|  |
| --- |
| Rapport |

**Artificiell intelligens –­framtid och säkerhet**

|  |
| --- |
| *Författare:* Agust Eliasson  *Termin: Ht 23*  *Kursnamn: Teknisk kommunikation*  *Kurskod: 1ZT010* |



Sammandrag

Artificiell Intelligens (AI) har revolutionerat dagens programmering. Även om teorin funnits sedan 1950-talet, blev AI först nyligen relevant med den ökande digitaliseringen. Intresset för AI ökade starkt efter ChatGPT, men den nuvarande AI-teknologin, inklusive ChatGPT, har visat imponerande förmågor men också begränsningar. AI kan tolka språk men har svårt att generera egna idéer och koppla information till situationer som inte har behandlats förut.

Denna rapport utforskar AI:s påverkan på individer och föreslår åtgärder för att hantera dess effekter för änsklighetens säkerhet och välbefinnande. Diskussionen om AI:s framtid inkluderar dess potentiella fördelar inom områden som sjukvård och utbildning, men också oron för ökad arbetslöshet och minskad personlig integritet. För att hantera detta krävs regleringar, etiska riktlinjer och juridiska åtgärder för att möjliggöra en säker och balanserad utveckling av AI.

Etiska regler är avgörande för att skapa en pålitlig och rättvis AI-teknologi. Detta innebär att man minimera partiskhet och sårbarheter, skydda dataintegritet och inkludera mångfald i utvecklingen. En integrerad strategi som kombinerar teknisk innovation och etiska principer är nödvändig för att säkra en hållbar framtid för artificiell intelligens.  
  
Abstract  
Artificial Intelligence (AI) has revolutionized today's programming. Although the theory has existed since the 1950s, AI only recently became relevant with the increasing digitization. Interest in AI surged after ChatGPT, but current AI technology, including ChatGPT, has exhibited impressive capabilities but also limitations. AI can interpret language but struggles to generate its own ideas and connect information to situations not previously encountered.

This report explores AI's impact on individuals and proposes measures to address its effects on humanity's safety and well-being. Discussions about AI's future encompass its potential benefits in areas like healthcare and education, alongside concerns about increased unemployment and reduced personal privacy. Addressing this requires regulations, ethical guidelines, and legal measures to enable a safe and balanced development of AI.

Ethical regulations are crucial in establishing reliable and fair AI technology. This involves minimizing bias and vulnerabilities, safeguarding data integrity, and incorporating diversity in development. An integrated approach that combines technical innovation and ethical principles is necessary to secure a sustainable future for artificial intelligence.

Nyckelord

Artificiell Intelligens, Säkerhet, Framtid, Mjukvara, Etik

Keywords

Artificial Intelligence, Security, Future, Software, Ethics

Innehåll

1 Inledning 3

1.1 Syfte och frågeställningar 3

1.2 Metod och material 3

2 Resultat 3

2.1 Introduktion 3

2.2 Framtidens AI påverkan 4

2.3 Reglerad AI auktoritet 4

2.4 Hur säkerställs utvecklingen av AI 4

3 Diskussion 5

4 Slutsats 6

Referenser 7

# Inledning

Artificiell Intelligens (AI) är ett relativt nytt fenomen som har startat ett nytt kapitel i programmeringsbranschen. Teorin bakom Artificiell Intelligens, d.v.s mänskligt maskinlärande-har funnits sen 1950-talet men inte varit aktuell fram till dagens digitalisering där allt kan göras genom mjukvara online. Innan Generative Pre-trained Transformers (GPT)var AI sedd som framtidens teknologi av många. När ChatGPT en dag blev tillgängligt för allmänheten ökade inte bara intresset utan även rädslan för framtidens AI-teknologi bland många individer. Efter ChatGPT:s dramatiska utveckling ökade sökningar med AI som sökterm drastiskt på Google. Många fruktar nu att vår framtid kommer att influeras av teknologin och är oroliga för kommande säkerhetsproblem.

