

Tekst til video med avatars

Henrik Fladkjær, 2026

Working paper

Abstract

Artiklen undersøger, hvordan tekst-til-video-teknologi (T2V) med AI-genererede avatarer opleves af professionelle praktikere på tværs af brancher, og hvilke pædagogiske, organisatoriske og etiske implikationer teknologien har i professions- og erhvervsrettede kontekster. Udgangspunktet er et fokusgruppeinterview med deltagere fra blandt andet finanssektoren, produktionsvirksomheder, uddannelsessektoren og konsulentbranchen, koblet til aktuel forskning i videobaseret læring med AI-avatarer.

Metodisk bygger studiet på et semistruktureret fokusgruppeinterview med en heterogent sammensat gruppe af deltagere, hvor data analyseres med henblik på at identificere centrale mønstre i deltagernes erfaringer og vurderinger.

Resultaterne viser, at T2V opleves som en teknologi med betydeligt effektivitetspotentiale, særligt i organisationer der allerede arbejder systematisk med video, men hvor produktion og opdatering tidligere har været tids- og ressourcekrævende. Deltagerne fremhæver især mulighederne for at omdanne eksisterende skriftligt materiale til korte, målrettede mikrovideoer, der kan tilpasses forskellige målgrupper og hurtigt opdateres, når regler, systemer eller procedurer ændres. Samtidig peger både fokusgruppe og litteratur på, at T2V ikke introducerer nye læringsprincipper, men skalerer kendte videodidaktiske greb, når teknologien anvendes som del af blended-forløb, PBL-aktiviteter og andre fælles studieformer.

Et centralt fund er spændingsfeltet mellem effektivitet og autenticitet. Avatar-baserede videoer vurderes som velegnede til standardiseret informations- og instruktionskommunikation, men som utilstrækkelige og til tider problematiske i sårbare, værdiladede og relationelt krævende situationer, hvor kropssprog, improvisation og empati er afgørende. Fokusgruppen peger desuden på væsentlige juridiske og etiske udfordringer knyttet til datasikkerhed, deepfakes, transparens og bias, som fordrer klare organisatoriske politikker for, hvornår mennesker kan og ikke kan erstattes af digitale "mennesker". Artiklen konkluderer, at T2V's potentiale primært ligger i at fungere som fleksibel komponent i bredere video- og læringsstrategier, mens teknologien i højt relationelle forløb bør ses som supplement til, ikke erstatning for, menneskelig tilstedeværelse.

Introduktion

Tekst-til-video-teknologi med AI-genererede avatarer er på få år gået fra teknologisk kuriosum til et tilgængeligt produktionsværktøj i både private virksomheder og uddannelsesinstitutioner. I takt med at værktøjer kan omdanne konkrete tekster til færdig video på minutter, opstår nye muligheder for skalering af undervisnings- og informationsmateriale, men også nye spørgsmål om autenticitet, læringskvalitet og styring af data og etik. For professions- og erhvervsrettede uddannelser, hvor

praksisnær viden, rollemodeller og dialogiske læreprocesser er centrale, er det særligt afgørende at forstå, hvad tekst-til-video (T2V) faktisk kan bidrage med – og hvor teknologien bør begrænses eller indlejres i mere klassiske pædagogiske formater.

Denne artikel undersøger, hvordan professionelle praktikere på tværs af brancher oplever potentialer og begrænsninger ved T2V og AI-avatarer i deres daglige arbejde og i professionsrettet efter- og videreuddannelse. Udgangspunktet er et fokusgruppeinterview med deltagere fra blandt andet bank, a-kasse, produktionsvirksomheder, uddannelsessektor og konsulentbranchen, som alle har interesse for eller erfaring med at anvende video i forskellige sammenhænge. Analysen kobler deres perspektiver til aktuel forskning i videobaseret læring, mikrovideo, multimodale læremidler og AI-baserede avatarer i undervisning, service og marketing (Sidi et al., 2021; Gijzen et al., 2024; Ramos et al., 2022; Zhang et al., 2022; Schulz et al., 2023).

Baggrund

Teknologisk set bygger moderne T2V-systemer på generative modeller, der kombinerer tekst-til-billede-arkitekturer med store mængder video- og billeddata (Ge et al., 2023; Zheng et al., 2024). Denne arkitektur gør det muligt at generere relativt realistiske avatarer og scenarier ud fra korte tekstprompter. I praksis opleves teknologien af fokusgruppens deltagere primært som et værktøj til hurtig produktion og opdatering af instruktions- og informationsvideoer.

Inden for videobaseret læring peger forskningen på at korte, velstrukturerede videoer kan understøtte kognitiv bearbejdning og selvreguleret læring, især hos personer der ikke er eksperter på området (Gijzen et al., 2024; Sidi et al., 2021; Zhang et al., 2022). Studier viser samtidig, at erfarne eksperter ofte foretrækker tekst og selektiv læsning, mens video formater har størst effekt for mindre erfarne målgrupper (Sidi et al., 2021). Denne differentiering genfindes i fokusgruppen, hvor T2V vurderes særligt relevant til trainees, studerende og medarbejdere, der står foran nye systemer og regler, mens erfarne fagpersoner først og fremmest ser teknologien som et formidlingsværktøj, der kan supplere allerede etablerede læringsmetoder.

