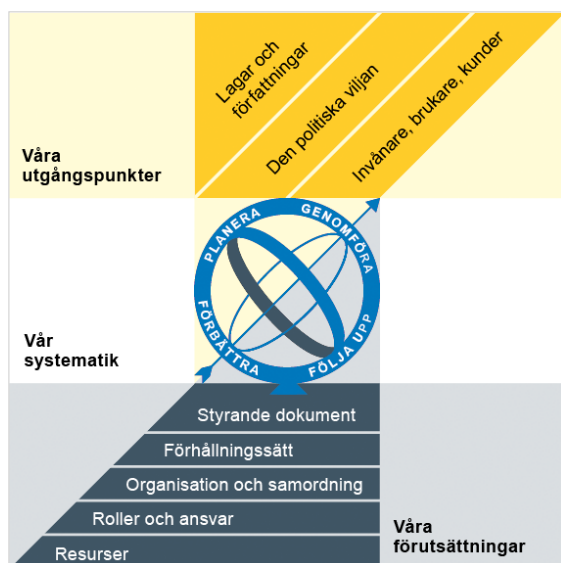


Grundskoleförvaltningens anvisning för ansvarsfull, säker och värdeskapande användning av artificiell intelligens (AI)

Reglerande styrande dokument

Policy
Riktlinje
Regel
► Anvisning
Rutin

Göteborgs Stads styrsystem



Utgångspunkterna för styrningen av Göteborgs Stad är lagar och författningar, den politiska viljan och stadens invånare, brukare och kunder. För att förverkliga utgångspunkterna behövs förutsättningar av olika slag. Stadens politiker har möjlighet att genom styrande dokument beskriva hur de vill realisera den politiska viljan. Inom Göteborgs Stad gäller de styrande dokument som antas av kommunfullmäktige och kommunstyrelsen. Därutöver fastställer nämnder och bolagsstyrelser egna styrande dokument för sin egen verksamhet. Kommunfullmäktiges budget är det övergripande och överordnade styrande dokumentet för Göteborgs Stads nämnder och bolagsstyrelser.

Om Göteborgs Stads styrande dokument

Göteborgs Stads styrande dokument är våra förutsättningar för att vi ska göra rätt saker på rätt sätt. De anger vad nämnder/styrelser och förvaltningar/bolag ska göra, vem som ska göra det och hur det ska göras. Styrande dokument är samlingsbegreppet för dessa dokument.

Stadens grundläggande principer såsom demokratisk grundsyn, principer om mänskliga rättigheter och icke-diskriminering omsätts i praktisk verksamhet genom att de integreras i stadens ordinarie beslutsprocesser. Beredning av och beslut om styrande dokument har en stor betydelse för förverkligandet av dessa principer i stadens verksamheter.

De styrande dokumenten ska göra det tydligt både för organisationen och för invånare, brukare, kunder, leverantörer, samarbetspartners och andra intressenter vad som förväntas av förvaltningar och bolag. De styrande dokumenten ligger till grund för att utkräva ansvar när vi inte arbetar i enlighet med vad som är beslutat.



Beslutad av: Avdelningschef Planering och Grundskoleförvaltningen utveckling	Gäller för:	Diarienummer: -	Datum för beslutet: 2025-12-08
Dokumentsort: Anvisning	Giltighetstid: Tillsvidare	Senast reviderad: 2025-12-08	Dokumentansvarig: Utveckling och transformation

Innehåll

Inledning	6
Syftet med denna anvisning	6
Vem omfattas av anvisningen	6
Bakgrund	7
En ny teknologisk era för grundskoleförvaltningen.....	7
Bakgrund och teknologisk kontext	7
Koppling till andra styrande dokument	7
Anvisning	8
Etiskt och juridiskt ramverk.....	8
1. Princip 1: Mänskligt agentskap och tillsyn	8
1.1 Definition och rättslig grund	8
1.2 Innebörd för grundskoleförvaltningen	8
1.3 Krav på medarbetaren	8
1.4 Krav på organisationen och ledarskapet.....	9
1.5 Scenarier och exempel	9
2. Princip 2: Teknisk robusthet och säkerhet.....	9
2.1 Definition och rättslig grund	9
2.2 Innebörd för grundskoleförvaltningen	10
2.3 Krav på medarbetaren	10
2.4 Krav på organisationen och ledarskapet.....	10
2.5 Scenarier och exempel	10
3. Princip 3: Dataskydd och datakvalitet.....	11
3.1 Definition och rättslig grund	11
3.2 Innebörd för grundskoleförvaltningen	11
3.3 Krav på medarbetaren	11
3.4 Krav på organisationen och ledarskapet.....	12

3.5 Scenarier och exempel	12
4. Princip 4: Transparens	12
4.1 Definition och rättslig grund	12
4.2 Innebörd för grundskoleförvaltningen	12
4.3 Krav på medarbetaren	13
4.4 Krav på organisationen och ledarskapet	13
4.5 Scenarier och exempel	13
5. Princip 5: Mångfald, icke-diskriminering och rättvisa	14
5.1 Definition och rättslig grund	14
5.2 Innebörd för grundskoleförvaltningen	14
5.3 Krav på medarbetaren	14
5.4 Krav på organisationen och ledarskapet	14
5.5 Scenarier och exempel	15
6. Princip 6: Samhällelig och miljömässig hållbarhet	15
6.1 Definition och rättslig grund	15
6.2 Innebörd för grundskoleförvaltningen	15
6.3 Krav på medarbetaren	15
6.4 Krav på organisationen och ledarskapet	16
6.5 Scenarier och exempel	16
7. Princip 7: Ansvarsskyldighet	16
7.1 Definition och rättslig grund	16
7.2 Innebörd för grundskoleförvaltningen	16
7.3 Krav på medarbetaren	17
7.4 Krav på organisationen och ledarskapet	17
7.5 Scenarier och exempel	17
Praktisk tillämpning och riskhantering	18
8. Den riskbaserade modellen - Grön, Gul och Röd zon	18
9 Grön zon: Tillåten användning med låg risk	18
9.1 Definition	18
9.2 Godkända syften	19
9.3 Krav på användaren	19
9.4 Detaljerade exempel på tillåten användning (Grön zon)	19
10 Gul zon: Användning som kräver försiktighet och strikt avidentifiering	20
10.1 Definition	20
10.2 Hög risk för användarfel	20
10.3 Obligatorisk process för avidentifiering	21

