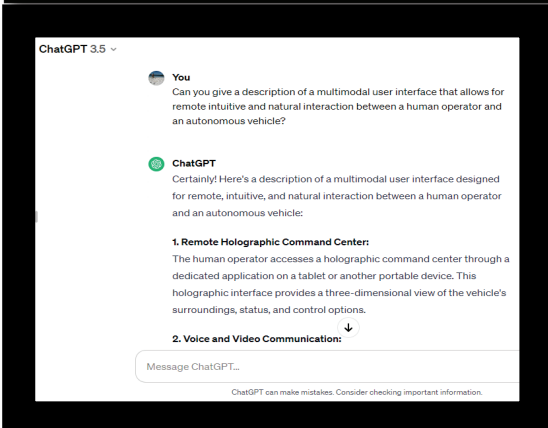


BRITTA LEVIN, ZACKARIAS ALENLJUNG,
JONAS HERMELIN, JOHAN KARLSSON,
VICTOR LINDHOLM, AGNES TEGEN,
SUSANNA NILSSON



Britta Levin, Zackarias Alenljung, Jonas Hermelin,
Johan Karlsson, Victor Lindholm, Agnes Tegen,
Susanna Nilsson

Framtida gränssnitt

Statusrapport 2021-2023

Titel	Framtida gränssnitt – Statusrapport 2021-2023
Title	Future human-machine interfaces: 2021-2023
Rapportnr/Report no	FOI-R--5568--SE
Månad/Month	12
Utgivningsår/Year	2023
Antal sidor/Pages	43
ISSN	1650-1942
Uppdragsgivare/Client	Försvarsmakten
Forskningsområde	Ledningsteknologi
FoT-område	Ledning och MSI
Projektnr/Project no	E38514
Godkänd av/Approved by	Emil Hjalmarsen
Ansvarig avdelning	Cyberförsvar och ledningsteknik

Bild/Cover: Britta Levin och Susanna Nilsson, FOI

Detta verk är skyddat enligt lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk, vilket bl.a. innebär att citering är tillåten i enlighet med vad som anges i 22 § i nämnd lag. För att använda verket på ett sätt som inte medges direkt av svensk lag krävs särskild överenskommelse.

This work is protected by the Swedish Act on Copyright in Literary and Artistic Works (1960:729). Citation is permitted in accordance with article 22 in said act. Any form of use that goes beyond what is permitted by Swedish copyright law, requires the written permission of FOI.

Sammanfattning

Gränssnitten mellan människa och teknik utvecklas ständigt liksom sätten vi interagerar med dem. Nya genombrott inom AI och maskininläring har lett till en snabb utveckling inom ett flertal forskningsområden. AI-assistenter som Open AIs chatGPT skapar stora möjligheter (och utmaningar) inom många områden, till exempel för beslutstöd, marknadsanalyser och programmering. Inom displayområdet har tekniker som XR (extended reality) och VR (virtual reality) blivit alltmer framträdande både som konsumentprodukter och som verktyg för forskning. Numera finns allt oftare gest-, tal- och blickstyrd interaktion inbyggt i XR-och VR-system.

Projektet Framtida Gränssnitt har följt utvecklingen inom gränssnittstekniker vilket utmynnat i en omvärldsbevakning av, för Försvarmakten, relevanta tekniker och gränssnitt. Det har genomförts arbete med att utforska gester som interaktionssätt och det har tagits fram exempel på hur XR-teknik kan nyttjas i en ledningskontext. Utöver detta har studier påbörjats kring hur AI i form av maskininläring kan nyttjas dels för analys och klassificering av data från sensorer (accelerometrar och gyroskop), dels som en del av ett koncept för hur en textbaserad generativ AI-assistent kan stödja vid taktisk planering.

Rapporten ger en översikt av det arbete som bedrivits inom projektet de senaste tre åren och som kommer ligga till grund för fortsatta studier och konceptutveckling under kommande år.

Nyckelord: människa-maskin gränssnitt, gränssnitt, gester, gestbaserad interaktion, obemannad farkost, UGV, motion capture handske, AI, maskininläring, ChatGPT, generativ AI, språkmodeller, XR, extended reality, MR, mixed reality, VR, virtual reality.

Summary

The interfaces between human users and technology are constantly evolving, as well as the ways we interact with them. New breakthroughs in AI and machine learning have led to rapid development in a number of research areas. A prominent example is AI assistants, such as Open AI's chatGPT, that has created new possibilities (and challenges) in many different domains, such as business intelligence, decision support and programming. In the display area, technologies such as XR (extended reality) and VR (virtual reality) have become increasingly available both as consumer products and as tools for research. Currently a wide array of functionalities, from gesture and speech input to eye-controlled interaction, are often built directly into the XR interfaces.

The project described in this report has monitored this development, which has resulted in survey of technologies and interfaces of relevance to the Swedish Armed Forces. The project has explored gestures as a way of interaction and developed an example of how XR technology can be used in a C2 context. In addition to this, studies have been initialised on how AI, in the form of machine learning, can be used for the analysis and categorization of data from sensors (accelerometers and gyroscopes), as well as part of a concept for how a text-based generative AI assistant can support tactical planning.

The report provides an overview of the work conducted within the project over the past three years, which will form the basis for further studies and concept development in the coming years.

Keywords: human-machine interface, gestures, gesture based interaction, unmanned ground vehicle, UGV, motion capture glove, AI, machine learning, ChatGPT, language models, generative AI, XR, extended reality, MR, mixed reality, VR, virtual reality.

Innehållsförteckning

1	Inledning	7
1.1	Bakgrund.....	7
1.2	Om projektet Framtida gränssnitt	8
1.3	Syfte med rapporten	9
1.4	Översikt genomförd verksamhet.....	9
2	Gestbaserad interaktion	11
2.1	Gestbaserade gränssnitt.....	11
2.2	AI för att klassificera kroppsrörelser	13
2.3	Genomförd verksamhet	15
2.4	Pågående verksamhet.....	18
3	Interaktion i en utökad verklighet	21
3.1	Extended Reality	21
3.2	Genomförd verksamhet	22
3.3	Pågående verksamhet.....	24
4	Gränssnitt för generativ AI.....	26
4.1	Vad är generativ AI?	26
4.2	Pågående verksamhet.....	27
5	Omvärldsbevakning.....	32
5.1	Automotiva tillämpningar	32
5.2	Kommersiell XR-teknik	33
5.3	AR-teknik för militära tillämpningar.....	35
6	Fortsatt verksamhet.....	37
6.1	Gester	37
6.2	XR	38
6.3	Gränssnitt för generativ AI.....	39
7	Slutsats	40
	Referenser	41

1 Inledning

Tekniken och gränssnitten mellan tekniken och användare utvecklas ständigt. Nya tekniska system kräver nya former av gränssnitt, samtidigt som nya typer av gränssnitt kan möjliggöra andra tillämpningar av tekniken än det som den var avsedd för från början. När ny teknik leder till radikala förändringar i dess användning sker så kallad *disruptiv innovation* (Christensen m. fl., 2015). Disruptiv innovation innebär en stor förändring av förutsättningarna, i ekonomiska termer en omställning av marknaden, och ur ett användningsperspektiv en förändring av hur tekniker används och i vilket syfte.

Ett av de mest välkända exemplen på en disruptiv innovation är Apples utveckling av Iphone. Innan Iphone lanserades var datorer det främsta mediet för användare att koppla upp sig mot internet. Mobiltelefoner användes främst för samtal, att skicka meddelanden och att spela enklare spel, men i och med den nya telefonens design och interaktionsmöjligheter blev Iphone (och konkurrenternas motsvarigheter) den främsta tekniken för användare att ta sig ut på internet. Den smarta mobiltelefonen gjorde snart den bärbara datorn överflödigt för många av de uppgifter den tidigare varit helt nödvändig för att utföra, till exempel surfa på nätet, boka resor, titta på film med mera. Androidtelefonen är ett exempel på hur en smart telefon, med en stor display och navigation baserat på GPS, blivit basen för ett soldatstödsystem i den militära kontexten.

En annan teknik som nu pekas ut (av bl.a. Gartner¹) som disruptiv är 5G som möjliggörare för en mängd tekniker som kräver högre bandbredd och snabbare dataöverföring (till exempel olika former av AI-stöd, virtuell och förstärkt verklighet). I och med att bärbara datorer i stor utsträckning nu kan ersättas av mobiltelefoner med hög beräkningskapacitet har också interaktionsprinciperna förändrats – en dator styrs med mus och tangentbord, men det krävs andra interaktionsätt och principer för en mobiltelefon. Det förändrade behovet av interaktion är en viktig del i utvecklingen av framtidens gränssnitt.

1.1 Bakgrund

Även om system har god användbarhet, det vill säga är enkla att använda och går snabbt att lära sig, finns det skäl att beakta utformningen av gränssnitten med målgruppen som utgångspunkt. En del system anskaffas för bred användning medan andra är förbehållna en smalare användargrupp med ett mer specialiserat syfte och kommer därför att användas av få personer med specifik utbildning. I det första fallet innebär antalet individer som behöver utbildas och tränas att utbildningsinsatsen blir omfattande. Detta utgör i sig en drivkraft för att satsa på tydlighet och enkelhet vid utformningen av gränssnitten för att förkorta utbildningstiden. Det som är lätt att använda är i regel också lätt att lära sig. I det andra fallet innebär anpassningen till operationsområde, miljö, och uppgift att särskilda krav ställs på vikt, utformning och funktion. Den funktion systemet ska fylla påverkar designen av användargränssnittet – komplexa system kan inte alltid hanteras genom enkla användargränssnitt. Inte bara komplexiteten i systemet påverkar utformningen av gränssnittet utan även de säkerhetskrav som ställts påverkar hur användaren kan och bör interagera med systemet.

För att designa ett användargränssnitt krävs både förståelse för systemet som sådant (inklusive användarens behov och förmågor) och förståelse för vad som utgör själva användargränssnittet. En utmaning kan ligga i att dra skiljelinjen mellan själva användargränssnittet och bakomliggande funktioner sett ur användarens synvinkel. Det är till exempel inte så enkelt att det är pekskärmen som utgör gränssnittet, utan gränssnittet består av kombinationen av touch-funktionen och den information som presenteras. Under projektets första år inventerades gränssnitt i egenskap av presentationsyta, sinnesmodalitet (syn, hörsel, känselsinne med flera) och biosensorer (Levin m.fl., 2021). Ett gränssnitt kan

¹ <https://www.gartner.com/en/articles/7-disruptive-technologies-you-might-not-see-coming>

dock vara mycket mer – det kan vara försett med ”intelligent” funktionalitet, det kan vara adaptivt och förändra sig efter användarens tillstånd och kapacitet eller vara anpassningsbart till användarens preferenser.

Medan ett gränssnitt kan ha många olika funktioner är det som syns och interageras med också det som upplevs. Adaptiva gränssnitt och ”intelligens” i form av talassistenter (Figur 1) har funnits en tid, men med många begränsningar, till exempel att en fråga som ställs måste vara formulerad på ett specifikt sätt för att talassistenten ska kunna ge ett relevant svar (som i många talsvarsfunktioner).



Figur 1. Illustration av holografisk talassistent. Bilden kan tolkas som ett exempel på en svårighet att illustrera något som inte syns (som till exempel en talassistent) eller så illustrerar den möjligtvis ett behov av att associera en assistent med något som kan ses, även om det bara är ett hologram.
Bildkälla: Shutterstock.

På senare tid har applikationer inom artificiell intelligens (AI) seglat upp som potentiella lösningar på en mängd svårigheter och problem. Förhoppningen är att det skapas möjligheter till mer ”naturlig interaktion”, med till exempel talassistenter, som gör att användaren kan prata som vanligt snarare än enligt en formaliserad mall.

Introduktionen av AI i gränssnitt skapar förutsättningar för funktionalitet som ligger bortanför det som tidigare varit tillgängligt för flertalet användare. AI har fått mycket uppmärksamhet under det senaste året genom Open AIs lansering av ChatGPT och andra språkmodeller och liknande tekniker, till exempel för bildgenerering. Det har också spekulerats i vad AI kan medföra på längre sikt, och nästa generations AI-teknik har beskrivits som en svart låda där ingen egentligen har kontroll över hur den utvecklas och vilka förmågor den kan nå och huruvida dessa förmågor kommer vara positiva (Hern, 2023). Trots dessa farhågor används dock redan varianter av AI i stort sett överallt och lagstiftningen kämpar för att komma ikapp². AI i olika former (språkmodeller, bildanalys, sensorfusion etc) kommer sannolikt utgöra en viktig komponent av många framtida gränssnitt.

1.2 Om projektet Framtida gränssnitt

Projektet Framtida gränssnitt är ett FoT³-projekt inom ramen för Försvarens samlingsbeställning till FOI och har pågått under åren 2021 till 2023. Syftet med projektet var att ta ett framåtblickande och utforskande grepp om gränssnittsområdet för att kartlägga och bedöma vilka typer av idag kända gränssnitt som i framtiden kan bli relevanta för militära tillämpningar och på vilket sätt dessa kan användas inom Försvarens makt. Under projektets sista år tillkom uppdraget att utforska framtida gränssnitt i relation till AI.

² EU Artificial Intelligence Act (2023). Official Journal of the European Union, L110, 1-30 (2023).

³ Forskning- och teknikutveckling

1.3 Syfte med rapporten

Syftet med rapporten är att beskriva den verksamhet som bedrivits under åren 2021 till 2023 samt att indikera vilken fortsatt verksamhet som planeras.

1.4 Översikt genomförd verksamhet

Årens verksamhet har omfattat omvärldsbevakning såväl som studier inriktade mot att undersöka potentialen med teknik som finns tillgänglig på marknaden. Avrapporteringen är utformad som en kort introduktion av respektive forskningsspår (gestbaserad interaktion, XR, samverkan människa och AI, samt omvärldsbevakning) följt av beskrivningar av genomfört och planerat framtida arbete.

Arbetet som handlar om gränssnitt baserade på gester har varit inriktat mot hur dessa kan tolkas med hjälp av artificiell intelligens (maskininlärning) och hur dessa bland annat kan användas för att styra farkoster (Karlsson m.fl., 2022). Verksamheten beskrivs utförligare i kapitel 2.

I forskningsspåret *utökad verklighet* (eng. Extended Reality, XR) har fokus varit på möjligheter med nya kommersiella produkter. Bland annat har ett examensarbete med syfte att utforska möjligheter med *förstärkt verklighet* (eng. Augmented reality, AR) för ett integrerat soldatsystem genomförts (Alenljung, 2022). Under 2023 påbörjades deltagande i NATO STO⁴-verksamheten genom projektet: HFM-IST-ET-200 med titeln ”Extended Reality (XR) technology for mediated communication”. Projektet är inriktat mot Social XR, vilket är en gren av XR med fokus på hur dessa tekniker kan användas för kommunikation mellan människor och mellan människor och avatarer. En frågeställning är hur kommunikation, interaktion och samarbete på distans kan förbättras genom upplevelsen av att vara på samma plats som en avatar eller en holografisk representation. Projektet är ett så kallat Exploratory Team (ET) med syfte att undersöka möjligheter och intresse för försvarstillämpningar av avancerad XR-teknik. Verksamheten kan bana väg för ytterligare forskningsverksamhet inom en uppföljande Research Task Group (RTG). Övrig verksamhet relaterat till XR redovisas i kapitel 3.