Teoretisk ramme

Dette afsnit samler fire komplementære perspektiver, som danner udgangspunkt for både design og analyse af undersøgelsen. For det første præsenteres en multimedielæringsteoretisk ramme med fokus på kognitiv belastning og målgruppers forkundskaber. For det andet introduceres en pædagogisk-didaktisk ramme med modeller for videounderstøttet kollaborativ læring. For det tredje

udfoldes en socio-teknisk og etisk ramme om autenticitet, data og deepfakes, og endelig beskrives en teknologisk ramme, der tydeliggør T2V-teknologiens aktuelle muligheder og begrænsninger.

Multimedielæringsteoretisk ramme

Gijsen m.fl. (2024) bygger på Cognitive Theory of Multimedia Learning (CTML) og kognitiv belastningsteori og viser, hvordan interaktive videoer kan understøtte dyb versus overfladisk bearbejdning, afhængigt af opgavekrav og kontrolmuligheder. Song m.fl. (2023) demonstrerer expertise reversal effect i en systematisk sammenligning af tekstbaserede og multimodale dokumenter. Personer der ikke er eksperter inden for området, har fordel af dokumenter der indeholder video og illustrationer, mens eksperter ofte ignorerer disse elementer eller finder dem forstyrrende. Zhao m.fl. (2023) viser i en klinisk kontekst (burn surgery), at kombinationen af problem-baseret læring (PBL) og mikro-videoer (korte, fokuserede sekvenser) i et blended forløb øger både teoretiske testresultater, praktiske færdighedsniveauer og evnen til kritisk, klinisk dømmekraft.

Pædagogisk-Didaktisk Ramme:

Ramos m.fl. (2021) gennemgår i en systematisk review 120 empiriske studier om video-understøttet kollaborativ læring (Video-Supported Collaborative Learning, VSCL) og identificerer fire hovedmodeller for, hvordan video kan integreres i socialt, fælles studieaktivitet design:

1. Observation og analyse af fælles studieaktiviteter
2. Fælles studieaktiviteter baseret på video udvikling
3. Læring gennem fælles studieaktiviteter, der er baseret på videoindhold
4. Video-understøttet samarbejde

Det er min vurdering, at T2V-avatar-videoer kan placeres ind i alle fire modeller:

- Som praksisnære "video-cases".
- Som -studenterproducerede T2V-projekter hvor studerende designer egne teoretiske og praktiske diskussioner
- Som korte, struktureret indholdslevering i forhold til en konkret undervisning
- Som forberedelse til både fysisk og online undervisning.

- optakt eller illustrativt supplement til live online-rådgivnings- eller undervisningssessioner (model 4)

Socio-teknisk og etisk ramme

Forskning i AI-avatarer understreger en forskel mellem effektivitet og kvalitet. (Schulz et al. 2023) fremhæver, at avatarer kan øge tilgængelighed, differentiering og feedback, men at de samtidig udfordrer oplevelsen af autenticitet og social tilstedeværelse, hvis de præsenteres som direkte erstatning for menneskelige undervisere. I et business-ethics-kursus oplevede mange studerende en AI-avatar som klar og struktureret, men beskrev samtidig undervisningen som følelsesmæssigt flad og mindre engagerende, især når det ikke var tydeligt, at der var tale om en syntetisk figur (Vallis et al., 2023).

Teknologisk ramme

Afsnittet om teoretisk ramme samler fire komplementære perspektiver, som danner udgangspunkt for både design og analyse af undersøgelsen. For det første præsenteres en multimedielæringsteoretisk ramme med fokus på kognitiv belastning og målgruppers forkundskaber. For det andet introduceres en pædagogisk-didaktisk ramme med modeller for videounderstøttet kollaborativ læring. For det tredje udfoldes en socio-teknisk og etisk ramme om autenticitet, data og deepfakes, og endelig beskrives en teknologisk ramme, der tydeliggør T2V-teknologiens aktuelle muligheder og begrænsninger.

De fire rammer

T2V teknologien befinder sig aktuelt på et stadie, hvor den kan automatisere store dele af produktionskæden fra manuskript til færdige korte videoer. Avancerede modeller som Sora og Veo 3 kan generere relativt realistiske scener, komplekse bevægelser i høj opløsning, men de bruges primært eksperimentelt. I praksis er det derfor især platforms-løsninger som Synthesia, HeyGen og lignende, der definerer, hvad der er muligt at udføre med teknologien f.eks. i uddannelsesmæssige sammenhænge.

Synthesia er en softwareløsning, der illustrerer, hvad den nuværende standard kan bruges til. Brugeren skriver et manuskript, vælger en avatar og en baggrund, og systemet genererer en video med relativt flydende tale og mundbevægelser uden behov for kamera, lys eller redigering. Synthesia understøtter hundredvis af sprog, et stort bibliotek af avatarer samt mulighed for at skabe en personlig avatar baseret på en kort videooptagelse, hvilket gør det muligt at anvende den samme person på tværs af mange videoer og målgrupper. Synthesia er ikke så scenisk kreativ, men har derimod fokus på hurtigt at være produktionsklar.

