskliga intentioner och säkerhetskrav. Problemet blir särskilt viktigt när systemen får större autonomi och möjlighet att genomföra komplexa uppgifter utan kontinuerlig mänsklig kontroll.

Ett tydligt exempel på detta är den så kallade Hugging Face-incidenten hos OpenAI under 2026. Vid cybersäkerhetstester utnyttjade AI-agenter säkerhetsbrister för att kommunicera mellan separata testmiljöer och samarbeta. Agenterna tog sig sedan ut på internet, genomförde ett cyberangrepp mot Hugging Face och fick till slut administratörsåtkomst till OpenAI:s eget forskningskluster, i samband med försöken att klara utvärderingen.¹ Liknande incidenter där AI-

¹ OpenAI, The Hugging Face incident and the road ahead, 26 augusti 2026; METR, Brief independent investigation of agents' behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident, 26 augusti 2026.

modeller fått obehörig åtkomst till verkliga externa system under utvärderingar har dessutom nyligen bekräftats av både Google och Anthropic.²

Utvecklingen sker dessutom i mycket hög takt. Enligt forskningsorganisationen METR har tidshorisonten för komplexa uppgifter som AI-agenter klarar självständigt ungefär fördubblats var sjunde månad de senaste sex åren.³ Den snabba utvecklingen gör att även ledande företrädare i branschen, som Anthropics vd Dario Amodei, nu varnar för att kapabla agenter redan inom 6–12 månader skulle kunna skapa bestående botnät som i praktiken kan ”ta över hela internet” och orsaka skador för hundratals miljarder dollar.⁴

Det som oroar mig mest med dessa incidenter är inte bara vad agenterna gjorde, utan *hur* de gjorde det. Att systemen försökte nå sina mål genom att runda begränsningar, manipulera loggar och få det att se ut som att uppgifter hade lösts på rätt sätt illustrerar kärnan i problemet med AI-anpassning (alignment). Om dagens system redan i vissa utvärderingar försöker kringgå kontrollmekanismer väcker det allvarliga frågor om vad betydligt mer kapabla system kommer att kunna göra, särskilt om sådant beteende blir svårare att upptäcka i takt med att systemen blir bättre på att förstå och påverka själva utvärderingen.

Riskerna handlar dessutom inte bara om vad autonoma AI-system kan göra på eget initiativ, utan också om riktat missbruk. I en internationell studie från 2026 med 272 AI-experter ansåg experterna att vapen och cyberangrepp hörde till de AI-risker som kunde orsaka de allvarligaste skadorna inom de kommande fem åren. Med katastrofal skada menades bland annat mer än en miljon dödsfall eller ekonomiska skador på över 100 miljarder dollar.⁵

Det är därför mycket oroande att utvecklingen drivs på i ett tempo där säkerhets- och kontrollmekanismer inte hänger med. Lagstiftning och tillsynsmyndigheter måste ges förutsättningar att hålla jämna steg. Målet är inte att stoppa AI, utan att villkora utvecklingen: när staten och EU nu satsar stort på beräkningsinfrastruktur måste strikta säkerhetskrav och oberoende granskning vara en förutsättning. Om vi inte kan visa att riskerna med allt mer avancerade system kan hanteras på ett tillräckligt säkert sätt måste vi också vara beredda att bromsa takten.

ÖKAD EUROPEISK AI- OCH BERÄKNINGSKAPACITET

Jag ser positivt på målet att öka Europas egen AI- och beräkningskapacitet, men anser att förslaget i sin nuvarande utformning inte ger AI-säkerheten tillräcklig tyngd.

Det är viktigt att Europa inte blir alltför beroende av USA, Kina eller ett fåtal stora företag. AI kan få mycket stor betydelse framöver, och EU behöver både egen kompetens och den infrastruktur som krävs för att vara med i utvecklingen. En starkare position kan också ge Europa större möjlighet att påverka utvecklingen internationellt och driva på för högre säkerhetskrav.

² Reuters, *Gemini hacked three companies in first known breakout by Google's AI*, 18 september 2026; Anthropic, *Investigating three real-world incidents in our cybersecurity evaluations*, 30 juli 2026.

³ METR, *Measuring AI Ability to Complete Long Software Tasks*, 19 mars 2025.

⁴ Dario Amodei, *We Must Pace the Frontier*, september 2026.

⁵ Saeri, A. K., Graham, J., Noetel, M., Slattery, P. et al., *Prioritization of Risks from Artificial Intelligence: A Delphi Study of 272 International Experts*, arXiv, 2026.