Under 2022 gjordes en översyn av området samverkan mellan människa och AI (Alenljung & Lindholm, 2022). Syftet var att studera förslag på utformningar av AI-baserade system med fokus på beslutsstöd och därigenom besvara den övergripande forskningsfrågan: hur kan en effektiv interaktion och samverkan erhållas mellan människor och AI-baserade system? Övrig verksamhet inom forskningsspåret gränssnitt till AI återfinns i kapitel 4.

Omvärldsbevakningen har resulterat i en bred genomgång av gränssnitt som befinner sig på teknikens framkant (Levin m.fl., 2021) med en begränsad uppdatering i kapitel 5 av denna rapport. Deltagande i olika externa sammanhang har bidragit till omvärldsbevakningen bland annat genom besök vid olika mässor: (1) IT2EC⁵ 24-26 april 2023 i Rotterdam, (2) Elbilmässan 2023-02-03 i Stockholm och (3) AUSA⁶ oktober 2022 och 2023 i Washington DC.

⁴ Science and Technology Organization

⁵ International Training Technology Exhibition and Conference 28-30 november 2023

⁶ Association of the United States Army, årlig mässa.

Kunskapsspridning har både skett genom deltagande i en nyligen instiftad Human factors-dag (september 2023) med redovisningar inom projektets huvudområden som gester, XR och gränssnitt mot AI, och två konferenser: (1) AI Partnership for Defence, 8-9 november 2023, en multilateral, digital konferens för särskilt inbjudna länder som hålls var sjätte månad där projektdeltagare presenterade ett koncept för hur generativ AI kan tillämpas för taktisk planering, och (2) 28th ICCRTS⁷, där presenterades ett koncept för militärt tillämpad generativ AI och metod för hur interaktion mellan människa och generativ AI kan studeras (Lindholm m.fl., 2023).

⁷ International Command and Control Research and Technology Symposium

2 Gestbaserad interaktion

Gester och kroppsspråk är en naturlig del av mänsklig kommunikation. Vi använder gester och kroppsspråk för att interagera med varandra, uttrycka oss och tolka intentioner mellan varandra. Dessa gester inkluderar allt från kroppsrörelser i form av känslomässigt och spontant gestikulerande, pekande gester, kulturellt definierade handtecken till fullt utvecklade teckenspråk med definierad morfologi och syntax. Gester omfattar alla *uttryckande, betydelsebärande kroppsliga rörelser där fingrar, händer, armar, huvud, ansikte eller kropp används i syfte att 1) förmedla meningsfull information eller 2) interagera med omgivningen* (Mitra & Acharya, 2007).

Tecken "överröstar" stridsbuller och "viskar" tystare än vinden.

SoldF (2001)

Teknikutvecklingen inom området gestbaserad interaktion mellan människor och teknik har varit omfattande sedan början av 2000-talet. Med smarta telefoner anammades en ny form av gränssnittsdesign. Istället för att indirekt kontrollera en muspekare via styrdon möjliggjorde touchdisplayer en mer direkt interaktion, att med själva fingret "röra" och påverka den digitala världen. Hur människor interagerar med verkliga objekt kunde nu återspeglas i interaktionen med det digitala, genom gester som härmar en verklig interaktion. Fysiska knappar och reglage återskapas nu digitalt med tryckande eller svepande gester, val av objekt sker genom pekande eller tryckande gester och digitala objekt kan flyttas genom att bokstavligen putta på dem med dragande gester. Även mer abstrakta gester används, där antalet fingrar som används i en svepande gest avgör vad som aktiveras.

Gestbaserad interaktion kan särskiljas baserat på om interaktionen sker på en yta i två rumsliga dimensioner, likt touchdisplayer, eller om gesten görs i luften och därmed nyttjar tre dimensioner (Wigdor & Wixon, 2011). Medan den förra, beröringsbaserade tekniken, är vanlig, är det mer ovanligt med den senare formen med beröringsfri teknik. Utrustning för att detektera denna typ av kroppsrörelser har däremot sen 2010-talet varit kommersiellt tillgängligt i form av till exempel produkter som Microsoft Kinect. Utvecklingen inom AI-baserad bildanalys och spridning av kunskap om detta via öppen källkod har dessutom inneburit att det har varit möjligt att implementera snarlik teknisk förmåga med billiga webbkameror. En öppen fråga är dock vilken påverkan denna typ av gester kommer att ha på framtida gränssnitt och vilka nya möjligheter det innebär för människa-maskininteraktionen.

Ett fokusområde i forskningsprojektet Framtida gränssnitt har varit att studera vilka eventuella militära tillämpningar som möjliggörs med beröringsfria gestgränssnitt. Kapitlet summerar vad denna typ av gränssnitt innebär, vilka studier som bedrivits och vilka fortsatta experiment som planeras.

2.1 Gestbaserade gränssnitt

Gestbaserade gränssnitt för människa-maskin interaktion baseras på att kroppsrörelser, hos en eller flera personer, registreras och omsätts till styrkommandon eller annan form av indata till teknisk utrustning.

Teknisk registrering av kroppsrörelser kan ske på ett flertal sätt. Dessa kan delas in i huruvida kroppsrörelsen mäts genom sensorer som påverkas direkt eller genom sensorer som på avstånd detekterar rörelsen. I det första fallet kan en person till exempel bära kläder eller handskar med inbyggda sensorer, som tröghetsmätare (eng. Inertial Measurement Unit, IMU), för att mäta olika kroppsdelars relativa acceleration och position. Sensorer som kan detektera gester på avstånd inkluderar kamerabaserade system, men det går även nyttja radar eller ultraljud. Varianter av registrering sträcker sig från att enbart fokusera på någon specifik kroppsdel, som till exempel händer eller ansikte, till att fånga hela kroppens rörelser. En mer omfattande översikt över olika tekniska lösningar för

registrering av gester återges i Levin m.fl. (2021). Beroende på vilket system och med vilket syfte interaktionen sker kan gestbaserade gränssnitt variera stort. Gestbaserade gränssnitt kan både vara uppbyggda kring explicit och implicit användning av gester. Explicit användning av gester innebär att en person gör ett aktivt val att utföra en viss gest i syfte att interagera med teknik. Implicit användning innebär att personen inte aktivt gör gester i syfte att interagera, utan att en gest görs av andra orsaker. Detta kan till exempel vara att personen sträcker sig efter ett objekt eller pekar på ett objekt som personen talar om. Ett försök att kategorisera olika typer av gestbaserade gränssnitt baserat på dess syfte visas i Tabell 1.

Tabell 1. Översikt av olika typer av system som gester kan vara ett gränssnitt mot.

Typ av system	Beskrivning
Interaktion med datorprogram och reglage	Gester används explicit av användaren och ersätter användningen av styrdon. Till exempel kan gester användas för att trycka på digitala knappar eller dragande gester med två-fingrar istället för scrollhjul osv.
Interaktion med autonoma enheter eller AI-assistenter	Gester används explicit eller implicit på liknande sätt som vid mellan-mänsklig kommunikation. Till exempel kan användare peka på objekt eller riktningar som refereras, kroppsspråk kan användas för att tolka intentioner eller för att underlätta turordning vid talgränssnitt.
Interaktion med andra människor via teknik	Gester används explicit för att kommunicera med en annan människa men via ett tekniskt system. Till exempel kan teckenspråk visualiseras hos en mottagare eller översätts till talat språk.
Interaktion med adaptiva gränssnitt	Gester används inte explicit som interaktionsmetod, utan främst för att situationsanpassa gränssnittet. Till exempel genom att en touchskärm tänds redan när man sträcker sig efter den.
Interaktion med värderande system	Gester används inte explicit som interaktionsmetod, utan registreras för att värdera hur en rörelse görs. Till exempel i rehabilitering eller träningssyfte kan systemet återkoppla så att användaren gör en viss rörelse korrekt.

Tidigare studier inom militära och närliggande domäner har berört en del av dessa typer av gestbaserade gränssnitt. Ett exempel på hur gränssnitt kan förstärka mänsklig kommunikation är identifiering av militära handtecken för att möjliggöra tyst kommunikation inom en grupp soldater utan fri siktlinje (Elliot m.fl., 2014; Hartnett m.fl., 2018). För militär användning har gestbaserad manuell styrning av obemannade farkoster testats för att ersätta dedikerade styrdon (Elliot m.fl., 2014; Hartnett m.fl., 2018), vilket också har demonstrerats inom det nuvarande projektet (Karlsson m.fl., 2022). Förutom direkt manuell styrning av obemannade enheter finns studier av hur gester kan användas för att instruera mer autonoma beteenden hos obemannade farkoster (Chiarella m.fl., 2018).

Militära tillämpningar av gestbaserad interaktion gäller dock inte bara interaktion med obemannade enheter. Baserat på indelningen ovan av olika system ges i Tabell 2 exempel på tänkbara militära tillämpningar för var och en systemtyperna.

Tabell 2. Möjliga militära tillämpningar av gestbaserade gränssnitt.

Typ av system	Exempel på tänkbara militära tillämpningar
Interaktion med datorprogram och reglage	<ul style="list-style-type: none"> • Överföra riktningar/bäringar till ett ledningsstödsystem genom att peka • Justera inställningar för utrustning, som mörkerhjälpmedel eller vapen⁸ • Interaktion med digitala objekt som presenteras i huvudburna AR-system
Interaktion med autonoma enheter eller AI-assistenter	<ul style="list-style-type: none"> • Styra obemannade enheter genom att peka ut riktningar eller mål • Autonoma enheter anpassar sitt uppträdande beroende på hur soldater i gruppen rör sig (smyger, springer osv)
Interaktion med andra människor via teknik	<ul style="list-style-type: none"> • Handtecken kan tekniskt registreras och tekniskt förmedlas inom en grupp för att möjliggöra tystkommunikation utan fri siktlinje
Interaktion med adaptiva gränssnitt	<ul style="list-style-type: none"> • Användarens rörelsemönster används för att automatiskt situationsanpassa andra gränssnitt, t.ex. displayer. Systemet skulle då kunna bete sig olika beroende på om personen ålar, springer eller reagera på hur personen håller i utrustning eller vapen
Interaktion med värderande system	<ul style="list-style-type: none"> • Stöd vid utbildning och träning, till exempel inläring av korrekt skjutställning

Med de studier som redovisas i detta kapitel har tre olika konceptuella spår utforskats: (1) *interaktion med andra människor via teknik*, (2) *interaktion med autonoma enheter* och (3) *interaktion med adaptiva gränssnitt*.

2.2 AI för att klassificera kropps rörelser

Att tekniskt kunna avgöra vad för gest en människa gör kräver dels sensorer som kan uppfatta relevanta rörelser, dels metoder för att tolka data från sensorerna till en identifierad och klassificerad gest. I huvudsak används kameror eller kroppsburna sensorer för registrering av gester. Tolkningen av data baseras i regel på maskininlärningsalgoritmer som kopplar ihop rörelsemönster med klassificerade gester.

Den vanligaste metoden för att registrera gester utgår från videoströmmar och använder bildanalysmetoder för att detektera rörelser. Metoden har en stor fördel jämfört med kroppsburna sensorer genom att det under många år funnits ett fokus på bildanalys inom AI-forskningen. Tröskeln för att utveckla enkla former av gestidentifiering är låg eftersom kostnaden för sensorn är låg (webbkamera) och att det finns öppet tillgängliga modeller för bildanalys, som dessutom ofta är bra på att klassificera innehåll i bilder och videoströmmar. Nackdelar med kamerabaserade lösningar är att bildanalys generellt sett är beräkningsintensivt och att det kan uppstå problem med skymd sikt som gör att kameran inte kan uppfatta alla rörelser som utförs.

⁸ Ett exempel är elpistolens Taser 7 där man med en snabb vridning i handleden medan man siktar kan växla läge på vapnet (<https://www.axon.com>).

Metoder för att mäta kroppsrörelser med kroppsburna sensorer varierar. Vanliga sensortyper är accelerometrar och gyroskop, samt flexsensorer som är en typ av sensor som mäter hur mycket något böjs. Kroppsburna sensorer har fördelar genom att de är oberoende av fri sikt och potentiellt sett är mindre beräkningskrävande. Nackdelarna är att sensorerna främst mäter relativa rörelser, samt att sensorerna måste placeras och fixeras på ett givet ställe på kroppen.

De kroppsrörelser som är av intresse för gestbaserade gränssnitt är inte bara stereotypa gester som specifika handgester, utan inkluderar även mer allmän aktivitetsigenkänning. Ett exempel på detta är en studie där maskininlärningsalgoritmer användes för att klassificera vilken aktivitet en person utförde (Cheng m.fl., 2017). De maskininlärningsalgoritmer som undersöktes var stödvektormaskin (SVM), dold Markovmodell (HMM) samt artificiella neurala nätverk (ANN). Datamängden som användes var inspelningar av fyra olika personer som utförde fem olika aktiviteter ("sitta", "sätta sig ned", "stå", "ställa sig upp", "gå") under två timmar var. Inspelningarna gjordes med accelerometrar placerade på vänster lår, höger arm, höger vrist och buken. Resultaten visade på 90% korrekthet om algoritmerna hade både tränats och testats på samma individ. Algoritmerna gav dock betydligt sämre resultat om modellerna tränades och testades på olika personer. I det senare fallet uppnådde författarna som bäst 61,9 % korrekthet. Författarna diskuterar att de sämre resultaten förmodligen beror på att personerna har olika rörelsemönster när de utför aktiviteterna.

I en översikt över området maskininläring för kroppsburen teknik presenterade Dai m.fl. (2022) vilka sensorer som används och hur insamling av data sker, hur data behandlas och hur maskininläring används samt olika applikationer. Maskininläringen som diskuterades kan kategoriseras som antingen traditionell maskininläring (såsom beslutsträd och SVM) eller djupinläring (olika varianter av ANN). Traditionell maskininläring fungerar typiskt bra för mindre datamängder, men kräver vanligen någon form av förbearbetning. Djupinläring har inte samma krav på förbearbetning av data, utan behöver istället justering av modellparametrar. En annan faktor som är viktig att beakta, speciellt för kroppsburen teknik, är huruvida beräkningarna ska göras lokalt eller ej. Modeller baserade på djupinläring kan vara så beräkningskrävande att de inte är går att exekvera på buren utrustning.

Det finns ett antal studier av prestationen hos olika maskininlärningsalgoritmer med avseende på klassificering av handgester. Le m.fl. (2019) utvärderade fem traditionella maskininlärningsalgoritmer (NaiveBayes, BayesNet, SVM, Adaboost, Random Forest) mot en datamängd av tolv olika handgester registrerade med accelerometrar och gyroskop. I deras jämförelse presterade en Random Forest-modell bäst. Modellen fick högst korrekt klassificering (87,61%) när data från både accelerometrar och gyroskop användes. En ytterligare jämförelse av algoritmer har rapporterats av Siddiqui m.fl. (2021). I deras studie jämfördes fyra olika maskininlärningsalgoritmer (K-nearest neighbour, beslutsträd, Random Forest och ANN). De använde en datamängd av 24 handgester registrerade med accelerometrar och gyroskop placerade på handleden. Resultatet indikerade en liknande prestation med över 91% korrekt klassificering med K-nearest neighbour, Random Forest och ANN när modellen tränats på nio personer och testats mot en tionde.

