



Pseudonymisering og anonymisering i praksis

En speed date med ulike teknikker

Normkonferansen 2026, 23. april

Robindra Prabhu // data scientist // Teknologiavdelingen

Sara Elise Wøllo // data scientist // Teknologiavdelingen



Det er ikke alltid så lett å vite hva
man skal gjøre...



Fødselsnummer	Navn	Alder	Bosted	Stilling	Favorittkake	Guilty pleasure
11223344556	Rita	28	Trondheim	Jurist	Sitronkake	Love Island (TV2)
12312312312	Per	39	Oslo	Jurist	Eplekake	Jakten på Kjærligheten (TV2)
32323232323	Michael	35	Kristiansand	Utvikler	Skolebolle	Mesternes Mester (NRK1)
99999999999	Emily	43	Molde	Jurist	Brownies	Farmen Kjendis (TV2)
09090909090	Sara	58	Bergen	Data scientist	Muffins	Luksusfellen (TV3)
84758475847	Nora	63	Alta	Utvikler	Kringle	71° nord (TVNorge)
39393939393	Pierre	38	Tromsø	Utvikler	Croissant	Forræder (TV2)
98877665544	Robindra	55	Gjøvik	Data scientist	Gulrotkake	Bakemesterskapet (NRK)
09876543211	Iris	49	Leikanger	Data scientist	Kanelboller	Big Brother (TVNorge)
77788776676	Jacob	33	Bodø	Data scientist	Eplekake	Sommerhytta (TV2)

Pseudonymisering

“ «pseudonymisering» [er] behandling av personopplysninger på en slik måte at personopplysningene ikke lenger kan knyttes til en bestemt registrert **uten bruk av tilleggsopplysninger**, forutsatt at nevnte tilleggsopplysninger lagres atskilt og omfattes av tekniske og organisatoriske tiltak som sikrer at personopplysningene ikke kan knyttes til en identifisert eller identifiserbar fysisk person

- Personopplysningsloven Artikkel 4, definisjoner nr. 5



Fødselsnummer	Navn	Alder	Bosted	Stilling	Favorittkake	Guilty pleasure
11223344556	Rita	28	Trondheim	Jurist	Sitronkake	Love Island (TV2)
12312312312	Per	39	Oslo	Jurist	Eplekake	Jakten på Kjærligheten (TV2)
32323232323	Michael	35	Kristiansand	Utvikler	Skolebolle	Mesternes Mester (NRK1)
99999999999	Emily	43	Molde	Jurist	Brownies	Farmen Kjendis (TV2)
09090909090	Sara	58	Bergen	Data scientist	Muffins	Luksusfellen (TV3)
84758475847	Nora	63	Alta	Utvikler	Kringle	71° nord (TVNorge)
39393939393	Pierre	38	Tromsø	Utvikler	Croissant	Forræder (TV2)
98877665544	Robindra	55	Gjøvik	Data scientist	Gulrotkake	Bakemesterskapet (NRK)
09876543211	Iris	49	Leikanger	Data scientist	Kanelboller	Big Brother (TVNorge)
77788776676	Jacob	33	Bodø	Data scientist	Eplekake	Sommerhytta (TV2)

Direkte identifiserende		Indirekte identifiserende			Konfidensielt	
Fødselsnummer	Navn	Alder	Bosted	Stilling	Favorittkake	Guilty pleasure
11223344556	Rita	28	Trondheim	Jurist	Sitronkake	Love Island (TV2)
12312312312	Per	39	Oslo	Jurist	Eplekake	Jakten på Kjærligheten (TV2)
32323232323	Michael	35	Kristiansand	Utvikler	Skolebolle	Mesternes Mester (NRK1)
99999999999	Emily	43	Molde	Jurist	Brownies	Farmen Kjendis (TV2)
09090909090	Sara	58	Bergen	Data scientist	Muffins	Luksusfellen (TV3)
84758475847	Nora	63	Alta	Utvikler	Kringle	71° nord (TVNorge)
39393939393	Pierre	38	Tromsø	Utvikler	Croissant	Forræder (TV2)
98877665544	Robindra	55	Gjøvik	Data scientist	Gulrotkake	Bakemesterskapet (NRK)
09876543211	Iris	49	Leikanger	Data scientist	Kanelboller	Big Brother (TVNorge)
77788776676	Jacob	33	Bodø	Data scientist	Eplekake	Sommerhytta (TV2)

