Adobe

Vändpunkten för AI

Hur AI kan implementeras på ett ansvarsfullt sätt i organisationen

Innehåll

Återvänd till denna sida genom att klicka på den här menyikonen

I. Introduktion: Behovet av att implementera AI på ett ansvarsfullt sätt redan idag.

I takt med att AI blir en omvandlingskatalysator i hela branschen upplever cheferna ett exempellöst behov av att kunna innovera snabbt, hantera konkurrens och effektivisera verksamheten. Behovet av att implementera AI är brådskande, vilket medför nya risker för företagen. Utan noggrann övervakning kan en rasande snabb AI-implementering leda till lagöverträdelser, verksamhetsstörningar och ett långvarigt skadat anseende. Att balansera behovet av snabbhet med kravet på ansvar är inte längre en fråga om avvägning, utan en strategisk nödvändighet.

Att kontrollera AI:n och hantera riskerna kan ofta framstå som ett herkulesarbete. Med nya riktlinjer, ramverk och policyer samt tillkomsten av nya lagar på internationell, nationell och lokal nivå kan dessa omständigheter ofta göra det svårt för organisationer att se var man ska börja, och skapa utmaningar för intressenterna redan från starten. På Adobe bygger våra erfarenheter av ansvarsfull innovation på våra etiska AI-principer: ansvar, ansvarsutkrävande och transparens. Dessa principer har gett oss djupgående insikter om hur utmaningarna ska hanteras.

Våra erfarenheter visar att även om vägen mot ansvarsfull AI-innovation kan framstå som utmanande kan framgång nås med rätt verktyg, strategi och inställning.

Ett av de stora besluten som organisationen behöver fatta när AI-strategierna tas fram är om man ska bygga, köpa eller anpassa en AI-lösning, eller använda en kombination av alla tre alternativen. Följande metod riktar sig till organisationer som tänker köpa AI-lösningar och bygga på befintliga värderingar och bästa affärspraxis för organisationer som skaffar AI-lösningar externt – och möta intressenterna där de befinner sig i stunden. Ramverket bygger på oberoende forskning och expertintervjuer om AI-styrning och erbjuder en framkomlig väg framåt som hjälper organisationer att bedöma sin aktuella position och tillhandahålla bästa praxis för implementeringen av ansvarsfulla AI-principer i hela företaget. Här ingår praktiska steg för upprättandet av riktlinjer för medarbetarnas användning av generativ AI, leverantörsutvärdering med robusta enkäter och uppdatering av AI-styrningsprocesser för att hålla jämna steg med utvecklingen.

Var ni än befinner er på AI-resan – oavsett om ni bedömer organisationens AI-beredskap eller förfinar befintliga strategier – tillhandahåller detta ramverk en dokumenterad metod som kombinerar mänsklig påhittighet och banbrytande AI-styrning för ansvarsfull skalning. Med denna färdplan kan organisationer bedöma, pilottesta, anpassa och övervaka AI-lösningar på ett effektivt sätt och bygga en stark grund som främjar tilliten, minskar riskerna och skapar ett hållbart affärsvärde.

II. Översikt över ramverket: Bygga en skalbar, etisk framtid för AI.

För en lyckad implementering av generativ AI krävs mer än att bara bocka av en rad åtgärder i en lista – här behövs en strategisk flerstegsmetod där varje steg bygger på det föregående, vilket skapar en grund för hållbar innovation och etisk AI-användning. Detta ramverk fungerar som en serie sammanhängande byggblock som är utformade för att integrera en ansvarsfull AI-strategi på varje nivå – från bedömningen av organisationens AI-beredskap till en effektiv skalning och fortlöpande övervakning av AI-systemen.

Snarare än att se AI-implementeringen som en processdriven övning fokuserar detta ramverk på att bygga system som utvecklas i harmoni med organisationens behov. Ramverket betonar den kritiska balansen mellan mänsklig övervakning och avancerad AI-teknik, vilket säkerställer att organisationer kan utnyttja AI-potentialen och samtidigt följa lagkrav och uppnå etiska mål och verksamhetsmål.

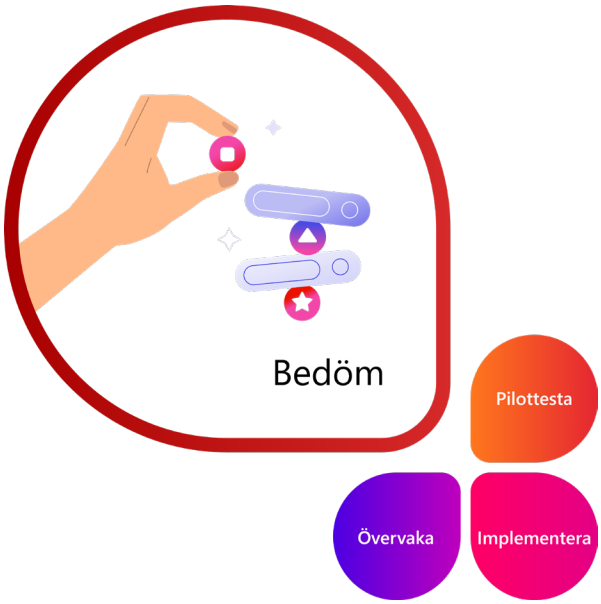
Varje fas i detta ramverk – beredskapsbedömning, ansvarsfulla pilottest, skalad implementering och fortlöpande övervakning – stöder framgången på lång sikt eftersom dessa integrerade pelarna förstärker varandra steg för steg. Genom att bygga in ansvarsfull AI-användning i varje fas kan företag navigera den komplexa AI-implementeringen och samtidigt främja tillit, transparens och ansvarighet.

Bygger på expertis och forskning.

Adobe anlitade ett oberoende forskningsföretag för att undersöka implementeringen av generativ AI, och man samlade in insikter från över 200 IT-chefer, företagschefer och efterlevnadsansvariga inom olika branscher. Undersökningen uppmärksammar nuvarande metoder, utmaningar och framgångsrika strategier för AI-implementering. Adobe har även gjort djupgående intervjuer med branschexperter och granskat globala standarder, däribland EU:s AI-lag, NIST AI Risk Management Framework, Singapores AI Verify, IEEE Standard 7000 och ISO 42001. Detta arbete säkerställer att ramverket kan användas i olika branscher, av stora och små företag och oavsett hur långt AI-implementeringen har hunnit.