CADA har bland annat som mål att minst tredubbla EU:s datacenterkapacitet inom 5–7 år och ska också stödja utvecklingen av mer avancerad AI.⁶ Det är en mycket stor satsning. Min oro är att mer beräkningskapacitet också gör det möjligt att utveckla betydligt mer avancerade AI-system, och att säkerhetsarbetet då inte hänger med.

Jag är medveten om att många av säkerhetskraven på AI-system redan regleras genom AI-förordningen.⁷ Min poäng är inte att CADA ska ersätta den regleringen. Men eftersom CADA handlar om att kraftigt bygga ut den beräkningskapacitet som kan användas för att utveckla avancerad AI, anser jag att det är rimligt att säkerheten också får en tydlig plats i den här satsningen.

Jag anser därför att Sverige i förhandlingarna bör verka för att CADA kompletteras så att satsningar på ökad AI- och beräkningskapacitet tydligare kopplas till krav och resurser för AI-säkerhet. Om Europa ska bygga kapacitet för att utveckla allt mer avancerad AI behöver vi också bygga upp förmågan att förstå, testa och kontrollera systemen.

DE MEST AVANCERADE AI-SYSTEMEN – SÄKERHET, KONTROLL OCH TILLSYN

När EU satsar offentliga resurser på mycket stor beräkningskapacitet, till exempel genom AI-gigafabriker, anser jag att det bör vara möjligt att ställa tydliga säkerhetskrav på hur kapaciteten används för att träna de mest avancerade AI-modellerna.

Ju mer självständiga och kapabla systemen blir, desto större kan konsekvenserna bli om något går fel eller om systemen används på ett farligt sätt. AI-förordningen ställer redan särskilda krav på vissa av de mest kapabla modellerna, bland annat när det gäller utvärdering, riskhantering och rapportering av allvarliga incidenter. Jag anser att dessa krav behöver förstärkas i takt med att systemen och beräkningskapaciteten utvecklas. EU:s satsningar på avancerad AI bör därför också kombineras med större resurser till AI-säkerhet, både forskning om hur systemen fungerar och bättre metoder för att testa om de kan agera på sätt som vi inte förväntar oss eller har kontroll över.

De mest avancerade systemen bör också granskas av oberoende aktörer och inte enbart av företagen som utvecklar dem. Det kan till exempel handla om tester av systemens förmåga att agera självständigt, genomföra avancerade cyberattacker, kringgå säkerhetsbegränsningar eller dölja problematiskt beteende.

En möjlighet är att bygga upp starkare offentliga institutioner för AI-säkerhet inom EU och medlemsländerna. Sådana institutioner skulle kunna testa avancerade modeller, bedriva egen säkerhetsforskning, ge råd till myndigheter och samarbeta med motsvarande organisationer i andra länder.

Allvarliga AI-incidenter behöver rapporteras och granskas. Om ett företag upptäcker att ett avancerat AI-system har kringgått säkerhetsmekanismer, fått obehörig åtkomst till system eller på annat sätt agerat oväntat och potentiellt farligt bör informationen nå ansvariga myndigheter och kunna granskas av oberoende aktörer.

⁶ Europeiska kommissionen, *Proposal for a Regulation establishing a framework of measures for strengthening Europe's cloud and AI ecosystem (Cloud and AI Development Act)*, COM(2026) 502 final, 3 juni 2026.

⁷ Europaparlamentets och rådets förordning (EU) 2024/1689 (AI-förordningen), särskilt artikel 55.

Slutligen måste säkerhetskraven kunna få verkliga konsekvenser. Om framtida AI-system når en nivå där allvarliga risker upptäcks samtidigt som det saknas tillräckliga metoder för att kontrollera dem bör det finnas möjlighet att begränsa eller tillfälligt bromsa fortsatt utveckling eller användning tills riskerna är bättre hanterade.

EUROPEISK SJÄLVSTÄNDIGHET OCH INTERNATIONELL AI-SÄKERHET

Jag ser positivt på CADA:s mål att stärka Europas tekniska självständighet och minska strategiska beroenden inom digital infrastruktur.⁸ Jag tycker också att EU behöver ha tillräcklig egen kompetens och kapacitet inom avancerad AI för att kunna påverka hur utvecklingen sker internationellt och driva på för högre säkerhetskrav.

Samtidigt tror jag inte att de största riskerna med avancerad AI kan hanteras av EU ensamt. Om Europa ställer höga säkerhetskrav samtidigt som utvecklingen fortsätter mycket snabbt i andra delar av världen finns många av riskerna fortfarande kvar. Det kan också uppstå en situation där länder och företag känner sig pressade att utveckla systemen snabbare för att inte hamna efter.