För att träna upp modeller för gestidentifiering med maskininläring krävs det generellt sett mycket data, speciellt om modellen ska kunna generaliseras för att fungera för olika personer och med många olika gester. Det blir därmed snabbt kostsamt att samlas in tillräckliga mängder av annoterad data. En vanlig metod inom maskininläring är därför att nyttja automatisk insamling av träningsdata för att bygga tillräckligt stora träningsmängder, volymer som inte skulle vara görliga med manuell annotering. En möjlig metod för detta är att använda andra redan tränade maskininlärningsmodeller för annotering. Att i videoklipp avgöra vad som händer kallas inom området för datorseende för *action recognition* och det finns flera dataset, bland annat (Kay m.fl. 2017) och (Zhang m.fl. 2019) som har använts för att träna modeller som klassificerar vad som händer i en video. Ett exempel på hur detta kan användas är identifiering av en persons pose

(positioner på olika kroppsdelar) från WiFi-signaler (Geng m.fl. 2022). I denna studie användes en befintlig modell för att prediktera personers pose från videodata för att annotera data baserat på mätningar av samma kroppsposition i signalen som extraherats ur WiFi-signaler. Det är omöjligt att i efterhand annotera enbart WiFi-signalen med korrekta poser eftersom den typen av signal inte går att förstå. Men genom att både registrera rörelse med videokamera och mäta WiFi-signaler kunde insamlad träningsdata automatiskt annoteras i samband med mätningen.

2.3 Genomförd verksamhet

Under projektets gång har ett flertal olika koncept för gestbaserade gränssnitt utforskats med avseende på tekniska utmaningar i relation till de olika tänkta tillämpningarna. Fokus för arbetet har dock varit på att använda kropps-buren teknik där sensorer såsom accelerometrar och gyroskop placerats ut på specificerade delar på kroppen och signalerna har sedan använts för att klassificera rörelserna. I de experimentella studier som hittills genomförts har kroppsburna sensorer använts, initialt i form av sensorer monterade i en handske och i senare skede även separata IMU:er som placerats på olika kroppsdelar.

I Karlsson m.fl. (2022) rapporteras om två olika studier av gestbaserade gränssnitt med olika fokus. I den första studien utvärderades hur väl det tekniskt var möjligt att klassificera olika typer av handtecken med maskininlärning. I den andra studien utvärderades de tekniska möjligheterna att manuellt styra mindre obemannade markgående fordon genom rörelser hos fingrar och handled. I Figur 2 illustreras dessa olika studiespår, tillsammans med en pågående studie om gestadaptivt gränssnitt som fokuserar på att identifiera aktivitet och poser istället för specifika handtecken.



Figur 2. Illustration över genomförda och pågående studier inom gestbaserade gränssnitt.

2.3.1 Klassificering av handgester

I den första studien som presenterades i Karlsson m.fl. (2022) utforskades möjligheterna att klassificera 29 handtecken från svenskt teckenspråk, samt elva handtecken som används i kommunikation inom avsutten trupp. Figur 3 ger exempel på militära handtecken som ingick. I studien jämfördes tre olika modeller för maskininlärning. En modell baserad på logisk regression och två modeller baserade på ANN. I studien påträffades flera utmaningar vad det gäller identifiering av militära handtecken. Medan handtecken från teckenspråk har en stor variation i själva handens form, särskiljer sig militära handtecken främst genom olika typer av rörelser och var gester utförs relativt kroppen (Figur 3). Detta ställer andra krav på typen av sensorer som behövs jämfört med att enbart klassificera handtecken som i högre grad särskiljer sig genom handens form.



Figur 3. Exempel på tre militära handtecken. Notera hur de två första har samma handform och skiljer sig åt endast genom handens rörelse. Bilderna publiceras med tillstånd av Försvarsmakten.

I studien användes en kommersiellt tillgänglig motion capture-handske (Manus II Haptic⁹) med ett gyroskop och flexsensorer monterade. Sensorerna gav data för handens och enskilda fingrars relativa orientering och fingrarnas grad av böjning med en frekvens av 60 Hz. För att förenkla insamlingen av träningsdata (dvs. data om handens form för varje gest) utvecklades ett eget gränssnitt som instruerade en person att göra en viss gest under en viss tid och under den tiden lagra annoterad data från handskarna.

Gestigenkänning definierades i studien som ett klassificeringsproblem med en på förhand definierad mängd gester. Data från handskarna vid en given tidpunkt (datapunkt) associerades därför med en av de fördefinierade gesterna. Klassificering kan antingen utgå från en enskild datapunkt eller en serie av datapunkter. Dynamiska gester som består av rörelse kräver att en serie datapunkter används vid klassificering. En problematik vid klassificering av rörliga gester är att olika gester momentant kan råka likna varandra. En hand som pendlar fram och tillbaka i höjd med axeln kommer i vissa ögonblick ge samma mätvärden som en hand som hålls stilla vid axeln (jämför Figur 3 till vänster och i mitten). Om enbart en datapunkt används för att klassificera gester kommer det därmed vara svårt att särskilja de två gesterna för vissa handpositioner. Detta kan hanteras genom att beakta en längre tidsserie av datapunkter och på så sätt identifiera handens relativa rörelser och varierande orientering.

Studien visade på ett flertal utmaningar med att ta fram en modell för tillförlitlig klassificering av gester, särskilt för gester som inkluderar rörelse och därmed inte går att klassificera momentant. Utförligare genomgång av resultatet av den genomförda datainsamlingen ges av Karlsson m.fl. (2022). En utmaning är att säkerställa tillräcklig mängd träningsdata som möjliggör klassificering av ny data. Om modellen ska klassificera gester från en ny person kan det vara svårt om datapunkterna från den nya personen skiljer sig avsevärt från träningsdata.

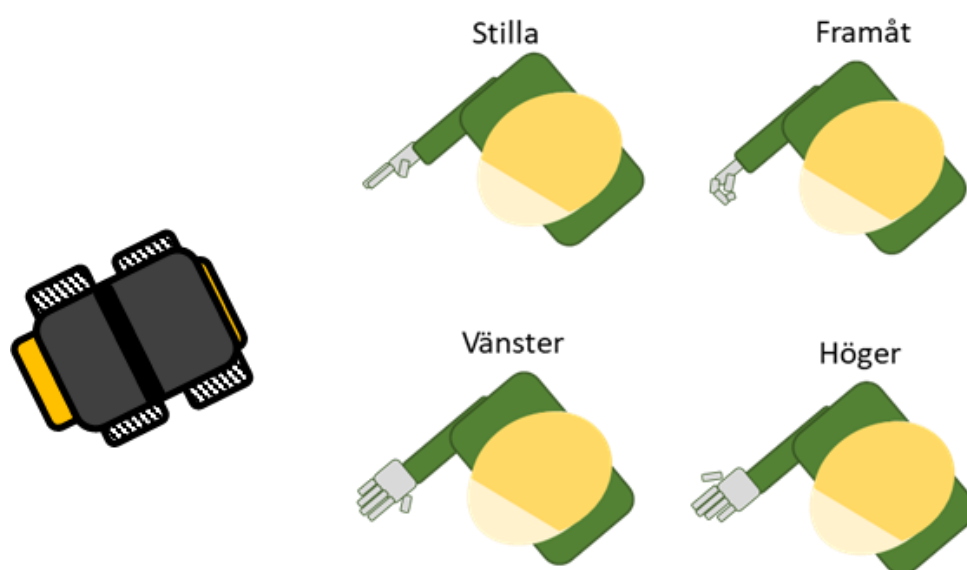
Det förekom också svårigheter med att få ut vissa mätpunkter från handskarna på ett tillförlitligt sätt. En annan identifierad utmaning är hur träningsdata ska samlas in för de situationer/perioder när inga gester genomförs. För ett flertal gester var det dessutom svårt att uppnå en tillförlitlig klassificering när modellen tränats på andra personer än den som utförde gester. En orsak till svårigheterna att identifiera gester med rörelse kan härledas till att det inte var möjligt att extrahera data för linjär acceleration från sensorerna. Gester som endast genomförs rakt uppåt, neråt eller åt sidan går inte att detektera då det endast innebär linjär acceleration.

⁹ <https://www.manus-meta.com/>

2.3.2 Manuell styrning

Den andra studien som rapporterats i Karlsson m.fl. (2022) hade fokus på gestbaserad interaktion med en obemannad markgående farkost (eng. Unmanned Ground Vehicle, UGV). För denna studie implementerades ett gränssnitt för manuell styrning av en UGV i form av hastighets- och riktningsskontroll. I detta fall användes data från sensorerna direkt utan att behandlas med maskininlärning. Handens rotation var direkt kopplad till rattutslag och fingrarnas böjning kopplades till fartkontrollen.

Givet att personer redan är utrustade med gränssnitt för gestdetektion kan separata styrdon undvikas vid basal kontroll av en UGV. Detta innebär att samtliga soldater i en grupp/pluton dynamiskt skulle kunna överta kontrollen över en UGV vid behov. Ett exempel på hur sådan styrning kan gå till illustreras i Figur 4. Rattutslaget bestämdes av handledens rotation medan hastigheten bestämdes av hur mycket fingrarna böjdes inåt. Tummens läge användes som reglage för att aktivera körning framåt respektive bakåt. När tummen inte var lyft rörde sig UGV:n framåt medan den backade när tummen var lyft.



Figur 4. Illustration över de gester som användes för att manuellt styra UGV:n (som är illustrerad till vänster i bilden).

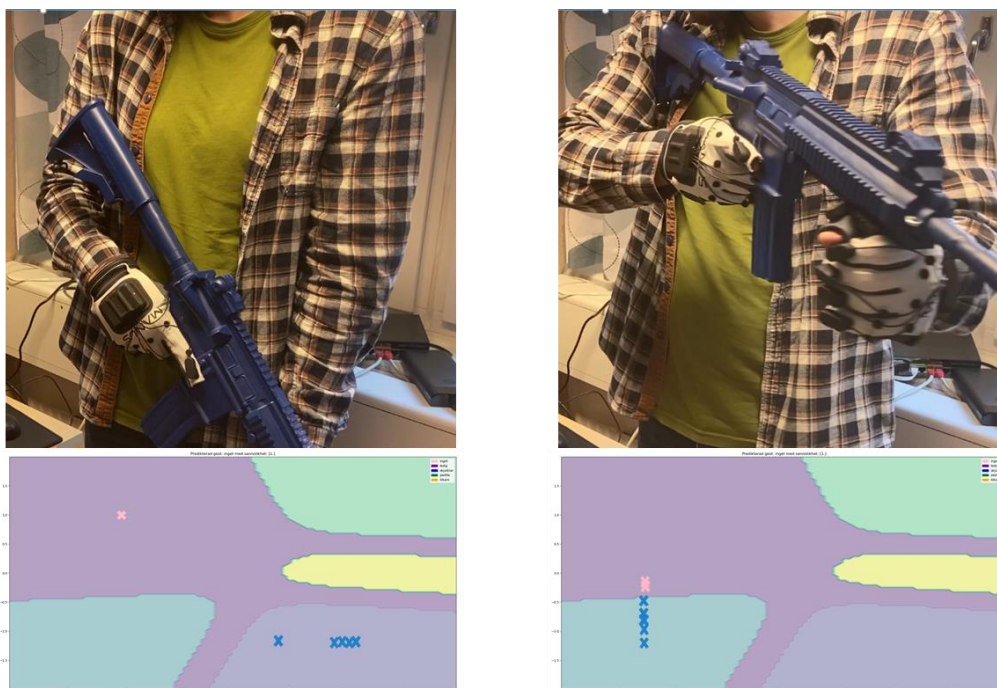
Ett mindre experiment genomfördes i vilket personer fick styra fordonet genom en hinderbana och slutsatserna som drogs var att styrningsförfarandet fungerar men att det är svårt att göra precisa förflyttningar såsom mindre korrigeringar vid till exempel smala passager (Karlsson m.fl., 2022). Dedikerade styrdon är antagligen mer effektiva för manuell styrning när det krävs hög precision. Geststyrning kan dock vara av intresse för enklare behov av styrning. De genomförda försöken hade fokus på manuell styrning, snarare än interaktion för att styra mer autonoma beteenden. Enklare exempel på styrning av autonomt beteende kan vara gester för att snabbt stanna ett fordon, att instruera det att följa efter en person eller att hålla längre avstånd. Att designa lämpliga gester för de typer av instruktioner som behövs är något som bör utforskas vidare. Det finns även flera utmaningar med denna typ av styrning, bland annat med hantering av de tillfällen då ingen gest genomförs.

2.4 Pågående verksamhet

Fokus för 2023 har varit på att nyttja fler sensorer för att kunna registrera fler typer av mätvärden och studera hur aktivitetsklassificering kan genomföras baserat på hela kroppens position eller rörelse (Figur 5). Aktivitetsidentifiering innebär här att baserat på rörelsemönster kunna identifiera vad för aktivitet en person utför, till skillnad från att klassificera specifika och explicita gester som tidigare studier fokuserat på.

Aktivitetsidentifiering baserat på rörelsemönster är ett exempel hur gester implicit kan användas i interaktionen med teknisk utrustning. Gestidentifieringen i detta fall skulle vara ett transparent gränssnitt där användaren inte medvetet gör specifika gester, utan agerar som vanligt för att utföra vissa uppgifter. Genom att tekniskt identifiera vad en person gör (vilken aktivitet/rörelse som utförs) kan interaktionen med ett system situationsanpassas dynamiskt och automatiskt.

För att exemplifiera konceptet med aktivitetsidentifiering studerades möjligheterna att särskilja på fyra olika typer av aktiviteter: (1) att stå i grundställning med automatkarbin, (2) att stå i färdigläggning med automatkarbin, (3) att stå med kikare mot ögonen och (4) att hålla i en surfplatta med två händer. I Figur 5 visas hur skillnaden mellan grundställning och färdigläggning kan särskiljas. Den konceptuella idén med att särskilja olika aktiviteter kan innebära att mängden eller typen av information som presenteras via andra gränssnitt, som huvudburna displayer eller sikten, förändras givet den uppgift som utförs, med ett så kallat adaptivt gränssnitt.



Figur 5. Exempel på hur aktivitetsklassificering kan visualiseras. Placering av ett kryss indikerar hur algoritmen identifierat ett visst rörelsemönster vid en viss tidpunkt. Kryss placerade inom de fyra mindre inringade områdena motsvarar klassificering av fyra olika typer av positioner: *grundställning*, *färdigställning*, *kikare* och *surfplatta*. I den vänstra bilden illustreras klassificering av *grundställning*, medan den högra bilden visar hur positionen klassificeras som *färdigställning* när personen lyfter vapnet. Rosa kryss utanför de markerade områdena indikerar att det inte gick klassificera rörelsemönstret som någon av de fyra positionerna.

Identifiering av de två fallen ovan, grundställning och färdigställning, skulle också kunna göras med hjälp av sensorer som mäter vapnets vinkel. I det utforskade konceptet har analysen bara utgått från mätning av kroppens rörelse, för att inte vara knutet till specifik utrustning.

2.4.1 Träningsdata med fler sensorer

I de studier som redovisats i Karlsson m.fl. (2022) om projektets första försök med gestbaserade gränssnitt användes ett par kommersiella motion capture-handskar som sensorer. En identifierad brist i studien om att klassificera olika handtecken var att handskena inte gav tillgång till data gällande acceleration hos handen. Detta innebar som tidigare nämnts att det tekniskt inte gick att mäta rörelser som var av relevans för att särskilja vissa typer av gester. För att förbättra mätvärdena från handskena har det således under året utförts arbete för att komplettera datainsamlingen med fem stycken externa IMU:er med både accelerometer och gyroskop. IMU:er monterades på respektive handled, överarmar samt en centralt placerad enhet på bröstet. Målet med detta var att samla in data om hela armens rörelse och få en referenspunkt gällande bålens orientering.

