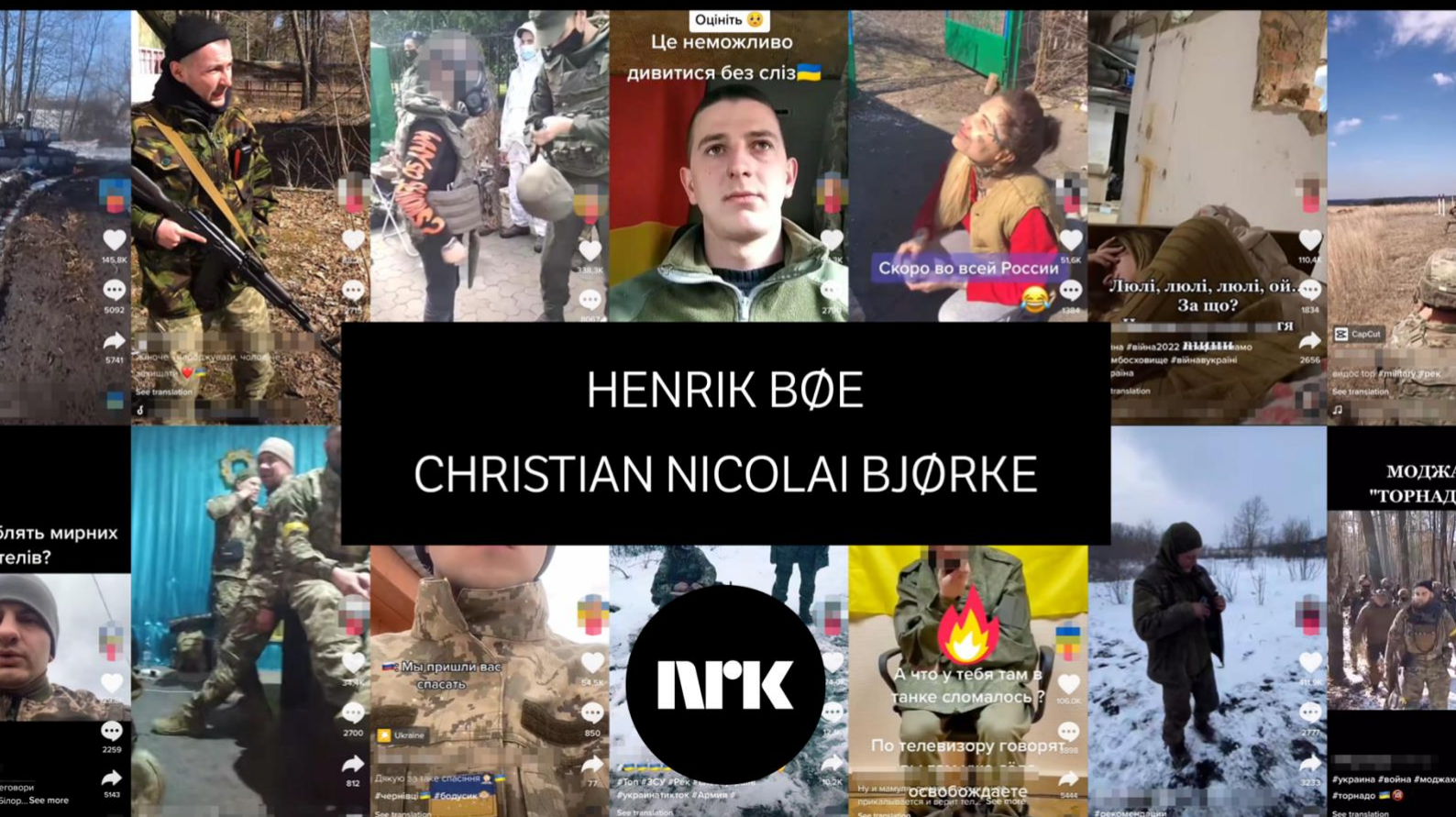




Én app – to verdener

NRK UNDERSØKENDE GRUPPE REGION SØR-ØST

Metoderapport Data-SKUP 2022



Innhold

1. Innsendere	1
2. Bakgrunn: Krig i Europa	2
3. Slik startet prosjektet	2
4. Tiden og veien.....	3
5. Metode	3
5.1 Å bygge en TikTok-bot	3
5.2 Plassere bot i Russland og Ukraina	4
5.3 Opprette russisk og ukrainsk TikTok-bruker	4
5.4 Innhenting av data.....	5
5.5 Analyse av data	5
6. Historiefortelling og presentasjon.....	7
6.1 Hvordan fortelle en datahistorie?	7
6.2 Navn og bilde på bot-ene	8
6.3 TikTok-universet.....	8
6.4 Verifisering av videoer	8
7. Etikk.....	8
7.1 Bruk av videoer fra TikTok	8
7.2 Falske brukere.....	9
7.3 Bruk av residential proxies	9
8. Spesielle erfaringer.....	9
8.1 Web-versjonen av TikTok	9
8.2 Jeg er ikke en robot	9
9. Konsekvenser.....	10
9.1 Ut i verden.....	10
9.2 Inspirerer flere norske journalister.....	10
9.3 TikTok viser nye videoer.....	10

1. Innsendere

Innsendere:

Henrik Bøe. Tlf. 902 56 884. Epost: henrik.boe@nrk.no

Christian Nicolai Bjørke. Tlf. 932 266 41. Epost: christian.nicolai.bjorke@nrk.no

Redaktør:

Anne Vinding. Tlf. 975 939 44. Epost: anne.vinding@nrk.no

Takk til Anne Vinding, Hilde Martine Lindgren, Rina Therese Blekkerud, Tuva Skei Tønset, Maria Pedko Lindbäck, Norunn Reinertsen, Thomas Nikolai Blekeli, alle i Undersøkende Gruppe i Sør-Øst og de som er på «hjemmekontorene» våre i Fredrikstad og Tønsberg.

Lenke til hovedsak: [Én app – to verdener](#)

Lenke til engelsk versjon: [Worlds Apart](#)

2. Bakgrunn: Krig i Europa