10.4 Detaljerade exempel (Gul zon)	22
11. Röd zon: Absolut förbud.....	22
11.1 Definition	22
11.2 Kategorier av förbjuden information.....	22
12. Styrningsmodell för AI-initiativ	24
12.1 AI-rådet	24
13. Livscykelhantering: från behov till avveckling	24
14. Implementering och uppföljning	24
14.1 Utbildning	25
14.2 Process för revidering av anvisningen	25
Länkar	27

Inledning

Syftet med denna anvisning

AI ska vara ett strategiskt verktyg för grundskoleförvaltningen i dess strävan att vara en modern och effektiv förvaltning som använder digitaliseringens möjligheter för att skapa maximalt värde för våra elever och deltagare.

Denna anvisning syftar till att:

1. Etablera ett bindande och heltäckande regelverk för all användning av AI.
2. Skydda information som omfattas av personuppgiftslagstiftning och sekretess.
3. Minimera juridiska, finansiella, operativa och förtroendemässiga risker.
4. Främja en kultur av ansvarsfull innovation och medvetenhet kring AI.

Vem omfattas av anvisningen

Denna anvisning omfattar samtliga anställda, förtroendevalda, konsulter, praktikanter och andra som utför arbete för grundskoleförvaltningen, oavsett anställningsform.

Anvisningen gäller all användning av AI-teknologi i tjänsten. Den gör en skarp åtskillnad mellan:

- **Kategori A - Externa, publika AI-tjänster:** Webbaserade verktyg som inte är specifikt upphandlade av organisationen. För dessa gäller de striktaste restriktionerna.
- **Kategori B - Godkända och upphandlade AI-system:** Särskilda AI-lösningar som har genomgått organisationens process för riskbedömning, säkerhetsgranskning och upphandling. Dessa system regleras av denna anvisning samt av specifika användarinstruktioner och avtal.

Ansvar för en klok och etisk användning av artificiell intelligens åligger alla medarbetare. Genom att följa de principer och regler som fastställs här kan AI användas för att skapa verkligt värde för de vi är till för, samtidigt som förvaltningen upprätthåller och stärker allmänhetens förtroende.

Bakgrund

En ny teknologisk era för grundskoleförvaltningen

Grundskoleförvaltningen står inför ett teknologiskt paradigmskifte drivet av artificiell intelligens (AI), i synnerhet utvecklingen inom stora språkmodeller (LLM) och generativ AI. Denna teknologi har potentialen att i grunden omforma hur offentlig förvaltning arbetar, interagerar med medborgare och arbetar med beslut. Den erbjuder en historisk möjlighet att förbättra servicen, automatisera komplexa processer och frigöra mänskliga resurser till de mest värdeskapande uppgifterna. Samtidigt medför denna kraftfulla teknologi betydande risker relaterade till dataskydd, sekretess, informationssäkerhet, diskriminering och allmänhetens förtroende. En ansvarslös eller oreglerad användning av AI riskerar att underminera grundläggande principer om: legalitet, objektivitet och proportionalitet.

Denna anvisning är framtagen för att etablera ett robust och tydligt ramverk. Den syftar till att möjliggöra innovation på ett kontrollerat och säkert sätt, och säkerställa att förvaltningens användning av AI är ansvarsfull, laglig, etisk och i linje med uppdraget att tjäna allmänheten.

Bakgrund och teknologisk kontext

AI är ett samlingsnamn för system som kan uppvisa intelligent beteende genom att analysera sin omgivning och fatta beslut för att uppnå specifika mål. Generativ AI, som denna anvisning särskilt adresserar, är en gren av AI som kan skapa nytt innehåll såsom text, bilder, ljud och data, baserat på de mönster den lärt sig från enorma datamängder. Det är avgörande att förstå att de flesta publika, webbaserade AI-tjänster verkar utanför EU/EES. All data som matas in i dessa system ska betraktas som överförd till tredje land och kan användas av leverantören för att vidare träna sina modeller.

Detta faktum omöjliggör i praktiken användningen av all form av information som är skyddsvärd enligt svensk och europeisk lagstiftning, såsom personuppgifter (GDPR) och sekretessbelagd information (offentlighets- och sekretesslagen). Denna anvisning är utformad i enlighet med befintlig och kommande lagstiftning, inklusive EU:s dataskyddsförordning (GDPR), offentlighets- och sekretesslagen (OSL), och de principer som fastställs i EU:s kommande AI-förordning (AI Act).

Koppling till andra styrande dokument

Styrande dokument	Koppling till denna anvisning
Göteborgs Stads regel för IT-användare	Fastställer allmänna ansvar och åtagande för IT-användare i Göteborgs Stad.
Göteborgs Stads policy för digitalisering och IT	Beskriver styrande principer för Göteborgs Stads arbete med verksamhetsutveckling med stöd av digitalisering.

Anvisning

Etiskt och juridiskt ramverk

1. Princip 1: Mänskligt agentskap och tillsyn

Detta kapitel definierar de sju grundläggande principer som ska genomsyra all identifiering, utveckling, upphandling och användning av artificiell intelligens inom förvaltningen. Dessa principer, härledda från EU-kommissionens högnivåexpertgrupp för artificiell intelligens, utgör grunden för förvaltningens AI-styrning. Varje princip är av lika stor vikt.

1.1 Definition och rättslig grund

Principen om mänskligt agentskap och tillsyn fastställer att AI-system ska fungera som verktyg för människan, förstärka mänskliga förmågor och agera i enlighet med mänsklig avsikt och kontroll. Människan, inte teknologin, är och förblir den slutgiltiga beslutsfattaren och den som bär ansvaret. Denna princip är djupt förankrad i förvaltningsrättsliga principer där krav på individuellt ansvar är centrala. Ett AI-system kan aldrig vara en juridisk person och kan därmed aldrig åläggas ett juridiskt ansvar.