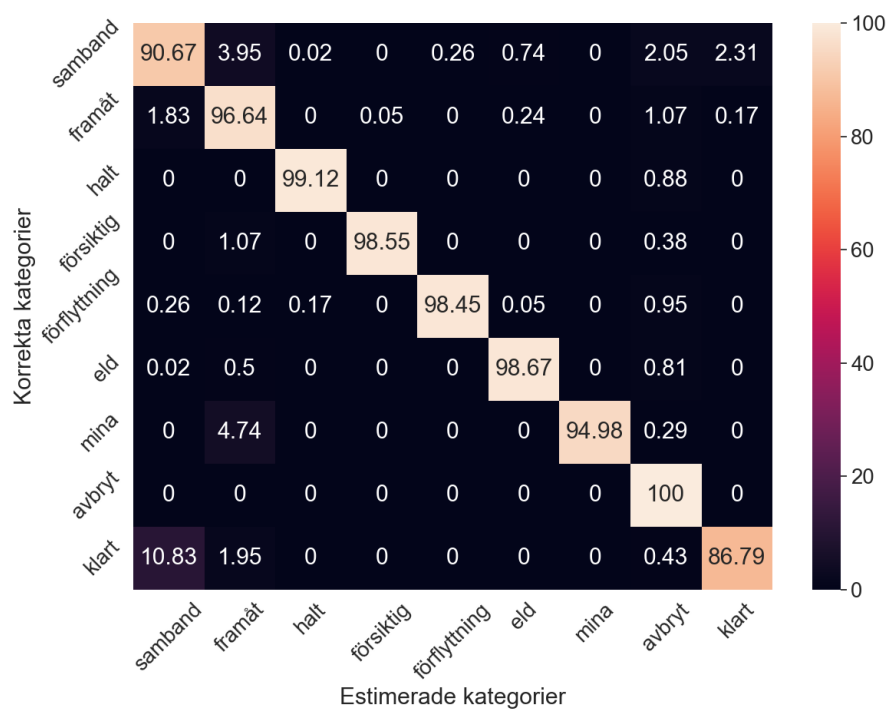
Att kombinera flera olika sensorer har inneburit nya krav på datainsamlingen och modellen som används för analys. Verktyget för att samla in datapunkterna har vidare utvecklats för att kunna sammanfoga data både från handskar och från de nya IMU:erna. En utmaning med detta är att säkerställa hur data från flera sensorer är i synk med varandra för varje mätpunkt. De genomförda testerna med sensorerna har också identifierat viktiga lärdomar gällande initialisering av sensorerna innan mätningar för att säkerställa jämförbar träningsdata mellan olika insamlingstillfällen. En felaktig initialisering kan till exempel innebära att det som en sensor indikerar är *uppåt* vid ett tillfälle är *nedåt* efter omstart av sensorn.

Baserat på antal sensorer och hur de placeras är det inte säkert att alla de mätvärden, eller särdrag, som de samlar in tillför ny eller relevant information givet ett klassificeringsproblem. Det kan till exempel bero på att vissa värden från flera olika sensorer samvarierar, vilket innebär att det samlas in redundant data. Då ett större antal särdrag leder till längre beräkningstider och tyngre beräkningar, är det önskvärt att prioritera de särdrag som ger unik information. Ett motsatt problem kan vara att valet eller placeringen av sensorer innebär att inget av de särdrag som genereras är tillräckliga för att göra en klassificering. För att identifiera vilka särdrag som är relevanta för en uppsättning gester som studeras behöver först variansen hos varje särdrag kontrolleras, då låg eller ingen varians innebär att värdet inte förändras oavsett gest. Därefter behöver särdrag som samvarierar identifieras eftersom det tyder på att de är redundanta och att endast en av dem behöver beaktas i beräkningarna. Denna typ av analys behöver göras för att öka kvaliteten såväl som informationstätheten på använd träningsdata.

2.4.2 Utvärdering av maskininlärningsalgoritmer

Under året har en teknisk utvärdering genomförts för att samla in träningsdata och undersöka olika maskininlärningsalgoritmer. Valet av algoritmer att utvärdera gjordes med utgångspunkt att de inte skulle vara mer beräkningstunga än att de går att exekvera lokalt. De algoritmer som utvärderades var Logisk regression, SVM, och K-Nearest Neighbour. Träningsdata samlades in från en person under totalt 14 sessioner. Varje session bestod av nio militära handtecken. De tecken som användes är definierade i SoldF som tecken för: *samband sökes, framåt, halt, försiktig, förflyttning i riktning, eld i riktning, mina i spåret, avbryt* samt till sist *allt klart*.

Eftersom viss mätdata baseras på magnetometer, vilket innebär att riktning spelar roll, genomfördes varje gest i fyra olika väderstreck. Vid utvärdering av algoritmer användes korsvalidering. Totalt användes 14 stycken sessioner, där 13 sessioner användes för träning och 1 session för att testa algoritmen. Detta repeterades för samtliga sessioner i ett roterande schema och medelvärdet av modellens prestation från varje separat test användes som resultat. Resultaten från de initiala experimenten visade att de flesta algoritmer som testades klarade av klassificeringssuppgiften på en granska bra eller bra nivå. Den modell som uppnådde bäst resultat var SVM-modellen, vars sammanblandningsmatris (eng. confusion matrix) ses i Figur 6. Figuren visar hur SVM-modellen presterade, dvs. hur väl den klarade av att identifiera de olika gesterna och att det fanns en viss skillnad mellan olika typer av gester.



Figur 6. Klassificering av handtecken. Resultat från initiala experiment med data från en person. Skalan visar hur stor andel i procent av respektive kategori är klassificerade till de olika kategorierna.

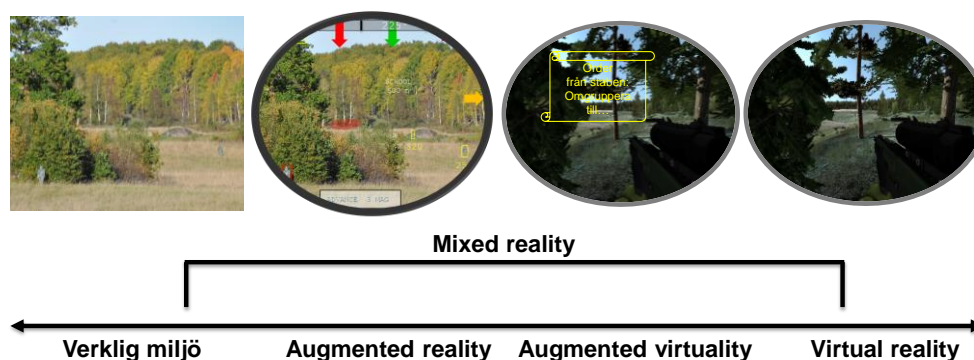
3 Interaktion i en utökad verklighet

Teknologier som förändrar, förstärker eller byter ut verkligheter har utvecklats under flera decennier. En av pionjärerna var Ivan Sutherland (1968) som utvecklade en av de första huvudmonterade displayerna (dock upphängd i taket). Idag har de utvecklats till robusta och kvalificerade huvudburna enheter som kan projicera högupplösta digitala objekt i användarens synfält. Tillämpnings- och användningsområdena sträcker sig alltifrån spel (Pallavicini m.fl., 2019), till utbildning och träning (Andrews m.fl., 2019), till skarp användning i form av hjälmmonterade displayer för stridsflygare (Webster, 2021). Kapitlet går igenom begreppet Extended reality, XR, samt redovisar genomförd och pågående verksamhet.

3.1 Extended Reality

Mixed Reality (MR), eller blandad verklighet, är ett begrepp som myntades redan på 1990-talet av Milgram & Kishino (1994) för att beskriva det spann av upplevelse som finns mellan en helt verklig värld och en helt simulerad, virtuell, värld (Figur 7). På den ena sidan av kontinuumet är användarens omvärld helt verklig (utan några lager av virtuell tillagd information), medan på den andra sidan är användarens omvärld helt virtuell.

Utökad verklighet (eng. Extended Reality, XR) är ett paraplybegrepp som dels innefattar MR-kontinuumet, men som även syftar till att fånga upp interaktionen mellan människa och teknik som möjliggörs av bland annat burna sensorer och kontrollenheter.



Figur 7. Illustration av Virtualitetskontinuumet (Milgram & Kishino, 1994). Bilder från vänster: den verkliga terrängen, den verkliga terrängen med överlagrad AR-symbolik, simulering i VBS3¹⁰ med inslag av verklig ordgivning, och till sist ren simulering i VBS3.

Inom XR-domänen och MR-kontinuumet finns även begreppet *Förstärkt verklighet* (eng. Augmented Reality, AR) vilket innebär att artificiella element tillförs till användarens uppfattning av den verkliga miljön, ofta med hjälp av datorgenererade bilder genom någon form av enhet. Azuma (1997) definierar AR utifrån tre grundläggande egenskaper: kombinationen av det verkliga och det virtuella, interaktivitet i realtid och registrering/spårning i tre dimensioner. Begreppet *Förstärkt virtualitet* (eng. Augmented virtuality, AV) beskriver applikationer som i huvudsak är virtuella men i vilka det finns fönster ut till den verkliga omvärlden.

Virtuell verklighet (eng. Virtual Reality, VR) är det kanske mest välkända begreppet inom XR-domänen, då det idag finns ett stort antal konsumentprodukter för att spela VR-spel, där användaren blir helt innesluten i en virtuell värld utan någon koppling till den fysiska

¹⁰ Virtual Battle Space 3 (VBS3) utvecklat av Bohemia Interactive Solutions

omvärlden (i en del applikationer kopplas dock virtuella objekt till fysiska för att öka immersionen, eller den upplevda verklighetskänslan).

XR-applikationer skapar nya unika utmaningar och möjligheter med avseende på såväl gränssnitts- som interaktionsdesignfrågor. Till skillnad från ”traditionell” interaktion med 2D-skärmar kan XR-applikationer innesluta användare i digital information som presenteras i en 360-graders vy. Detta i kombination med stereoseende (upplevt djup), exakt positionering och rörelsespårning förändrar och utökar användarens upplevelse av digitala applikationer (Vi m.fl., 2019).

I Levin m.fl. (2021) ges exempel och beskrivningar av olika applikationer och system som tillgängliggör XR-teknik för olika användningsområden. Ett exempel på AR-system är HoloLens 2 från Microsoft som baseras på optisk see-through, dvs. symbolik och objekt syns genom en genomskinlig display, där objekt kan placeras på olika avstånd från användaren. Objekten kan vara knutna till användaren så att de alltid följer med när denne förflyttar sig, men de kan också vara geografiskt bundna till en specifik plats eller föremål.

En viktig aspekt vid användning av överlagrad information, som i AR-teknik, är att säkerställa att informationen inte hindrar användaren att genomföra sina ordinarie uppgifter. Detta ställer krav på en design som undviker att i alltför stor utsträckning skymma användarens synfält. Gränssnittskoncepten som presenteras i detta kapitel utgick från den möjlighet som finns med HoloLens 2 att nyttja hela det spatiala rummet, dvs. att presentera information på olika ställen i omvärlden runt om användaren. Detta innebar att informationen kunde placeras utanför användarens synfält, men ändå vara lätt tillgänglig.

3.2 Genomförd verksamhet

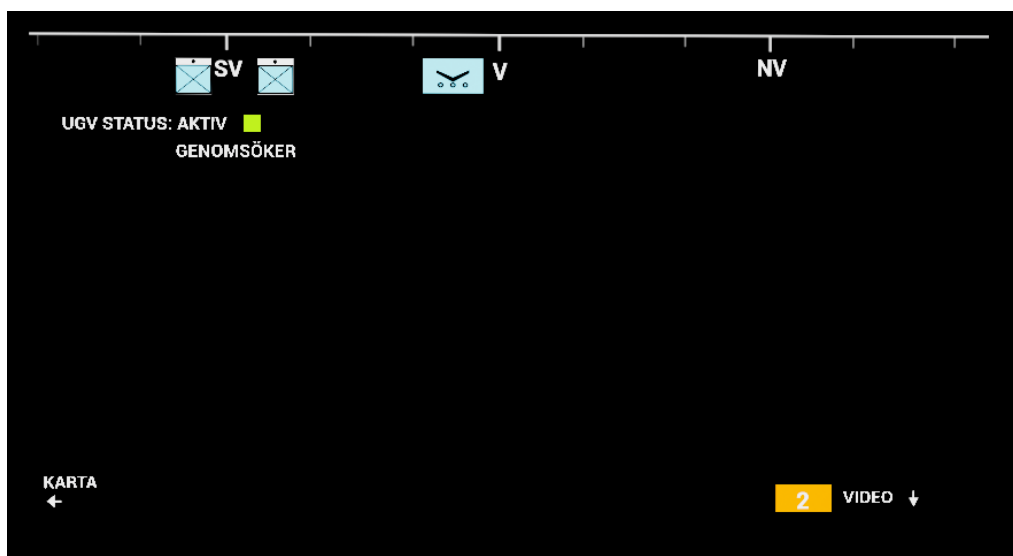
Inom ramen för projektet genomfördes ett examensarbete med syfte att utforska möjligheter med AR teknik för ett integrerat soldatsystem (Alenljung, 2022). För ändamålet utvecklades en gränssnittsprototyp i AR som utvärderades med försökspersoner från FOI. Syftet med studien var att undersöka prototypens användbarhet med fokus på tillfredsställelse och användaracceptans. Prototypen byggdes på Microsofts HoloLens 2 och utvecklades i spelmotorn Unreal Engine. Gränssnittsprototypen inkluderade både information som presenterades i det omedelbara synfältet (Figur 8) och information som återfanns vid sidan av detsamma (Figur 9).

Den information som en användare kunde tänkas behöva mer frekvent placerades i det omedelbara synfältet. Denna typ av information inkluderade egen riktning och förbands-symboler som indikerade riktning till egen trupp. Även statusinformation om gruppens obemannade farkost återfanns i synfältet. En risk med att placera information i det omedelbara synfältet är att störa och belasta användare med för mycket information. För att undvika detta nyttjade konceptet det spatiala rummet runtom användaren för att presentera olika typer av information. Den information som enbart behövdes vid vissa tillfällen kunde därmed förläggas till positioner vid sidan om användarens omedelbara synfält.

Gränssnittsprototypen bestod av två moduler¹¹, en karta och en videoströmningsmodul (Figur 9). Dessa moduler var låsta till det spatiala rummet, dvs. om en användare förflyttade sig så förblev de båda modulerna kvar på sin ursprungliga plats¹². Den information som kart- och videomodulerna innehöll kunde nås genom att vrida huvudet åt sidan respektive nedåt.

¹¹ Med modul avses virtuella objekt som innehar en specifik funktion.

¹² I ett ”verkligt” användningsfall var tanken att modulerna istället skulle vara bundna till en användare, dvs. om en användare förflyttar sig kommer modulerna att följa med och finnas kvar i samma relativa position till användaren.



Figur 8. Information som presenteras i det omedelbara synfältet. Det svarta området motsvarar den yta som i verkligheten är transparent. I övre delen av bilden visas kompassriktning och förbandssymboler för egen trupp. Under förbandssymbolerna presenteras status på gruppens obemannade farkost. I nedre vänster respektive höger hörn finns navigationspilar för att informera användaren om kart- och videomodulernas positioner. Skärmbild från spelmotorn Unreal Engine.