Begrænsningerne ved at anvende i dette eksempel, Synthesia, ligger især på tre områder: For det første er der stadig udfordringer med følelsesmæssig nuancering, stemmeføring og empati. Stemmerne beskrives som klare, men relativt neutrale uden den store dynamik, hvilket begrænser anvendelsen i sårbare eller værdiladede situationer. For det andet kan små fejl i synkroniseringen mellem mund og udtale, gestik og mimik skabe en oplevelse, som nogle brugere oplever som forstyrrende, især ved længere sekvenser. For det tredje er der tekniske og juridiske rammer. Der kan være licensrestriktioner på brug af visse avatarer, der er en datalagrings problematik, og uklare grænser for, hvor realistisk og personbåret indhold bør være. Det får som konsekvens, at teknologien ikke er så velegnet til en undervisningsform baseret på dialog, men mere på konkret vidensoverføring som gør, at teknologien i praksis primært anvendes til standardiseret, lavt-relationel kommunikation frem for egentlig dialogisk undervisning. (<https://magai.co/text-to-video-ai-tools-pros-and-cons/>)

Metoder

I forbindelse med fokusgruppeinterviewet er Synthesia brugt som referencerammer for T2V. Det er også anvendt til at udarbejde illustrative eksempler på, hvad teknologien kan præstere.

Formål og anvendelse af fokusgruppeinterview

I dataindsamling er der anvendt et fokusgruppeinterview med henblik på at indsamle nuancerede data i form af holdninger, perspektiver og oplevelser (Kvale & Brinkmann, 2015) gennem en struktureret gruppediskussion. Fokusgruppe interviewet er valgt som metode for diskussioner mellem deltagerne. Målet er at frembringe dybere refleksioner og nye indsigter gennem fælles dialog (Morgan, 1997).

Formålet med det konkrete fokusgruppeinterview er at identificere muligheder og ulemper ved anvendelse af T2V i forbindelse med virksomheders og organisationers distribution af viden internt i organisationen og eksternt rettet på tværs af brancher og sektorer.

Sammensætning af deltagergruppen

Ifølge (Krueger & Casey 2015) er en heterogen gruppe ofte fordelagtig, når formålet er at indhente forskellige perspektiver på en problemstilling. Det har været vigtigt i forbindelse med udvælgelsen af fokusgruppedeltagerne, at det var personer, der alle havde en relation til, hvordan de på optimal måde kan distribuere og præsentere viden for forskellige aftagergrupper. Den konkrete gruppe bestod derfor af deltagere med forskellige baggrunde. Der deltog følgende: En Senior HR læringspartner fra en bank, en salgs- og informationskonsulent fra en fagforening, en direktør og ejer i en virksomhed der udvikler it systemer til dyrlæger, en erhvervschef, en sektionschef i et stort revisionsselskab samt en influencer inden for privatøkonomisk rådgivning.

Denne tværfaglige sammensætning sikrede en bred vifte af synspunkter og indsigter, som kunne bidrage til en mere nuanceret forståelse af emnet. For at kunne skabe en dynamisk interaktion mellem deltagerne og skabe mulighed for at kunne komme i dybden med de forskellige emner blev der inviteret en gruppe på mellem 5-8 deltagere (Morgan, 1997; Halkier, 2010).

Facilitering og spørgeramme

Undertegnede fungerede som moderator i fokusgruppeinterviewet for at sikre, at vi kom omkring alle de udvalgte emner i spørgerammen, samt sikrede, at alle deltagere fik mulighed for at bidrage til diskussionerne (Stewart & Shamdasani, 2014). Det var en vigtig opgave for mig som moderator at skabe en god atmosfære, hvor alle følte sig trygge og havde lyst til at udtrykke deres meninger. Et vigtigt bidrag hertil avr, at alle deltagere på forhånd modtog et dokument med en kort biografi af alle deltagere.

Spørgerammen fulgte en semistruktureret tilgang, hvilket betød, at interviewet tog udgangspunkt i en række åbne spørgsmål, der dækkede de vigtigste temaer (Kvale & Brinkmann, 2015).

Dataindsamling og analyse

Jeg optog samtalerne i fokusgruppeinterviewet med deltagernes samtykke, så de efterfølgende kunne transskriberes med henblik på efterfølgende analyser. Analyserne vil relatere sig til de temaer der blev præsenteret i løbet af fokusgruppeinterviewet Braun og Clarke (2006). Det medførte en efterfølgende systematisk kodning af data fra interviewet med henblik på at identificere og kategorisere centrale temaer og mønstre i deltagernes udsagn. Formålet med fokusgruppeinterviewet var også, at undersøgelsen skulle bidrage med en nuanceret viden om, hvilket potentiale der var i anvendelsen af teknologien i forhold til at distribuere viden.

Etiske overvejelser

Der blev i forbindelse med fokusgruppeinterviewet taget højde for de etiske retningslinjer for kvalitative studier som beskrevet af (Kvale & Brinkmann 2015). Dette inkluderede:

- **Informeret samtykke:** Deltagerne ville på forhånd modtage information om interviewets formål, anonymisering af data og deres ret til at trække sig.
- **Anonymitet:** Deltagernes identiteter vil blive anonymiseret i den endelige analyse og rapportering.
- **Fortrolighed:** Lydoptagelser og transskriptioner vil blive opbevaret sikkert og udelukkende anvendt til det specificerede formål.

