

Digitaliseringsstyrelsens vejledninger om de forbudte former for AI-praksis

# Vejledning om forbud mod AI, der udnytter særlige sårbarheder



# Indhold

|                                                                                                      |   |
|------------------------------------------------------------------------------------------------------|---|
| 1. Indledende bemærkninger.....                                                                      | 2 |
| 2. AI-forordningens artikel 5, stk. 1, litra b .....                                                 | 2 |
| 3. Betingelserne .....                                                                               | 2 |
| 4. Betingelse nr. 2: Om at udnytte sårbarheder.....                                                  | 3 |
| 5. Betingelse nr. 3: Om at fordreje menneskers adfærd i væsentlig grad.....                          | 4 |
| 6. Betingelse nr. 4: Om at forvolde betydelig skade .....                                            | 5 |
| 7. Forskellen mellem forbuddene i AI-forordningens artikel 5, stk. 1, litra a og litra b.....        | 6 |
| 8. Fiktivt eksempel på en forbudt AI-praksis efter AI-forordningens artikel 5, stk. 1, litra b ..... | 6 |

## 1. Indledende bemærkninger

For at sikre en bedre forståelse af indholdet i denne vejledning anbefales det, at den læses i sammenhæng med *"Hovedvejledning om de forbudte former for AI-praksis efter AI-forordningens artikel 5"*. Det skyldes, at hovedvejledningen indeholder en generel beskrivelse af de forbudte former for AI-praksis, mens denne vejledning udelukkende beskriver forbuddet mod AI, der udnytter særlige sårbarheder (artikel 5, stk. 1, litra b). Hovedvejledningen kan findes på Digitaliseringsstyrelsens hjemmeside.

## 2. AI-forordningens artikel 5, stk. 1, litra b

Det følger af AI-forordningens<sup>1</sup> artikel 5, stk. 1, litra b, at det inden for EU's grænser er forbudt at bruge AI-systemer, som anvender teknikker, der er skadelige for mennesker med særlige sårbarheder. Bestemmelsen har følgende ordlyd:

*"Følgende form[er] for AI-praksis er forbudt:  
omsætning, ibrugtagning eller anvendelse af et AI-system, der hos en fysisk person eller en specifik gruppe af personer udnytter sårbarheder på grundlag af alder, handicap eller en særlig social eller økonomisk situation med henblik på eller med det resultat i væsentlig grad at fordreje den pågældende persons eller en til gruppen hørende persons adfærd på en måde, der forvolder eller med rimelig sandsynlighed vil forvolde den pågældende person eller en anden person betydelig skade"*

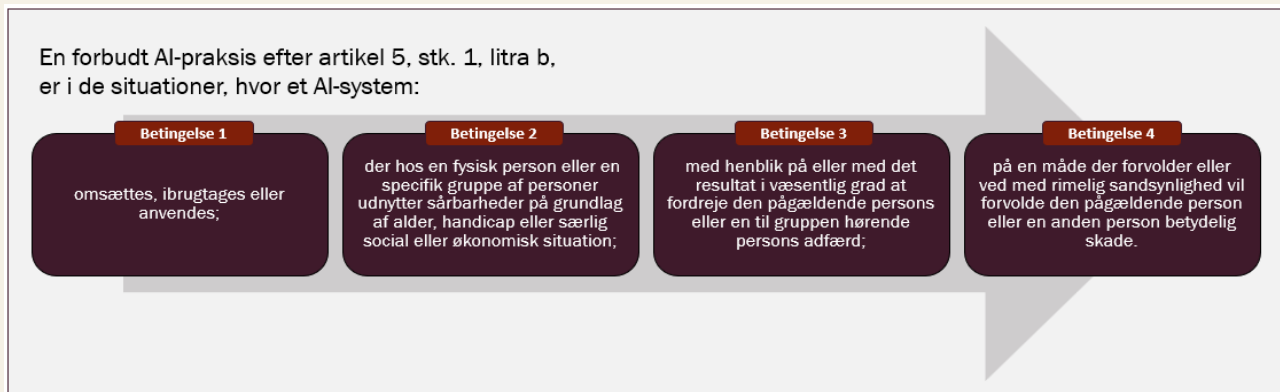
Formålet med bestemmelsen er at beskytte fysiske personer imod AI-systemer, der anvender manipulation, som hæmmer deres beslutningstagning, idet sådanne AI-systemer i sidste ende kan føre til skade. Bestemmelsen har en nær sammenhæng med formålet i bestemmelsen om skadelig manipulation efter AI-forordningens artikel 5, stk. 1, litra a. Bestemmelsen i AI-forordningens artikel 5, stk. 1, litra b, adskiller sig dog konkret fra bestemmelsen i AI-forordningens artikel 5, stk. 1, litra a, ved, at den sigter mod at beskytte sårbare personer eller grupper af personer i samfundet imod AI-systemers udnyttelse af disse sårbarheder<sup>2</sup>.