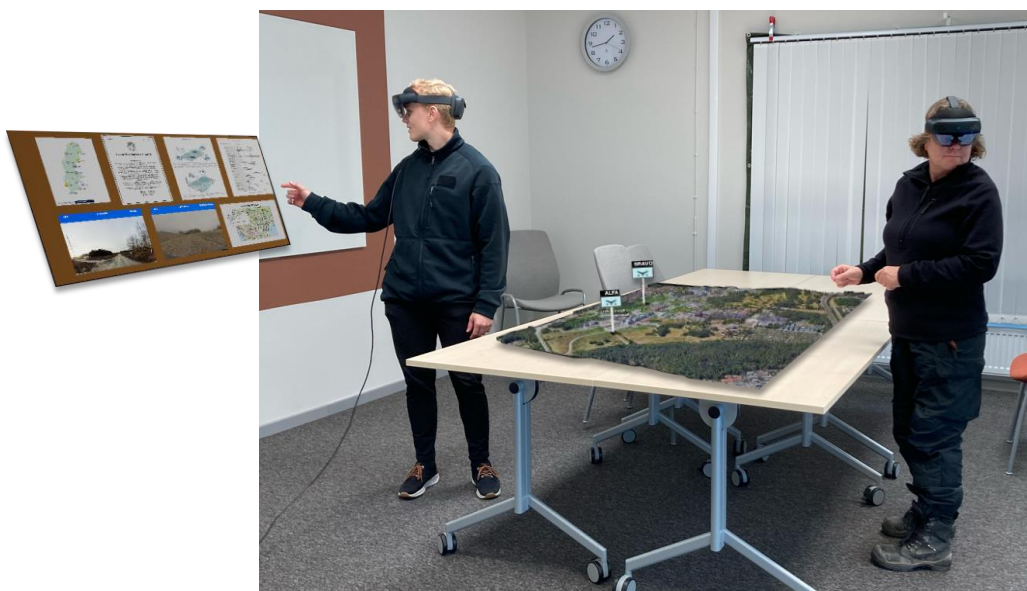
Figur 9. En överblick av prototypens moduler. Till vänster och nedanför användarens synfält presenteras kartbild respektive videoströmmar från andra enheter/källor. Skärmbild från spelmotorn Unreal Engine.

Utvärderingen av gränssnittsprototypen visade att applikationen har potential att nå hög användaracceptans och vara tillfredställande, dvs. hur väl prototypen möter användares behov och förväntningar. För att uppnå ännu högre användaracceptans och tillfredsställelse måste tekniken förbättras, eftersom användarna upplevde att kontrast och bildkvalité var för dålig för att enkelt kunna urskilja informationen i gränssnittet. Dessutom måste designen av gränssnittet utvecklas då användare upplevde det svårt att hitta information som låg utanför det direkt synfältet, speciellt videomodulen.

3.3 Pågående verksamhet

Årets verksamhet har fokuserat på att vidareutveckla det AR-koncept som togs fram inom projektet *Virtuella ledningsplatser* (Levin, m.fl. 2022). Konceptet utgörs av en enklare demonstration av hur en ledningsplats på en lägre taktisk nivå inom armén skulle kunna realiserats i AR-miljö. Arbetet har fokuserat på att förbättra detaljnivån på kartan, implementera huvud- och blickinteraktion samt utöka det virtuella rummet (Figur 10). Det nya konceptet inkluderar även virtuella UGV:er som baserats på en modell av HUSKY¹³. Dessa virtuella HUSKYs representerar en så kallad digital tvilling, dvs. en digital motsvarighet till den fysiska entiteten till vilken den är sammankopplad genom dataströmmar (Jones m.fl., 2020). De virtuella UGV:erna kopplades till videoströmmar, vyer, motsvarande en frontplacerad kamera.

Idén bakom konceptet bygger på att demonstrera en modulär, flexibel och dynamisk arbetsyta som kan generera och presentera en individuellt anpassad lägesbild (Figur 10). Tanken är att kartan ska vara försedd med olika informationslager, oleat, som kan anpassas efter individuella behov. Detta innebär att användarna förses med en gemensam lägesbild samtidigt som de kan välja att se de lager som de är intresserade av, dvs. den uppsättning symboler/detaljer som är relevant för enskilda roller (till exempel logistik, samband, sensortäckning). Förhoppningen är också att det ska gå att dela bilden med andra användare oberoende av vilken geografisk plats personerna befinner sig på.



Figur 10. Det nya konceptet bygger på att utöka det virtuella rummet, implementera huvud- och blickinteraktion och skapa en mer detaljerad kartbild som kan nyttjas av flera användare.

På anslagstavlan bakom användaren (Figur 10) återfinns flyttbara paneler med information som kan nyttjas för olika uppgifter. Anslagstavlan kan innehålla olika typer av information, och den framtida användaren kan individuellt anpassa den information som visas. I det framtagna exemplet försågs anslagstavlan med planerade operationer, väderleksrapport, uppdragsbeskrivning och videoströmmingar från UGV-enheter.

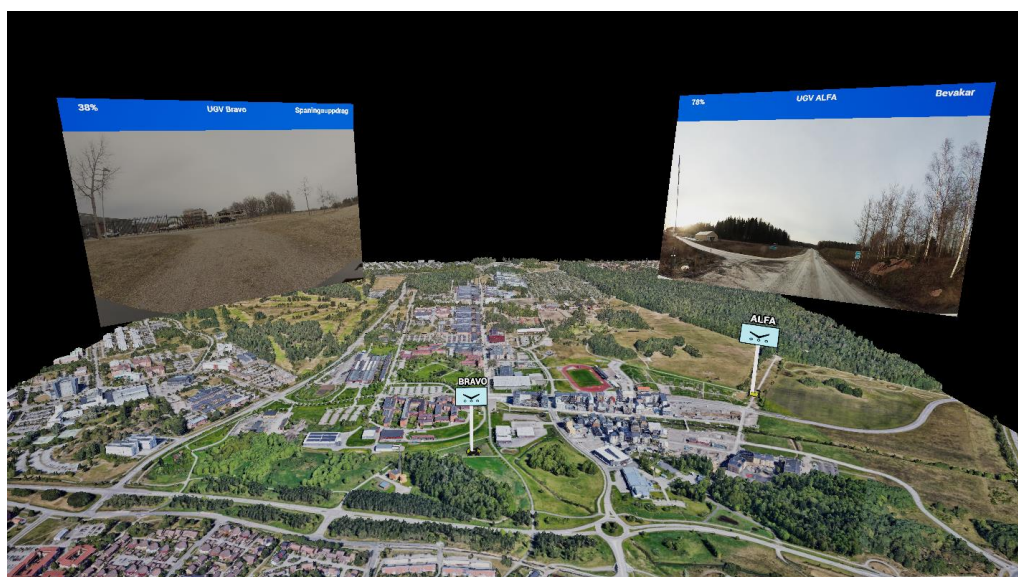
¹³HUSKY är en obemannad markgående robot från Clearpath Robotics (<https://clearpathrobotics.com/husky-unmanned-ground-vehicle-robot/>).

Konceptet inkluderar en del komponenter och funktioner som ska ge stöd till beslutsfattare för att uppnå en bättre lägesbild. De förbandssymboler som är placerade på kartan reagerar på användarens huvud- och blickriktning och presenterar mer detaljerad information om enheten, så som status, batteritid och hastighet (Figur 11).



Figur 11. Förbandssymboler på kartan. När en användare tittar på förbandssymbolen presenteras mer detaljerad information. I detta fall tittar en användare på HUSKY ALFA där information som batteritid, hastighet och status presenteras.

Ett användningsexempel med videoströmmar från respektive markgående fordon ses i Figur 12. Det går att placera vyerna på kartan för att få ett direktflöde med bildinformation från UGV-enheterna. På så sätt får användaren en sammanhållen överblick över enheternas aktuella position och vad de för tillfället ser.



Figur 12. Videoströmmar från respektive fordon har placerats ut på kartan i närheten av symbolen för respektive fordon.

4 Gränssnitt för generativ AI

Idén att låta AI generera text, ljud och andra konstnärliga verk är inte ny – AI har använts ända sedan 1970-talet för att generera konstverk (Frana & Klein, 2021). Sedan 2017 har flera viktiga tekniska genombrott skett, vilka kulminerade i OpenAI:s¹⁴ lansering av chattbotten ChatGPT under slutet av år 2022. Under 2023 har det lanserats flera språkmodeller och användandet av generativ AI spred sig över hela världen. Intresset för AI är inte nytt, men en sökning på Google Trends visar att sökningar på begreppet AI ökade markant efter lanseringen av ChatGPT (Figur 13).



Figur 13. Söktrenden för begreppet AI där intresset ses tydligt öka i samband med lanseringen av Chat GPT under 2022.

Hela spannet från privatpersoner till stora företag använder nu tekniken i syfte att effektivisera sin verksamhet, men även för nöjes skull. Konsultbolaget McKinsey genomförde i april 2023 en enkät där de fick svar från över 1600 representanter från olika typer av företag, myndigheter och industrier där över 900 svarade att deras organisation hade börjat använda AI (Chui, 2023). Samtidigt som enkätundersökningen visar på att allt fler använder AI (främst generativ AI) för allt från marknadsföring och marknadsanalyser till produktion, är det inte lika många som har initierat arbete kring eventuella risker och osäkerheter med de nya teknikerna.

I följande kapitel beskrivs generativ AI på en övergripande nivå, med nedslag i möjliggörande tekniker, följt av en beskrivning av genomfört och framtida arbete inom området.

4.1 Vad är generativ AI?

Generativ AI är en gren inom AI som använder djupinlärningsmodeller för att generera ny data. Dessa modeller har tränats på stora mängder data för att identifiera mönster och strukturer. Modeller kan både ta in och generera olika typer av data, t.ex. text, bild, ljud och programmeringskod. Modellerna kan vara multimodala, dvs. de förstår och/eller genererar flera olika typer av data. Generativ AI är en komplicerad teknik som bygger på flera tekniska framsteg som skett under de senaste åren. Tre av dessa framsteg bedöms som särskilt viktiga och beskrivs nedan: *Transformers*, *Reinforcement Learning from Human Feedback* och *Instruction following*.

År 2017 introducerade forskare vid Google den neurala nätverksarkitekturen Transformer, vilken effektiviserade behandlingen av data genom så kallad "self-attention" (Wasvani m.fl., 2017). Self-attention är en metod för att identifiera relevansen av enstaka data i förhållande till all data som matas in i den färdigtränade modellen. Förkortningen GPT, som återfinns i OpenAI:s ChatGPT, står för *Generative Pre-trained Transformer*. Denna

¹⁴ OpenAI är ett amerikanskt företag som forskar på och utvecklar artificiell intelligens (källa: <https://openai.com/about>)

arkitektur har visat sig vara särskilt användbar inom stora språkmodeller (eng. Large Language Models, LLMs), då det ger en språkmodell förmågan att förstå kontext bättre, vilket i sin tur förenklar möjligheten till interaktion mellan människa och modell avsevärt. Forskare vid Stanford Institute for Human-Centered Artificial Intelligence kallar transformers för ”foundation models” med motiveringen att de är en starkt bidragande faktor till det paradigmskifte som sker inom AI under början av 2020-talet (Bommasani m.fl., 2021).

The sheer scale and scope of foundation models from the last few years have stretched our imagination of what is possible...

Bommasani m.fl. (2021)

Reinforcement Learning (RL) är en beprövad teknik för att träna upp modeller genom en matematisk belöningsfunktion för att erhålla en bättre modell (Sutton & Barto, 2018). RL har fungerat bra vid väldefinierade problem, där denna belöningsfunktion har varit enkel att formulera. Vid komplexa, eller dåligt definierade problem, har det varit svårare att effektivt träna modeller med RL (Christiano m.fl., 2017). Dessa forskare vid OpenAI presenterade år 2017 en ny variant av RL, där mänsklig återkoppling användes som parameter, dvs. människor rankar den automatiska agentens svar och tränar den på vad som är rätt (eller bättre än andra alternativ). Denna variant kallas Reinforcement Learning from Human Feedback (RLHF). Genom RLHF kan modeller tränas effektivt på problem som var för komplexa för tidigare modeller, t.ex. att konversera genom text med en människa, likt en människa.

Instruction following är en viktig aspekt när fokus är på interaktion mellan människa och generativ AI, vilken bygger på framstegen inom RLHF. När människor interagerar med generativa AI-modeller sker det vanligtvis genom uppmaningar (eng. prompts). Dessa uppmaningar representerar indata, med vilken den generativa AI-modellen genererar utdata. Förmågan att effektivt följa instruktioner blir således central för att möjliggöra en kärnfull interaktion mellan användaren och AI-modellen. OpenAI:s forskare har även bidragit med nya insikter om hur modeller effektivt kan tränas att följa instruktioner på ett bättre sätt (Ouyang m.fl., 2022).

4.2 Pågående verksamhet

En avskanning av generativa språkmodeller samt bästa praxis (eng. best practices) vid användning av generativa språkmodeller har genomförts. Den låg till grund för ett koncept för användning av generativ AI. Inom ramen för detta koncept pågår just nu utveckling av ett prototypgränssnitt avsett för användarstudier.

4.2.1 Avskanning av generativa språkmodeller

Utvecklingen av språkmodeller och vad de använts till har gått och fortsätter att gå snabbt. I princip alla stora techbolag i världen lägger omfattande resurser på att utveckla språkmodeller och chattbotar. Det handlar dels om att ta fram kraftfulla generella modeller som till exempel GPT-4 (OpenAI, 2023) och Gemini (Gemini Team & Google, 2023), dels om att ta fram mer specialiserade modeller för till exempel hjälp med att skriva programmeringskod integrerat i utvecklingsverktyg¹⁵ eller som fristående verktyg (Rozière m.fl., 2023).

Generellt ges tillgång till LLMs och chattbotar på utvecklarnas molntjänster genom *Application Programming Interfaces* (API:er) och det är förenat med möjligheter och begränsningar. Att använda modellerna lokalt kräver ofta mycket datorresurser och ett API ger då vanliga användare möjligheten att skapa egna applikationer för att interagera med modellerna. Utvecklarna sätter samtidigt begränsningar på vilka typer av

¹⁵ <https://github.com/features/copilot> (hämtad 06/12-2023)

användningsområden som är tillåtna där till exempel deepfakes¹⁶ och pornografi oftast är förbjudna teman. Modellerna är tränade att inte följa instruktioner med förbjudna teman, men går att komma runt detta genom att interagera med dem på ett särskilt sätt (OpenAI, 2023). I de fall modellerna ligger bakom API:er har utvecklarna möjlighet modifiera interaktionen mellan användaren och modellen för att använda modellen på, enligt utvecklarna, bästa sätt. Det ger också möjligheten att övervaka vad modellerna används till (och därmed stänga av felaktig användning). Att utvecklarna inte lämnar ut de tränade modellerna beror framförallt på två anledningar. Dels vill de behålla kontrollen över vad modellerna används till och dels av affärsmässiga skäl eftersom så fort modellen är släppt kan andra organisationer kopiera förmågan hos dem. Möjligheter som att använda modellerna till att behandla känslig information samt för användningsfall som inte tillåts av utvecklarna är bara möjlig om modellerna drivs på egna resurser. För det krävs mycket resurser, i form av bland annat processorkraft och träningsdata. De nedladdningsbara modellerna är inte lika kraftfulla som modellerna bakom API:er. De klarar oftast bara engelska och presterar överlag sämre på dataset som används för att jämföra modeller. Det finns dock möjlighet att utgå från en fritt tillgänglig modell och omträna (eng. fine-tune) den för ett mer specifikt ändamål om träningsdata och beräkningsresurser finns att tillgå. Det krävs inte lika mycket resurser för att omträna modeller, men det är fortfarande en utmaning.