1. Bedömning: Organisationens beredskap och valet av ansvarsfullt byggd AI-teknik.

AI-implementeringsresan startar med personerna som ska leda den. **Bedömningsfasen** ger beslutsfattarna de verktyg, data och insikter som behövs för att utvärdera hur AI passar in med organisationens strategiska prioriteringar. Denna fas ger tvärfunktionella chefer möjlighet att undersöka organisationens tekniska infrastruktur, styrningsramverk och AI-kompetens, vilket hjälper dem att avgöra den övergripande beredskapen.



1.1 Utvärdera organisationens AI-beredskap.

Även om många organisationer har påbörjat sin AI-implementering hade bara 21 % av organisationerna i undersökningen fullt utvecklade prioriteringar för en ansvarsfull AI-implementering och 78 % arbetade fortfarande med dem eller befann sig på planeringsstadiet, vilket understryker behovet av beredskapsstrategier. De som ansvarar för IT, efterlevnad, riskhantering och strategi är oumbärliga när grunden för en ansvarsfull AI-användning ska skapas. Arbetet börjar med en heltäckande granskning av organisationens styrningsramverk och AI-kompetens för att identifiera luckor som kan påverka AI-implementeringen.

Organisationerna behöver ha en **helhetsstrategi** för utvärdering av AI-beredskapen, som kombinerar **ledningsinitiativ uppifrån och ner** med **feedback nerifrån och upp** från medarbetare som dagligen arbetar med AI.

Beredskapsåtgärder:

Gör en heltäckande granskning av beredskapen – utvärdera organisationens tekniska infrastruktur, styrningsstandarder, AI-relaterade policyer, ramverk för ansvarsfull innovation och efterlevnadstillämpningar för att identifiera styrkor och förbättringsmöjligheter – för att säkerställa inriktningen med strategiska mål såväl som kraven på en ansvarsfull AI-implementering.

Identifiera och hantera betydande luckor tillsammans – dokumentera ytterligare behov i AI-policyn avseende säkerhet, integritet, lagkrav och efterlevnads- och transparensstandarder i samarbete med tvärfunktionella team, inklusive team som jobbar med IT, juridiska frågor, efterlevnad och affärsmål – för att prioritera nästa steg framåt.

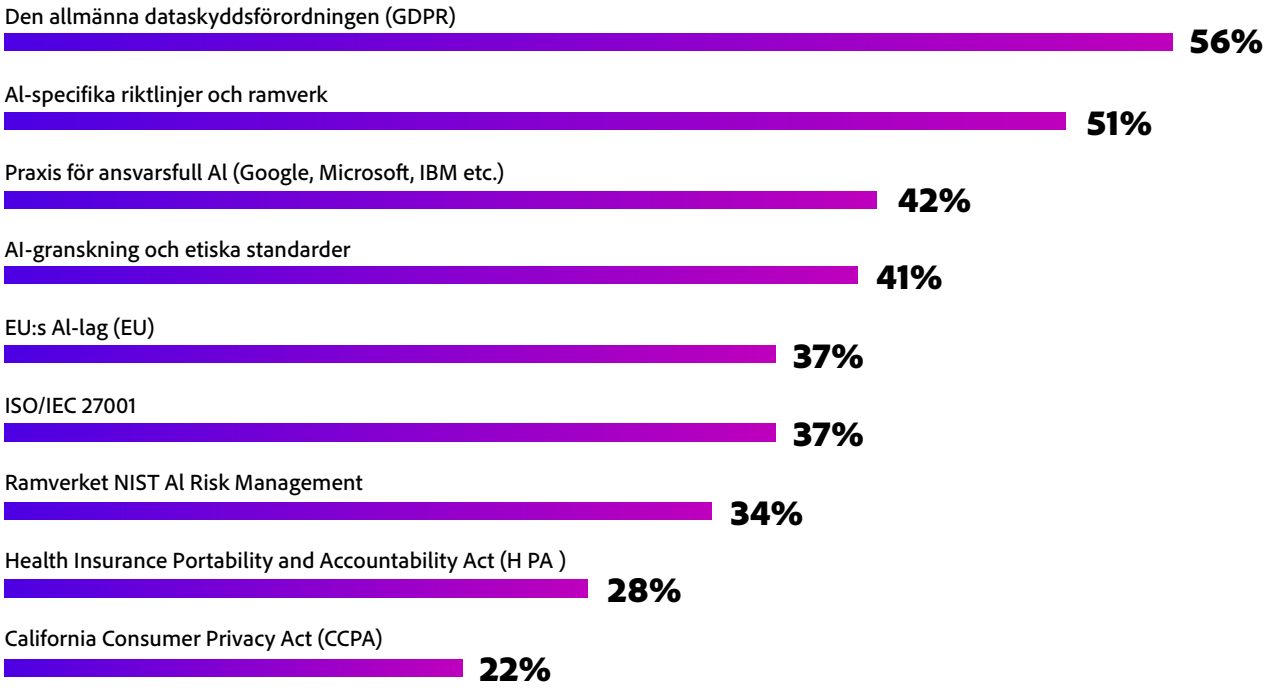
Etablera styrningsteam – utse team som ska övervaka AI-styrningen för att säkerställa efterlevnaden av interna standarder för ansvarsfull AI-användning och externa regelverk – och ge dessa team de befogenheter och resurser som behövs för att på ett proaktivt sätt hantera risker och göra anpassningar i enlighet med tillkomsten av nya krav.

1.2 Välj en AI-lösning som är byggd på ett ansvarsfullt sätt.

Börja med en grundlig genomgång av företagets befintliga styrningsstandarder. Dessa standarder omfattar förmodligen redan viktiga områden, till exempel **integritet, säkerhet, tillgänglighet** och **juridiska frågor**. Globala **riktmärken** som till exempel **den allmänna dataskyddsförordningen (GDPR)** och **AI-specifika ramverk** är en del av att upprätthålla efterlevnaden och riskövervakningen i många organisationer. Även **regionala policyer** och **branschspecifika standarder** – till exempel **AI-granskningar** och **ansvarsstandarder** – ska införlivas i styrningsstandarderna.

Nödvändiga säkerhets- och integritetsstandarder eller certifiering av AI-lösningar.

% av det totala antalet respondenter, sorterade i fallande ordning.