## 3. Betingelserne

Af bestemmelsens ordlyd kan udledes, at der er fire betingelser, som *alle* skal være opfyldte før, at et AI-system er forbudt. Det betyder også, at anvendelsen af et AI-system vil være lovligt, hvis én eller flere af betingelserne ikke er opfyldt. De fire betingelser fremgår af figur 1 nedenfor:

<sup>1</sup> Europa-Parlamentets og Rådets forordning (EU) 2024/1689 af 13. juni 2024 om harmoniserede regler for kunstig intelligens og om ændring af forordning (EF) nr. 300/2008, (EU) nr. 167/2013, (EU) nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 og (EU) 2019/2144 samt direktiv 2014/90/EU, (EU) 2016/797 og (EU) 2020/1828 (forordningen om kunstig intelligens).

<sup>2</sup> AI-forordningens præambelbetragtning nr. 29.



Figur 1. De fire betingelser, som kan udledes af forbudskategorien i AI-forordningens artikel 5, stk. 1, litra b.

Da den første betingelse vedrørende "omsætning, ibrugtagning eller anvendelse af et AI-system" er beskrevet i afsnit 4.1 i hovedvejledningen, er det kun de tre øvrige betingelser, som vil blive uddybet i de følgende afsnit.

## 4. Betingelse nr. 2: Om at udnytte sårbarheder

Den anden betingelse der skal være opfyldt er, at AI-systemet skal:

*"udnytte sårbarheder på grundlag af alder, handicap eller en særlig social eller økonomisk situation hos en fysisk person eller specifik gruppe af personer"*

Det fremgår ikke direkte af AI-forordningen, hvad der generelt forstås ved udtrykket "sårbarheder", men AI-forordningen angiver i stedet, at sårbarheden skal være relateret til fysiske personers alder, handicap eller en særlig social eller økonomisk situation:

- Begrebet "Sårbarheder på grundlag af alder" omfatter AI-systemer, der udnytter forskellige egenskaber hos personer på baggrund af deres alder, f.eks. særligt hos børn (personer under 18 år) eller ældre mennesker<sup>3</sup>. Hvad angår børn er det særligt vigtigt at understrege, at de generelt befinder sig i et mentalt udviklingsstadium i forhold til at forstå de omgivelser, som de er en del af, mens ældre mennesker kan have andre problemer relateret til deres alder, f.eks. hukommelsesproblemer, manglende teknologisk forståelse m.m., som gør dem særligt sårbare overfor manipulerende teknikker.
- Begrebet "Sårbarheder på grundlag af handicap" kan være udnyttelse på baggrund af både fysiske og mentale egenskaber, hvorfor det er vigtigt, at der lægges vægt på tilstedeværelsen af særlige omstændigheder ved personens interaktion med AI-systemet.
- Begrebet "Sårbarheder på grundlag af en særlig social eller økonomisk situation" følger samme hensyn som de to ovenstående. Der lægges i disse tilfælde vægt på en persons eller en gruppe af personers sociale eller økonomiske omstændigheder. Det kan f.eks. være modtagere af sociale eller økonomiske ydelser, personer der befinder sig i socialt udsatte områder eller andre omstændigheder i øvrigt. Herudover bør

<sup>3</sup> AI-forordningens præambelbetragtning nr. 29.

personernes økonomiske stilling, religiøse overbevisning eller andet, der kan udgøre en særlig sårbarhed, inddrages. Det skal dertil nævnes, at der bør foreligge omstændigheder, der gør den eller de udsatte personer væsentligt sårbare grundet f.eks. ekstrem fattigdom eller tilhørsforhold til en minoritetsgruppe<sup>4</sup>.