De AI-verktyg som allmänheten har tillgång till idag är inte helt felfria, men deras brister döljs av deras revolutionerande funktioner. ChatGPT hör till de extraordinära AI-program som kan läsa mänskligt språk och hitta lösningar utan kommandon. Men ChatGPT kan inte komma på egna idéer eller koppla information till logik. Detta har sin grund i programmets mjukvara, en språkmodell som har sparat en enorm mängd artiklar i en databas och svarar med respons från dessa artiklar. Därför kan programmet inte förstå information och använda den i andra sammanhang. Det är anledningen till varför ChatGPT inte är så användbar inom teoretiska ämnen som matematik och fysik, där formler och text inte är allt för att lösa en uppgift.

## Syfte och frågeställningar

Syftet med denna rapport är att undersöka hur AI kommer att påverka individen, samt hur det bör hanteras för att inte reducera mänsklighetens säkerhet och välmående. För att ha möjlighet till detta bör följande frågor besvaras:

RQ1: Hur kommer AI verktyg påverka oss i framtiden?

RQ2: Behöver AI begränsas, i så fall av vem?

RQ3: Vilka andra åtgärder behövs för att garantera en säker utveckling av AI?

## Metod och material

Denna rapport bygger på en insamling av information från vetenskapliga artiklar.

Materialet består främst av andra vetenskapliga källor inom ämnet med likartade nyckelord, som hämtats från IEEE databas -främst Onesearch.

# Resultat

## Introduktion

Framtida påverkan av artificiell intelligens (AI) har blivit ett centralt samtalsämne när samhället står inför en omvälvande teknologisk transformation. Diskussionerna kretsar kring de potentiella fördelar och risker som AI medför. AI förväntas förbättra livet genom att öka effektiviteten inom områden som sjukvård, utbildning. Samtidigt lyfts frågor om arbetslöshet och integritet fram.

## Framtidens AI påverkan

Dagens AI har skapat spekulationer om hur framtiden kommer påverkas av teknologin, och en omformning av samhället är oundviklig. Debatten handlar främst om mänsklighetens framtid med AI. Det talas om en förbättrad värld med fördelar inom flera delar av samhället, men nackdelar som arbetslöshet och minskad integritet är lika viktiga att spekulera om [1]. De senaste årtionden har AI betraktats på olika sätt. Å ena sidan kommer sjukvård, utbildning och annan samhällsservice bli mer effektiv och tillgänglig. Teknologiskutveckling kommer att accelerera med framtida AI modeller och algoritmer. Å andra sidan finns det många hot. Det vanligaste hotet från djup maskininlärning är det att intelligens och makt samspelar med bristande moral [2]. Mänskligheten kan utrota sig själv med teknologi.

Det finns alltså många spekulationer om hur artificiell intelligens kommer påverka världen i framtiden. Assisterande chatt-botar inom sjukvård och utbildning är ett vanligt exempel inför framtiden. Detta kommer hjälpa en enorm mängd människor då bristen på personal inom just dessa delar av samhället kommer att bestå[3].