När en organisation har fastställt sina förväntningar på ansvarsfull AI och styrningsramverk är nästa steg att fastställa kriterier för ansvarsfullt byggda AI-lösningar. Dessa kriterier ska integrera befintliga standarder och fokusera på unika element som förknippas med generativ AI, till exempel transparens avseende ursprung, resultatens korrekthet, licensiering av träningsdata, minskning av fördomar och kulturell lokalisering.

Enligt forskningsresultaten är de främsta kriterierna för organisationer som utvärderar generativa AI-lösningar de följande:

1. **Utvärdering av träningsdata (72 %)**
2. **Upptäckter om AI-användning (63 %)**
3. **Skadebegränsning (60 %)**
4. **Transparens avseende ursprung (55 %)**
5. **Minskning av fördomar (50 %)**

Dessa faktorer säkerställer att de valda AI-lösningarna uppfyller både **affärsbehoven** och **de etiska kraven** för att stödja organisationens framgång på lång sikt.

Organisationer bör ta fram inriktade urvalskriterier som samordnar AI-lösningar med strategiska affärsmål såväl som ansvarsfulla AI-principer. Dessa kriterier poängterar följande:

Transparens Säkerställer att AI-processerna kan förklaras och spåras.	**Kulturell lokalisering** Implementerar AI-system som respekterar olika kulturella och regionala sammanhang.
Korrekthet Upprätthåller höga standarder för datans fullständighet och förutsägbara tillförlitlighet.	**Minskning av fördomar** Minskar på ett aktivt sätt fördomar för att stödja rättvisa och opartiska AI-resultat.

Att dokumentera varje nivå av bedömnings- och urvalsprocessen stärker anpassningsförmågan och ansvarigheten och skapar en flexibel styrningsmodell som kan utvecklas i takt med AI-framsteg och lagändringar. Nedan följer en sammanfattning av stegen som tas i bedömningsfasen:

Bedömning

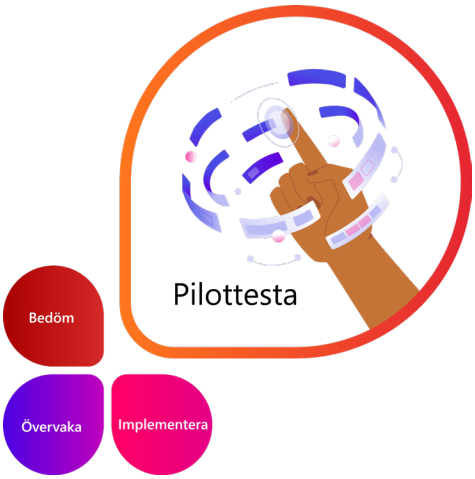
Steg 1: Utvärdera organisationens AI-beredskap.

- Definiera och kommunicera företagets standarder för ansvarsfull teknikanvändning, inklusive AI-användning.
- IT-chefer och/eller en företagskommitté granskar nuvarande system och affärsprocesser för att identifiera de områden som gynnas mest av en ansvarsfull AI-implementering.
- Sammanställ insikter från interna företagschefer och funktionella ledare om ytterligare användningsfall som kan övervägas för en ansvarsfull AI-implementering.

Steg 2: Välj en AI-lösning som är byggd på ett ansvarsfullt sätt.

- Granska befintliga styrningsstandarder avseende integritet, säkerhet, tillgänglighet och juridiska frågor i samband med AI.
- Ta fram urvalskriterier som kan integrera tidigare etablerade standarder som uppfyller förväntningarna på ansvarsfull AI genom att fokusera på transparens, riktighet, fördomar, kulturell lokalisering och efterlevnad.
- Utvärdera och välj AI-lösningar som på bästa sätt uppfyller fastställda kriterier och affärsbehov, och dokumentera beslutsfattandeprocessen.

2. Pilottestning: Identifiera och pilottesta användningsfall med stor genomslagskraft.



Pilotfasen förenar AI-experimenten med verksamhetens realiteter. I denna fas kan de främsta intressenterna utvärdera lösningens prestanda avseende hur den förhåller sig till affärsmålen och målen för ansvarsfull AI. Här går man längre än till att testa den tekniska genomförbarheten och fokuserar på att ge de mest involverade ledarna och intressenterna möjlighet att interagera direkt med lösningen på ett meningsfullt sätt. Det handlar om att ge medarbetarna möjlighet att jobba med AI, fatta välgrundade beslut om lösningens skalbarhet och säkerställa att den uppfyller etiska, verksamhets- och reglemässiga standarder. I pilottestfasen kan organisationer stresstesta AI-systemen i rätt kontext och lättare förstå var ansvarighetsbedömningar och transparensdokumentation kan behövas, och även testa hur väl de nya funktionerna motsvarar förväntningarna. Genom att dokumentera insikter och samla in användbara lärdomar kan organisationer ta fram en färdplan för ansvarsfull AI-skalning och skapa en grund som stöder både omedelbara och långsiktiga mål.

2.1 Identifiera och förbereda prioriterade användningsfall.

För att utveckla övertygande AI-affärsfall ska viktiga intressenter och medarbetare i frontlinjen inkluderas för att skapa en helhetsbild av AI-lösningens potential. Genom att involvera alla som kommer att interagera direkt med lösningen i ett tidigt skede kan organisationer identifiera användningsfall med stor genomslagskraft där AI:n levererar påtagliga fördelar, till exempel för marknadsföringsinnehåll, kodning, arbetsflödesautomatisering och datahantering.

Var specifika – fokusera på processer, inte på roller: Snarare än att bilda användningsfall kring specifika roller (till exempel ”AI för utvecklare”) bör ni fokusera på processer som kan effektiviseras med hjälpa av AI, till exempel ”AI-assisterad kodning för automatisering av rutinmässig kodgranskning och feldetektion”.

Ta fram mätbara mätvärden för användnings- och kostnadsbesparingar: Även om investeringsavkastningen är viktig bör AI-pilottester också betona avkastningen i en bredare mening, till exempel avseende produktivitet, time-to-market, arbetstillfredsställelse och förbättrade kundupplevelser – mätvärden som kan beskrivas som ”upplevelseavkastning”.