Indirekte identifiserende

Konfidensielt

ID	Navn	Alder	Bosted	Stilling	Favorittkake	Guilty pleasure
01292-A36D	Reidar	28	Trondheim	Jurist	Sitronkake	Love Island (TV2)
86949-B32C	Pernille	39	Oslo	Jurist	Eplekake	Jakten på Kjærligheten (TV2)
29874-I45F	Michael	35	Kristiansand	Utvikler	Skolebolle	Mesternes Mester (NRK1)
34116-P49T	Emil	43	Molde	Jurist	Brownies	Farmen Kjendis (TV2)
98891-B12U	Sara	58	Bergen	Data scientist	Muffins	Luksusfellen (TV3)
76894-Y55M	Nora	63	Alta	Utvikler	Kringle	71° nord (TVNorge)
46786-K98S	Per	38	Tromsø	Utvikler	Croissant	Forræder (TV2)
00282-V81L	Robindra	55	Gjøvik	Data scientist	Gulrotkake	Bakemesterskapet (NRK)
08845-Z35G	Iver	49	Leikanger	Data scientist	Kanelboller	Big Brother (TVNorge)
69847-J49E	Johan	33	Bodø	Data scientist	Eplekake	Sommerhytta (TV2)

Når skal jeg pseudonymisere?

Når skal jeg ikke pseudonymisere?

Verdien ligger i forkledningen

Pseudonyme opplysninger er
fremdeles personopplysninger

Kravene i forordningen gjelder fortsatt

Hva gjør vi i Nav?

Mye data og bred portefølje av oppgaver

Hva gjør vi i Nav?

Mye data og bred portefølje av oppgaver

Skjema

i Dette er et pseudonymisert view

Name	Mode	Type	Description	Personopplysning ?
person_ident	NULLABLE	BYTES	Kan være indirekte identifiserende	SHA256 hashed
id	NULLABLE	STRING	Kan være indirekte identifiserende	
bruker_id	NULLABLE	STRING	Kan være indirekte identifiserende	
gjennomforing_id	NULLABLE	STRING	Kan være indirekte identifiserende	
start_dato	NULLABLE	DATE	Kan være indirekte identifiserende	
slutt_dato	NULLABLE	DATE	Kan være indirekte identifiserende	

Legg til nytt dataprodukt

Legg til ny datafortelling

Legg til nytt innsiktsprodukt

Koble sammen
pseudonymiserte tabeller

Mine produkter

Mine fortellinger

Mine innsiktsprodukter

Tilgangssøknader til meg

Mine tilgangssøknader

Mine tilganger

Mine team tokens

Min Knast **beta**

Anonymisering

WP 29: Er det mulig å...

1

...plukke/skille ut et individ?

2

...koble til andre data knyttet til et individ?