1.2 Innebörd för grundskoleförvaltningen

Inom grundskoleförvaltningens verksamhet innebär detta ett absolut förbud mot att delegera myndighetsutövning eller andra beslut som har rättslig eller på annat sätt betydande påverkan på enskilda till ett AI-system. AI kan användas för att förbereda, sammanställa, analysera och presentera underlag, men det slutgiltiga ställningstagandet, bedömningen och beslutet ska alltid fattas av en behörig befattningshavare. Medarbetare får aldrig hänvisa till "AI:ns bedömning" som grund för ett beslut, det finns alltid ett personligt tjänstemannaansvar.

1.3 Krav på medarbetaren

- Du ska alltid agera som den slutgiltiga granskaren och redaktören av allt AI-genererat material som du använder i ditt arbete.
- Du ska utöva ett kritiskt förhållningssätt och aktivt ifrågasätta de resultat, sammanfattningar och slutsatser som ett AI-system presenterar.
- Du får aldrig okritiskt kopiera och klistra in AI-genererad text i officiella dokument, beslut, protokoll eller extern kommunikation utan att först ha verifierat dess korrekthet samt säkrat kvalitet och tonfall.
- Du är personligen ansvarig för riktigheten och lämpligheten i det material du producerar, oavsett vilka verktyg du har använt för att ta fram det.

1.4 Krav på organisationen och ledarskapet

- Organisationen ska utforma arbetsprocesser så att de innehåller obligatoriska mänskliga gransknings- och godkännandesteg i de fall AI används för att ta fram beslutsunderlag.
- Ledarskapet ska främja en kultur där kritiskt tänkande och ifrågasättande av teknologiska resultat uppmuntras.
- Organisationen ska säkerställa att medarbetare får tillräcklig utbildning i att förstå och utvärdera de AI-verktyg som ställs till deras förfogande.

1.5 Scenarier och exempel

- **Felaktig tillämpning av AI verktyg:** En specialpedagog som utreder en anmälan om kränkande behandling använder en AI-tjänst för att sammanfatta flera elevers och lärares skriftliga redogörelser. AI:ns sammanfattning lyfter fram en specifik konflikt, men missar att flera elever beskriver ett upprepat mönster av social utfrysning, vilket indikerar systematisk mobbning. Specialpedagogen läser endast sammanfattningen och bedömer att eleven inte blivit utsatt för kränkande behandling, rektor fastställer bedömningen. Ansvaret för den felaktiga bedömningen ligger här på specialpedagogen och rektor.
- **Korrekt tillämpning av AI verktyg:** En utvecklingsledare använder ett godkänt, internt AI-verktyg för att transkribera en lång ljudinspelning från ett möte. Utvecklingsledaren använder sedan transkriptionen som ett sökbart register för att snabbt hitta relevanta passager, lyssnar på originalinspelningen för att bekräfta citatens exakthet, och skriver därefter sitt eget protokoll baserat på sin professionella bedömning och anteckningar. AI:n har här fungerat som ett effektivt arbetsverktyg under full mänsklig kontroll.

2. Princip 2: Teknisk robusthet och säkerhet

2.1 Definition och rättslig grund

Principen om teknisk robusthet och säkerhet kräver att AI-system är motståndskraftiga, tillförlitliga och reproducerbara. De ska fungera som avsett och ha en säkerhetsmarginal för att hantera fel, inkonsekvenser och oväntade situationer. Denna princip är direkt kopplad till lagkrav på informationssäkerhet samt dataskyddsförordningens krav på inbyggt dataskydd och dataskydd som standard (Artikel 25).

2.2 Innebörd för grundskoleförvaltningen

Medarbetare måste vara medvetna om de inneboende svagheter i dagens AI-teknologi. Det gäller i synnerhet generativa AI-modellers tendens att "hallucinera", det vill säga att generera övertygande men helt felaktiga svar. För medarbetaren innebär detta en absolut skyldighet att faktagranska all information som genereras av AI i all fortsatt användning. Förvaltningen ska ha en rigorös process för att utvärdera, testa och validera AI-system innan de godkänns för användning.

2.3 Krav på medarbetaren

- Du ska utgå från att all faktainformation (årtal, namn, statistik, laghänvisningar, etc.) som genereras av en AI är felaktig tills du har verifierat den mot en tillförlitlig, oberoende källa.
- Du ska aldrig använda AI för uppgifter där exakthet är kritisk och där du saknar möjlighet eller kompetens att självständigt verifiera resultatet.

2.4 Krav på organisationen och ledarskapet

- Organisationen ska upprätta och underhålla en förteckning över godkända AI-tjänster som har genomgått en teknisk säkerhetsgranskning och riskbedömning.
- Organisationen ska i möjligaste mån blockera åtkomst till icke-godkända AI-tjänster via nätverk och utrustning.
- Vid upphandling av AI-system ska det ställas krav på leverantören avseende systemets robusthet, säkerhet, prestanda och processer för sårbarhetshantering.

2.5 Scenarier och exempel

- **Felaktig tillämpning av AI verktyg:** En kommunikatör får i uppdrag att snabbt skriva en intern artikel om en ny lagstiftning. Personen ber en publik AI-tjänst att "förklara de viktigaste förändringarna i den nya skollagen". AI:n genererar en text som felaktigt refererar till paragrafer som inte finns och hittar på ett sista datum för implementering. Artikeln publiceras utan faktagranskning, vilket leder till att personal får felaktig information och förvaltningen tvingas publicera en rättelse.
- **Korrekt tillämpning av AI verktyg:** En analytiker använder ett godkänt AI-verktyg för att skriva ett Python-skript (som ett enklare datorprogram) för att analysera en stor mängd oidentifierad statistik. AI:n genererar en fungerande kod. Analytikern, som har grundläggande kunskaper i Python, granskar koden rad för rad för att säkerställa att den gör korrekta beräkningar och inte innehåller några säkerhetsmässiga svagheter innan den används.