Få en högre effekt än bara omedelbara vinster: Positionera AI-initiativet som en drivkraft för långsiktig omvandling. Användningsfallen ska inte bara avse omedelbara verksamhetsbehov utan även samordnas med strategiska mål, till exempel digitalisering eller differentiering gentemot konkurrenterna.

2.2 Pilottestning mot affärsresultat och kriterier för ansvarsfull AI.

Pilottesterna ska utvärderas från två håll – affärsresultat och kriterier för ansvarsfull AI – för att säkerställa att AI-initiativen uppfyller både affärsmål och riktmärken för ansvarsfull AI. Över hälften (54 %) av organisationerna i undersökningen har fastställt en acceptabel risknivå för sina prioriterade användningsfall. Organisationer bör dokumentera dessa utvärderingar systematiskt och inhämta lärdomar som kan ligga till grund för framtida AI-projekt. Denna strukturerade strategi skapar en robust grund för skalbara AI-implementeringar.

Åtgärder vid pilottestning:

Ange riktmärken för affärsmål och ansvarsfull AI: Definiera både verksamhetsmål (till exempel produktivitet, kostnadsbesparingar) och mätvärden för ansvarsfull AI (till exempel transparens, rättvisa).

Fastställ risktrösklar: Ange riskparametrar och skapa ett ramverk för fortlöpande bedömningar för att på ett effektivt sätt hantera och minska de AI-relaterade riskerna.

Inhämta och dela lärdomar: Utveckla en standardiserad process för dokumentation av resultat från pilottesterna som stöd för transparensen och vägledning för framtida skalningsprojekt.

Pilottestning

Steg 1: Identifiera prioriterade användningsfall.

- Identifiera 2–3 pilottester bland de befintliga affärsanvändningsfallen där en etisk och ansvarsfull AI-användning är viktig.
- För dessa användningsfall fastställer ni mätvärden och trösklar för att spåra AI-prestandan avseende både affärsmål och ansvar.

Steg 2: Pilottestning mot affärsresultat och kriterier för ansvarsfull AI.

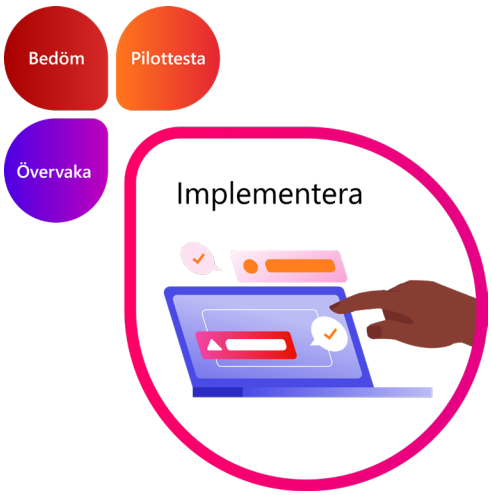
- Genomför pilottester med ytterligare validering och testning av teknik, affärsmål och ansvar efter behov.
- Utvärdera testresultaten gentemot förhandsdefinierade mätvärden och trösklar för affärsförväntningar och förväntningar på ansvarsfull AI, och dokumentera lärdomarna för framtida bedömnings- och testningsstrategier.
- Fortsätt till inköp/implementering baserat på resultat och insikter från pilottesterna.

---

3. Implementering: Integrera AI på ett ansvarsfullt sätt i hela organisationen.

Implementeringsfasen kännetecknar övergången från pilottestning till integrering i hela organisationen. Denna fas fokuserar på övergången från experimentella tillämpningar till fullt fungerande AI-system. AI:n driftsätts på ett ansvarsfullt sätt, samtidigt som lärdomarna från pilottesterna byggs in i den verkliga användningen.

I denna fas blir medarbetarna aktiva ägare av AI-lösningens roll i de befintliga arbetsflödena. Genom att bygga på deras praktiska erfarenheter från pilotfasen är de rustade att driva AI-implementeringen.



3.1 Utbilda organisationen.

För en effektiv AI-skalning krävs en kompetent arbetskraft som förstår både AI-funktionerna och det etiska ansvar som AI-användningen innebär. Skräddarsydda utbildningsprogram hjälper medarbetarna i olika roller och på olika avdelningar att utnyttja AI-verktygen. Många organisationer (89 %) är medvetna om utbildningens betydelse, och nära två tredjedelar inkluderar riktlinjer för ansvarsfull AI-användning. Utbildningen ska integrera tekniska funktioner och principer för ansvarighet, transparens och regelefterlevnad.

Utbildningsåtgärder:

Samordna utbildningen med styrning: Införliva riktlinjer för ansvarsfull AI i utbildningsmaterialet för att säkerställa att medarbetarna är medvetna om kraven på efterlevnad, riskhantering och transparens.

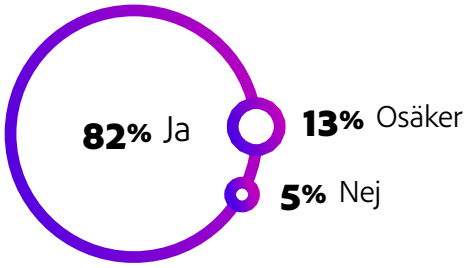
Anpassa utbildningen efter rollerna: Ta fram skräddarsydda utbildningsmoduler som inriktar sig på specifika funktioner, inklusive bästa praxis för både affärsmål och ansvarsfull AI.

3.2 Implementera med ansvarighet i åtanke.

En storskalig AI-implementering kräver att man upprättar ett styrningsramverk som säkerställer en ansvarsfull användning. Organisationer bör samordna sina AI-initiativ med befintliga styrningspolicyer och samtidigt fortsätta att förfina sina policyer så att man kan följa nya AI-standarder för regelefterlevnad, verksamhet och ansvarighet.

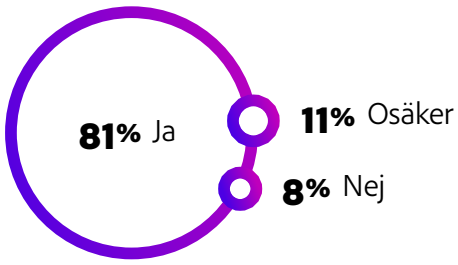
Planerar att införliva frågor om ansvarsfull AI i standarder för bred implementering.

% av det totala antalet respondenter, sorterade i fallande ordning.



Inkludering av frågor om ansvarsfull AI i IT-styrningen.

% av det totala antalet respondenter, sorterade i fallande ordning.