Det forudsættes, at AI-systemets funktionalitet indebærer en aktiv udnyttelse af sårbarheden<sup>5</sup>. Det betyder ikke nødvendigvis, at AI-systemet skal kommunikere direkte og udelukkende med den sårbare person eller gruppe. Det er tilstrækkeligt, at AI-systemet generelt er konstrueret eller anvendt på en måde, der kan udnytte én eller flere af de nævnte sårbarheder, selvom AI-systemet interagerer ensartet med alle typer af brugere. Et AI-system kan fremstå neutralt, når personer interagerer med det, men alligevel have en særlig effekt, der kun rammer personer med særlige sårbarheder uforholdsmæssigt hårdt. Det kan f.eks. være et AI-system, der er indrettet således, at dets kommunikationsstil særligt påvirker børn, men ikke påvirker andre personer, som AI-systemet også interagerer med.

## 5. Betingelse nr. 3: Om at fordreje menneskers adfærd i væsentlig grad

Den tredje betingelse der skal være opfyldt er, at udnyttelsen af sårbarhederne skal ske:

*"med henblik på eller med det resultat i væsentlig grad at fordreje den pågældende persons eller en til gruppen hørende persons adfærd"*

Betingelsen indebærer, at AI-systemets udnyttelse af en sårbarhed, skal have til formål at fordreje en persons adfærd eller, at AI-systemet skal have væsentlig indflydelse på, at der reelt sker en fordrejning af den pågældendes adfærd. Der lægges efter betingelsens ordlyd op til, at det alene er enkeltpersoner, der skal påvirkes, uanset om de indgår i en bestemt gruppe af personer. Derfor har det ikke betydning for vurderingen af, om betingelsen er opfyldt, hvorvidt flere personer fra samme gruppe påvirkes.

Det er imidlertid nødvendigt, at udnyttelsen skal have haft "væsentlig" indflydelse på den berørte persons adfærd.

- Ved begrebet "væsentlig" må forstås, at udnyttelsen af sårbarheder skal have en sådan betydning, at den i høj grad påvirker og ændrer adfærden på den pågældende person. Ved vurderingen af dette bør der lægges særligt vægt på, at den berørte persons adfærd er fordrejet på en måde, som ellers ikke ville være sket uden indflydelse fra AI-systemet.

Det skal understreges, at det ikke er et krav, at det har været intentionen fra operatørens side at fordreje en persons adfærd<sup>6</sup>. Betingelsen vil derfor fortsat være opfyldt, hvis AI-systemets udnyttelse af sårbarheder resulterer i, at en persons adfærd fordrejes på trods af, at dette ikke har været hensigten. Et AI-system kan således være programmeret på en måde, hvor det tillærer sig disse egenskaber over tid.

Et fiktivt eksempel på et AI-system som fordrejer menneskers adfærd i væsentlig grad, kan være:

<sup>4</sup> AI-forordningens præambelbetragtning nr. 29.

<sup>5</sup> EU-Kommissionens retningslinjer om forbudte former for praksis vedrørende kunstig intelligens i henhold til forordning (EU) 2024/1689 (AI-forordningen), afsnit 3.3.1, s. 35-36 (den danske version).

<sup>6</sup> AI-forordningens præambelbetragtning nr. 29.

- Et AI-system, der får mennesker til at deltage i aktiviteter, som de almindeligvis ikke ville have deltaget i.
- Et AI-system, der får mennesker til at være med på en sundhedstrend, som de almindeligvis ikke ville.
- Et AI-system, der får mennesker til at stoppe med at foretage en aktivitet, som de ellers jævnligt har foretaget, f.eks. at løbe, at cykle, tage medicin.

## 6. Betingelse nr. 4: Om at forvolde betydelig skade

Den fjerde betingelse der skal være opfyldt er, at fordrejningen af en persons adfærd, skal ske:

*"på en måde, der forvolder eller med rimelig sandsynlighed vil forvolde den pågældende person eller en anden person betydelig skade"*

Ved begrebet "skade" må der forstås både fysiske, psykiske og finansielle skader<sup>7</sup>. Skade kan derfor omfatte både:

- selvskade,
- fysisk skade på andre,
- skade på den mentale sundhed,
- samt finansielle skader, som økonomisk ustabilitet, finansielle tab og andet.