24. februar 2022 gikk Russland til krig mot Ukraina. Journalister ble evakuert ut for å komme seg i sikkerhet. I informasjonsvakuemet som fulgte ble plutselig det sosiale mediet TikTok, som så langt var mest kjent for dansevideoer og gøyale challenges, et populært sted å finne førstehåndsskildringer fra krigen. Flere medier døpte det «First TikTok War».

4. mars innførte Russland svært strenge lover mot «falske nyheter» om militæret. En rekke medier trakk seg ut av landet og Facebook, Twitter og etter hvert også Instagram ble blokkert. Kinesisk-eide TikTok var et av få sosiale medier som fortsatt fungerte.

6. mars skrudde TikTok av muligheten til å laste opp nytt innhold i Russland for å beskytte sine ansatte og brukere mot den nye loven. Men tusenvis av videoer om Ukraina-krigen ble fortsatt lastet opp i resten av det globale TikTok-samfunnet.

3. Slik startet prosjektet

Høsten 2021 ble Undersøkende gruppe i NRK region Sør-Øst opprettet. Gruppen består av rundt ti journalister på NRKs kontorer i Oslo og Viken, Vestfold og Telemark og på Sørlandet.

Under et møte før jul diskuterte vi [Wall Street Journals](#) undersøkelse av TikToks algoritmer. Det amerikanske mediehuset hadde opprettet TikTok-brukere som automatisk skrollet seg gjennom det sosiale mediet. Funnet var at disse brukerne fort havnet i «kaninhull» der de ble servert mer og mer av det de var interessert i.

Vi ønsket å gjøre en lignende undersøkelse, men med norske temaer og brukere. Flere medlemmer i gruppen hadde nylig fått et kompetanseløft innenfor datajournalistikk. Spørsmålet var om vi kunne klare å sette opp automatiserte TikTok-brukere. Henrik Bøe fikk i oppgave å undersøke dette.