3. Princip 3: Dataskydd och datakvalitet

3.1 Definition och rättslig grund

Denna princip omfattar rätten till personlig integritet, skydd av personuppgifter och robust dataförvaltning. Den baseras direkt på principerna i EU:s dataskyddsförordning (GDPR), särskilt Artikel 5, som kräver att personuppgifter behandlas lagligt, korrekt, öppet, för specifika ändamål, med dataminimering, korrekthet, lagringsbegränsning samt integritet och konfidentialitet. Principen omfattar även skydd av all annan sekretessbelagd och verksamhetskritisk information.

3.2 Innebörd för grundskoleförvaltningen

Denna princip utgör grunden för den riskbaserade modellen som beskrivs senare i anvisningen. Den innebär ett absoluta förbud mot att använda personuppgifter eller sekretessbelagd information i externa, publika AI-tjänster. Även vid användning av internt godkända system som behandlar personuppgifter, ska en dataskyddskonsekvensbedömning (DPIA) genomföras och principerna om inbyggt dataskydd och dataskydd som standard tillämpas rigoröst.

Kvaliteten på den data som används för att träna eller interagera med AI-system är avgörande för resultatets tillförlitlighet och rättvisa.

3.3 Krav på medarbetaren

- **Du ska** strikt och utan undantag följa den riskbaserade modellen (Grön, Gul, Röd zon) som beskrivs i denna anvisning.
- **Du får aldrig**, under några som helst omständigheter, mata in personuppgifter, sekretessbelagd information eller annan verksamhetskritisk data i en extern, publik AI-tjänst. Detta inkluderar även indirekta personuppgifter som i kombination kan identifiera en person.
- **Du är** skyldig att genomföra en noggrann avidentifiering av all text innan den används i enlighet med "Gul zon". Om du är osäker på om avidentifieringen är tillräcklig, ska du avstå.
- **Du ska** hantera all data i enlighet med organisationens befintliga informationssäkerhets- och dataskyddsriktlinjer.

3.4 Krav på organisationen och ledarskapet

- Organisationen ska tillhandahålla tydlig och återkommande utbildning för all personal gällande dataskydd, sekretess och informationsklassificering.
- Organisationen ska säkerställa att det finns ett giltigt personuppgiftsbiträdesavtal med alla leverantörer av AI-tjänster som kommer att behandla personuppgifter.
- Dataskyddsombudet ska involveras i ett tidigt skede vid alla initiativ som rör upphandling eller utveckling av AI-system som hanterar personuppgifter.

3.5 Scenarier och exempel

- **Felaktig tillämpning av AI verktyg:** En chef vill sammanfatta medarbetarsamtalen för sitt team. Chefen kopierar sina digitala anteckningar, som innehåller namn, personliga omdömen och diskussioner om hälsa och framtida karriär, och klistrar in de i en extern, webbaserad AI-tjänst (kategori A) med prompten "Sammanfatta de viktigaste utvecklingsområdena för mitt team". Detta utgör ett brott mot denna anvisning och en personuppgiftsincident.
- **Korrekt tillämpning av AI verktyg:** En registrator vill förbättra sökbarheten i äldre protokoll. Efter att ha säkerställt att protokollen är offentliga handlingar som inte innehåller några personuppgifter eller sekretessbelagd information som inte är godkänd att publiceras, används en AI-tjänst för att automatiskt generera sökord (taggar) för varje protokoll. Detta effektiviserar arbetet utan att kompromissa med dataskyddet.

4. Princip 4: Transparens

4.1 Definition och rättslig grund

Transparensprincipen innebär att medborgare och användare har rätt att veta när de interagerar med ett AI-system. Detta omfattar även spårbarhet och förklarbarhet (Explainable AI "XAI"), det vill säga möjligheten att på en begriplig nivå få förklarat hur ett AI-system har kommit fram till ett visst resultat eller beslut. Principen är nära kopplad till offentlighetsprincipen och rätten till insyn i offentliga organisationers verksamhet.

4.2 Innebörd för grundskoleförvaltningen

Förvaltningen ska vara öppna och ärliga med vår användning av AI. I extern kommunikation ska vi inte dölja om AI har använts. Internt dokumenteras användningen av AI-system så att beslut och processer är spårbara och kan granskas. Även om de mest komplexa AI-modellerna kan vara svåra att förklara i detalj, ska vi alltid kunna redogöra för systemets syfte, vilken data det använder, och de övergripande principerna för dess funktion.

4.3 Krav på medarbetaren

- **Du ska** på lämpligt sätt indikera när AI har använts som ett väsentligt verktyg för att skapa ett dokument eller annat material. Detta kan göras genom en fotnot, en kommentar i dokumentets metadata, eller enligt annan anvisning från organisationen.
- **Du får inte** medvetet framställa ett i huvudsak AI-genererat verk som ditt eget originella arbete, särskilt inte i sammanhang där originalitet och egen analys är av vikt.
- Vid frågor från medborgare eller media ska du på ett korrekt och öppet sätt kunna redogöra för hur du använder AI i ditt arbete, i enlighet med denna anvisning.

4.4 Krav på organisationen och ledarskapet

- Organisationens ska upprätta och underhålla ett internt register över godkända AI-system (kategori B), som beskriver varje systems syfte, dataanvändning och ansvarig funktion.
- Organisationens ska utveckla tydliga interna riktlinjer för när och hur AI-genererat innehåll ska märkas eller deklarerats.
- Alla AI-drivna tjänster som riktar sig direkt till medborgare (t.ex. chattbotar) ska tydligt identifiera sig som AI-system.

4.5 Scenarier och exempel

- **Felaktig tillämpning av AI verktyg:** En enhet publicerar en stor rapport om framtida demografiska utmaningar på förvaltningens webbplats. Stora delar av den analyserande texten är direkt genererad av en AI, men detta nämns ingenstans. En journalist som granskar rapporten upptäcker att texten liknar standardutdata från en känd AI-modell, vilket leder till en artikel om att förvaltningen "*låter robotar skriva sina rapporter*" och en efterföljande förtroendekris.
- **Korrekt tillämpning av AI verktyg:** Samma avdelning publicerar rapporten. I förordet eller metodavsnittet skriver de: "*För att effektivt sammanställa och analysera de stora datamängder som ligger till grund för denna rapport har förvaltningen använt ett AI-baserat analysverktyg för att identifiera trender och generera initiala textutkast. Alla slutsatser, bedömningar och den slutgiltiga texten har formulerats och kvalitetssäkrats av avdelningens analytiker.*" Detta är ett transparent och förtroendeingivande förhållningssätt.