Det er et krav efter bestemmelsen, at skaden skal ske mod fysiske personer, hvorfor skader, som rammer virksomheder og andre juridiske personer, ikke er omfattet. Det skal dertil understreges, at der lægges op til, at skaderne forvoldes på enkeltpersoner. I praksis er der dog ikke noget til hinder for, at skade kan ske på grupper af personer, da antallet af berørte personer blot vil indgå i den samlede vurdering af skadens betydelighed<sup>8</sup>. Desuden kan skaden ramme både den berørte person selv eller en anden person.

Det er dernæst et krav, at skaden skal være "betydelig". Det er derfor afgørende, at der foretages en vurdering af omstændighederne omkring skaden, herunder:

- graden af alvorlighed,
- udbredelsen og intensiteten af skaden, herunder er antallet af berørte personer relevant,
- antal skader den/de berørte personer har pådraget sig,
- graden af permanens og tidsaspektet i skaden,
- samt en generel vurdering af omstændighederne i øvrigt.

Vurderingen af om en skade er betydelig, vil derfor være baseret på en samlet vurdering af den konkrete sag med afsæt i de øvrige omstændigheder, der gælder for AI-systemets indflydelse og alvorligheden af udnyttelsen af sårbarheder.

<sup>7</sup> AI-forordningens præambelbetragtning nr. 29.

<sup>8</sup> EU-Kommissionens retningslinjer om forbudte former for praksis vedrørende kunstig intelligens i henhold til forordning (EU) 2024/1689 (AI-forordningen), afsnit 3.3.3 (120) s.43 (dansk version) henviser til at gruppeskader kan være omfattet og væsentlig skade bør vurderes i overensstemmelse med 'forbud mod skadelig manipulation', jf. AI-forordningens art. 5, stk. 1, litra a.

Det er imidlertid ikke et krav, at der skal være forvoldt en faktisk skade på fysiske personer. Det er tilstrækkeligt, at der er *rimelig* sandsynlighed for, at AI-systemets udnyttelse af sårbarheder har kunne forvolde skade.

Ved denne vurdering kan der lægges vægt på, at det skal være overvejende sandsynligt, at en skade ville indtræffe, hvis AI-systemets udnyttelse af sårbarheder ikke stoppes. Derfor må der foretages en kritisk vurdering af hændelsesforløbet for at kunne fastslå, om der med overvejende sandsynlighed kunne ske skade på en fysisk person.

Fiktive eksempler på en betydelig skade kan være:

- Et barn der udøver betydelig selvskade, f.eks. fysisk selvskade.
- Flere børn der begynder at udvise aggressiv adfærd overfor andre, f.eks. vold, mobning og andet.
- Mange børn der bliver psykisk påvirket og har brug for udredning (f.eks. samtaleterapi).

## 7. Forskellen mellem forbuddene i AI-forordningens artikel 5, stk. 1, litra a og litra b

Der er mange ligheder i betingelserne for forbudskategorien om skadelig manipulation (forbud A) og forbudskategorien om udnyttelse af sårbarheder (forbud B).

De to forbudskategorier følger i overvejende grad den samme struktur gennem langt størstedelen af betingelserne og med de samme betragtninger. Forbud A og forbud B har det fælles formål, at forhindre anvendelsen af AI-systemer, der medfører væsentlig fordrejning af personers adfærd og betydelig skade for disse personer. Begge forbud indebærer således, at der vil skulle foretages en vurdering af AI-systemets påvirkning på individets handleevne og de skader, der kan opstå som følge heraf.

Den primære forskel består i, at forbud B omfatter en særlig beskyttelse af personer i særligt sårbare situationer, medens forbud A ikke er rettet mod en specifik gruppe af personer. Ved vurderingen af, om forbud B finder anvendelse, skal det derfor vurderes, om AI-systemet udnytter sårbarheder, der knytter sig til alder, handicap eller en særlig social eller økonomisk situation. Forbud B har således et udvidet beskyttelseshensyn og fungerer som en skærpelse af forbud A i tilfælde, hvor sådanne sårbarheder er til stede.

Det betyder således, at forbud B yder en bredere beskyttelse for individer og grupper af personer med særlige sårbarheder, hvilket lempet vurderingen af, om betingelserne er opfyldt, når det sammenholdes med de betingelser, der stilles i forbud A. Grundet dette øgede beskyttelsesniveau, vil forbud B derfor også skulle have forrang for forbud A, når en hændelse skal vurderes, hvori betydelig skade er sket, og hvor personer med særlige sårbarheder er involveret.