Da Russland invaderte Ukraina noen måneder senere ble Undersøkende gruppe bedt om å bidra til krigsdekningen i NRK. Christian Nicolai Bjørke kom opp med en hypotese: Unge russiske og ukrainske TikTok-brukere får servert helt forskjellige nyheter fra krigen.

På dette tidspunktet hadde vi funnet en løsning for å skrolle gjennom og lagre TikTok-poster automatisk. Vi fikk dermed ideen om å sende bot-ene våre til en russisk og en ukrainsk grenseby for å se om hypotesen stemte.

Det endte med en blank avsløring: TikTok viste ikke krigen til russiske brukere.

4. Tiden og veien

Å jobbe med både en pågående krig og algoritmene til et sosialt medium er en utfordring. Ting endret seg fort på bakken i Ukraina, og algoritmene til TikTok kunne endres over natta. Vi satte derfor en kort tidsfrist for arbeidet for å være så oppdaterte og aktuelle som mulig.

Vi kom på ideen om ettermiddagen torsdag 10. mars og jobbet fulltid med saken. 25 dager senere (3. april) publiserte vi hovedsaken, med god støtte fra kollegaer og ledere. Innenfor denne perioden ferdigstilte Henrik Bøe og Christian Nicolai Bjørke programmeringen av botene, høstet data, bygde historiefortellingen, utformet visuelle elementer, hentet inn tilsvarende og sydde det hele sammen.

5. Metode

5.1 Å bygge en TikTok-bot

Løsningen vi kom fram til, er skrevet i programmeringsspråket Python. Den er basert på følgende komponenter:

- nettleserautomatiseringsverktøyet Selenium
- biblioteket BeautifulSoup som brukes til å «skrape» nettsider
- TikToks nettversjon (tiktok.com)
- Azure Computer Vision API
- databasen og søkeverktøyet Elasticsearch
- visualiseringsverktøyet Kibana

Boten konfigureres i et eget script. Der kan vi legge inn:

- Navnet informasjon fra boten skal indekseres under i databasen
- En liste med emneknagger den er interessert i
- En liste med stikkord den er interessert i
- Hvor lenge den skal se på videoer den er interessert i, og hvor lenge den skal se videoer den ikke er interessert i
- Antall videoer boten skal se før operasjonen avbrytes

Vi har skrevet et script som gjør følgende:

- Åpner <https://tiktok.com> i Google Chrome
- Skraper følgende informasjon om den første videoen:
 - Brukernavn
 - Tekst
 - Emneknagger
 - Musikk
 - Bilde-URL til coverbilde
 - Video-URL
 - Antall likes, kommentarer og delinger
- Sender coverbilde-URL til Microsoft Azure Computer Vision API
- Får tilbake følgende informasjon fra API-et:
 - Beskrivelse av bildet

- Stikkordliste med hva API-et mener bildet inneholder
- Eventuell tekst på bildet blir «transkribert» med OCR (Optical Character Recognition)
- Sammenligner botens interesser med emneknagger og tekst som hører til videoen, i tillegg til tekst og stikkordliste fra API-et
- Dersom noe av dette stemmer med botens interesser, vil den se lenge på videoen. Hvis ikke ser den kortere på videoen.
- Sender all info vi har skrapet til en Elasticsearch-database
- Scroller videre til neste video, og gjentar de samme operasjonene
- Når boten har sett antall videoer vi spesifiserte i konfigureringsscriptet, lukkes nettleseren igjen

5.2 Plassere bot i Russland og Ukraina

Da vi bestemte oss for å plassere TikTok-bot-ene våre i Russland og Ukraina, hadde vi det meste av løsningen beskrevet i forrige avsnitt allerede på plass. Nå trengte vi en metode for å få TikTok til å tro at vi fysisk var plassert i Russland og Ukraina på en best mulig måte.

Etter litt research og samtaler med kolleger, kom vi fram til at den beste løsningen ville være å bruke såkalte «residential proxies», mellomtjenere som lar oss velge hvor i verden trafikken vår skal rutes gjennom, helt ned på bynivå.