Åtgärder för ansvarsfull implementering:

Främja en kultur av att ta ansvar: Uppmuntra teamen till att förstå vilken inverkan som AI har både på verksamhetens arbetsflöden och intressenternas tillit, så att ni väcker ansvarskänslan på varje nivå.

Fortlöpande utbildning: I takt med att ramverken för AI-styrning utvecklas uppdaterar ni utbildningsprogrammen så att de alltid reflekterar aktuell bästa praxis och nya regler.

Implementering

Steg 1: Utbilda organisationen.

- Utveckla specifika riktlinjer för olika användningsfall avseende när och hur AI ska användas med principer för ansvarsfull AI.
- Ge relevanta grupper en heltäckande utbildning, inklusive bästa praxis, i takt med att lösningarna rullas ut.
- Fira och dela framgångar med hela organisationen.

Steg 2: Implementera med ansvarighet i åtanke.

- För varje teknisk lösning ska de huvudsakliga implementeringskraven kodifieras (till exempel affärspåverkan, integreringens enkelhet och riskminskning).
- Jobba tätt ihop med företagsledare för att vid behov samordna kompromisser.
- Bygg in viktiga frågor som gäller AI och ansvar i de befintliga styrningsramverken (till exempel åtkomst, kontroll, roller).

4. Övervakning: Fortlöpande övervakning och förbättring.

När AI-systemen implementeras fullt ut blir en fortsatt övervakning och förbättring väsentlig. **Övervakningsfasen** betonar realtidsspårning, rigorösa prestandagranskningar och proaktiv riskhantering för att säkerställa att AI-systemen fortsätter att vara effektiva, följa reglerna och vara i linje med organisationens mål. Genom att bygga in mätvärden för ansvarsfull AI och etablera en strukturerad granskningsprocess kan organisationer anpassa sig till nya lagar och risker, och samtidigt främja tillit och verksamhetsvärde.

4.1 Övervaka prestandan avseende riktmärken för affärsmål och ansvarsfull AI.

Klassens bästa övervakningslösningar kombinerar automatiserad prestandaspårning med mänsklig expertis. Även om många organisationer har implementerat verktyg för realtidsövervakning för att bedöma AI-systemens prestanda blir dessa verktyg effektivare i kombination med mänsklig övervakning. Mänskliga team är bättre utrustade för att analysera data, upptäcka risker och fatta välgrundade beslut om nödvändiga justeringar.

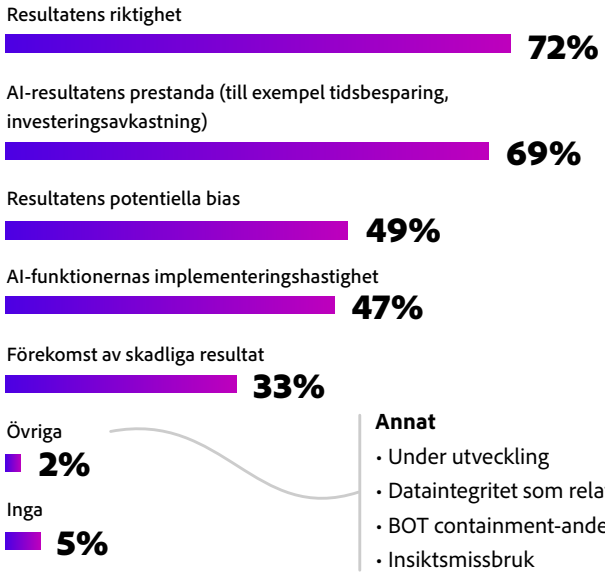
Även om 69 % av organisationerna använder verktyg för realtidsövervakning blir dessa betydligt effektivare när de kombineras med mänskligt omdöme. Många organisationer prioriterar tekniska mätvärden – 72 % fokuserar på riktighet och 69 %

på investeringsavkastning – men en ansvarsfull skalning kräver att även etiska dimensioner beaktas. Inbyggd mänsklig övervakning säkerställer transparens och förutsägbarhet, vilket skapar tillit både internt och externt. Vidare möjliggör övervakning en proaktiv fördomsdetektering, och 49 % av organisationerna spårar detta mätvärde medan 33 % övervakar skadlig output. Utan en fortlöpande, proaktiv övervakning kan AI-systemen äventyra både integritet och tillit. Genom att förfina prestandaövervakningen för att hantera både tekniska och etiska risker skyddar företagen sitt varumärke, skapar tillit hos användarna och bygger en stabil grund för ansvarsfull AI-skalning.

Samarbetet mellan teknik och människa ger organisationer möjlighet att identifiera potentiella risker i ett tidigt skede, till exempel felaktiga data, nya fördomar eller efterlevnadsavvikelser.

AI-specifika frågor om spårningsteknikens prestanda och effektivitet.

% av det totala antalet respondenter



Annat
- Under utveckling
- Dataintegritet som relaterar till träningsläge
- BOT containment-andel
- Insiktsmissbruk

Närmare en tredjedel av organisationerna saknar personal som kan stödja fortlöpande förbättringar av de tekniska lösningarnas prestandamätvärden och affärsresultat

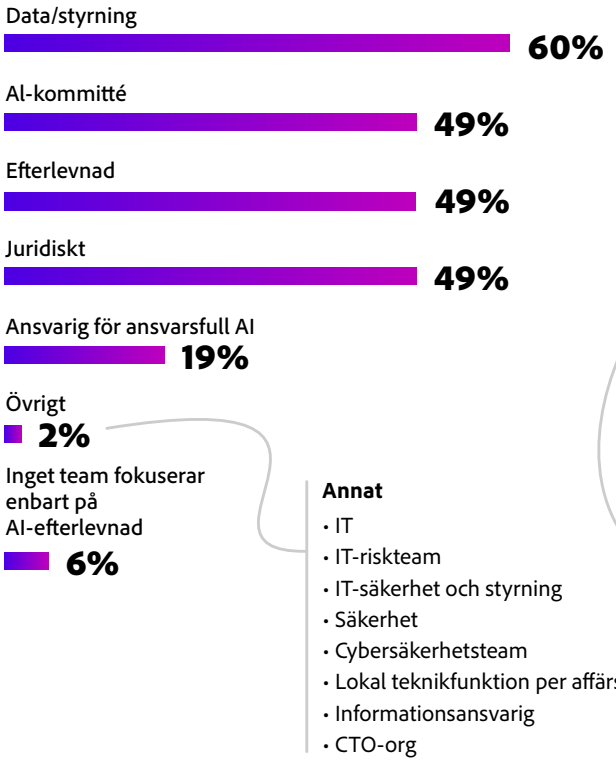
4.2 Fortlöpande riskhantering.