5. Princip 5: Mångfald, icke-diskriminering och rättvisa

5.1 Definition och rättslig grund

Denna princip kräver att AI-system utformas och används på ett sätt som är tillgängligt för alla, undviker orättvis partiskhet (bias) och motverkar alla former av diskriminering. Principen är grundad i de mänskliga rättigheterna och diskrimineringslagstiftningen.

5.2 Innebörd för grundskoleförvaltningen

AI-system lär sig från data, och om den data reflekterar historiska eller samhälleliga fördomar, kommer AI:n att reproducera och potentiellt förstärka dessa fördomar. Förvaltningen har ett juridiskt och etiskt ansvar att säkerställa att vår användning av AI inte leder till direkt eller indirekt diskriminering av individer eller grupper. Detta kräver en aktiv medvetenhet om riskerna med bias/partiskhet i både data och algoritmer, samt processer för att identifiera och motverka detta.

5.3 Krav på medarbetaren

- **Du ska** vara medveten om att AI-genererat innehåll kan återspegla stereotyper eller fördomar. Du har ett ansvar att granska och korrigera innehåll så att det är inkluderande och neutralt.
- **Du ska** vara särskilt uppmärksam när du använder AI för att analysera data som rör människor, och reflektera över om resultaten kan vara påverkade av underliggande partiskhet.
- **Du ska** rapportera om du upptäcker att ett godkänt AI-verktyg systematiskt genererar partiskt, stereotrypt eller på annat sätt diskriminerande innehåll.

5.4 Krav på organisationen och ledarskapet

- Vid upphandling och utveckling av AI-system ska organisationen ställa krav på leverantören att redogöra för hur de har arbetat för att minimera bias i sin modell och träningsdata.
- Organisationens ska, där det är relevant, genomföra konsekvensbedömningar avseende mänskliga rättigheter och diskriminering innan ett AI-system med potentiellt stor påverkan på individer tas i bruk.
- Organisationens ska aktivt arbeta för att den data som används i AI-sammanhang är så representativ och av så hög kvalitet som möjligt.

5.5 Scenarier och exempel

- **Felaktig tillämpning av AI verktyg:** HR-avdelningen använder ett AI-verktyg för att sälla bland jobbansökningar för en chefsposition. Eftersom verktyget har tränats på historisk data där de flesta chefer var män, har det felaktigt "lärt sig" att manliga attribut är att föredra. Systemet nedprioriterar systematiskt ansökningar från kvinnor med motsvarande meriter. Detta utgör indirekt diskriminering.
- **Korrekt tillämpning av AI verktyg:** HR-avdelningen använder ett AI-analysverktyg för att hitta mönster i incidentrapporter från alla skolor. Förvaltningen har säkerställt att verktyget inte bara räknar antal incidenter, utan också identifierar underliggande risksituationer för att kunna förbättra arbetsmiljön på djupet. Medarbetaren använder AI:ns analys som ett av flera underlag, men gör den slutgiltiga bedömningen baserat på en helhetssyn som inkluderar dialoger med relevanta funktioner.

6. Princip 6: Samhällelig och miljömässig hållbarhet

6.1 Definition och rättslig grund

Denna princip fastställer att AI ska användas för att främja positiv samhällsförändring och en hållbar framtid. AI-system bör utformas och användas på ett sätt som är socialt, ekonomiskt och miljömässigt ansvarsfullt. Detta inkluderar att beakta demokratiska processer, social sammanhållning och miljömässig påverkan, såsom den energi- och resursförbrukning som krävs för att träna och driva storskaliga AI-modeller.

6.2 Innebörd för grundskoleförvaltningen

Offentliga aktörer har ett särskilt ansvar att säkerställa att teknologianvändning bidrar till det allmänna goda. Förvaltningen ska prioritera AI-initiativ som har en tydlig potential att skapa positivt samhällsvärde och är i linje med våra hållbarhetsmål. Vi bör vara medvetna om AI-teknologins miljöavtryck och, när det är möjligt, välja de mest resurseffektiva lösningarna.

6.3 Krav på medarbetaren

- **Du ska** i ditt användande av AI alltid överväga hur det kan bidra till organisationens övergripande mål och uppdrag.
- **Du ska** använda AI-verktyg på ett ansvarsfullt och resurseffektivt sätt. Undvik onödigt komplexa eller dataintensiva AI-operationer för enkla uppgifter.
- **Du uppmuntras** att identifiera och föreslå nya, innovativa sätt att använda AI för att lösa utmaningar inom ditt verksamhetsområde.

6.4 Krav på organisationen och ledarskapet

- Organisationen ska i sin strategiska planering prioritera AI-projekt som adresserar viktiga samhällsutmaningar och har en positiv inverkan på de verksamheten är till för.
- Vid upphandling av AI-tjänster ska miljömässiga aspekter, såsom leverantörens redovisning av energiförbrukning och klimatkompensation, inkluderas som en del av utvärderingskriterierna där det är relevant och proportionerligt.
- Organisationens ska främja en ansvarsfull innovationskultur där AI ses som ett medel för att uppnå en mer hållbar och effektiv offentlig service.

6.5 Scenarier och exempel

- **Felaktig tillämpning av AI verktyg:** En avdelning investerar stora resurser i att utveckla en komplex AI-modell för en intern process med marginell nytta, utan att ha gjort en ordentlig behovsanalys eller utvärderat enklare, mindre resurskrävande lösningar. Projektet förbrukar betydande data- och energiresurser utan att leverera motsvarande verksamhetsvärde.
- **Korrekt tillämpning av AI verktyg:** Förvaltningen gör en noggrann behovsanalys och upphandlar ett enkelt, resurseffektivt AI-stött system för en intern process med hög nytta. Systemet ersätter en ineffektiv manuell process, vilket leder till betydande tidsbesparingar för administratörer och ökad rättssäkerhet för elever. Investeringen är socialt och ekonomiskt hållbar med ett tydligt verksamhetsvärde.