Vi valgte å bruke selskapet Oxylabs, fordi dette er en av de største leverandørene av residential proxies. Løsningen fungerer ved at husstander i byer rundt omkring i diverse land har en avtale med selskapet om å selge deler av internett-trafikken sin. Dermed kan man ned på bynivå velge hvilken by internett-trafikken skal gå gjennom, så lenge dette er en by med brukere som har avtale med Oxylabs. Det er etiske utfordringer knyttet til bruk av slike mellomtjenere, som vi diskuterer nærmere i avsnittene om etikk.

Etter å ha gjort en del testing, fikk vi koblet proxyløsningen fra Oxylabs til scriptene våre. Vi valgte å plassere bot-ene i byene Belgorod i Russland og Kharkiv i Ukraina, fordi vi mente det var det beste for historiefortellingen. Det beskriver vi nærmere senere i metoderapporten.

5.3 Opprette russisk og ukrainsk TikTok-bruker

Vi opprettet først to nye epost-adresser. En for den russiske boten og en for den ukrainske. For å gjøre det mest mulig realistisk var vi «plassert» i henholdsvis Russland og Ukraina da vi opprettet e-postadressene. For den russiske brukeren brukte vi epost-tjenesten Yandex.ru, fordi det er en av de mest brukte epost-tjenestene i Russland. For den ukrainske brukte vi Gmail.

Deretter opprettet vi TikTok-brukere for hver av bot-ene på samme måte. Nå hadde vi et system som lot oss automatisk skrolle gjennom TikTok som en russisk eller en ukrainsk bruker.

5.4 Innhenting av data

Vi startet innhenting av data 16. mars. Til å begynne med lot vi de to bot-ene se på hver TikTok-video omtrent like lenge. Vi ville finne ut av hva TikTok presenterte for en helt ny russisk eller ukrainsk bruker som ikke hadde noen spesielle preferanser. Vi gjorde ingen søk og likte ingen videoer.

Det ble raskt klart at den ukrainske bot-en fikk opp krigshandlinger fra første video, mens den russiske ikke så snurten av en eneste tanks. Fra og med 22. mars prøvde vi derfor å «hjelp» den russiske bot-en til å se mer av krigen. Vi programmerte begge bot-ene til å se lengre på videoer med krigsinhold. På dette tidspunktet hadde bot-ene sett på rundt 650 videoer hver.

Vi brukte følgende stikkord og emneknagger for å detektere krigsinhold:

- Stikkord: "weapon", "military uniform", "military", "officer", "soldier", "camouflage", "rifle", "military vehicle", "military organization", "army", "military rank", "ballistic vest", "military camouflage", "military person", "gun", "firearm", "infantry", "ranged weapon", "marines", "uniform", "shooting", "gun barrel", "military vehicle", "assault rifle"
- Hashtager: "#украина", "#славаукраїні", "#україна", "#україна", "#война", "#ukraine", "#українаUA", "#UA", "#зеленский", "#зсу", "#украинаUA", "#війна", "#война2022", "#путин", "#путінхуйло 🇺🇦", "#нетвойне"

Vi hentet inn data til og med 22. august. Da hadde hver av bot-ene skrollet seg gjennom rundt 2000 videoer.

5.5 Analyse av data

Planen var å analysere ved hjelp av databasen i Elasticsearch og visualiseringsverktøyet Kibana. Vi hadde derimot ikke mye erfaring med verktøyene, og situasjonen endret seg raskt både på bakken i Ukraina og med TikToks algoritmer. Derfor valgte vi å eksportere mesteparten av dataene over til Microsoft Excel for å spare tid, siden vi var kjent med programmet fra før. Vi brukte likevel Elasticsearch og Kibana til å lage ordskyer som viste hvilke emneknagger som dukket opp oftest hos bot-ene.

Den første analysen var den enkleste: Å se hva slags videoer som dukker når en helt nylaget konto i Russland og Ukraina tar sine første sveip. Som nevnt fungerer algoritmen til TikTok slik at den starter å servere deg videoer, og så tilpasser den hvilke nye videoer som dukker opp ut ifra hvordan du reagerer på de første.