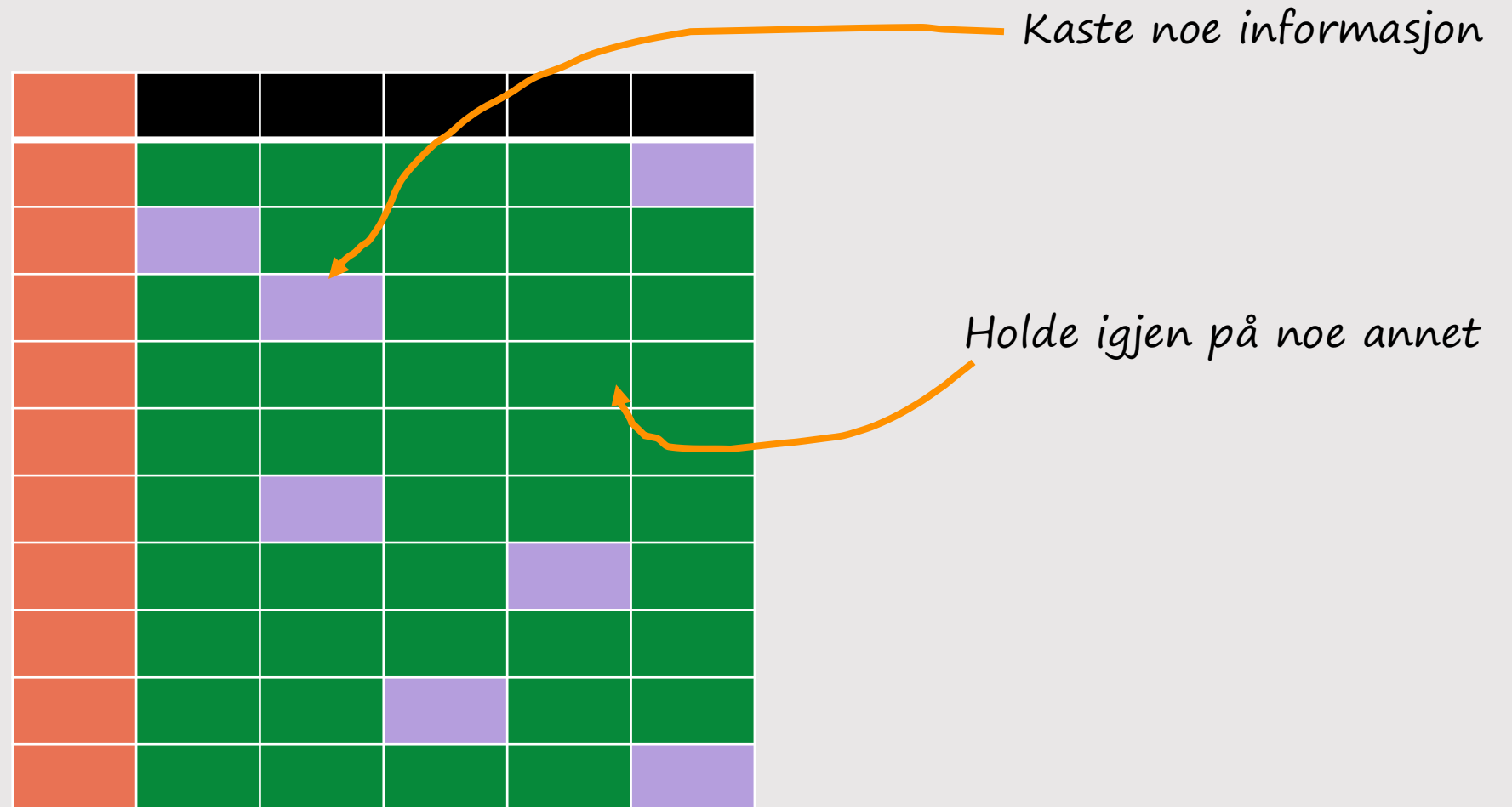
3

...avlede informasjon om et individ?



Pseudonymer		Indirekte identifiserende			Konfidensielt	
ID	Navn	Alder	Bosted	Stilling	Favorittkake	Guilty pleasure
		28	Trondheim	Jurist	Sitronkake	Love Island (TV2)
		39	Oslo	Jurist	Eplekake	Jakten på Kjærligheten (TV2)
		35	Kristiansand	Utvikler	Skolebolle	Mesternes Mester (NRK1)
		43	Molde	Jurist	Brownies	Farmen Kjendis (TV2)
		58	Bergen	Data scientist	Muffins	Luksusfellen (TV3)
		63	Alta	Utvikler	Kringle	71° nord (TVNorge)
		38	Tromsø	Utvikler	Croissant	Forræder (TV2)
		55	Gjøvik	Data scientist	Gulrotkake	Bakemesterskapet (NRK)
		49	Leikanger	Data scientist	Kanelboller	Big Brother (TVNorge)
		33	Bodø	Data scientist	Eplekake	Sommerhytta (TV2)