Parallellt med att nya modeller släppts bakom API:er har även större företag som till exempel Meta släppt Large Language Model Meta AI (LLaMA) (Touvron m.fl., 2023a) och Llama-2 (Touvron m.fl., 2023b) vilka är särskilt intressanta eftersom en del av syftet med utvecklandet och tillgängliggörandet av modellerna var att de skulle demokratisera tillgången på LLMs. Vid introducerandet av LLaMA lämnades den tränade modellen enbart ut till individer och organisationer som skrivit på ett licensavtal från fall till fall, dvs. den fanns inte publikt tillgänglig. Det tog dock inte lång tid innan den tränade modellen läckts med följd att vem som helst kunde ladda ner den. Llama-2 gjordes tillgänglig i större grad direkt vid lansering. LLaMA var bland de första lanseringarna av en fritt tillgänglig modell och efter det har det följt flertalet andra. Många applikationer har utvecklats till följd av lanseringen. Bland annat applikationer som möjliggör att använda modellen med mycket mindre resurser, även på mobiltelefoner¹⁷.

Fritt tillgängliga modeller har även släppts av AI Sweden¹⁸ med GPT-SW3. GPT-SW3 är en familj av modeller tränade på framförallt svensk data¹⁹. Dessa går att ladda ner och använda genom en användarlicens. De flesta fritt tillgängliga modeller har begränsad prestanda, särskilt vid användning på andra språk än engelska. Det beror bland annat på att de är tränade på stora delar av internet, vilka huvudsakligen består av engelska texter, vilket gör dem bättre på att ge mer nyanserade och detaljerade svar på engelska jämfört med till exempel svenska. Svenska varianter kommer sannolikt att ge bättre resultat vid användning i en svensk kontext, eftersom dess träningsdata i större utsträckning består av texter på svenska.

En chattbot följer de instruktioner som den ges, vilket inkluderar instruktioner om hur den ska uppträda. När en användare ger en uppmaning eller skriver en fråga till en chattbot kommer den att generera en respons, vilken präglas av förutsättningarna chattbotten har för att svara. Med rätt förutsättningar är det mer sannolikt att responsen är i linje med vad användaren förväntar sig. Genom att instruera chattbotten med en text i stil med ”du är en programmeringsexpert” kommer svaret vara starkt influerat. På engelska kallas detta för *priming* (sv. att prima) och används för att styra chattbottens beteende i en viss riktning.

¹⁶ Deepfakes är AI-genererade bilder, ljud- eller filmklipp, m.m., som på ett verklighetstroget sätt liknar en existerande människa.

¹⁷ <https://github.com/ggerganov/llama.cpp> (hämtad 06/12-2023)

¹⁸ AI Sweden är ett nationellt centrum för tillämpad AI, finansierat av svenska staten och deras partners, både privata och offentliga. <https://www.ai.se/sv/om-ai-sweden> (hämtad 18/12-2023).

¹⁹ <https://www.ai.se/en/project/gpt-sw3> (hämtad 06/12-2023)

Om två instanser av samma chattbot primas på två olika sätt för att därefter ställas samma fråga är det sannolikt att deras svar markant skiljer sig åt. Att formulera sig ”på rätt sätt” kallas på engelska *prompt engineering* och är en viktig komponent i alla typer av generativ AI. Rätt ställda frågor kan också hjälpa modellen att undvika hallucinering²⁰ och få fram bra svar. Eftersom många modeller är placerade bakom ett API är det svårt att avgöra exakt hur modellerna tar sådana inledande beskrivningar i beaktning. Eftersom tekniken gör snabba framsteg är priming och prompt engineering områden med ständigt föränderlig bästa praxis vid användning. Effektiv användning av en chattbot är något av en personlig färdighet som användare utvecklar genom att använda AI-modellen i särskilda kontexter. Vissa typer av formuleringar är nämligen bara effektiva i vissa typer av användningsområden.

Generativa språkmodeller har olika möjligheter för användare att påverka hur chattbotten ska svara genom användningen av s.k. priming och prompt engineering. OpenAI släppte i november 2023 möjligheten för betalande användare att bygga sina egna GPT:er genom att erbjuda en utvecklingsmiljö där utvecklare kan bygga och distribuera sina egna versioner av ChatGPT för särskilda ändamål²¹. På så sätt kan utvecklare till exempel instruera språkmodellen att svara på ett särskilt sätt eller ge den specifik data att utgå ifrån vid sina svar.

Utvecklingen inom datorseende har kantats av stor öppenhet, både i form av släppta tränade modeller och tillgång till träningsdata. Samma öppna förhållningssätt ses inte inom generativ AI. Stora språkmodeller och chattbotar kan anses vara mer användbara av illasinnade individer varför det kan vara värt att eftersträva begränsad tillgång. Samtidigt är det en utmaning att förstå hur dessa modeller kan användas och kommer att kunna användas utan bättre tillgång till dem.

4.2.2 Koncept för tillämpad generativ AI

Under 2023 har det inom projektet tagits fram ett koncept för militär användning av generativ AI där en textbaserad generativ AI-assistent kan stödja vid taktisk planering. Konceptet har presenterats vid flera tillfällen för bl.a. Försvarmakten, FMV och Förvarsdepartementet i olika forum för kunskapsspridning. En övergripande metod för hur konceptet kan användas vid användarstudier har formulerats i ett konferensbidrag som presenterats och publicerats vid 28th International Command and Control Research & Technology Symposium (Lindholm m.fl., 2023). Denna metod avses ligga till grund för framtida användarstudier med generativ AI.

Vid framtagning av detta koncept har OpenAI:s olika versioner av ChatGPT varit huvudsaklig experimentplattform. Däremot är val av modell ej centralt i konceptet, så länge det är en textbaserad generativ AI-modell. Konceptet syftar till tre saker:

- 1) Ge inspiration och idéer till hur militära tillämpningar av generativ AI kan se ut och användas.
- 2) Bygga en teoretisk och metodologisk grund från vilken användarstudier med fokus på interaktion mellan människa och AI kan bedrivas.
- 3) Skapa användarcentrerad förståelse för teknikens möjligheter genom att använda textbaserad generativ AI inom ett avgränsat område med en definierad metod.

Framgångsfulla militära operationer karaktäriseras av att kunna fatta välgrundade beslut snabbare än fienden. Ett vanligt förekommande arbetssätt vid taktisk planering inom Försvarmakten är Planering Under Tidspress (PUT) (Thunholm, 2003). Inom ramen för

²⁰ När en chattbot hallucinerar innebär det att den genererar ett svar som innehåller falsk eller missvisande information, presenterat som fakta.

²¹ <https://openai.com/blog/introducing-gpts>

konceptet undersöks hur ChatGPT kan användas för att stödja vid olika moment i PUT genom en modifierad PUT-mall (Tabell 3).

Användaren av planeringsassistenten går igenom fyra övergripande steg, inklusive flertalet delsteg. Under planeringen med AI-assistenten kan användaren genom priming och prompts påverka vilken information användaren vill ha, hur den vill få information presenterad, be om ytterligare förklaringar, etc. För att få modellen att svara på ett bra sätt har olika former av priming och prompt engineering undersökts. Tydliga instruktioner om hur modellen ska svara, t.ex. att den ska svara kort och koncist samt att den ska svara enligt en mall underlättar och hjälper modellen att hålla sig till ämnet. Genom att prima AI-assistenten med en beskrivning av PUT går det att hänvisa till de olika delmomenten för att t.ex. be om en sammanfattning av ett specifikt moment. Konversationen har varit på engelska, då modellen formulerat sig mer precist på engelska jämfört med svenska.

Tabell 3. Modifierad PUT-mall som används i konceptet.

1. Uppfatta situationen	1.1 Vårt läge och stridsvärde 1.2 Fiendens läge och stridsvärde 1.3 Terräng, väder och sikt 1.4 Styrkejämförelse
2. Generera fiendens handlingsmöjligheter	2.1 Lista möjliga handlingsmöjligheter 2.2 Besluta om mest sannolika handling
3. Generera egna handlingsmöjligheter	3.1 Lista möjliga handlingsmöjligheter 3.2 Definiera framgångsfaktorer
4. Utveckla en plan	4.1 Förberedelser 4.2 Inledningsvis, därefter, slutligen 4.3 Styrkor och svagheter med planen, ”spela på planen”

Eftersom utvecklingen av generativ AI går snabbt, och bättre modeller släpps för var månad, har det centrala intresset i konceptet varit *hur* en användare interagerar genom priming och prompt engineering för att få önskvärda svar från chattbotten. För att förenkla arbetet och måla upp en tydlig situation för AI-modellen undersöks PUT-metoden applicerat på ett *capture the flag*-spel med militärtaktiska likheter. Figur 14 visar en interaktion mellan användaren och modellen där tidigare delar av konversationen introducerade problemet, styrkeförhållanden och generella förutsättningar. När användaren ber om förslag på områden med kanaliserande terräng, givet det aktuella området, får användaren förslaget att övergången till vattendraget som går över spelplanen sannolikt är kanaliserande. Det här görs i punkt 3.2 i den modifierade PUT-metoden. Modellens svar är ett stöd vid fortsatt planering. Utöver att få stöd enligt planeringsmallen som chattbotten följer, har användaren möjlighet att be modellen om förtydliganden, ändra på förutsättningar, be om sammanfattningar, med mera.

<p>You</p> <p>Can you give me an example of a choke point, given our playing field?</p> <p>ChatGPT</p> <p>Certainly, Captain. Considering the playing field is a 300-meter-wide by 500-meter-long forested area in Sweden with a stream running diagonally, one potential choke point could be:</p> <p>Stream Crossing: Given that the stream runs diagonally across the playing field, it naturally narrows the effective width of the terrain at specific crossing points. You can expect the opponent to use these crossings as part of their movement.</p> <p>Setting up the Third Group at this stream crossing, particularly if it is narrow and surrounded by forest, would create a highly defensible position. The water obstacle slows down enemy movement, making them easier targets for your more heavily armed players, while the surrounding forest provides concealment for your troops and complicates enemy targeting.</p> <p>By controlling this choke point, we force the opponent to either engage us under unfavourable conditions or seek alternative, likely slower, routes. Over.</p>

Figur 14. Utdrag från konversation med egenjusterad ChatGPT-modell som visar hur AI:n kan stötta användaren vid terränganalys.

Under hösten 2023 utvecklade en student från Linköpings universitet en prototyp för ett gränssnitt inom ramen för projektet. Prototypen användes för att arbeta med de första stegen av PUT modellen. Syftet var att integrera OpenAIs språkmodell med ett API och bygga upp och utvärdera interaktionen med en gränssnittsprototyp. Utvärderingen av gränssnittet genomfördes med försöksdeltagare från FOI. Projektet resulterade i insikter som vikten av att AI assistenten kan ge ett stöd till användare i *hur* den ska interagera med AI assistenten, då flera försöksdeltagare uppfattade det som svårt att veta vad de ska skriva i specifika steg. Tillitsproblemet till assistenten var en framträdande diskussionspunkt i huruvida användare faktiskt skulle nyttja denna i skarpa lägen. Det framkom även tekniska aspekter som var viktiga att beakta för framtida utvecklare, vilket var delvis hanterandet av *tokens*, dvs. antalet ord som går att mata in och få ut av OpenAIs språkmodell, då det inte räckte till. Ett annat problem, delvis kopplat till antalet tokens, var att få språkmodellen att hantera kontexten utan att överbelastas med för mycket information som inte är relevant för en användare.

5 Omvärldsbevakning

Omvärldsbevakningen har riktats mot en uppdatering av läget inom några av de områden och applikationer som togs upp i översikten 2021 (Levin, 2021). Marknaden för framtida gränssnitt inte är helt okomplicerad – hajpad teknik och stora förväntningar på nya produkter publiceras samtidigt som planer på nedläggningar har annonserats. En beskrivning av området ska därför ses som en ögonblicksbild av läget just nu.

Kapitlet innehåller ett nedslag i vad som är på gång sett till automotiva tillämpningar, det redovisar en översiktlig avskanning av burna XR-system som har eller är på väg att lanseras på marknaden och redogör kort för militära tillämpningar inom AR.

5.1 Automotiva tillämpningar

I februari 2023 besöktes elbilsmässan *eCar Expo* i Friends Arena, Stockholm. Vid mässan visade biltillverkare upp sina senaste elbilar och konceptbilar. Besökarna kunde även provköra vissa bilar. Målet med besöket var att identifiera trender inom dagens gränssnitt och se hur biltillverkarna visualiserar framtidens gränssnitt i sina konceptbilar.

Mässan gav en inblick i förarmiljöer i vilka föraren kan interagera med fysiska och digitala knappar, heads up- och touchdisplayer samt röst-, gest- och blickstyrning. Samtidigt möts föraren av ett konstant informationsflöde från bilen genom ljud, haptiska och visuella signaler. Det kan till exempel vara visuell information om adaptiv hastighetsbegränsning, baserad på GPS-kartor, filhållningsassistenten som får ratten att vibrera och en avståndsmätare som reglerar hastighet och avstånd till framförvarande bil. Genom tekniska framsteg har mängden information till föraren ökat under de senaste åren. Vid mässan var det tydligt att det inte längre enbart är flaggskeppsmodellerna som innehåller dessa funktioner, utan allt fler av biltillverkarnas modeller utrustas med allt mer teknik. Utmaningen hos biltillverkarna ligger framförallt i att använda rätt typ av gränssnitt för rätt typ av behov, med det huvudsakliga målet att kunna styra funktioner eller ta emot information utan att tappa fokus från körningen.

Blickföljning och geststyrning är gränssnitt som har fått en större betydelse i förarmiljön hos bilar som implementerar tekniken. Flera tillverkare vid mässan visade upp teknik där vissa bildskärmars innehåll förändrades beroende på om föraren eller passageraren tittade på skärmen. Beroende på var föraren fäster blicken anpassar dessa bilar informationsflödet genom att förstärka eller sluta visa viss information. Mercedes visade upp en teknik där föraren kan skapa personliga gester för särskilda funktioner, vilket kan göra geststyrningen mer tillgänglig för användaren. En annan trend som identifierades var att personlig teknik, såsom smarta klockor, integreras i fordonets system för att samla in fysiologisk data från föraren. Genom att till exempel kombinera förarens förändring i puls med data om förarens ögonrörelser kan bilen med större säkerhet mäta hur pass trött en förare är och om det är dags att ta en rast. Utveckling av bilars strålkastare möjliggör bland annat att projicera information på vägen framför bilen till stöd för föraren och att avblända särskilda delar av vägen samtidigt som andra delar behåller helljuset på, för att öka säkerheten²².

Biltillverkare bedöms vara en drivande aktör inom utveckling och utformning av framtida gränssnitt. Att studera deras lösningar och koncept kan leda till idéer om hur militära gränssnitt kan utformas.

²² Flera olika tillverkare har eller planerar att ha denna teknik i sina bilar, exempelvis Mercedes (länk: <https://group.mercedes-benz.com/innovation/specials/geneva-2018/digital-light.html>, hämtad 19/12-23) och Audi (länk: <https://www.audi-mediacycenter.com/en/press-releases/how-audis-light-digitization-is-pointing-the-way-toward-the-future-14624>, hämtad 19/12-23).