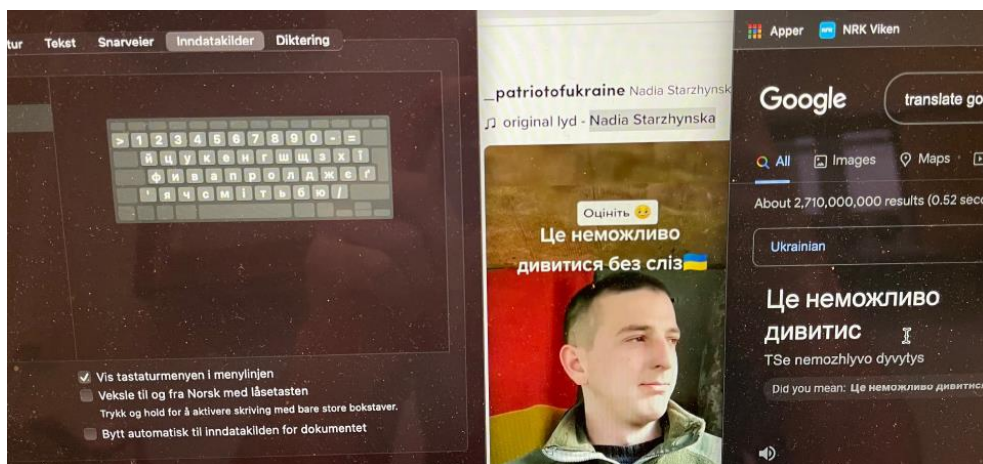
For å ikke legge noen føringer for hva vi ble servert i starten, hadde vi latt bot-ene se omtrent like lenge på alle videoer den første dagen. Vi la kun inn en viss randomisering for å ikke bli oppfattet som bot-er.

Vi tok deretter for oss de 25 første videoene fra hvert land og opprettet to nye kolonner i Excel-arket:

- Hva handler videoen om?
- Viser videoen krigshandlinger eller referer den til krigen i Ukraina?

Før vi kunne svare på det første spørsmålet, måtte vi løse et annet problem: Språkbarrieren. De fleste videoene var naturlig nok på russisk eller ukrainsk, noe ingen av oss behersket. Det gikk greit å bruke Google Translate når vi kunne lime inn teksten som fulgte med videoen.

Men ofte var budskapet skrevet på selve videoen. Dessverre ble ikke den automatiske OCR-transkriberingen vellykket, siden vi ikke hadde spesifisert at den skulle kjenne igjen kyrilliske bokstaver. Vi har i etterkant funnet ut hvordan det kan gjøres, men løsningen i kampens hete ble å skrive inn teksten i Google Translate manuelt. For å få til det måtte vi endre tastaturinnstillingene på datamaskinen til det kyrilliske alfabet og legge tre vinduer ved siden av hverandre: Et med Google Translate, et med teksten vi skulle oversette og et med et bilde av hvor på tastaturet de kyrilliske bokstavene var.



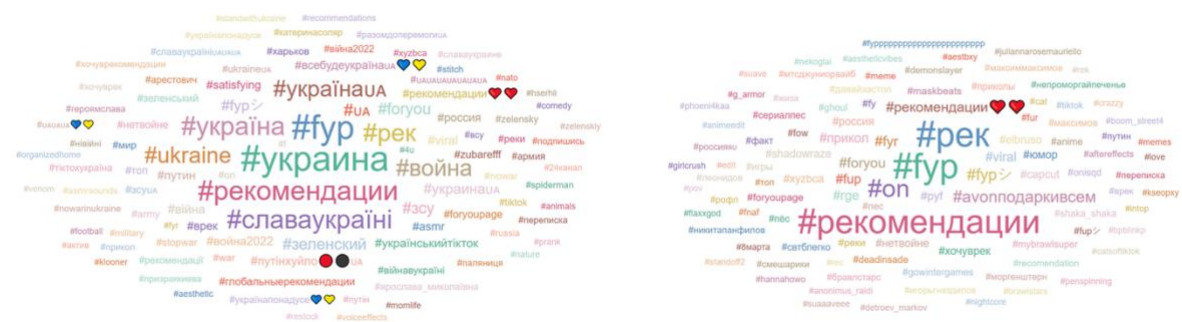
I de videoene der det kun var prating eller den oversatte teksten ikke ga mening for oss, måtte vi ha litt ekstra hjelp. Vi kom da i kontakt med Maria Pedko Lindbäck, som på det tidspunktet var praktikant i NRK Super. Hun har ukrainsk mor og snakker russisk. Dessuten kunne hun som ung TikTok-bruker hjelpe oss å forstå kulturelle nyanser som kunne forsvinne i oversettelsen. På et tidspunkt var hennes mor også med via FaceTime for å identifisere en ukrainsk vise som ble sunget av en soldat.