7. Princip 7: Ansvarsskyldighet

7.1 Definition och rättslig grund

Ansvarsskyldighet är den övergripande princip som säkerställer att de övriga sex principerna efterlevs. Den kräver att det finns robusta mekanismer för att kunna utkräva ansvar för AI-system och deras resultat. Detta inkluderar spårbarhet, granskningsbarhet samt tydligt definierade roller och ansvarsområden. Principen är fundamentet för att åtgärda om något går fel samt för att upprätthålla förtroende.

7.2 Innebörd för grundskoleförvaltningen

Denna anvisning och dess tillhörande processer utgör grunden för ansvarsskyldighet inom förvaltningen. Genom att definiera tydliga regler, roller (se RACI-matris under rubrik 12) och processer för incidenthantering (rubrik 15.3), skapas en struktur där beslut och handlingar kan spåras och ansvar kan utkrävas. Det innebär att varje medarbetare är ansvarig för sin användning av AI, och att organisationen är ansvarig för att tillhandahålla ramverket för att göra detta på ett säkert sätt.

7.3 Krav på medarbetaren

- **Du är** skyldig att känna till, förstå och efterleva denna anvisning när du använder dig av AI.
- **Du ska** kunna redogöra för och motivera din användning av AI om så efterfrågas av chef eller vid intern eller extern granskning.
- **Du ska** utan dröjsmål rapportera misstänkta eller konstaterade incidenter genom de kanaler som beskrivs under rubrik 15.

7.4 Krav på organisationen och ledarskapet

- Organisationens ska säkerställa att denna anvisning är kommunicerad till och tillgänglig för all personal.
- Chefer ska agera som föredömen och säkerställa att anvisningen efterlevs inom sina respektive ansvarsområden.
- Organisationen ska etablera och underhålla de processer för styrning, riskhantering, incidentrapportering och uppföljning som beskrivs i denna anvisning.
- Organisationen ska säkerställa att det finns effektiva och tillgängliga kanaler för att anmäla incidenter och att dessa utreds på ett objektivt och rättssäkert sätt.

7.5 Scenarier och exempel

- **Felaktig tillämpning av AI verktyg:** En medarbetare använder regelbundet en icke-godkänd AI-tjänst för att sammanfatta känsliga interna dokument. När detta upptäcks visar det sig att varken medarbetaren eller dennes chef kände till anvisningen, då den endast publicerats på en undangömd sida på intranätet utan någon aktiv kommunikationsinsats. Ansvar ligger tungt på organisationen för att inte ha säkerställt implementeringen av anvisningen.
- **Korrekt tillämpning av AI verktyg:** En medarbetare råkar av misstag klistra in en oidentifierad text som fortfarande innehåller en indirekt personuppgift i en publik AI-tjänst. Medarbetaren inser sitt misstag, och tack vare tydliga instruktioner i anvisningen vet hen hur man omedelbart rapporterar händelsen som en potentiell dataskyddsincident. Organisationen kan snabbt agera för att bedöma och minska skadan. Även om ett misstag har begåtts, fungerar ansvarskedjan som avsett.

Praktisk tillämpning och riskhantering

Detta kapitel utgör den operativa kärnan i anvisningen. Det fastställer den obligatoriska, riskbaserade modell som varje medarbetare ska tillämpa för att säkerställa att principerna efterlevs i praktiken. Modellen är utformad för att vara enkel att förstå och tillämpa i det dagliga arbetet samt för att minimera risken för oavsiktliga felsteg.

8. Den riskbaserade modellen - Grön, Gul och Röd zon

All användning av externa, publika AI-tjänster (Kategori A) ska föregås av en medveten och dokumenterbar kategorisering av den information som avses användas.

Klassificeringen sker enligt en trestegsmodell: Grön zon (låg risk), Gul zon (förhöjd risk) och Röd zon (hög/oacceptabel risk).

Modellen är obligatorisk. Det åligger varje enskild medarbetare att inför varje användningstillfälle bedöma vilken zon informationen tillhör.

Vid minsta osäkerhet ska informationen kategoriseras i den högre riskzonen. Om tveksamhet råder mellan Gul och Röd zon, ska informationen undantagslöst behandlas som Röd zon, det vill säga den får inte användas i en AI-tjänst.

Riskmodellen är enbart baserad på karaktären på den data som medarbetaren avser att mata in i AI-tjänsten.

9 Grön zon: Tillåten användning med låg risk

9.1 Definition

Grön zon omfattar information som är offentlig, allmänt känd eller harmlös.

Informationen innehåller inte personuppgifter så som namn och annat som kan identifiera en individ och saknar helt koppling till enskilda individer, känslig verksamhet, specifika ärenden eller annan skyddsvärd data. Att denna information skulle komma till allmän kännedom utgör ingen som helst risk för organisationen eller för tredje part.

9.2 Godkända syften

Användning inom Grön zon syftar till att nyttja AI som ett allmänt kunskaps- och kreativt stöd. Exempel på godkända syften är:

- Inhämta allmän kunskap och få komplexa ämnen förklarade.
- Brainstorma idéer, få inspiration och skapa utkast.
- Förbättra språk, grammatik och struktur i texter som inte innehåller känslig eller skyddsvärd information.
- Generera formler för kalkylblad, generisk kod eller andra tekniska instruktioner.
- Översätta enstaka ord eller harmlösa texter.

9.3 Krav på användaren

Även inom Grön zon kvarstår ansvaret att förhålla sig kritiskt till AI-tjänstens resultat, i enlighet med de sju principerna. All faktainformation som genereras ska verifieras och eventuellt innehåll som används ska förhålla sig till principerna.