Resultater

Effektivitet, tidsforbrug og økonomi

Fokusgruppen beskrev tekst-til-video som en teknologi med betydeligt effektivitetspotentiale, især i organisationer, der allerede arbejder systematisk med video, men de oplevede produktionen som tids- og ressourcekrævende. En deltager fra banksektoren fremhævede, at en instruktionsvideo tidligere kunne tage omkring 30 timer at producere, mens tilsvarende materiale nu kan udvikles på cirka 2 timer, når storyboard og manus er på plads, fordi optage- og redigeringsleddet i høj grad automatiseres. Tidsbesparelsen opleves særligt tydelig ved opdateringer, hvor små ændringer i regler, satser eller skærmbilleder tidligere krævede næsten fuld genoptagelse, mens T2V gør det muligt hurtigt at rette i manus, udskifte enkelte skærmbilleder og regenerere videoen uden logistiske omkostninger.

Denne erfaringsbaserede tilgang genfindes i studier af generativ AI i produktion af indhold, hvor værktøjer til tekst, billede og video beskrives som teknologier, der sænker adgangsbarrierer, reducerer marginalomkostninger og øger outputvolumen. Luy et al. (Lyu et al. 2024) viser, hvordan indholdsproducenter kæder sprogmodeller, billed- og videoprocessorer sammen fra idé og manus til færdig video, hvor hurtige ændringer er centrale. Fokusgruppens deltagere beskrev tilsvarende sammenhænge, hvor de så T2V som et yderligere lag oven på eksisterende kommunikations- og undervisningsstrategier: "det samme som før – bare hurtigere, billigere og i flere versioner".

Flere deltagere understregede dog, at den største tidsomkostning fortsat ligger i udvikling af indhold i et manuskript. Tidsgevinsten viser sig især, når der allerede findes eksisterende skriftlige materialer, som kan omformes til video, der efterfølgende også uden det store tidsforbrug kan vises i nye versioner eller forskellige versioner rettet mod forskellige modtagergrupper.

Økonomisk fremstår T2V i fokusgruppen som en relativt attraktiv investering, selv når licensprisen opleves som høj, fordi den sammenlignes med omkostninger til professionelle filmhold, intern tidsbinding og behovet for løbende opdateringer. En deltager beskrev, hvordan organisationen tidligere ofte undlod at opdatere ældre videoer på grund af omkostninger, mens T2V gør det mere realistisk at holde materialet ajour med værdi både for læring og compliance. Dermed ændre T2V ikke kun tempoet i produktionen, men også de reelle muligheder for at arbejde systematisk med kvalitet gennem aktualitet. Når der ikke længere kræver kamera, optagelokaler og ekstern redigering at opdatere en video, sænkes tærsklen for at rette fejl, tilføje præciseringer og tilpasse indhold til nye regler, systemer eller målgrupper. Hvor organisationer tidligere "levede med" forældede videoer, fordi omkostningerne ved at optage om var for høje, bliver det nu realistisk at tænke i løbende vedligehold frem for flerårige kampagner. Det betyder, at T2V kan understøtte en mere iterativ og feedback-drevet praksis.

Fokusgruppen pegede samtidig på, at effektivitet og kvalitet kun går hånd i hånd under bestemte betingelser. Teknologien kræver en læringskurve. Effektivitet kan derfor ikke reduceres til simpel

tids- og omkostningsberegning, men må forstås i samspil med krav til faglig kvalitet, målgrupper og didaktisk integration.

Autenticitet, empati og relationel kommunikation

Fokusgruppen udtrykte en tydelig ambivalens over for T2V-videoer, når kommunikationen bevæger sig væk fra rene instruktioner og over mod interpersonelle og værdiladede temaer. Teknologien blev vurderet som velegnet til korte, faglige indslag og simple tutorials. Avatarbaserede videoer blev oplevet som stive og mindre overbevisende, når der er behov for rådgivning, usikkerhedshåndtering og kropsligt nuanceret kommunikation, særligt i sårbare økonomiske situationer. Her anses situationsforståelse, empati og fleksibel respons som uadskillelige fra den professionelle rolle, og en digital rådgiver vurderes derfor som utilstrækkelig og til tider upassende.

Forskning i AI-baserede avatarer i undervisning bekræfter, at AI-baserede undervisere ofte opleves som klare og lette at forstå, men samtidig "for glatte" og følelsesmæssigt flade, hvilket reducerer engagement og relationel kvalitet. På den baggrund skelner flere deltagere mellem viden- og regelkommunikation og lærings- og udviklingsforløb. I førstnævnte dominerer behovet for neutralitet, klarhed og stabilitet, og her fremstår en avatar, der leverer samme forklaring ensartet hver gang, som en mulig styrke. I lærings- og udviklingsforløb opleves dialog, improvisation og underviserens egen stemme derimod som uerstattelige, fordi deltagerne også skal orientere sig i faglig dømmekraft, værdikonflikter og professionelle rolleforventninger.