Når vi hadde en god oversikt over de 25 første videoene fra både den russiske og den ukrainske bot-en, kunne vi oppsummere starten:

- 18 av videoene som den ukrainske bot-en hadde sett handlet om krigen. Det var soldater, bomberom, utbrente stridsvogner og krigsfanger.
- 4 av videoene fra den russiske bot-en kunne knyttes opp mot krigen, og det på en langt mer abstrakt måte. Som at noen blandet blå og gul (det ukrainske flagget) og fikk rødt, hvitt og blått (det russiske flagget). Eller et klipp fra en tegneserie der fem veltrente menn går gjennom flere ulike verdener, ledsaget av teksten «POV: Når du og vennene dine går på tur om vinteren i Russlands nye territorier».

Det neste vi gjorde var å samle alle emneknagger for de neste 1300 videoene. Med Kibana lagde vi ordskyer, som er en visuell presentasjon der størrelsen på ordene øker jo hyppigere de forekommer. I den ukrainske ordskyen var det mange varianter av #україна (Ukraina),

#война (krig) og #зсу (Ukrainas forsvar). Den russiske ordskyen var derimot dominert av generiske TikTok-emneknagger som #viral, #fyp (For You Page) og ##рек (anbefalt).



Som tidligere nevnt forsøkte vi etter dette å «hjelp» den russiske bot-en til å se mer av krigen. Vi var derfor spente på om noe hadde endret seg når den hadde bladd gjennom 1000 nye videoer noen dager senere.

I Excel filtrerte vi ut alle videoer som bot-ene hadde sett lenger på enn normalt og la dem over i et nytt dokument. Deretter tok vi en manuell gjennomgang av videoene for å luke ut de tilfellene der bot-en hadde misforstått og sett lenge på ved en feil.

Resultatet var tydelig:

- Den ukrainske bot-en hadde stoppet 300 ganger for å se på blant annet soldater som flosset, raketter som ble skutt ut og demonstranter som ble arrestert.
- Den russiske bot-en fant derimot bare én krigshandling: En overvåkningsvideo som viste et missil treffe en bygning i Kharkiv. Klippet var delt av en russisk konto og inneholdt en gaming-referanse til å sette spill på pause.

Vi brukte også Excel-funksjonen Xoppslag for å kryssjekke hvilke videoer som begge bot-ene hadde sett. Det var kun 19 videoer – og ingen hadde noe med krigen å gjøre.

Deretter forsøkte vi via vår proxy i Russland å se videoene med krigshandlinger som den ukrainske bot-en hadde sett. Men det var ikke mulig. Altså viste ikke TikTok krigen til sine russiske brukere.

Dermed hadde vi dekning for å si at geografisk ligger ikke byene Belgorod og Kharkiv langt fra hverandre – men på TikTok lever de i to forskjellige verdener.