9.4 Detaljerade exempel på tillåten användning (Grön zon)

Roll	Exempel på godkänd prompt/indata	Kommentar
Alla roller	Förklara principen om subsidiaritet inom EU.	Inhämtning av allmän kunskap.
Kommunikatör	Ge mig fem förslag på en rubrik till en nyhet om att stadens bibliotek har fått nya öppettider.	Kreativt stöd utan känslig information.
Utreddare	Sammanfatta de huvudsakliga argumenten i John Rawls "En teori om rättvisa".	Bearbetning av offentligt känd information.
Administratör	Hur skriver jag en Excel-formel som räknar antalet tomma celler i kolumn B?	Generering av generisk teknisk information.
IT-tekniker	Skriv ett Bash-skript som hittar alla filer större än 100MB i en katalog och dess underkataloger.	Generering av generisk kod.
Lärare	Skapa en lektionsplan för årskurs 8 om den industriella revolutionens påverkan på Sverige.	Pedagogiskt stöd utan elevdata.
Chef	Lista vanliga teorier om situationsanpassat ledarskap.	Kompetensutveckling och generell research.

10 Gul zon: Användning som kräver försiktighet och strikt avidentifiering

10.1 Definition

Gul zon omfattar information som tillhört Röd zon, men som genomgått en medveten process för avidentifiering eller anonymisering. Informationen kan inte klassas som Grön zon då en fullständig och oåterkallelig anonymisering är svår att tekniskt garantera vilket innebär en kvarstående risk för återidentifiering eller att känslig information kan härledas. Informationen ska därför alltså hanteras som förhöjd risk.

10.2 Hög risk för användarfel

OBS. Gul zon är där risken för användarfel är som störst.

Ansvaret för att säkerställa att avidentifieringen är fullständig och inte går att göra ogjord vilar uteslutande på medarbetaren. En ofullständig avidentifiering kan innebära en personuppgiftsincident med de konsekvenser det kan medföra.

10.3 Obligatorisk process för avidentifiering

Innan information från Röd zon får behandlas som Gul zon, ska följande process genomföras:

1. **Identifiera och radera personuppgifter:**

Avlägsna metodiskt all information som direkt kan peka ut en person. Detta inkluderar, men är inte begränsat till: namn, personnummer, adresser, telefonnummer, e-postadresser, foton, registreringsnummer, ärendenummer.

2. **Identifiera och radera indirekta identifierare:**

Avlägsna all information som i kombination med annan data kan göra en person identifierbar. Detta är det mest kritiska steget. Exempel är: specifika och ovanliga yrkestitlar, namn på små arbetsplatser eller skolor, exakta datum för händelser i små sammanhang, kombinationer av ålder, postort och yrke.

3. **Generalisera och anonymisera sammanhanget:**

Ersätt specifik information med generiska platshållare.

Exempel: "*Anna Persson, enhetschef på stadsbyggnadskontoret i Gävle*" blir "*[Medarbetare A], [Befattning] på [Avdelning] i [Stad]*".

Projektet "*Utbyggnad E4 Södra Länken*" blir "*[Infrastrukturprojekt]*".

4. **Kontrollera metadata:**

Säkerställ att filens metadata (författare, bolagsnamn etc.) är borttagen om du laddar upp ett dokument.

5. **Tillämpa "Fyra ögons principen":**

Vid osäkerhet låt en kollega, som inte har förkunskap om det specifika ärendet, läsa den avidentifierade texten. Ställ frågan: "*Kan du på något sätt gissa vilken person, plats eller händelse detta handlar om?*" Om svaret är ja, eller om kollegan tvekar, har avidentifieringen misslyckats och texten ska behandlas som Röd zon.

10.4 Detaljerade exempel (Gul zon)

- **Misslyckad avidentifiering:** En medarbetare vill sammanfatta en intern rapport om en incident på “den enda grundskolan i en stadsdel”. Medarbetaren tar bort namnen på de inblandade eleverna och lärarna, men lämnar kvar namnet på skolan och stadsdelen. Eftersom kontexten är så specifik går det att identifiera individer. Detta är en överträdelse och texten skulle ha behandlats som Röd zon.
- **Lyckad avidentifiering:** En chef vill analysera mönster från flera incidentrapporter, inklusive en från “den enda grundskolan i en stadsdel”. Chefen tar bort alla personnamn och generaliserar kontexten (t.ex. “den enda grundskolan i kommundelen” ersätts med [Skola A]). Informationen är därmed inte indirekt identifierbar. Detta är en korrekt avidentifiering och texten kan användas i en AI-tjänst för att identifiera mönster.

11. Röd zon: Absolut förbud

11.1 Definition

Röd zon omfattar all information som är förbjuden att mata in i externa, publika AI-tjänster. Förbudet är absolut och medger inga undantag. Att mata in information klassificerad som Röd zon utgör en allvarlig överträdelse av denna anvisning och kan, beroende på informationens art, även utgöra ett lagbrott.

11.2 Kategorier av förbjuden information

11.2.1 Personuppgifter

Rättslig grund: Dataskyddsförordningen (GDPR).

Förvaltningen är skyldig att säkerställa att all behandling av personuppgifter sker lagligt, säkert och för specifika ändamål. Att överföra personuppgifter till en extern AI-tjänst utan laglig grund och personuppgiftsbiträdesavtal är ett direkt brott mot GDPR.

Konsekvenser: Risk för höga sanktionsavgifter från Integritetsskyddsmyndigheten (IMY), skadeståndsanspråk från den/de registrerade samt allvarlig förtroendskada för förvaltningen.

11.2.2 Känsliga personuppgifter

Rättslig grund: Dataskyddsförordningen (GDPR) Artikel 9.

Behandling av uppgifter om till exempel hälsa, religion, etniskt ursprung osv. är som huvudregel förbjuden enligt GDPR.

Konsekvenser: Som ovan, men med förhöjd risk för både sanktionsavgiftens storlek och skadans omfattning.

11.2.3 Sekretessbelagd information

Rättslig grund: Offentlighets- och sekretesslagen (OSL).

Information som omfattas av sekretess får inte röjas för obehöriga. När data matas in i en extern AI-tjänst som inte omfattas av sekretessavtal anses den vara röjd.

Konsekvenser: Kan utgöra ett brott mot tystnadsplikten, vilket kan leda till åtal, skada för enskilda och ett allvarligt skadat förtroende för förvaltningen.

11.2.4 Verksamhetskritisk information

Rättslig grund: Interna säkerhetsföreskrifter, offentlighets- och sekretesslagen (OSL), upphandlingslagstiftning (LOU).