Riskhanteringen i samband med AI är en fortlöpande process som bör utvecklas samtidigt som AI-systemen. Etableringen av en strukturerad, tvärfunktionell strategi för AI-riskhantering ger organisationer möjlighet att på ett proaktivt sätt hantera risker avseende både affärsmål och anseende. Schemalagda granskningar bör involvera intressenter från hela organisationen, inklusive datavetare, affärsledare och jurister/efterlevnadsansvariga för att säkerställa en noggrann utvärdering av målen för både teknisk prestanda och ansvarsfull AI.

AI-riskhanteringen är en fortlöpande process som utvecklas i takt med de teknologiska framstegen. Sextio procent av organisationerna involverar data- och styrningsteamen, och 49 % inkluderar AI-kommittéer och efterlevnads- och juridiska team – vilket betonar behovet av ett tvärfunktionellt samarbete. Denna proaktiva strategi ger organisationer möjlighet att samordna både interna värderingar och externa förväntningar. ”Varför” handlar om att bygga en styrka som kan anpassas till nya regler. När 68 % av organisationerna betonar ansvarsfull AI i riskhanteringen är heltäckande dokumentation och fortlöpande riskbedömning grundläggande.

AI-specifika frågor om spårningsteknikens prestanda och effektivitet.

% av det totala antalet respondenter

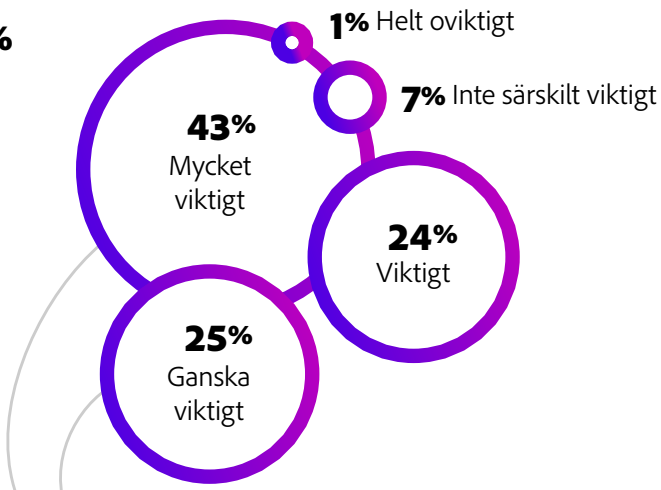


Annat

- IT
- IT-riskteam
- IT-säkerhet och styrning
- Säkerhet
- Cybersäkerhetsteam
- Lokal teknikfunktion per affärsenhet
- Informationsansvarig
- CTO-org

Vikten av att teamen använder AI på ett ansvarsfullt och etiskt sätt för att säkerställa efterlevnaden.

% av dem som indikerar att ett särskilt team fokuserar på att övervaka AI-regleringar och standarder



68%
Mycket viktigt
+
Ganska viktigt

Genom att upprätta rigorösa prestandamätvärden och främja en kultur av fortlöpande riskhantering kan organisationer säkerställa att AI-implementeringar förblir samordnade med både affärsmål och prioriteringar för ansvarsfull AI. Proaktiv övervakning i kombination tvärfunktionella granskningar och noggrann dokumentation positionerar organisationer för att leda AI-driven omvandling med både självsäkerhet och ansvarighet.

Nyckelåtgärder för riskhantering:

Etablera tvärfunktionella riskgranskningar: Skapa regelbundna riskbedömningar som involverar datavetare, efterlevnadsansvariga och juridiska experter för att identifiera nya risker baserat på realtidsdata.

Spåra och rapportera fortlöpande: Samla fortlöpande in feedback från medarbetare och slutanvändare för att detektera användningsproblem, fördomar och oväntade beteenden.

Övervakning

Steg 1: Övervaka prestandan.

- Definiera och spåra mätvärden för AI-prestanda på längre sikt, inklusive affärs- och ansvarsmål.
- Etablera fortgående granskningar och diskutera resultaten för att fortlöpande förbättra affärsresultaten och samtidigt garantera organisationens ansvarighetsprinciper.

Steg 2: Implementera med ansvarighet i åtanke.

- Tilldela och stärk roller för att spåra kommande AI-regleringar och standarder (till exempel Singapore AI-Verify och förslag från USA:s kongress) och säkerställa att företagets standarder uppdateras i enlighet med dem.
- Utveckla en process för fortlöpande identifiering och minskning av risker som förknippas med AI-användning.
- Uppdatera dokumentationen om hur organisationen följer företagsstandarderna.

ADOBES FALLSTUDIE

Intern användning av generativ AI.

På Adobe ser vi generativ AI som en omvälvande tekniklösning med kraft att öka den mänskliga kreativiteten, inte ersätta den. Vi stöder en intern ansvarsfull utforskning av generativ AI, som samordnas med våra egna AI Ethics-principer avseende ansvar, ansvarsutkrävande och transparens.

I juni 2023 etablerade Adobe en tvärfunktionell intern arbetsgrupp, som sponsrades av vår CIO och CHRO, för att hjälpa medarbetarna att navigera utforskandet och använda generativ AI inom Adobe på ett säkert, ansvarsfullt och flexibelt sätt. Denna grupp, som samarbetar med ledare och ämnesexperter från hela företaget, fokuserar på att leda vägen för en eftertänksam strategi för hur medarbetare på gräsrotsnivå experimenterar med AI genom att förstå de hypotetiska användningsområdena för generativ AI, skapa lämpliga riktlinjer och effektivisera experimenten. Initiativet har lett till att fyra arbetsgrupper har bildats, som baseras på kundtyper och representerar användningsfall med generativ AI på Adobe. Man har upprättat en inhämtningsprocess, ett ramverk för vilka risker som kan tolereras avseende generativ AI och en mall för granskning av användningsfall som tar hänsyn till nya frågor som rör etik, säkerhet och integritet samt andra juridiska frågor. En lista över godkända verktyg för generativ AI, modeller som baseras på specifika användningsfall samt riktlinjer för medarbetarnas användning av generativ AI finns också. Utrullningen av leverantörsriktlinjer för generativ AI i mars 2024 inkluderade utbildningstillfällen för användning av generativ AI och funktionerna i de valda produkterna.