6. Historiefortelling og presentasjon

6.1 Hvordan fortelle en datahistorie?

Det er fint med mye data, men for at mange skal få det med seg må man også klare å formidle det på en god måte. Derfor tenkte vi på historiefortellingen helt fra starten av. Grunnen til at vi sendte bot-ene til nettopp Belgorod og Kharkiv var at vi hadde saumfart den russisk-ukrainske grensa i Google Maps for å finne krigsrammede byer i Ukraina som

hadde en russisk by like over grensa. Denne geografiske nærheten (bare 8 mil mellom byene) ga en sterkere nerve enn om vi hadde tatt for oss hele landet.

Vi valgte også å la vår egen arbeidsprosess spille en viktig rolle saken. På den måten kunne leseren være med og se hvordan vi bygget bot-ene, hva slags egenskaper vi ga dem, hvordan ting endret seg når vi endret innstillingene og hvor forskjellige videoer bot-ene Alexei og Nykolai fikk.

6.2 Navn og bilde på bot-ene

For å gjøre det nærere for leseren og lettere å henge med (og slippe å skrive «den ukrainske bot-en» og «den russiske bot-en» gang på gang), ga vi bot-ene navn og bilde. Navnene Alexei Fedorov og Nykolai Semenchenko er satt sammen fra en liste over de vanligste for- og etternavnene i de to landene.

Bildene er laget av kunstig intelligens via nettsiden generated.photos. De har et nettbasert program der man selv kan skrive inn hvilke parametere man vil ha, og så lager siden et unikt bilde av et menneske som ikke finnes. Dette er ikke uproblematisk etisk. Vi kunne også ha løst det med bilder av ekte mennesker, men vurderte at det ikke ville være riktig gjengivelse av journalistikken da det ikke er ekte menneskers TikTok-kontoer vi har brukt, men boter. Men selv om AI i stor grad kan lage portretter som ser helt ekte ut, var det viktig for oss at leserne umiddelbart forstod at dette ikke var ekte mennesker. Illustrasjonene er derfor laget slik at de ivaretar også presseetiske hensyn.

6.3 TikTok-universet

Det var et poeng at leseren skulle føle at han eller hun var i TikTok-universet. Derfor valgte vi å bruke en NRK-presentasjonsmal som kalles Visuell kortstokk, der man blar seg ett og ett kort nedover – akkurat som i TikTok-appen.

6.4 Verifisering av videoer

Vi fikk hjelp av Norunn Reinertsen fra researchsenteret i NRK til å verifisere videoer vi brukte i presentasjonen. Hun brukte OSINT-metoder (Open Source Intelligence) som for eksempel Google Streetview, nyhetskilder, solens plassering, metadata og andre metoder i arbeidet.

7. Etikk

7.1 Bruk av videoer fra TikTok

Vi ville vise leserne et utvalg av de faktiske videoene TikTok serverte til den russiske og ukrainske brukeren. Spesielt de aller første videoene bot-ene så, fordi vi allerede der så en enorm forskjell. Vi mener vi kan bruke nyhetsretten for å vise videoene, fordi det er svært viktig å få folk til å forstå hvordan algoritmene bidrar til å gi folk i Russland og Ukraina et ulikt bilde av krigen. Videoene var dessuten offentlig tilgjengelig for alle.

Det har imidlertid vært viktig å gå nøye gjennom hver video som skal vises, og vurdere hvor mye av blant annet ansikter, tekst og gjenkjennende trekk/klesplagg som vi kan vise. For eksempel viser en av de første videoene et barn som får på seg militærutstyr. Her har vi sladdet ansiktet, og dette har vi også gjort i mange av de andre videoene. Vi har gått nøye

gjennom alle videoene som vises i saken, også de som bare vises som små elementer i en montasje.

7.2 Falske brukere

Vi bryter TikToks retningslinjer når vi oppretter falske brukere og lager programvare som automatisk scroller gjennom TikTok. Vi gjorde det likevel, fordi vi mener det er viktig at noen kontrollerer og viser hvordan algoritmene fungerer. TikTok er dessuten en vesentlig nyhetskilde for spesielt unge mennesker – i vårt tilfelle unge menn som skal ta stilling til om de skal delta i en krig. Det er kritisk for demokratiet og samfunnet at innbyggerne, også i Norge, forstår hvordan algoritmene bestemmer hvilket innhold du får i feeden din.