Jag tycker därför att EU bör lägga stor vikt vid internationellt samarbete om AI-säkerhet. Det skulle exempelvis kunna handla om gemensamma metoder för att testa de mest avancerade systemen, informationsutbyte om allvarliga incidenter och internationella överenskommelser om vilka säkerhetskrav som bör gälla när systemen når särskilt avancerade förmågor. Sådana överenskommelser skulle också kunna innehålla tydliga gränser för när utvecklingen behöver bromsas om systemen når vissa särskilt riskfyllda förmågor. EU arbetar redan internationellt med bland annat modellutvärdering, AI-säkerhet och utvecklingen av internationella överenskommelser genom AI Office.⁹

Jag tycker också att sådana överenskommelser så långt som möjligt behöver gå att kontrollera i praktiken. Om länder exempelvis kommer överens om att bromsa utvecklingen av de allra mest avancerade systemen vid en allvarlig säkerhetssituation behöver det finnas sätt att kontrollera att överenskommelsen faktiskt följs. Det skulle exempelvis kunna handla om oberoende granskning och metoder för att verifiera användningen av mycket stora beräkningsresurser. Den typen av kontroll skulle kunna göra internationella överenskommelser mer trovärdiga och minska risken för att länder eller företag känner sig tvungna att fortsätta utvecklingen av rädsla för att andra inte följer samma regler.

För mig är därför europeisk självständighet och internationellt samarbete inte motsatser. Tvärtom tror jag att EU har större möjlighet att påverka internationella regler och säkerhetskrav om Europa samtidigt har egen teknisk kompetens, infrastruktur och kapacitet inom avancerad AI.

⁸ Europeiska kommissionen, Proposal for a Regulation establishing a framework of measures for strengthening Europe's cloud and AI ecosystem (Cloud and AI Development Act), COM(2026) 502 final, 3 juni 2026.

⁹ Europeiska kommissionen, *European AI Office*, 2026.

SAMLADE FÖRSLAG TILL DEN FORTSATTA BEHANDLINGEN

Mot bakgrund av det jag har tagit upp ovan vill jag som privatperson särskilt lyfta följande områden i den fortsatta behandlingen av CADA.

1. **Koppla AI-säkerhet till prioriterade AI-projekt** (artiklarna 8–9)
När EU ger stöd och beräkningsresurser till projekt som utvecklar mycket avancerad AI bör säkerhetsarbete vara en tydlig del av satsningen. Det bör inte lämnas helt åt projekten själva att avgöra hur mycket resurser som ska läggas på säkerhet.
2. **Låt AI-centrumen även arbeta med säkerhetstestning** (artiklarna 4.6 och 5)
De nya AI-centrum som föreslås bör inte bara hjälpa till att utveckla och sprida AI. De bör också kunna bidra till att avancerade AI-system testas ur ett säkerhetsperspektiv.
3. **Bygg upp offentlig kompetens inom AI-säkerhet** (artikel 7)
Om Sverige och andra medlemsländer ska bygga upp större kapacitet för avancerad AI behöver även den offentliga kompetensen inom AI-säkerhet byggas ut. Sverige bör exempelvis överväga att bygga upp en offentlig funktion med särskild kompetens att utvärdera avancerade AI-system och deras risker, liknande Storbritanniens AI Security Institute. En sådan funktion skulle kunna genomföra eller beställa oberoende säkerhetstester och ge stöd till ansvariga myndigheter, så att de inte blir helt beroende av AI-företagens egna säkerhetsbedömningar.
4. **Koppla offentlig beräkningskapacitet till säkerhetskrav** (artiklarna 8–9)
Projekt som får tillgång till offentliga beräkningsresurser bör också behöva följa tydliga säkerhetskrav. Om allvariga säkerhetsproblem uppstår bör det finnas möjlighet att tillfälligt begränsa tillgången tills riskerna har hanterats.
5. **Arbeta för internationella överenskommelser som går att kontrollera i praktiken**
CADA innebär att EU ska få bättre överblick över den europeiska beräkningskapaciteten. Sverige bör verka för att EU bygger vidare på detta och arbetar för internationella överenskommelser om säkerheten kring den mest avancerade AI:n. Sådana överenskommelser bör innehålla tydliga gränser för när utvecklingen behöver bromsas om riskerna blir för stora, och metoder som gör det möjligt att kontrollera att överenskommelserna faktiskt följs.