Konsekvenser: Röjande av upphandlingsinformation, IT-sårbarheter eller fysiska säkerhetsrutiner (t.ex. larm/passersystem) kan skada förvaltningens ekonomiska intressen eller möjliggöra dataintrång och obehörigt tillträde till lokaler.

11.2.5 Upphovsrättsskyddat material

Rättslig grund: Lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk.

Konsekvenser: Att ladda upp eller bearbeta material som organisationen inte äger eller har licens för, kan leda till skadeståndskrav från rättighetsinnehavaren.

12. Styrningsmodell för AI-initiativ

12.1 AI-rådet

Grundskoleförvaltningen har ett permanent **AI-råd** som centralt organ för strategisk och taktisk styrning av AI-frågor.

Mandat:

- Äga, förvalta och ansvara för revidering av denna anvisning.
- Granska och fatta beslut om godkännande av nya AI-system (kategori B).
- Ge råd till ledningsgruppen i strategiska AI-frågor.
- Övervaka efterlevnaden av anvisningen.
- Följa utvecklingen inom AI, juridik och etik.

Sammansättning:

- Ordförande: Utvecklingschef
- Informationssäkerhet: Säkerhetschef
- Teknisk expertis: IT-Strateg
- Dataskydd: Dataskyddssamordnare

Rådet kan adjungera experter vid behov.

Mötesfrekvens:

Minst kvartalsvis samt vid behov. Samtliga beslut ska protokollföras.

13. Livscykelhantering: från behov till avveckling

Varje initiativ som rör införandet av ett nytt AI-system (kategori B) ska följa den ordinarie processen för anskaffande för verksamhetsövergripande system och med beaktande av AI verktygs särskilda förutsättningar. Vid behov ska AI-rådet konsulteras.

14. Implementering och uppföljning

En anvisning är endast verksam om den är känd, förstådd och konsekvent tillämpad i hela organisationen. Detta avsnitt fastställer kommunikation, utbildning, mätning och revidering som ska säkerställa att anvisningen får fullt genomslag och förblir relevant över tid.

14.1 Utbildning

All personal inom grundskoleförvaltningen som använder AI-verktyg ska genomgå en anpassad och obligatorisk utbildning. Utbildningsprogrammet är indelat i tre nivåer:

Nivå 1: Allmän medvetenhet

Obligatorisk för samtliga medarbetare som använder AI-tjänster. E-utbildning som täcker anvisningens syfte, de sju etiska principerna, riskmodellen (Grön, Gul, Röd zon) samt en tydlig beskrivning av de absoluta förbuden och processen för incidentrapportering.

Nivå 2: Fördjupad kunskap

Obligatorisk för chefer och nyckelroller. Utbildningen fokuserar på ledarskapets och organisationens ansvar, inklusive riskbedömning, processen för att godkänna nya system, hantering av incidenter inom det egna teamet och metoder för att främja en ansvarsfull AI-kultur.

Nivå 3: Specialistkompetens

För utpekade specialister. Externa och interna fördjupningskurser efter behov.

14.2 Process för revidering av anvisningen

AI är ett område under extremt snabb utveckling. Denna anvisning måste därför vara ett levande dokument som anpassas i takt med att teknologi, lagstiftning och omvärldsförutsättningar förändras.

14.2.1 Årlig revidering

AI-rådet ansvarar för att genomföra en fullständig och dokumenterad revidering av denna anvisning minst en (1) gång per kalenderår. Revideringsprocessen ska omfatta:

- Analys av insamlade nyckeltal definierade i 5.2.1.
- Genomgång av samtliga incidentrapporter från det gångna året och de lärdomar som dragits.
- Omvärldsbevakning av ny lagstiftning (särskilt EU AI Act), nya rättsfall och rekommendationer från tillsynsmyndigheter.
- Insamling av feedback och förbättringsförslag från verksamheten.

14.2.2 Extraordinär revidering

Utöver den årliga cykeln kan en extraordinär revidering av anvisningen initieras av AI-rådets ordförande om till exempel en eller flera av följande omständigheter inträffar:

- Ny, tvingande lagstiftning med direkt påverkan på anvisningens innehåll träder eller kan komma att träda i kraft.
- En synnerligen allvarlig incident inträffar som påvisar fundamentala brister i anvisningen.
- Ett teknologiskt genombrott sker som radikalt förändrar risklandskapet för AI-användning.

14.2.3 Beslutsprocess

Förslag till ändringar i anvisningen bereds av AI-rådet. Samtliga väsentliga ändringar ska genomgå samma formella beslutsprocess som vid anvisningens ursprungliga fastställande och beslutas av beslutsfattare.

Länkar

- **Lagstiftning och förordningar**

- Europaparlamentets och Rådets Förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter och om upphävande av direktiv 95/46/EG (allmän dataskyddsförordning). Tillgänglig via: <https://eur-lex.europa.eu/legal-content/SV/TXT/?uri=CELEX:32016R0679>
- Offentlighets- och sekretesslag (2009:400). Tillgänglig via: https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/offentlighets-och-sekretesslag-2009400_sfs-2009-400/
- Lag (1960:729) om upphovsrätt till litterära och konstnärliga verk. Tillgänglig via: https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-1960729-om-upphovsratt-till-litterara_sfs-1960-729/
- Förvaltningslag (2017:900). Tillgänglig via: https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/forvaltningslag-2017900_sfs-2017-900/

- **EU-dokument och riktlinjer**

- High-Level Expert Group on AI, *Ethics guidelines for trustworthy AI*, Europeiska Kommissionen, 2019. Tillgänglig via: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Europeiska Kommissionen, *Ethics by design and ethics of use approaches for Artificial Intelligence*, 2021. Tillgänglig via: https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en

- **Nationella riktlinjer och strategier**

- DIGG – Myndigheten för digital förvaltning, *Vägledning för användning av generativ AI i offentlig sektor*. Tillgänglig via: <https://www.digg.se/ai-for-offentlig-forvaltning/riktlinjer-for-generativ-ai>
- AI Sweden – Sveriges nationella center för artificiell intelligens, *Riktlinjer för ansvarsfull AI*. Tillgänglig via: <https://www.ai.se/ansvarsfull-ai>