7.3 Bruk av residential proxies

Residential proxies gjør det mulig å «lure» en nettside eller tjeneste til å tro at du er et annet sted i verden enn du egentlig er. Det er vanskeligere å gjennomskue residential proxies enn for eksempel enn VPN-løsning, fordi internett-trafikken går gjennom IP-adressen til en vanlig internettbruker. Dette skaper imidlertid også etiske utfordringer.

En internettbruker kan potensielt dele internett-trafikk med andre uten å vite om det. Det hadde vært svært uheldig dersom NRK bidro til noe slikt. Derfor var det viktig for oss å bruke en av de største leverandørene av slike tjenester, selv om dette også betød en av de dyreste (vi betalte 300 euro for en måneds tilgang). Oxylabs garanterer at alle som deler internett-trafikk vet om det, og har inngått en avtale med dem (lenke til håndbok [her](#)). Selskapet er lokalisert i Litauen.

8. Spesielle erfaringer

8.1 Web-versjonen av TikTok

Et ankepunkt mot metoden vår kan være at vi har brukt TikTok gjennom nettsiden [tiktok.com](#), og ikke via en mobil-app. De aller fleste bruker TikTok gjennom en app. Vi vet ikke om, og eventuelt hvordan, TikTok behandler brukere annerledes dersom de ser videoer gjennom nettsiden og ikke appen.

Til denne saken var imidlertid hovedpoenget å vise hvilket innhold TikTok serverer til helt nye brukere i Russland og Ukraina, og at videoer den ukrainske brukeren får servert er sperret i Russland. Vi har ingen grunn til å tro at vi ville fått vesentlig annerledes resultater dersom vi hadde basert undersøkelsen på mobilversjonen.

8.2 Jeg er ikke en robot

TikTok og andre sosiale medier har sofistikerte metoder for å hindre at bot-er herjer rundt på plattformene deres. Vi måtte derfor en rekke ganger justere metoden vi brukte for å styre nettleseren for å unngå å bli avslørt. Vi opplevde for eksempel at TikTok ga oss små puslespill som vi måtte løse for å sjekke om vi var bot-er. Etter hvert fant vi løsninger som gjorde at vi stort sett unngikk slike forsvarsmekanismer.

9. Konsekvenser

9.1 Ut i verden

Saken gikk viralt og ble oversatt av ulike brukere i hele verden. Derfor lagde vi også en engelsk versjon for å sikre at oversettelsen var korrekt. Det gjorde at vi nådde enda lenger ut, og fikk blant annet en anbefaling i nyhetsbrevet til New York Times.



Det at saken fikk så stor oppmerksomhet i utlandet førte også til at flere tok kontakt for å lære mer om hvordan vi jobbet eller tilby samarbeid. Vi har holdt foredrag for ARD, som er en sammenslutning av tyske allmennkringkastingsselskaper, blitt [intervjuet](#) av Global Investigative Journalism Network (GIJN) og fått tilbud om samarbeid med den europeiske NGO-en Tracking Exposed, som ser på hvordan algoritmer påvirker samfunnet.

9.2 Inspirerer flere norske journalister

Vi har opplevd at denne saken også har satt spor blant norske datajournalister, og vi har blitt invitert til å holde foredrag i datajournalistikknettverket MBC, Nordiske Mediedager og nå Data-Skup. Vår erfaring er at når vi i denne saken *viser* i stedet for å *snakke om* hvordan algoritmer styrer innholdet, så har det enorm kraft og bevisstgjør både kolleger og andre mennesker.

9.3 TikTok viser nye videoer

Siden publisering har TikTok i det stille begynt å vise nye videoer fra både russiske og internasjonale brukere – stikk i strid med det som har vært selskapets holdning utad. Vi laget en [oppfølger](#) om hvordan TikTok «skyggepromoterer» innhold som TikTok selv har sagt er forbudt. I juli annonserte de også [nye funksjoner](#) som skal hjelpe brukerne med å oppdage nye interesser sammen med et nytt sikkerhetssystem.