## Reglerad AI auktoritet

Den nuvarande AI teknologin skiljer sig från framtidens genom att dagens maskininlärning endast fungerar inom uppbyggda områden. ChatGPT är exempel på en språkmodells AI som endast sparat information för att hjälpa användaren [4]. Det finns fler sorters AI som t.ex. kan förstå vad en bild skildrar eller skapa en egen bild från en beskrivning. Forskare säger att i framtiden kan AI koppla samman olika områden och förstå allt material och sammansättningen. När en sådan AI är skapad så kommer den att kunna agera nästintill som en människa, eftersom dess sinnen som syn och hörsel kan skapa en värld den kan förstå [5]. Det finns dock fortfarande problem som bör lösas innan utvecklingen går så långt. Det viktigaste är att lära AI att förstå etiska värderingar och regelverk som en människa, så att den kan äga olika faktorer och se den större bilden vid beslutsfattande [6]. Om en oreglerad AI bara har direktivet att minska miljöförstöring kan den tolka det som att eliminera allt som kan bidra till det, utan att tänka på de sociala, ekonomiska eller humana konsekvenser som är kopplade till transporter eller produktion. Människors lösningar kan vara mer nyanserade. Människor kan utforska alternativa sätt att minimera miljöpåverkan utan att helt eliminera nödvändiga funktioner. De kan utveckla miljövänliga teknologier, förändra produktionsmetoder eller hitta sätt att minska utsläppen och samtidigt tillgodosesamhällets behov. Därför är det viktigt att reglera och utveckla AI med etiska riktlinjer och mekanismer för att främja en balanserad och samhällsorienterad användning. Vi måste säkerställa att AI inte bara agerar på ett sätt som uppfyller enskilda mål utan tar hänsyn till olika aspekter av människors välbefinnande och hållbarhet [7].

När vi överväger framtida regleringar för AI är en viktig fråga vem som bör ha kontroll över dessa system. Det är avgörande att inte AI själv tar beslut om vad som är möjligt eller inte. För enframtida kontroll av AI-system ligger fokus på en kollektiv övervakningsmekanism [8]. Det är viktigt att förhindra enskilda eller begränsade grupper från att ha för mycket makt, som kan leda till missbruk. Det krävs att en bredare grupp, som representerar en majoritet av människor, gemensamt kontrollerar dessa system. Målet är att skydda mot potentiellt missbruk och säkerställa att åtgärder för AI överensstämmer med samhällets intressen.

## Hur säkerställs utvecklingen av AI

Väsentligt i skapandet av AI. Är skapandet av etiska regler om rättvisa, öppenhet och integritet. Genom att säkerställa dessa principer från början kan vi göra teknologin mer pålitlig och rättvis för alla. Att utveckla ett opartiskt system är viktigt för att skapa ett rättvist samhälle i framtiden. Om framtida AI-system bygger på samma metoder som dagens ChatGPT kan de bli politiskt eller till och med etniskt påverkade, eftersom maskininlärningen tar in information från en mängd olika källor för att lära sig [4]. För att skapa teknologi som verkligen gynnar samhället måste vi aktivt arbeta för att minimera risker och sårbarheter. Det handlar om att säkerställa att den information som används är pålitlig och inte är påverkad av partiska åsikter [9].

Dataintegritet och sekretess är viktiga instrument för att säkerställa att informationen som AI-systemen använder är korrekt och att individens integritet skyddas. När vi strävar efter att göra systemen opartiska är det viktigt att se till att algoritmerna inte lär sig från partiska källor. Genom att medvetet välja och kvalitetssäkra datakällor, undvika partiskt material samt regelbundet granska och uppdatera modellerna kan vi minska risken för att skapa system som förstärker diskriminering eller exkludering. Det är viktigt att balansera nyttan med AI-systemen med skyddet av människors rättigheter och säkerhet [10]. En av de mest effektiva metoderna för att skapa opartiska system är att aktivt involvera en mångfald av perspektiv och kompetenser i utvecklingsprocessen. Att kontinuerligt inkludera olika tankesätt och kunskaper är viktigt för att bygga AI som gynnar alla på ett rättvist sätt [11].