Syntaktisk (SDC) anonymisering



Indirekte identifiserende

Konfidensielt

Alder	Bosted	Stilling	Favorittkake	Guilty pleasure
28	Trondheim	Jurist	Sitronkake KAKE	Love Island (TV2)
39	Oslo	Jurist	Eplekake KAKE	Jakten på Kjærligheten (TV2)
35	Kristiansand	Utvikler	Skolebolle BAKVERK	Mesternes Mester (NRK1)
43	Molde	Jurist	Brownies KAKE	Farmen Kjendis (TV2)
58	Bergen	Data scientist	Muffins KAKE	Luksusfellen (TV3)
63	Alta	Utvikler	Kringle BAKVERK	71° nord (TVNorge)
38	Tromsø	Utvikler	Croissant BAKVERK	Forræder (TV2)
55	Gjøvik	Data scientist	Gulrotkake KAKE	Bakemesterskapet (NRK)
49	Leikanger	Data scientist	Kanelboller BAKVERK	Big Brother (TVNorge)
33	Bodø	Data scientist	Eplekake KAKE	Sommerhytta (TV2)

Alder	Bosted	Stilling	Favorittkake	Guilty pleasure
28	Trondheim	Jurist	KAKE	Love Island (TV2)
39	Oslo	Jurist	KAKE	Jakten på Kjærligheten (TV2)
35	Kristiansand	Teknolog	BAKVERK	Mesternes Mester (NRK1)
43	Molde	Jurist	KAKE	Farmen Kjendis (TV2)
58	Bergen	Teknolog	KAKE	Luksusfellen (TV3)
63	Alta	Teknolog	BAKVERK	71° nord (TVNorge)
38	Tromsø	Teknolog	BAKVERK	Forræder (TV2)
55	Gjøvik	Teknolog	KAKE	Bakemesterskapet (NRK)
49	Leikanger	Teknolog	BAKVERK	Big Brother (TVNorge)
33	Bodø	Teknolog	KAKE	Sommerhytta (TV2)

Indirekte identifiserende

Konfidensielt

Alder	Stilling	Favorittkake	Guilty pleasure
25-45	Jurist	KAKE	Love Island (TV2)
25-45	Jurist	KAKE	Jakten på Kjærligheten (TV2)
25-45	Jurist	KAKE	Farmen Kjendis (TV2)
30-60	Teknolog	BAKVERK	Mesternes Mester (NRK1)
30-60	Teknolog	BAKVERK	71° nord (TVNorge)
30-60	Teknolog	BAKVERK	Forræder (TV2)
30-60	Teknolog	BAKVERK	Big Brother (TVNorge)
30-60	Teknolog	KAKE	Luksusfellen (TV3)
30-60	Teknolog	KAKE	Sommerhytta (TV2)
30-60	Teknolog	KAKE	Bakemesterskapet (NRK)



$k_1 = 3$

$k_2 = 4$

$k_3 = 3$

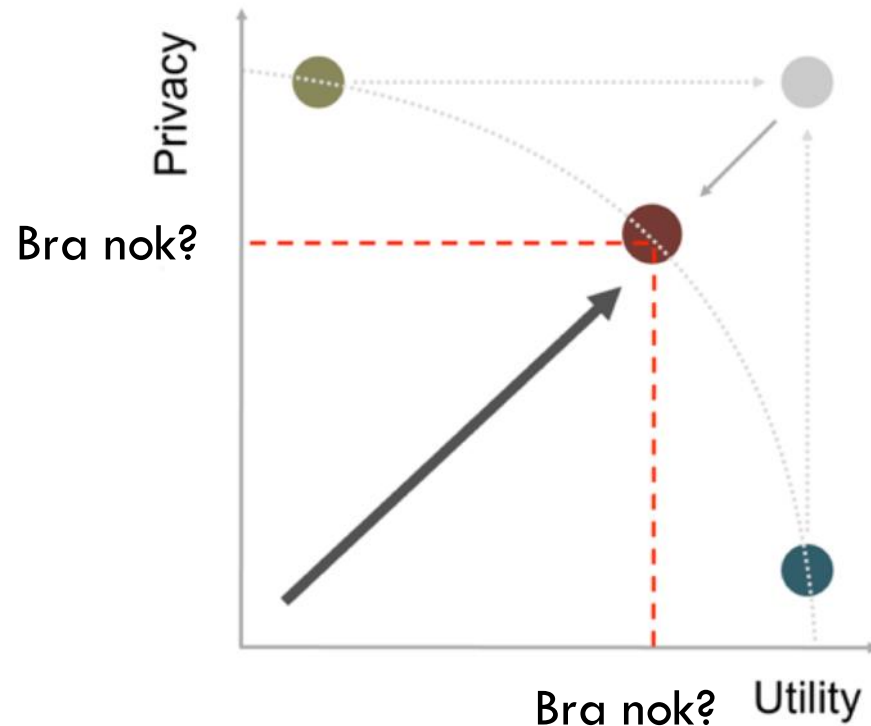
Re-identifiseringsrisiko i hver ekvivalensklasse:

$$\frac{1}{k}$$

Datasettet har blitt “3-anonymt”:

- Max risiko: 1/3
- Snittrisiko: 30,5 %

Solid anonymisering er krevende håndverk...



- Iterasjon: prøving og feiling...
 - Smartere teknikker krever spesialiserte verktøy
- Ingen *a-priori* garanti for at det er “bra nok”

Pros



Mye litteratur og etablert beste praksis.
Mange verktøy du kan anvende.



Tallfeste restrisiko.

Cons



Krever mye iterasjon – mange behandlinger!
Du vet ikke om det er godt nok
før du prøver.



Du endrer på datasettet, både i
innhold og struktur



Mange studier viser at
enkle teknikker ikke er gode nok

Differential privacy

“Vi kan verken bekrefte eller avkrefte at...”



Navn	Alder
Rita	28
Per	39
:	:
:	:
Kari	49
Jacob	33



Individ-
database



Per vet at Sara røyker.
Han ser statistikken
som relaterer røyking
og lungekreft.

Han konkluderer at
Sara har forhøyet risiko
for lungekreft.



Har Per avledet en
ny personopplysning
om Sara?



Har Per avledet en
ny personopplysning
om Sara?

Nei, hele poenget med statistikk
er jo å få innsikt i generelle
sannheter!

...men hvordan formelt skille
mellom “generelle sannheter” og
informasjon vi ikke bør avsløre?

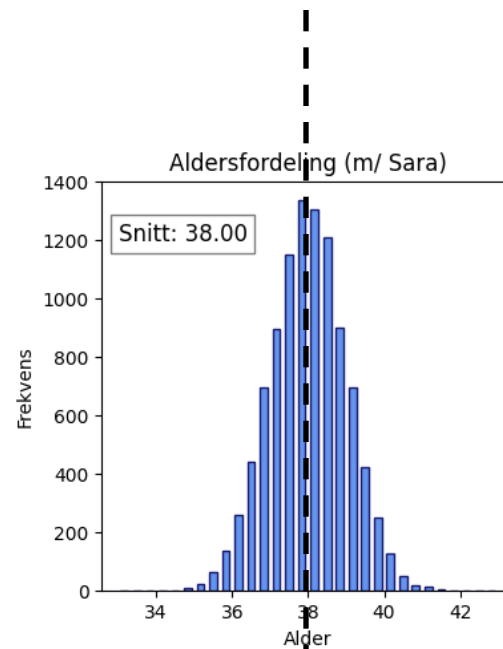
“ Du skal ikke kunne lære noe **om et individ**
fra den statistiske databasen

...som du ikke uansett kunne lært hvis
individet **ikke** var i databasen.