Et centralt spor i fokusgruppen handler om, hvordan "kunstighed" forstås og vurderes. Nogle deltagere accepterer uden større problemer en tydeligt digital figur, når dette er gennemsigtigt og forventningsafstemt, mens andre reagerer stærkt på uoverensstemmelser mellem mundbevægelser, mimik og tale, hvor næsten-menneskelighed opleves som mere forstyrrende end en stileret figur. Dette harmonerer med Vallis' fund af, at studerende overordnet accepterer avatarer som undervisere, men samtidig beskriver dem som mindre engagerende og følelsesmæssigt flade end menneskelige undervisere (Vallis, 2023).

Fokusgruppen så samtidig muligheder i mere personaliserede avatarer, der repræsenterer en kendt leder, en gennemgående figur i et læringsforløb eller en tydelig organisationsbærer, snarere end en generisk underviser. Disse avatarer kan dog ikke erstatte de menneskelige relationer, men kun supplere dem.

Læring, kognition og videoformatet

Fokusgruppen fremhævede særligt korte videoer, der præcist gennemgår et regneark, et digitalt system eller et regelsæt, som støtte til, at den der ser videoerne kan styre egen progression og vende tilbage til komplekse passager efter behov.

Når T2V kombineres med AI-genererede avatarer, ændres både det læringsmæssige udtryk og de kognitive krav til de lærende. Både forskning og fokusgruppe peger på, at unaturlig udtale, begrænset gestik og små synkroniseringsfejl kan fungere som kognitive distraktioner, der flytter

opmærksomheden fra fagligt indhold til at lede efter fejl i videoerne, hvilket medfører at der anvendes tid herpå i stedet for at tilegne sig den konkrete læring (Gijzen, 2024).

Brugsscenerier på tværs af brancher

Selv om fokusgruppen havde forskellige holdninger og krav til T2V samt, hvordan avatarer blev opfattet var der en vis enighed om, at T2V primært kan anvendes til en standardiseret informations- og instruktionskommunikation, mens mere relationelle og værdiladede interaktioner forbeholdes menneskelige aktører, i tråd med litteraturens skelnen mellem lav- og højt relationelle opgaver.

I uddannelse og efteruddannelse beskrives T2V-mikrovideoer som supplement i blended-forløb, fx til at introducere litteratur, dække emner man ikke når på seminarer, eller understøtte softwarebrug. Dette svarer til modeller, hvor video fungerer som content, der efterfølgende bearbejdes i dialogiske aktiviteter, så avatar-videoer bliver grundlag i PBL-workshops, cases og refleksionsopgaver frem for selvstændige forløb (Zhao et al., 2023).

Når video skal anvendes i forbindelse med interaktion mellem mennesker, som i bankrådgivning, jobcentre eller influencer-kommunikation – er fokusgruppen markant mere forbeholden. Her vurderes det, at T2V ikke kan erstatte menneskelig dialog i komplekse, sårbare rådgivnings- og sagsbehandlingssituationer.

Samlet tegnes T2V som en teknologi til standardiseret, skalerbar kommunikation og læring, hvor ensartethed, tilgængelighed og hurtig opdatering er de primære gevinster. Den vurderes derimod som mindre egnet i situationer, hvor interaktion mellem mennesker og improvisation er centrale, bør T2V suppleres af menneskelig tilstedeværelse.

Juridiske, etiske og governance-dimensioner

Fokusgruppen pegede på datasikkerhed, hosting og organisatoriske restriktioner som centrale barrierer for skalering af T2V, særligt i regulerede brancher som finans. Krav om EU-hosting, stramme databehandleraftaler og klare grænser for, hvilke tekst- og sagsoplysninger, der må uploades føles som en stor udfordring. Samtidig pegede fokusgruppen på risici og barrierer, som ikke nødvendigvis er lige relevante for alle typer af data og anvendelser. Bekymringer om datasikkerhed, og forveksling mellem menneskelig og syntetisk kommunikation synes især at knytte sig til personfølsomme oplysninger mens en stor del af de scenarier, deltagerne faktisk beskrev, handlede om generisk fagligt indhold, offentligt tilgængelige regler eller interne procedurer. Her kan de samme krav til EU-hosting og stramme databehandleraftaler risikere at fungere som en slags generel teknologisk skepsis, snarere end en differentieret vurdering af risiko.

Set i et socioteknisk perspektiv kan fokusgruppens forbehold derfor også læses som udtryk for en blanding af legitim professionsetik og mere diffust formulerede forestillinger om "farlig" AI. Det rejser spørgsmålet, om der i nogle tilfælde er tale om en overforsigtighed, der begrænser eksperimenter med relativt ufarlige brugsscenerier, eller om mangel på konkret viden om dataflows og

modelarkitektur gør det vanskeligt for deltagerne at skelne mellem reelle og hypotetiske risici. Dermed bliver en central governance-udfordring ikke kun at regulere teknologien, men at understøtte en informeret risikovurdering, hvor etisk og juridisk omtanke balanceres mod potentialet for at anvende T2V på en afgrænset og fagligt meningsfuld måde.