# Diskussion

Teknologiska framsteg och integrationen av artificiell intelligens (AI) är delar av en intressant men komplicerad utveckling. Diskussionen om AI:s framtida påverkan inehåller både optimistiska och oroande scenarier. Å ena sidan förväntas AI förbättra effektiviteten inom sjukvård, utbildning och andra områden, vilket skulle kunna gynna samhället. Å andra sidan lyfts arbetslöshet och integritetsproblem fram som potentiella konsekvenser. En lösning är att reglera AI-teknologin för att säkerställa dess etiska användning och förhindra risker. Det är avgörande att balansera fördelar med risker samt att bevaka etiska riktlinjer och säkerhetsprotokoll för att främja en balanserad samhällsorienterad användning av AI. Framtidens AI-utveckling kräver inte bara tekniskt avancerade system utan också en ansvarsfull utveckling. Det är nödvändigt att etablera globala normer och samarbeta över nationsgränserna för att skapa en säker och hållbar framtid för AI. En helhetsstrategi som kombinerar tekniska innovationer, etiska principer och juridiska regleringar är nödvändig för att hantera de komplexa utmaningar som AI-säkerhet innebär.

# Slutsats

Sammanfattningsvis är diskussionen kring AI:s framtida påverkan tvåsidig, med både fördelar och möjliga risker. För att styra denna utveckling på ett ansvarsfullt sätt är det avgörande att inkludera etiska riktlinjer och säkerhetsprotokoll. Det krävs en sammanhållen strategi som balanserar teknisk innovation med etiska principer för att skapa en säker och hållbar framtid för artificiell intelligens. Att säkerställa en opartisk och pålitlig utveckling av AI genom att minimera risken för partiskhet och sårbarheter är också av stort intresse för samhällets välbefinnande och fortsatta framsteg.

# Referenser

[1] S. Grinschgl and A. C. Neubauer, “Supporting cognition with modern technology: Distributed Cognition today and in an AI-Enhanced Future,” *Frontiers in Artificial Intelligence*, vol. 5, 2022. doi:10.3389/frai.2022.908261

[2] S. Makridakis, “The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms,” *Futures*, vol. 90, pp. 46–60, Jun. 2017. doi:10.1016/j.futures.2017.03.006

[3] J.-A. Moldt *et al.*, “Chatbots for future Docs: Exploring medical students’ attitudes and knowledge towards artificial intelligence and Medical Chatbots,” *Medical Education Online*, vol. 28, no. 1, 2023. doi:10.1080/10872981.2023.2182659

[4] Praveen, S.V., and S. Vijaya. “Exploring Infection Clinicians’ Perceptions of Bias in Large Language Models (LLMs) like ChatGPT: A Deep Learning Study.” *The Journal of infection* 87.6 (2023): 579–580. Web.

[5] Jarrahi, Mohammad Hossein. “Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making.” Business horizons 61.4 (2018): 577–586. Web

[6] P. Dempsey, "Is ai good or bad - and who decides?: Deployment of ai is pushing ahead faster than the ethical framework needed to control it," in Engineering & Technology, vol. 16, no. 8, pp. 1-6, Sept. 2021, doi: 10.1049/et.2021.0809.

[7] G. Gauglitz, “Artificial vs. human intelligence in analytics,” *Analytical and Bioanalytical Chemistry*, vol. 411, no. 22, pp. 5631–5632, Jun. 2019. doi:10.1007/s00216-019-01972-2

[8] M. Christen, T. Burri, S. Kandul, and P. Vörös, “Who is controlling whom? reframing ‘meaningful human control’ of AI systems in security,” *Ethics and Information Technology*, vol. 25, no. 1, Feb. 2023. doi:10.1007/s10676-023-09686-x

[9] Y. S. Chaudhry, U. Sharma and A. Rana, "Enhancing Security Measures of AI Applications," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 713-716, doi: 10.1109/ICRITO48877.2020.9197790.

[10] A. Ndrejaj and M. Ali, "Artificial Intelligence Governance: A Study on the Ethical and Security Issues that Arise," 2022 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, United Kingdom, 2022, pp. 104-111, doi: 10.1109/iCCECE55162.2022.9875082.

[11] E. Kartal, “A comprehensive study on bias in Artificial Intelligence Systems,” *International Journal of Intelligent Information Technologies*, vol. 18, no. 1, pp. 1–23, Jan. 2022. doi:10.4018/ijiit.309