5.2 Kommersiell XR-teknik

Syftet med avskanningen var att uppdatera projektet med information om ny eller befintlig teknik som kan vara värdefull för framtida forskning. Avskanningen begränsades till två olika typer av tekniker, smarta glasögon och XR-headsets, och inriktades mot enheter som just lanserats eller som var på väg att lanseras på marknaden. Smarta glasögon ingår i begreppet huvudburen XR men skiljer sig från XR-headset framförallt genom formfaktorn. Medan ett XR-headset i princip kan se ut hur som helst är ambitionen för smarta glasögon att de ska uppfattas som just glasögon, både av den som bär dem och andra runt omkring. Utveckling av smarta glasögon har gått långsamt framåt och det har varit svårt att nå intresse från en bredare marknad. Smarta glasögon har kanske främst lockat de mest teknikintresserade och vad som på engelska refereras till som *early adopters*.

De smarta glasögon som ingick i avskanningen fungerar främst som externa bildskärmar till anslutna enheter som mobiler eller laptops. Presentationsytan kan konfigureras av användaren till att omfatta ett flertal ”fönster” och kräver inget fysiskt utrymme eftersom den uppstår i de smarta glasögonen. Som en del av XR-upplevelsen kan smarta glasögon kan vara försedda med hifi-ljud och interaktionen med presenterade fönster baseras ofta på geststyrning eller knapptryckning.

Tidigare var huvudburen enheter konstruerade för att förmedla en upplevelse antingen i VR eller AR. Numera har dock antalet headset som lanseras som MR ökat markant. MR-headset är försedda med kameror/sensorer som möjliggör video see-through, vilket innebär att användaren samtidigt kan se både den verkliga och den virtuella miljön. Jämfört med AR-system ger dessa system större möjlighet att hantera olika ljusförhållanden, men det finns också utmaningar. Till exempel medför kamerornas distans från ögonen en förskjutning av synfältet som skapar ett parallaxfel, och det förekommer även distorsion av objekt i synfältet. Mjukvaran i de olika headseten på marknaden har varierande förmåga att hantera dessa utmaningar.

Företaget Varjo var tidiga med att få fram en robust MR-upplevelse och har nu lanserat en version kallad Focal Edition med betydligt högre upplösning. Andra stora företag som tar fram kommersiella XR-produkter är Meta och Apple som delvis fokuserar på konsumentsegmentet och marknadsför sina produkter som något som ska användas till vardags. Tabell 4 redovisar resultatet av avskanningen.

Tabell 4. Översikt av kommersiella XR-enheter.

Smarta glasögon	Sammanfattning
XREAL AIR/ Rokid Max/ TCL NXTWEAR S – XR Glasses	XREAL AIR, Rokid Max och TCL NXTWEAR fungerar som en portabel skärmdisplay, dvs. det saknas inbyggd dator och strömkälla.
Lenovo ThinkReality A3	Lenovos smarta glasögon innehåller samma egenskaper som de ovanstående, men är betydligt dyrare. En skillnad är att presentationsytan är mer anpassningsbar än för ovanstående glasögon.
TCL RayNeo X2 - AR Glasses	TCL RayNeos smarta glasögon har både inbyggd dator och strömkälla. Glasögonen fungerar som en röststyrd smart assistent med inbyggda funktioner som navigering, direktöversättning av både tal och skrift, samt inspelning av bild och video. Användaren erhåller även information i det omedelbara synfältet så som tid, väder och meddelanden m.m. Glasögonen har i skrivande stund inte släppts på marknaden.

XR-headsets	Sammanfattning
Varjo XR-4 focal edition	Varjo XR-4 med focal edition är en tekniskt avancerad XR-enhet och återfinns inom marknads dyraste segment. Enheten har ett flertal integrerade sensorer med blickdrivet autofokus och en upplösning som ligger nära ögats prestanda.
Apple Vision Pro	Apples nya produkt kommer ut på marknaden 2024 och ska vara byggd för en sömlös XR upplevelse. Apples Vision Pro nyttjar blickstyrning för att interagera med virtuella objekt och fingernyp-gest för att selektera. Produkten tillhandahåller funktioner så som videoinspelning i 3D vilket även kan återskapas i 3D. Vid ansiktssamtal återskapas en visuell digital representation av eget ansikte och axlar i motpartens enhet.
VIVE XR Elite	VIVE XR Elite är, i jämförelse med ovanstående enheter, en relativt billig XR-lösning som både kan nyttjas med eller utan extern dator. VIVE XR Elite tillhandahåller ingen unik funktion eller förmåga som får den att sticka ut i mängden.
Lynx R1 Headset	Lynx R1 har varit i utveckling sen 2021 som kickstarterprogram. Enheten fokuserar på en effektiv MR upplevelse genom att tillhandahålla ett bra djupseende och en bra kvalitet i videoupplösningen.
Meta Quest 3	Meta Quest 3 är Metas nya XR produkt som är skapad för konsumentproduktsegmentet och är de billigaste XR lösningen på marknaden.

Tillverkarna som återfinns i Tabell 4 lanserar regelbundet nya utgåvor och ett flertal nya produkter kommer att presenteras på CES²³ i januari 2024. En generell målbild för de kommersiella XR-systemen är ett fokus på en delad XR-upplevelse där två eller flera användare kan interagera med samma virtuella objekt i en verklig värld. Det är en utveckling som är relevant att följa då det kan skapa nya förutsättningar och möjligheter. Meta har också tillsammans med Rayban²⁴ släppt ett par smarta glasögon som dock saknar ansats till att överlagra information visuellt. Dessa glasögon har inbyggd kamera, mikrofon och högtalare som möjliggör för användaren att ta bilder och i framtiden även kunna få svar från Metas AI-assistent om till exempel vad det är som finns på bilden, eller översätta text på skyltar etc. En målsättning med denna typ av smartglasögon är att bygga in tekniska funktioner och möjligheter i saker som används dagligen, som glasögon och solglasögon, snarare än att användaren behöver bära med sig ytterligare en teknisk pryl.

Simulatormässan i Rotterdam april 2023 (se avsnitt 1.4) hade ett stort fokus på olika tillämpningar av XR teknik, primärt i form av VR-headset med olika former av video see-through för att möjliggöra interaktion med verklig utrustning. Det fanns ett exempel på tillämpning av AR-headset Microsoft HoloLens av BAE Systems, i vad de kallar Project OdySSEy²⁵. Syftet var att skapa en sammanhållen syntetisk militär träningsmiljö för

²³ Consumer Electronic Show, konsumentelektronikmessa som går i Las Vegas i januari varje år.

²⁴ <https://www.meta.com/se/smart-glasses>

²⁵ <https://www.baesystems.com/en/feature/project-odyssey>

multidomän där AR användes för att stödja genomförandet av *After Action Review*-sessioner. Två utställare (HaptX²⁶ och Sensegloves²⁷) demonstrerade varsin variant på haptiska handskar för att ge fysisk återkoppling i händer och fingrar vid hantering av virtuella objekt. Det som är speciellt med handsken från HaptX är att den förutom mekanisk återkoppling för fingrar, även ger taktill återkoppling i fingertoppar och handflata. Funktionen baseras på ett hundratal små luftblåsor placerade i fingertoppar och handflata som kan expanderas eller dras ihop med hjälp av tryckluft. Systemet drivs av en kompressor som är monterad i en ryggsäck. En återkommande punkt under mässan var hur teknik som burna displayer och AI kan användas för att realisera mer adaptiva lärmiljöer för effektiv och ändamålsenlig träning. Det uttrycktes en förhoppning att burna displayer kan innebära att träning kan ske nära, både tidsmässigt och geografiskt, till skarp verksamhet.



Figur 15. Haptiska handskar, till vänster Senseglove och till höger HaptX.

5.3 AR-teknik för militära tillämpningar

AR i form av överlagrad symbolik har nått en hög teknisk mognadsgrad och förekommer numera i mer avancerade varianter av sikten, bildförstärkare och termiska sensorer. Exempel på symbolik är kompass, riktning till andra förband och enheter samt olika typer av tilläggsinformation om dessa enheter. För att inte klottra ner synfältet, eller riskera att kognitivt överbelasta användaren, finns funktioner för att reglera den mängd information som visas i olika nivåer. Många tillverkare presenterar standardiserad symbolik och har anpassat sina system mot ATAK²⁸. ENVG-B²⁹ från L3Harris och ARTIM³⁰ från Thermoteknix är exempel på produkter som tillhandahåller överlagrad AR-symbolik. Den relativt höga prislappen innebär att systemen i regel är förbehållna förband med specialfunktioner.

Gemensamt för ovan nämnda system är att synfältet är begränsat (oavsett om enheterna är avsedda för det ena eller båda ögonen). Det amerikanska IVAS³¹ programmet är en ansats till att ta fram ett system som medger ett brett synfält med överlagrad symbolik, objekt och sensorbilder som dessutom kan användas runt hela dygnet. Systemet utvecklas av Microsoft baserat på den civila produkten HoloLens. Bland dess funktionalitet återfinns stöd för planering, navigering och träning. Satsningen som påbörjades 2018 är ambitiös

²⁶ <https://haptx.com/>

²⁷ <https://www.senseglove.com/>

²⁸ Android Team Awareness Kit är ett stödsystem för ökat situationsmedvetande. Det tillhandahåller geospacial information och skapar möjligt att samverka på distans.

²⁹ <https://www.l3harris.com/all-capabilities/enhanced-night-vision-goggle-binocular-envg-b>

³⁰ <https://www.thermoteknix.com/products/defence-security/artim>

³¹ Integrated Visual Augmentation System är ett AR-baserat system framtaget för den amerikanska armén.

och med 5000 förbandssatta system har användarnas feedback inte låtit vänta på sig. Det har bland annat rapporterats om rörelsesjuka, huvud- och nackvärk och svårighet att lösa uppgifterna (Roque, 2022). Med nya iterationer har förbättringar införts såsom uppfällbar display, förändrad formfaktor och justerad placering av presenterad information. Utvecklingen har dessutom kantats av en rad tekniska problem som lett till förseningar. Det har bland annat varit svårt att åstadkomma mörkerförmågan utgående från digitala bildförstärkare och den höga ljusnivån utomhus under dagtid är en utmaning som adresserats genom att förse visiret med ett färgat filter. I nuläget är IVAS uppe i version 1.2 och utvecklingen fortsätter med en flerårig utvecklingsplan där innehållet i kommande utgåvor redan är fastställt.

6 Fortsatt verksamhet

En målsättning med arbetet i kommande projekt är att kombinera flera olika tekniker för att åstadkomma ett fungerande multimodalt gränssnitt med förmåga till anpassning av både den information som presenteras och interaktionen. Det som hittills inte behandlats i projektet är på vilket sätt som tekniken kan anpassa den presenterade informationen till användarens förutsättningar, aktuellt behov och pågående aktivitet. Kommande arbete inom projektet kommer därför att ges ett ökat fokus mot adaptiva gränssnitt och hur de kan tillämpas och nyttjas för att öka användarens förmåga att lösa sin uppgift. Ett grundläggande syfte med ett adaptivt gränssnitt är att undvika kognitiv överbelastning.

Adaptionen kan ta hänsyn till användarens aktivitet (baserat på kroppsnära sensorer som mäter till exempel blickriktning och kroppsposition), omgivningens förutsättningar (till exempel geografisk position) och planerad uppgift. Den inhämtade informationen bearbetas och analyseras med hjälp av maskininlärningsmetoder och andra typer av algoritmer. Sammantaget skapas ett underlag för vilken information som är lämplig att förmedla, genom vilken modalitet och vid vilket tillfälle.

Det här avsnittet beskriver det planerade fortsatta arbetet i det kommande projektets huvudsakliga forskningsspår.

6.1 Gester

Det finns flera relevanta spår att undersöka utgående från det arbete som hittills har genomförts. I första steget finns det behov av att utveckla metoder och teknik för insamling av träningsdata och modeller för klassificering, för att därefter ta fram demonstratorer av koncept för hur gester kan användas för interaktion.

Utvecklingen under året har skapat bättre tekniska förutsättningar för att systematiskt och kontrollerat samla in träningsdata med högre kvalitet än tidigare. En utmaning med nuvarande metod för insamling av träningsdata är att varje gest har genomförts enligt ett förutbestämt schema under en specifik tid. Detta innebär att datainsamlingen har tagit förhållandevis lång tid då den genomförts gest för gest. Det vore av intresse att undersöka möjligheterna till en mer flexibel automatisk annotering av träningsdata genom att nyttja redan tidigare tränade dataset för gestklassificering av videodata. Genom att spela in gester både med kroppsbruna sensorer och videokamera kan modellerna för video eventuellt användas för att annotera data från sensorerna löpande och insamlingen skulle inte behöva vara så tidsmässigt uppstyrd. En synkroniserad videoinspelning möjliggör även att i efterhand annotera insamlad sensordata, vilket inte är möjligt i dagsläget.

En annan fördel med att komplettera datainsamling med video är, att det i efterhand blir möjligt att manuellt justera annoteringen av datamängden, vilket inte är fallet med data enbart från bruna sensorer. Insamling gjord på detta sätt skulle potentiellt kraftigt förenkla insamlingen av träningsdata, men skulle eventuellt innebära en lägre kvalitet jämfört med insamling där personerna blir direkt instruerade. Förhoppningen är dock att en kraftigt ökad volym på träningsdata kan kompensera för en eventuell minskning i datakvalitet. En förutsättning för denna typ av maskininlärningsbaserad annotering är att modellerna för videobaserad gestidentifiering kan klassificera de typer av gester eller aktiviteter som är av intresse för detta projekt. Om så inte är fallet behöver det undersökas om det är möjligt att förbättra modellerna.

En önskvärd egenskap för gestdetekteringen är att den inte ska behöva vara tränad på samma person som den sedan används av. Det är enklare för en modell att klassificera gester från en person om personen ingår i träningsdata men det är önskvärt att modellen generaliserar till nya personer. En viktig fortsättning är att undersöka och utveckla generaliserbarheten vidare, så att modellen kan identifiera gester oavsett vilken person som utför dem. För detta behövs mer träningsdata från flera olika personer samt justering av klassificeringsmodellen.

Det finns ett fortsatt behov av att undersöka vilka särdrag i registrerad data från burna sensorer som bidrar med relevant information för en klassificering. Denna analys syftar till att säkerställa vilka typer av sensorer och dess placering som är viktiga för att mäta en lämplig uppsättning särdrag och deras påverkan på inläringen hos maskininlärningsmodellerna. Det är också viktigt att beakta hanteringen av både statiska och dynamiska gester och behovet av att ta hänsyn till den temporala aspekten vid klassificering.

Med en förbättrad teknisk plattform för gestidentifiering vore det intressant att fortsatt utveckla konceptet med att kombinera gestbaserade och adaptiva gränssnitt, förslagsvis realiserade med XR-glasögon. Ett tänkbart förslag på sätt att demonstrera detta är att förändra vad för information som visas beroende på *hur* eller *om* användaren håller i en automatkarbin eller annan utrustning. Ett annat utvecklingsspår skulle vara att förbättra möjligheterna att identifiera explicita handtecken i en mer dynamisk och rörlig situation, tillsammans med att i XR-headset visualisera handtecken som andra i gruppen gör.