Et gennemgående etisk tema er risikoen for forveksling mellem menneskelig og syntetisk kommunikation og konsekvenserne for autenticitet, tillid og digital dannelse. Studier anbefaler, at det markeres tydeligt, at det er en AI underviser. Dette var fokusgruppen enige i ud fra en bekymring for, at brugere distraheres af et digitalt "menneskes" kunstige træk.

Surveyartikler om tekst-til-video og synthetic media fremhæver, at teknologien kan blive en del af et accelererende deepfake problem. Det bliver sikkert inden for en relativ kort periode vanskeligt eller umuligt at se forskel på videoer med autentiske personer eller AI genererede personer. Det skærper behovet for politikker for, hvornår mennesker er erstattet med digitale versioner og hvornår mennesker ikke må erstattes (*Cho et al. 2024*). Samtidig viser litteraturen omkring AI avatarer, at køn, race og social position i avatarer påvirker, hvordan et indhold opfattes og hvilken tillid der er til det fremførte. Fremadrettet bliver T2V ikke kun et model- og designvalg. Det bliver i endnu højere grad et spørgsmål om professionel etik og ikke kun teknik (*Hanus et al. 2015*).

Perspektiver for uddannelse og blended learning

Indsigterne fra fokusgruppeinterviewet viser, hvordan korte T2V-klip kan fungere som mikrolæringsselementer i blended-forløb, særligt i forskellige uddannelsesforløb. Denne brug ligger tæt på forskning i mikrovideo kombineret med PBL, hvor korte, fokuserede videoer, der understøtter forberedelse og problemløsning, kan øge både præstationer, kritisk tænkning og tilfredshed (*Zhao et al. 2024*). T2V foreslås tilsvarende brugt til at introducere nøglebegreber, demonstrere værktøjer og koble teori til praksis før PBL-workshops, så video ses som input til dialog og problemløsning frem for afsluttende vidensleverancer.

Ramos et al.'s review (*Ramos et al. 2022*) placerer T2V primært som videobaseret indhold, der efterfølgende bearbejdes i fælles studie aktiviteter som PBL-workshops, casearbejde og refleksionsopgaver. Video har størst pædagogisk effekt, når den kobles til strukturerede processer, hvor mennesker samarbejder og ikke blot bruges som "talking head". Dette stemmer overens med fokusgruppens skelnen mellem traditionel vidensoverførsel, hvor T2V er oplagt, og dialogisk læring, hvor den menneskelige underviser er central. Både fokusgruppe og litteratur peger på behovet for forskellige design for de mindre erfarne brugere versus de mere erfarne brugere.

På den baggrund tegner der sig praktiske designprincipper for T2V i PBL- og blended-forløb. Det er vigtigt at bruge korte forberedelsesvideoer ("hvad du skal vide, før vi mødes") og at sikre, at videoerne understøtter flere spor med forskellig kompleksitet. Fokusgruppens opfattelse er desuden, at integration af T2V bør ske iterativt og erfaringsbaseret – gennem få, velvalgte brugscases, hvor undervisere og studerende løbende evaluerer læringsudbytte, format og anvendelighed.

Samtidig peger diskussionen på, at T2V-materialer i en erhvervsrettet kontekst ikke blot handler om effektiv produktion, men om at skabe aktuelle og fagligt præcise ressourcer, der er tæt koblet til branchens praksisser og systemer. T2V's potentiale ligger netop i at muliggøre hurtige opdateringer, så læringsmateriale, der omhandler regler, processer eller digitale værktøjer, altid afspejler virkelighedens ændringer. Det betyder, at T2V ikke kun bidrager til effektivitet, men kan løfte kvaliteten gennem aktualitet og relevans, hvilket styrker materialernes værdi i professionsrettede uddannelser og intern kompetenceudvikling.

Diskussion

Fokusgruppeundersøgelsen pegede samlet på, at tekst-til-video med AI-avatarer tilbyder et betydeligt effektivitetspotentiale, men at denne gevinst kun kan forstås meningsfuldt, hvis den ses i sammenhæng med læringskvalitet, målgrupper og kontekstuelle krav. Deltagerne beskrev teknologien som "det samme som før – bare hurtigere, billigere og i flere versioner", hvilket harmonerer med litteraturens beskrivelser af generativ AI som en infrastruktur, der sænker adgangsbarrierer og marginalomkostninger i indholdsproduktion. Samtidig udfordrede fokusgruppen en instrumentel forståelse af effektivitet, fordi tidsgevinsterne primært ligger i produktion og opdatering, mens den faglige og didaktiske udvikling af manuskript fortsat er den mest ressourcekrævende del. (Lyu et al., 2023)

Set i lyset af multimedielæringsteorien og forskning i mikrovideo understøttede fokusgruppen antagelsen om, at korte, fokuserede videoer særligt gavner mindre erfarne målgrupper, der har brug for guidet progression og mulighed for gentagelse (*Gijzen et al., 2024; Sidi et al., 2021; Zhang et al., 2022*). T2V bidrager ikke umiddelbart med et helt nyt læringsprincip, men snarere med en skalering af allerede kendte videodidaktiske greb, forudsat at videoerne designes som korte, målrettede mikrovideoer tæt koblet til konkrete opgaver og cases (Cho et al., 20024).