Implementeringen av detta initiativ har bidragit till att effektivisera processen för snabbare experimentering och skalad tillämpning när så är möjligt, och möjliggjort en bedömning av landskapet för generativ AI i hela företaget. Adobe fortsätter befrämja att lärdomar och insikter delas inom företaget för att genom samarbete skapa ett ekosystem för kollektivt utforskande. Programmet fortsätter att utvecklas i takt med den ökande användningen av generativ AI i våra egna produkter och spridningen av generativ AI-teknologi och generativa AI-modeller såväl som framväxten av vägledande regelverk. Experimentgranskningen övervakas – teamet utvecklar experimentspårning efter godkännande, inklusive för sådana experiment som går i storskalig produktion.

III. Bygga in bästa praxis i organisationen.

För att implementera, övervaka och optimera AI-systemen bör organisationer fokusera på flera verksamhetsområden: ge medarbetarna heltäckande vägledning, utvärdera leverantörer på ett rigoröst sätt och upprätta robusta medel för AI-styrning. På så sätt kan företag säkerställa att deras AI-initiativ inte bara uppfyller nya regelverk utan även bygger på befintligt styrnings- och riskhanteringsarbete och samtidigt samordnas med praxis som främjar tillit, transparens och ansvarighet.

I detta avsnitt beskriver vi de praktiska stegen för att bygga in bästa praxis i den dagliga verksamheten.

1. Vägledning för medarbetarnas användning:

Att skräddarsy riktlinjerna för AI-användning så att de överensstämmer med de specifika behoven och riskerna i er organisation är helt väsentligt för att säkerställa en ansvarsfull implementering. Dessa riktlinjer ska hjälpa medarbetarna att navigera regelverken och styrningsprotokollen, och samordna AI-lösningarna med åtaganden för datasäkerhet, transparens och ansvarighet.

1.1 Känsliga data:

Specificera tydligt när databehandlingen ska ske lokalt eller under strikt åtkomstkontroll för att förhindra obehörig åtkomst. Detta innebär också att man avstår från att använda beskrivningar som kan generera eller manipulera känsliga resultat. Dessa riktlinjer skyddar egen information och upprätthåller efterlevnaden av dataintegritetslagarna.

1.2 Transparens i AI-användningen:

Visa tydligt att AI har använts, till exempel när interna dokument, kundgränssnitt eller externa budskap har skapats med hjälp av AI. På så sätt främjar ni ansvarstagande och skapar tillit till det AI-genererade innehållets autenticitet och tillförlitlighet och skyddar företagets varumärke och anseende.

1.3 Policyer för kontohantering:

Skapa uttryckliga policyer för användningen av sådana verktyg för generativ AI som kräver kontoregistrering. Definiera även huruvida organisationens e-postkonton kan användas och specificera vilka verktyg som är godkända för affärssyften, och förhindra att privata konton används för arbetsrelaterat innehåll. På så sätt förebygger ni obehörig användning och bidrar till att upprätthålla samordningen med organisationens bredare informationssäkerhetsprogram.

Genom att skräddarsy dessa riktlinjer efter organisationskontexten kan medarbetarna navigera användningen av verktygen för generativ AI med gott självförtroende och på ett ansvarsfullt sätt, vilket bidrar till att skapa en miljö där innovation och integritet går hand i hand.

2. Utvärdering av leverantörer: Exempel på frågor.

Utvärderingen av AI-leverantörer kräver att ni ställer informativa frågor och förstår vilka svar ni vill ha, så att ni kan säkerställa att systemen stöder ansvarsfull AI och följer juridiska standarder och regelverk. De följande frågorna är utformade för att tillhandahålla en grundläggande utvärdering som ger organisationer möjlighet att fatta välgrundade beslut om potentiella samarbeten och minska riskerna som förknippas med AI-implementering.

Tema	Frågor till leverantörer	Framställning	Adobe-exempel
Dataproveniens och-användning	*"Vilka specifika datatyper har använts när AI-systemet utvecklades och tränades?"*	Denna fråga ger insikter om källorna, arten och omfattningen avseende de data som användes när AI-modellerna tränades. Denna fråga säkerställer att leverantörens datapraxis överensstämmer med köparens standarder för ansvarsfull AI och juridiska mål.	**Firefly:** Adobe inkluderar inte innehåll från företagsanvändare (inklusive input och output från Firefly) i datamängder som används för att träna Firefly-grundmodeller.
Efterlevnad av immaterialrätt	*"Har datamängder som kan ha begränsningar avseende upphovsrätt, immaterialrätt eller licensiering använts?"*	Denna fråga verifierar att alla datakällor är lagliga och har erhållits med godkända mekanismer, så att potentiella rättstvister kan undvikas.	**Firefly:** Kunderna kan när som helst granska licenshistoriken på stock.adobe.com/Dashboard/LicenseHistory när de är inloggade med sina kontouppgifter från Adobe.
Transparens avseende träningsdata och logik	*"Kan ni tillhandahålla en detaljerad förklaring av vilka träningsdata och vilken logik som tillämpades när AI-systemet utvecklades?"*	Transparensen avseende dessa aspekter identifierar potentiell bias och tillhandahåller en förståelse av modellens slutledningsprocess, vilket är väsentligt för modellens tillförlitlighet och rättvisa.	**AEP:s AI-assistent:** Adobe använder inga kunddata för att träna eller finjustera tjänsten Azure OpenAI.
Tydlig output	*"Kan ni i klartext beskriva AI-systemets output?"*	Denna fråga säkerställer att all output är begriplig för granskare utan teknisk kompetens, vilket möjliggör ett effektivt beslutsfattande och en ansvarsfull användning.	**Firefly**: Adobe genererar automatiskt innehållsreferenser för vissa Firefly-genererade resurser för att bidra till att skapa transparens om att resursen skapades med generativ AI.
Mänsklig övervakning	*"Om mänsklig övervakning är en del av AI-systemet, i vilken mån och på vilket sätt är människor i så fall involverade?"*	Genom att veta hur automatiserade processer och mänskligt omdöme balanseras kan ni utvärdera systemets driftsmässiga dynamik och identifiera områden där mänsklig interaktion kan behövas för att upprätthålla standarderna för kvalitet och ansvar.	**AI-assistenten i Acrobat:** Adobe begränsar på ett strikt sätt informationsåtkomsten till ett litet antal utbildade Adobe-medarbetare som är direkt involverade i utvecklingen av Adobes tjänst för generativ AI.
Utvärdering av rättvisa och fördomar	*"Hur har AI-systemet utvärderats avseende fördomar, och vilka resultat gav dessa utvärderingar?"*	Denna fråga klargör leverantörens åtagande för rättvis AI-användning och vilka metoder som använts för att detektera och minska fördomar som kan påverka olika demografiska grupper på ett oproportionerligt sätt.	**Acrobats AI-assistent**: Adobe-teamen utför tester som minskar risken för fördomsfulla och skadliga resultat från våra produkter för generativ AI. Se översikten över generativ AI för företag.
Riskminskning	*"Har ni utvärderat potentiella skadliga utfall och vilka åtgärder har i så fall implementerats för att minska dessa risker?"*	En utvärdering av leverantörens proaktiva åtgärder för att identifiera och hantera potentiella negativa utfall visar hur skaderisker hanteras. Detta innebär att AI-systemet har testats för operativ säkerhet och ansvarighet.	**AEP:s AI-assistent:** Adobe använder internt utvecklade innehållsfilter för att (a) fastställa huruvida inmatningen (beskrivningen) i AEP:s AI-assistent följer Adobes användarriktlinjer för generativ AI och (b) filtrerar bort eventuella svar som bryter mot dessa riktlinjer (till exempel hat och svordomar).