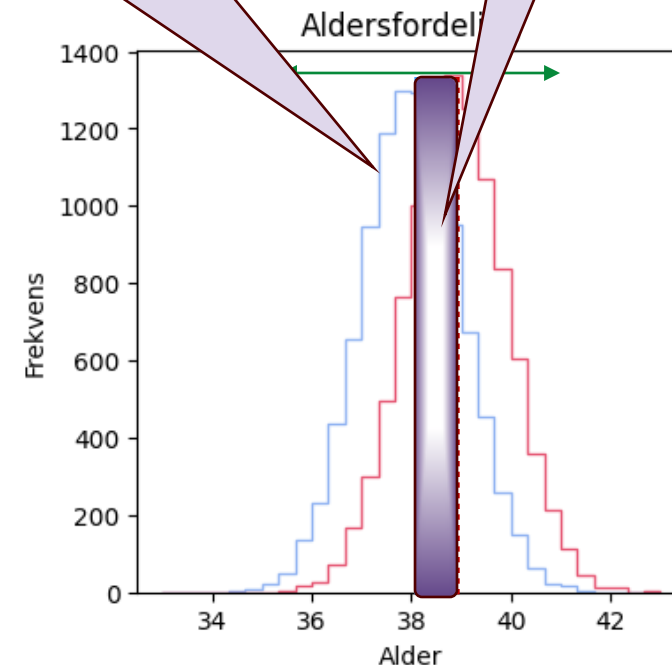
Sara er *med*

Navn	Alder
Rita	28
Per	39
:	:
Sara	28
:	:
Kari	49
Jacob	33



Legge på akkurat så mye randomisert støy at det er umulig å vite om Sara er med eller ikke!

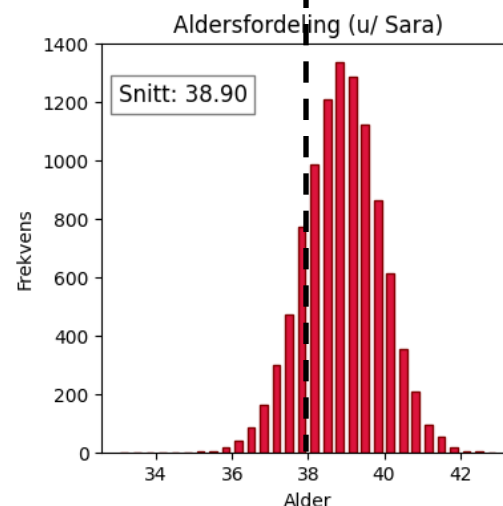
f.eks. 38.5



Svar () $\approx$ Svar ()

Randomisering beskytter den unike informasjonen som Sara (og alle andre individer) tilfører .

Navn	Alder
Rita	28
Per	39
:	:
:	:
Kari	49
Jacob	33



Sara er *ikke med*

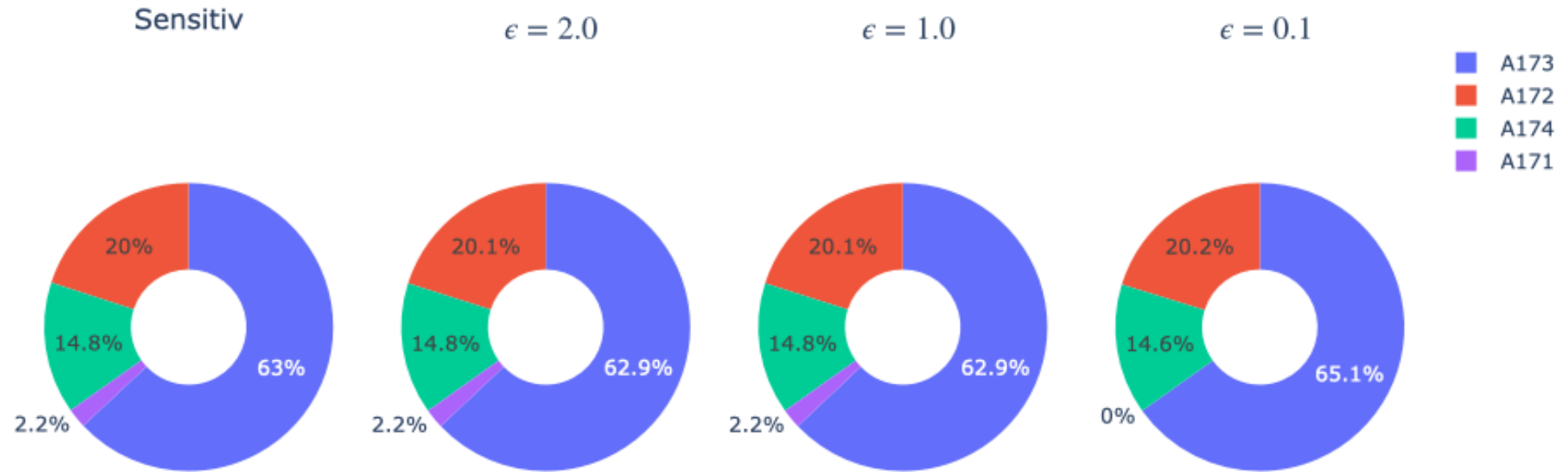


Mindre ϵ , mer støy,
mer beskyttelse



ϵ
(epsilon)

Mindre ϵ , mer støy, mer beskyttelse...