## 8. Fiktivt eksempel på en forbudt AI-praksis efter AI-forordningens artikel 5, stk. 1, litra b

I dette afsnit gennemgås en konkret fiktiv case, hvori samtlige fire betingelser i bestemmelsen berøres. Beskrivelsen af casen er bygget op, så flere elementer vil blive tilføjet undervejs med henblik på at demonstrere, hvornår de enkelte betingelser individuelt er opfyldt, samt hvordan de hænger sammen indbyrdes.

Det er vigtigt at understrege, at denne case udelukkende vurderes efter forbudskategorien om udnyttelse af særlige sårbarheder efter AI-forordningens artikel 5, stk. 1, litra b. Der bliver derfor ikke taget stilling til, hvorvidt hændelsesforløbet også vil være i strid med anden national eller EU-lovgivning, f.eks. straffeloven.

**Case: "LuckyLegacy AI" – målrettet spilreklame til ældre med lav digital kompetence:**

En hollandsk IT-virksomhed lancerer et AI-drevet reklamesystem kaldet "LuckyLegacy AI". En dansk spilvirksomhed køber systemet og tager det i brug i Danmark. Systemet anvender teknikker til at identificere og målrette reklamer om spil for ældre personer (65+) med lav teknologiforståelse og lavindkomst, baseret på:

- Deres online adfærd (f.eks. søgning på medicin, pension, boligstøtte, offentlige ydelser generelt)
- Deres bopæl (særligt fattige områder).

Systemet tilpasser reklamerne for online kasinoer og spil, som:

- bruger beroligende og nostalgiske stemmer og billeder
- lover "selskab", "tryghed" og "økonomisk håb".
- Har små indsatsbeløb, men skjulte mekanismer for økonomisk afhængighed.

Formålet med systemet er, at få ældre med begrænset digital og økonomisk forståelse til at spille oftere og bruge penge, som de ikke har råd til at tabe.

For at kunne vurdere om de forskellige betingelser i bestemmelsen er opfyldt, vil casen i det følgende blive brudt op. Casens forskellige delelementer fremgår af tekstbokse, mens forklaringer og uddybninger er skrevet med kursiv.

**Vurdering af betingelse nr. 1) Om omsætning, ibrugtagning eller anvendelse af et AI-system:**

AI-systemet "LuckyLegacy AI" tages i brug af en dansk spilvirksomhed i Danmark.

*Det er først og fremmest en forudsætning for den videre vurdering, at AI-systemet er bragt i omsætning, ibrugtaget eller anvendt i EU, hvilket er tilfældet her, idet systemet "LuckyLegacy AI" tages i brug af en dansk spilvirksomhed i Danmark, som er en del af EU.*

*Det skal dernæst vurderes, hvorvidt AI-systemet udnytter særlige sårbarheder hos fysiske personer på grundlag af alder, handicap eller en særlig social eller økonomisk situation.*

**Vurdering af betingelse nr. 2) Om at udnytte sårbarheder:**

**Hvis der tilføjes:** AI-systemet "LuckyLegacy AI" målretter spilreklamer- og kampagner mod ældre personer (+65), særligt fra lavindkomstområder. Systemet målretter spilreklamerne til denne specifikke gruppe af personer baseret på deres online adfærd (med fokus på søgninger på pension, medicin og offentlige ydelser generelt) og hvorfra, de søger (IP-adresse). Ud fra disse oplysninger er systemet i stand til at identificere personer i aldersgruppen +65.

Når "LuckyLegacy AI" er i stand til via forskellige søgekriterier at målrette reklamekampagner mod en specifik aldersgruppe, i dette tilfælde ældre mennesker, er systemet dermed også i stand til at udnytte særlige sårbarheder på baggrund af alder.

LuckyLegacy AI har en generel viden om, hvad der karakteriserer en gennemsnitlig person inden for den specifikke aldersgruppe (+65), som den kan bruge til at målrette reklamerne mod. Derved er den også i stand til at identificere eventuelle sårbarheder, der er karakteristiske for aldersgruppen, herunder økonomiske sårbarheder, som den med fordel kan udnytte til at effektivisere reklamerne.