6.2 XR

För framtida studier inom XR-spåret finns det flera inriktningar att utforska. Ny kommersiell teknik finns tillgänglig på marknaden som tar nya kliv i utvecklingen genom mer robusta virtuella upplevelser som kommer behöva värderas. XR området är brett och innefattar därmed MR, AR och VR teknologier, något som projektet ämnar studera i olika utsträckning. En studie som planeras inom ramen för XR-spåret är utvärdering av upplevelse som funktion av ergonomiska och tekniska faktorer vid användning av huvudburna enheter.

Mixed Reality har fått ett upplyft på den kommersiella sidan med nya huvudburna enheter som har släppts eller ämnas släppas under 2024, så som Meta Quest 3 eller Apples Vision Pro. Teknologierna förväntas möjliggöra nya potentiella förmågor eller inneha egenskaper som kan tillämpas i en militär kontext. För att studera och hitta tillämpningsområden för MR-spåret är tanken att genomföra användarstudier med Försvarsmakten för att tillsammans hitta värdefulla egenskaper med tekniken som kan vara till nytta i Försvarsdomänen.

Augmented Reality har varit ett fokus under 2022-2023 där utveckling av en virtuell ledningsplats genomfördes under åren inom ramen för XR spåret. Tanken för kommande period är att vidareutveckla den befintliga ledningsplatsprototypen för att skapa en mer dynamisk, skalbar och interaktiv prototyp. Hur saker och ting kan/bör visualiseras i AR är fortfarande ett relevant område som behöver studeras vidare och även utvärderas tillsammans med användarrepresentanter. Det planeras även för ett exjobb som rör utveckling och studie av en gränssnittsprototyp i AR med fokus på interaktion med huvud- och blickriktning.

Virtual Reality avses integreras i VBS4 det kommande året vilket ger nya möjligheter att studera scenarier i VR för träning- och utbildningssyften. Frågeställningar som berör övningar i form av mekaniserad strid eller avsuttan soldat i den virtuella världen är intressant att studera och jämföra med nuvarande bildskärmlösningar. Ett exempel är: Vad betyder ökad inlevelse i striden för träningseffekten?

6.3 Gränssnitt för generativ AI

Fortsatta studier i linje med konceptet som presenterats i 4.2.2 kräver en gränssnittsprototyp där interaktionen mellan användaren och chattbotten kan styras i högre grad än i de fritt tillgängliga verktygen. Det finns flera frågeställningar kopplade till gränssnittet men också till hur chattbotten bäst ska användas; till exempel hur en chattbott ska kunna sammanfatta flera delar av planeringsmetoden Planering Under Tidspress (PUT) samtidigt som den kanske behöver primas på olika sätt för att bäst hjälpa användaren i enskilda PUT-steg.

Med en gränssnittsprototyp på plats är tanken att börja utvärdera mot användare av PUT-modellen. Eftersom verktyget används för taktisk planering är lämpliga personer att utvärdera mot användare med erfarenhet av detta. Användarstudien syftar till att undersöka hur deras upplevelse och prestation påverkas av att interagera med en textbaserad generativ AI-modell som stöd vid planering.

Faktorer som är av intresse att undersöka är:

- 1) *Prestation*, om det är tidseffektivt och användbart i jämförelse mot dagens metod.
- 2) *Arbetsbelastning*, hur påverkas användare vid interaktion med en textbaserad generativ AI.
- 3) *Upplevelse*, vad tycker slutanvändare om att använda generativ AI vid militär planering.

Utöver konceptet kring AI för PUT planeras även arbete för att ta fram en möjlig virtuell assistent för telekrigsapplikationer för att skapa en lättillgänglig metod för användare att hantera telekrigsdata. Målsättningen är att undersöka om det är möjligt att med dagens AI-modeller, samt programmerings-API:er skapa en sådan virtuell assistent för analyser av data som liknar telekrigsdata.

Ett centralt begrepp inom telekrig är emitterbeskrivning. Det kan betyda lite olika saker, men i sin grundform har ett system sänt ut information som har snappats upp. Traditionellt har denna information sparats i tabellform som skapar möjlighet till indexering och snabb sökning, men det genererar också komplexitet när en emitter till exempel kan byta mellan ett par moder eller vågformer. Då finns det flera olika beskrivningar som har en relation. När en emitter är ännu mer dynamisk är så, och kan byta mellan ett 20-tal moder, får man ett praktiskt problem – det blir väldigt mycket att kartlägga och hålla reda på. Liknande problem uppkommer när nya versioner av vågformer och/eller protokoll tas fram i så hög takt att när en version precis är kartlagd har en ny version lanserats.

Baserat på antagandet att ett spaningssystem kan skatta en uppsättning parametrar, några kontinuerliga parametrar likt frekvens och bandbredd samt några icke-kontinuerliga parametrar som modulation, modulationsordning och bandspridningskod, är det då möjligt att med några exempel förklara problemet för assistenten och få hjälp att lösa detta med data i form av

- strukturerat dataset (bibliotek) samt
- ostrukturerat dataset (tidigare inmätning)
- ny ostrukturerad observation (ny mätning)?

Syftet med delstudien är att svara på frågan om det, med dagens teknik och AI-modeller är möjligt att skapa en virtuell telekrigsassistent, samt i så fall – hur bör interaktionen med denna assistent utformas?

7 Slutsats

Även om tekniken utvecklas och förändras med ökade inslag av autonomi, är människan fortsatt med i loopen som en viktig aktör genom sin förmåga att analysera, agera och fatta beslut. Väl utformade gränssnitt är en förutsättning för att öka människans förutsättningar att verka och därmed möjligheten att lösa förekommande uppgifter. Det finns ett värde i att fortsatt bevaka hur tekniken utvecklas för att ligga på framkant med att identifiera möjliga framtida försvarsmaktstillämpningar.

Behovet av enkla och trådlösa sätt att interagera med systemen har drivit på teknikutveckling inom området, och det går att se hur till exempel gestinteraktion fått en given roll för system inom XR-sfären. Det är möjligheten att kombinera separata tekniker på ett effektivt sätt som utgör grunden för kommande interaktionsprinciper – men det är också viktigt att beakta hur informationen kan flöda på ett sömlöst sätt mellan olika enheter och presenteras för användaren när och där den behövs.

Det är av intresse att studera hur flera olika tekniker kan kombineras för att åstadkomma ett fungerande multimodalt gränssnitt med förmåga till anpassning, både av den information som presenteras och interaktionen, utgående från kontexten och användarens status (fysisk och mental arbetsbelastning). För att nå denna målsättning behövs vidare arbete med dels gränssnitt för att presentera information (XR-teknik), dels gränssnitt för att förmedla input till de tekniska systemen (till exempel via blickriktning och gester). Metoder (maskininlärning) behövs även för att automatiskt anpassa informationen och sättet den presenteras på för användaren. I detta ligger också ambitionen att ta fram exempel på gränssnitt som kan öka användarens förmåga att uppfatta relevant information i omgivningen.

Referenser

- Alenljung, Z. (2022). *Augmented Reality and Remote Interaction with Military Unmanned Ground Vehicles*. Master's thesis. Linköping University. Department of Computer Science. Cognitive Science.
- Alenljung, Z., & Lindholm, V. (2022). *Samverkan mellan Människa och AI i militärkontext* FOI Memo 8035. Totalförsvarets forskningsinstitut.
- Andrews, C., Southworth, M. K., Silva, J. N., & Silva, J. R. (2019). Extended reality in medical practice. *Current treatment options in cardiovascular medicine*, 21, 1-12.
- Azuma, R. T. (1997). A Survey of Augmented Reality. In *Presence: Teleoperators and Virtual Environments* (Vol. 6). <http://www.cs.unc.edu/~azuma>
- Bommasani, R., Hudson, D. A., Ehsan Adeli, Altman, R. B., Arora, S., Sydney von Arx, Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N. S., Chen, A. T., Creel, K., Davis, J., Demszky, D., & Donahue, C. (2021). *On the Opportunities and Risks of Foundation Models*. ArXiv (Cornell University). <https://doi.org/10.48550/arxiv.2108.07258>
- Cheng, L., Guan, Y., Zhu, K., & Li, Y. (2017). Recognition of human activities using machine learning methods with wearable sensors. *IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, 2017, pp. 1-7. <https://doi.org/10.1109/CCWC.2017.7868369>
- Chiarella, D., Bibuli, M., Bruzzone, G., Caccia, M., Ranieri, A., Zereik, E., Marconi, L., & Cutugno, P. (2018). A novel gesture-based language for underwater human-robot interaction. *Journal of Marine Science and Engineering*, 6(3), 1–20. <https://doi.org/10.3390/jmse6030091>
- Christensen, C., Raynor, M. E., & McDonald, R. (2013). *Disruptive innovation*. Brighton, MA, USA: Harvard Business Review.
- Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2017). *Deep reinforcement learning from human preferences*. <https://doi.org/10.48550/arxiv.1706.03741>
- Chui, M. (2023) The state of AI in 2023: Generative AI's breakout year. McKinsey & Company. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-AIs-breakout-year>
- Dai, N., Lei, I-M., Li, Z., Li, Y., Fang, P., & Zhong, J. (2022). Recent advances in wearable electromechanical sensors—Moving towards machine learning-assisted wearable sensing systems. *Nano Energy*, 105. <https://doi.org/10.1016/j.nanoen.2022.108041>
- Elliott, L., Skinner, A., Pettitt, R., Vice, J., & Walker, A. (2014). *Utilizing glove-based gestures and a tactile vest display for covert communications and robot control*. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2014. Report No.: ARL-TR-6971.
- Frana, P. & Klein, M. (2021). *Encyclopedia of Artificial Intelligence: The Past, Present and Future of AI*. Bloomsbury Publishing USA.
- Gemini Team & Google. (2023). *Gemini: A Family of Highly Capable Multimodal Models*. https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf
- Geng, J., Huang, D., & De la Torre, F. (2022). *DensePose From WiFi*. Hämtad från <https://arxiv.org/abs/2301.00250>

- Hartnett, G., Elliott, L., Baraniecki, L., Skinner, A., Riddle, K., & Pettitt, R. (2018) *Wearable concepts for gesture-based robot control, direction localization, and communication to a tactile belt display*. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2018. Report No.: ARL-TR-8597.
- Hern, A. (2023, Mar 29). *Elon Musk joins call for pause in creation of giant AI 'digital minds'*. The Guardian. Hämtad december 2023.
<https://www.theguardian.com/technology/2023/mar/29/elon-musk-joins-call-for-pause-in-creation-of-giant-ai-digital-minds>
- Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. *CIRP journal of manufacturing science and technology*, 29, 36-52.
- Karlsson, J., Hermelin, J., Levin, B., & Nilsson, S., (2022). *Att styra med gester: Kunskapsöversikt och utvärdering av gester som gränssnitt*. FOI-R--5381--SE. Totalförsvarets forskningsinstitut.
- Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., m.fl. (2017). The kinetics human action video dataset. Hämtad från <https://arxiv.org/abs/1705.06950>
- Le, T-H., Tran, T-H., & Pham, C. (2019) The Internet-of-Things based hand gestures using wearable sensors for human machine interaction. *International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*. IEEE. s. 1-6.
- Levin, B., Nilsson, S., Herkevall, J., Alenljung, Z. & Granåsen, M. (2022). *Virtuella ledningsplatser – Slutrapport 2022*. (FOI-R—5406—SE). Totalförsvarets forskningsinstitut.
- Levin, B., Nilsson, S., Hermelin, J., Svensson, E., & Stenius, C. (2021). *Teknisk utveckling inom människa-maskininteraktion – Applikationer för framtida gränssnitt*. FOI-R--5245--SE. Totalförsvarets forskningsinstitut.
- Lindholm, V., Alenljung, Z., & Karlsson, J. (2023). *Bridging the Gap between Humans and AI in Military Planning: Officers' Response to a Text-Based AI Assistant*. FOI-S--6767--SE. Totalförsvarets forskningsinstitut.
- Milgram, P. & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems*, Vol E77-D, No.12 December 1994
- Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 37(3), 311–324.
<https://doi.org/10.1109/TSMCC.2007.893280>
- OpenAI. (2023). *GPT-4 Technical Report*. ArXiv (Cornell University).
<https://doi.org/10.48550/arxiv.2303.08774>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. K., Schulman, J., Hilton, J. K., Kelton, F., Miller, L. P., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., & Lowe, R. J. (2022). *Training language models to follow instructions with human feedback*. ArXiv (Cornell University). <https://doi.org/10.48550/arxiv.2203.02155>
- Pallavicini, F., Pepe, A., & Minissi, M. E. (2019). Gaming in virtual reality: What changes in terms of usability, emotional response and sense of presence compared to non-immersive video games?. *Simulation & Gaming*, 50(2), 136-159.
- Roque, A. (2022, Feb 01). *US Army soldiers experienced physical side-effects using IVAS, Pentagon report finds*. Janes. Hämtad december 2023.
<https://www.janes.com/defence-news/news-detail/us-army-soldiers-experienced-physical-side-effects-using-ivas-pentagon-report-finds>

- Roziere, B., Gehring, J., Gloeckle, F., Sootla, S., Gat, I., Tan, X.-Y., Adi, Y., Liu, J., Tal Remez, Rapin, J., Kozhevnikov, A. A., Evtimov, I., Bitton, J., Bhatt, M., Cristian Canton Ferrer, Grattafiori, A., Xiong, W., Alexandre Defossez, Copet, J., & Azhar, F. (2023). *Code Llama: Open Foundation Models for Code*. ArXiv (Cornell University). <https://doi.org/10.48550/arxiv.2308.12950>
- Siddiqui, U-A., Ullah, F., Iqbal, A., Khan, A., Ullah, R., Paracha, S., Shahzad, H., & Kwak, K-S. (2021). Wearable-sensors-based platform for gesture recognition of autism spectrum disorder children using machine learning algorithms. *Sensors*. 21(10):3319. <https://doi.org/10.3390/s21103319>
- Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I* (pp. 757-764).
- Sutton, R. S., & Barto, A. (2018). *Reinforcement learning: an introduction*. The MIT Press.
- Thunholm, P. (2003). Military decision making and planning: towards a new prescriptive model (Doctoral dissertation, Department of Psychology, Stockholm University). ISBN 91-7265-692-1.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023a). *LLaMA: Open and Efficient Foundation Language Models*. ArXiv (Cornell University). <https://doi.org/10.48550/arxiv.2302.13971>
- Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., & Fuller, B. (2023b). *Llama 2: Open Foundation and Fine-Tuned Chat Models*. ArXiv.org. <https://doi.org/10.48550/arXiv.2307.09288>
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
- Vi, S., da Silva, T. S., & Maurer, F. (2019). *User experience guidelines for designing hmd extended reality applications*. In Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part IV 17 (pp. 319-341). Springer International Publishing.
- Webster, E. (2021). *An inside look at F-35 pilot helmet fittings*. United States Air Force. <https://www.af.mil/News/Article-Display/Article/2719003/an-inside-look-at-f-35-pilot-helmet-fittings/> [hämtad 2023-11-15]
- Wigdor, D., & Wixon, D. (2011). *Brave NUI World. Designing Natural User Interfaces For Touch and Gesture*. Morgan Kaufmann Publishers Inc, USA. ISBN: 978-0-12-382231-4
- Zhang, J., Shen, F., Xy, X., & Shen, H.T. (2019) *Temporal Reasoning Graph for Activity Recognition*. Hämtad från <https://arxiv.org/abs/1908.09995>



FOI
Totalförsvarets forskningsinstitut
164 90 Stockholm

Tel: 08-55 50 30 00
Fax: 08-55 50 31 00

www.foi.se