I en pædagogisk-didaktisk optik bekræftede de personer i fokusgruppen, der har relation til undervisning, at T2V-avatarer primært bør forstås som ét element i video-understøttet fælles studieaktiviteter frem for en selvstændig undervisningsform. Deltagernes beskrivelser af T2V som forberedelse til PBL-workshops, som praksisnære cases viser, at teknologien er mest lovende, når avatar-videoer bruges som grundlag for efterfølgende dialogiske aktiviteter, hvor fælles analyse, refleksion og fælles studieaktiviteter problemløsning fortsat faciliteres af menneskelige undervisere særligt i en vejledningssituation, hvor særligt dialog, men også improvisation og kropssprog anses som uundværlige elementer i kommunikationen.

Et centralt spændingsfelt i materialet vedrører autenticitet, empati og relationel kommunikation, hvor fokusgruppen var markant mere forbeholden over for avatarer end ved rene instruktioner og regelkommunikation. Deltagernes beskrivelser af avatarer som "stive", følelsesmæssigt flade og til tider forstyrrende stemmer overens med studier, der dokumenterer reduceret oplevelse af social

tilstedeværelse, når syntetiske undervisere præsenteres som erstatning for menneskelige. Her tegner der sig en normativ grænse: T2V vurderes acceptabel og nyttig til standardiseret information, men problematisk i sårbare rådgivningssituationer, hvor kropssprog, improvisation og situeret empati anses som uadskillelige komponenter i professionel praksis (Vallis et al., 2023).

De socio-tekniske og etiske dimensioner, som træder frem i fokusgruppen, forstærker dette spændingsfelt ved at pege på både governance-udfordringer og dannelsesmæssige spørgsmål. Krav om datasikkerhed, EU-hosting og skarpe grænser for, hvilke tekster og sagsoplysninger der må uploades, fungerer som konkrete barrierer for skalering – særligt i regulerede sektorer. Samtidig rejser risikoen for forveksling mellem menneskelige og AI baserede afsendere samt de bias, der kan indlejres i avatarers køn, race og sociale position, spørgsmål om professionel etik, transparens og digital myndiggørelse af brugere. Det foreslås derfor både i litteratur og fokusgruppe, at brugen af T2V altid tydeligt markeres som syntetisk, og at organisationer udvikler politikker for, hvornår mennesker må og ikke må erstattes.

Metodisk gav dialogen medfokusgruppen rige, tværfaglige indsigter, men indebærer også væsentlige begrænsninger for generaliserbarheden. Deltagerne var relativt positive og interesserede i video og teknologi, hvilket kan betyde, at kritiske perspektiver fra mere skeptiske professionelle ikke er fuldt repræsenteret. Undersøgelsen producerer desuden erfarings- og holdningsbaserede udsagn snarere end læringsdata, hvilket betyder, at konklusionerne om T2V's pædagogiske effekt må forstås som hypoteser, der bør efterprøves i kontrollerede interventionsstudier i konkrete uddannelsesforløb. En oplagt videreudvikling vil derfor være at kombinere fokusgrupper med design-baserede studier, hvor T2V-scenarier afprøves og evalueres systematisk sammen med studerende og undervisere på professions- og erhvervsrettede uddannelser.

Samlet peger diskussionen på, at T2V-teknologiens potentiale ikke først og fremmest ligger i at erstatte menneskelige undervisere eller rådgivere, men i at integreres som en fleksibel komponent i bredere video- og PBL-strategier. Teknologien fremstår særlig relevant til standardiseret, skalerbar kommunikation og mikrolæring, hvor ensartethed, tilgængelighed og hurtig opdatering er centrale mål. I højt relationelle, værdiladede og dømmekrævende forløb bør T2V derimod ses som supplement til – og ikke erstatning for – menneskelig tilstedeværelse, med tydelige etiske retningslinjer for brugen. På den baggrund peger materialet mod nogle spændende muligheder for i efterfølgende studier at analysere, hvordan der kan udvikles praktiske kognitive designprincipper for T2V i professionsrettede forløb.

Konklusion

Tekst-til-video med AI-avatarer fremstår på baggrund af denne undersøgelse som en teknologi med markant potentiale for at effektivisere produktion og opdatering af instruktions- og informationsmateriale, men dens værdi afhænger afgørende af, hvordan den integreres didaktisk,

organisatorisk og etisk. Fokusgruppen viste, at T2V særligt giver mening, når eksisterende skriftligt materiale kan omsættes til korte, målrettede mikrovideoer, der understøtter modtagere, der ikke er eksperter i at tilegne sig nye systemer, regler og arbejdsgange på en fleksibel og selvreguleret måde. Teknologien tilføjer her en skalerbarhed og opdateringshastighed, som mange organisationer tidligere har manglet, fordi traditionelle videoproduktioner var for tids- og ressourcekrævende til løbende vedligehold.

Samtidig peger både litteratur og fokusgruppe på, at T2V ikke introducerer et fundamentalt nyt læringsprincip, men forstærker kendte videodidaktiske greb, når den anvendes som del af blended-forløb, PBL-arbejde og andre fælles studieaktiviteter. De mest lovende scenarier placerer avatar-videoer som forberedende input, cases eller korte demonstrationssekvenser, der efterfølgende bearbejdes i dialogiske og fælles studieaktiviteter faciliteret af menneskelige undervisere. I denne optik bør T2V forstås som en fleksibel komponent i en samlet video- og læringsstrategi snarere end som en selvstændig undervisningsform eller direkte erstatning for underviseren.