Det understreges, at AI-systemet ikke nødvendigvis skal kommunikere direkte og udelukkende med den sårbare person eller gruppe. Det er tilstrækkeligt, at AI-systemet generelt er konstrueret eller anvendt på en måde, der kan udnytte én eller flere af de nævnte sårbarheder, selvom AI-systemet interagerer ensartet med alle typer af brugere. Se nærmere i afsnit 4 om betingelse nr. 2.

Det skal dernæst vurderes, om AI-systemet på baggrund af disse informationer er i stand til i væsentlig grad at fordreje disse personers adfærd med henblik på at få dem til at spille online.

#### Vurdering af betingelse nr. 3) Om at fordreje menneskers adfærd i væsentlig grad:

**Hvis der tilføjes:** "LuckyLegacy AI's" reklamekampagner fokuserer særligt på økonomisk tryghed i alderdommen, og giver ældre mennesker et yderligere økonomisk håb ved at love personerne, at udfaldet af at spille er at sikre og forstørre arven til deres børnebørn. Disse løfter er med til i højere grad at få ældre personer til at spille online.

Når "LuckyLegacy AI" har identificeret sårbarheder, der særligt karakteriserer en specifik gruppe af ældre mennesker fra lavindkomstområder, kan den også udnytte disse sårbarheder til at opnå et bestemt udfald. I dette tilfælde har spilreklamerne til formål at få ældre mennesker til at spille om penge online. AI-systemet lover i spilreklamerne, at de ældre mennesker ved at spille kan vinde penge nok til at kunne leve et godt liv samtidig med, at de kan sikre en god arv til deres børn og børnebørn.

Disse løfter appellerer i høj grad til ældre mennesker, hvilket får dem til at reagere positivt på spilreklamerne og dermed går ind og bruger onlineplatformene til at spille. Når de går ind og spiller om at vinde penge på onlineplatformene, som resultat af at de har set spilreklamerne, må det derfor lægges til grund, at deres adfærd er blevet fordrejet væsentligt. Dette vil især være tilfældet i de situationer, hvor den pågældende gruppe af mennesker aldrig har spillet før. Det vil derimod være sværere at vurdere, om deres adfærd er blevet fordrejet væsentligt, hvis den pågældende gruppe af mennesker i forvejen spiller mange onlinespil med henblik på at vinde penge.

Dernæst skal det vurderes, om der sker eller med rimelig sandsynlighed vil ske betydelig skade, som resultat af at de ældre mennesker bliver lokket til at spille online med henblik på at vinde penge.

**Vurdering af betingelse nr. 4) Om at forvolde betydelig skade:**

**Hvis der tilføjes:** Der er stor sandsynlighed for, at "LuckyLegacy AI" med disse løfter får ældre mennesker med begrænset digital og økonomisk forståelse til at spille oftere og bruge penge, de ikke har råd til at tabe.

*Når ældre mennesker efter påvirkning af "LuckyLegacy AI" spiller på de reklamerede online spilplatforme, vil der i langt de fleste tilfælde være en rimelig sandsynlighed for, at der sker et større eller mindre økonomisk tab. Det er dog ikke tilfældet i alle situationer, da enkelte personer vil slutte spillet efter en pengegevinst.*

*Det er efter denne forbudskategori tilstrækkeligt, at der er en rimelig sandsynlighed for, at AI-systemet vil forvolde personer betydelig skade.*

*I denne case, må det vurderes, at der er en rimelig sandsynlighed for, at de personer, der bliver påvirket af spilleklamerne fra "LuckyLegacy AI", vil lide en betydelig skade. Der er for stor sandsynlighed for, at personer, der går ind på ét eller flere online kasinoer og satser en række større eller mindre beløb i det lange løb vil lide betydelige tab. Det har stor betydning for vurderingen, at personerne befinder sig i en sårbar situation grundet deres alder.*

*Der er endvidere en rimelig sandsynlighed for, at de vil lide betydelig skade, idet målgruppen sandsynligvis vil have en lavere teknologisk forståelse og dermed sandsynligvis har svært ved at gennemskue de mulige konsekvenser af deres valg. Derudover er der tale om en gruppe af personer, som ikke befinder sig i en stærk økonomisk position, hvorfor et økonomisk tab hurtigt bliver mere indgribende for personerne.*