Genom att ställa frågor med denna inriktning kan organisationer utvärdera AI-leverantörer och göra välgrundade val som är samordnade med standarder för AI-ansvarighet och verksamhetsprioriteringar. Denna strategi möjliggör riskhantering och säkerställer att relationerna med AI-leverantörer bygger på transparens, efterlevnad och ansvarsfull innovation.

3. Medel för AI-styrning.

Implementering av en robust AI-styrning säkerställer att AI-systemen utvecklas, implementeras och övervakas på ett sätt som är samordnat med organisationens värderingar och regelverk. Många regelverk förekommer, till exempel EU:s AI-lag och ramverk som Singapores AI Verify. I USA ska företag följa statligt övergripande integritetslagar och ramverket NIST AI Risk Management, som är den troliga grunden för framtida lagstiftning.

Följande styrmedel kan hjälpa organisationer att hantera AI-risker och öka transparensen, ansvarigheten och säkerheten.

AI-lagringsplatser	Upprätta lagringsplatser för AI-system och kategorisera dem i enlighet med riskprofiler och strategiska prioriteringar så att de fungerar som en central databas.	Har ni dokumenterat era AI-användningsfall och kategoriserat relevanta risker?
Feedback-mekanismer	Etablera robusta kanaler för feedback så att ni kan inhämta insikter utanför AI-utvecklingsteamet, till exempel från slutanvändare, kunder eller allmänheten.	Vilka feedback-kanaler används för att inhämta insikter från slutanvändare, kunder eller allmänhet?
Dokumentation av systemets begränsningar	Dokumentera systemets begränsningar, inklusive information om AI-modellens kunskapsluckor och i vilket sammanhang dess resultat kan användas på ett tillförlitligt sätt.	Har ni dokumenterat kända eller förväntade begränsningar för era AI-användningsfall?
Innehållsproveniens	Spåra och verifiera ursprung, historik och ändringar av AI-relaterade data, inklusive spårning av träningsdatakällor, använda algoritmer och omvandlingar.	Hur spårar ni ursprung, historik och ändringar av AI-relaterade data, inklusive spårning av datakällor och omvandlingar från skapandet till slutanvändningen?
AI-testning och red teaming	Utvärdera risker, till exempel oavsiktlig exponering av träningsdata, känslighet för baklängeskonstruktion och risker som förknippas med modellextrahering.	Vilka testprotokoll har följts, och hur hanterar de risker med specifika AI-användningsfall?
Säker mjukvaruutveckling	AI-system ska integreras med organisationens utvecklingslivscykel för säkerhetsmjukvara och följa etablerad bästa praxis för kodning och implementering.	Hur har era AI-användningsfall integrerats med befintliga protokoll för mjukvaruutveckling?
Utbildning	Utbildningen ska täcka relevanta policyer, procedurer och efterlevnadskrav och ge intressenterna kompetens att hantera AI-risker på ett effektivt sätt och agera i enlighet med organisationens standarder.	Har ni deltagit i en utbildning för AI-styrning och riskhantering?

Genom att bygga in all bästa praxis i befintliga processer säkerställer ni att AI-lösningar implementeras, hanteras och optimeras på ett ansvarsfullt sätt. Genom att integrera dessa aktiviteter, till exempel genom att tillhandahålla en robust vägledning för medarbetare, utvärdera leverantörer på ett rigoröst sätt och etablera heltäckande AI-styrning, kan organisationer skapa en stark grund för praxis som är samordnad med styrningsstandarder och regelkrav och ger dem möjlighet att inte bara möta dagens utmaningar utan även skala AI:n på ett ansvarsfullt sätt för framtida framgångar.

IV. Ansvarsfull implementering skapar ansvarsfull innovation.

För att maximera företagets AI-potential krävs ansvarsfullt byggda tekniska lösningar, tydliga riktlinjer för användning, skräddarsydd utbildning och stark styrningsimplementering. Denna strategi främjar affärsvärdet och säkerställer att AI-initiativen uppfyller förväntningarna på reglering, upprätthåller implementeringsstandarderna och bygger in en kultur av ansvarsfull AI-användning i hela organisationen.

Fortlöpande övervakning och anpassning håller AI-initiativen på rätt spår. Genom att fastställa prestandamätvärden, göra regelbundna utvärderingar och hantera risker på ett proaktivt sätt kan organisationer ligga steget före nya regleringar och upprätthålla integriteten i sina AI-projekt.

Med detta ramverk kan organisationer ta en ledande roll i en framtid med ansvarsfull AI. Strategin fokuserar på genomslagskraft, integrering och integritet och banar väg för hållbar innovation och varaktig framgång.