Undersøgelsen tydeliggør også en grænse for, hvor T2V vurderes som passende og legitimt, særligt i relation til autenticitet, empati og professionel dømmekraft. Avatar-baserede videoer opfattes som velegnede til standardiseret, lavt relationel kommunikation, men problematiske i sårbare rådgivnings- og udviklingsforløb, hvor kropssprog, improvisation og situeret relationel kompetence er centrale. Dertil kommer væsentlige juridiske og etiske udfordringer omkring datasikkerhed, deepfakes, transparens og bias, som gør det nødvendigt, at organisationer udvikler klare politikker for, hvornår mennesker må – og ikke må – erstattes af digitale repræsentationer. Samlet peger undersøgelsen derfor på, at T2V's fremtidige rolle i professions- og erhvervsrettede kontekster ikke blot er et spørgsmål om teknologi, men om pædagogisk design og professionel etik, der skal udvikles gennem iterative, praksisnære eksperimenter og systematisk evaluering.

Abstract in english

The article investigates how text-to-video technology (T2V) with AI-generated avatars is experienced by professional practitioners across sectors and the pedagogical, organisational, and ethical implications of this technology in professional and vocational contexts. The point of departure is a focus group interview with participants from, among others, the financial sector, manufacturing, education, and consulting, linked to current research on video-based learning with AI avatars. Methodologically, the study is based on a semi-structured focus group interview with a heterogeneous group of participants, and the data are analysed thematically, drawing on Braun and Clarke's framework, to identify central patterns in the participants' experiences and evaluations. The results show that T2V is perceived as a technology with considerable efficiency potential, particularly in organisations that already work systematically with video, but where production and updating have previously been time- and resource-intensive. Participants especially highlight the potential to transform existing written material into short, targeted micro-videos that can be adapted

to different target groups and rapidly updated when rules, systems, or procedures change. At the same time, both the focus group and the literature indicate that T2V does not introduce new learning principles but scales well-known video-didactic approaches when used as part of blended courses, PBL activities, and other collaborative study formats.

A central finding is the tension between efficiency and authenticity. Avatar-based videos are regarded as suitable for standardised information and instructional communication, but as insufficient and at times problematic in vulnerable, value-laden, and relationally demanding situations where body language, improvisation, and situated empathy are crucial. The focus group also points to significant legal and ethical challenges related to data security, deepfakes, transparency, and bias, which require clear organisational policies on when humans can and cannot be replaced by digital “humans”. The article concludes that T2V’s potential lies primarily in functioning as a flexible component within broader video and learning strategies, while in highly relational settings, the technology should be seen as a supplement to, rather than a replacement for, human presence.

Litteraturliste:

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.

Cho, J., Puspitasari, F.D., Zheng, S., Zheng, J., Lee, L.-H., Kim, T.-H., Hong, C.S. & Zhang, C. (2024). Sora as an AGI world model? A complete survey on text-to-video generation. *arXiv:2403.05131*.

Fiordelli, M., Diviani, N. & Schulz, P. (2013). Mapping mHealth research – A decade of evolution.

Gijssen, M., Catrysse, L., De Maeyer, S. & Gijbels, D. (2024). Mapping cognitive processes in video-based learning by combining trace and think-aloud data. *Learning and Instruction*, 90, 101851.

Halkier, B. (2010). *Fokusgrupper*. Frederiksberg: Samfundslitteratur.

Krueger, R.A. & Casey, M.A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). Thousand Oaks, CA: Sage.

Kvale, S. & Brinkmann, S. (2015). *InterViews: Learning the craft of qualitative research interviewing* (3rd ed.). Thousand Oaks, CA: Sage.

Kvale, S. & Brinkmann, S. (2015). *Interview: Introduktion til et håndværk*. København: Hans Reitzels Forlag.

Lyu, Y. et al. (2024) ‘A preliminary exploration of YouTubers’ use of generative-GAI in content creation’, in Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA ’24), Honolulu, HI: ACM

Morgan, D.L. (1997). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Ramos, J.L., Cattaneo, A.A.P., de Jong, F.P.C.M. & Espadeiro, R.G. (2022). Pedagogical models for the facilitation of teacher professional development via video-supported collaborative learning: A review of the state of the art. *Journal of Research on Technology in Education*, 54(5), 695–718.

Song, H., Healey, J., Siu, A., Wigington, C. & Stasko, J. (2023). Experts prefer text but videos help novices: An analysis of the utility of multi-media content. In *CHI EA '23: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*.

Stewart, D.W. & Shamdasani, P.N. (2014). *Focus groups: Theory and practice* (3rd ed.). Thousand Oaks, CA: Sage.

Zhao, Z., Lv, D. & Chen, L. (2024). Application of problem-based learning combined with micro-video teaching in burn surgery and its impact on satisfaction with teaching. *Journal of Investigative Surgery*, 37(1), 2403534.

Internetsider:

<https://magai.co/text-to-video-ai-tools-pros-and-cons/>