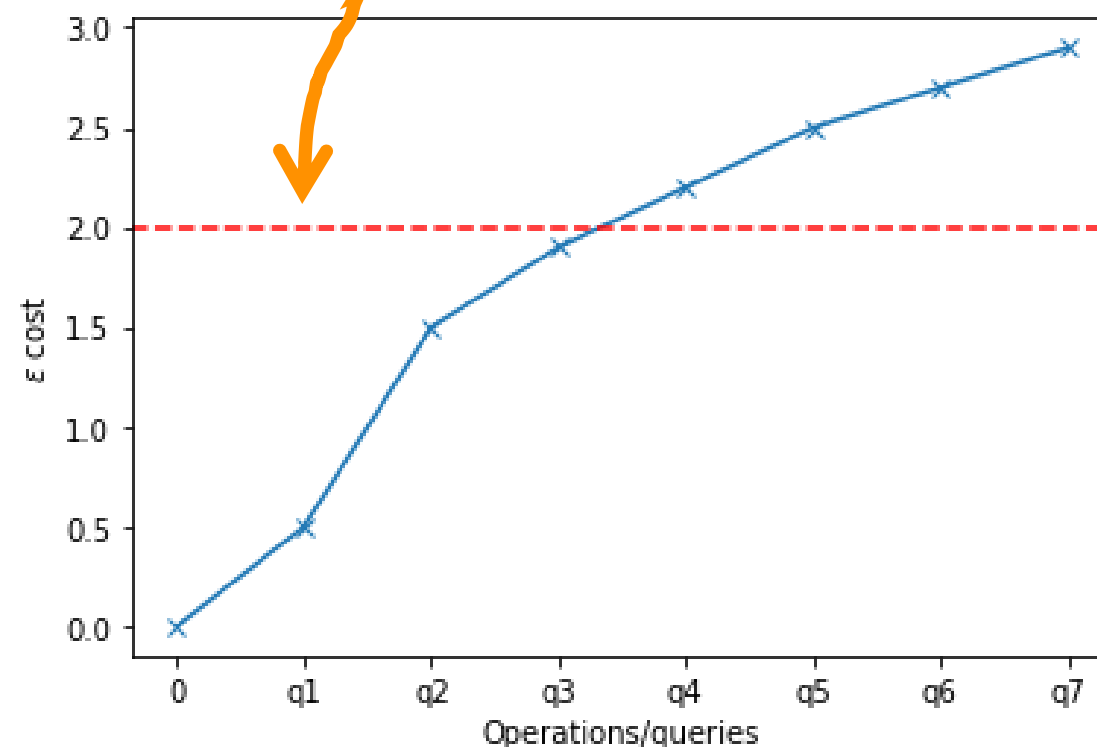


“Privacy budget” -

hjelper deg holde styr på hvor mye informasjon du lekker!

Eks: vi ønsker at statistikken vi deler skal ha en øvre grense $\epsilon < 2$

- **q1**: Snittalder
- **q2**: Snittvarighet
- **q3**: Standard avvik i varighet
- **q4**: Histogram av lånebeløp
- **q5**: Sum av lånebeløp over snitt
- **q6**: Andel av totale låneverdi gitt til "single women" (A95)
- **q7**: Andel av totale låneverdi gitt til "foreign worker" (A201)



Pros



Fint når du vil dele innsikt/statistikk
(og omtrentlige tall er ok)



Du bestemmer deg for ønsket risikonivå
før du går igang og får et “budsjett”.



Du trenger aldri se datasettet!
OK at datasettet endrer seg.



Sterke matematiske garantier for at
konfedensialitet er overholdt.

Cons



Kan være litt komplisert og forvirrende.
Krever øvelse. Lett å tråkke feil.



Usikkert om differentially private =
“anonymt”.

DP snakker ikke om re-
identifiseringsrisiko....det gjør PVL

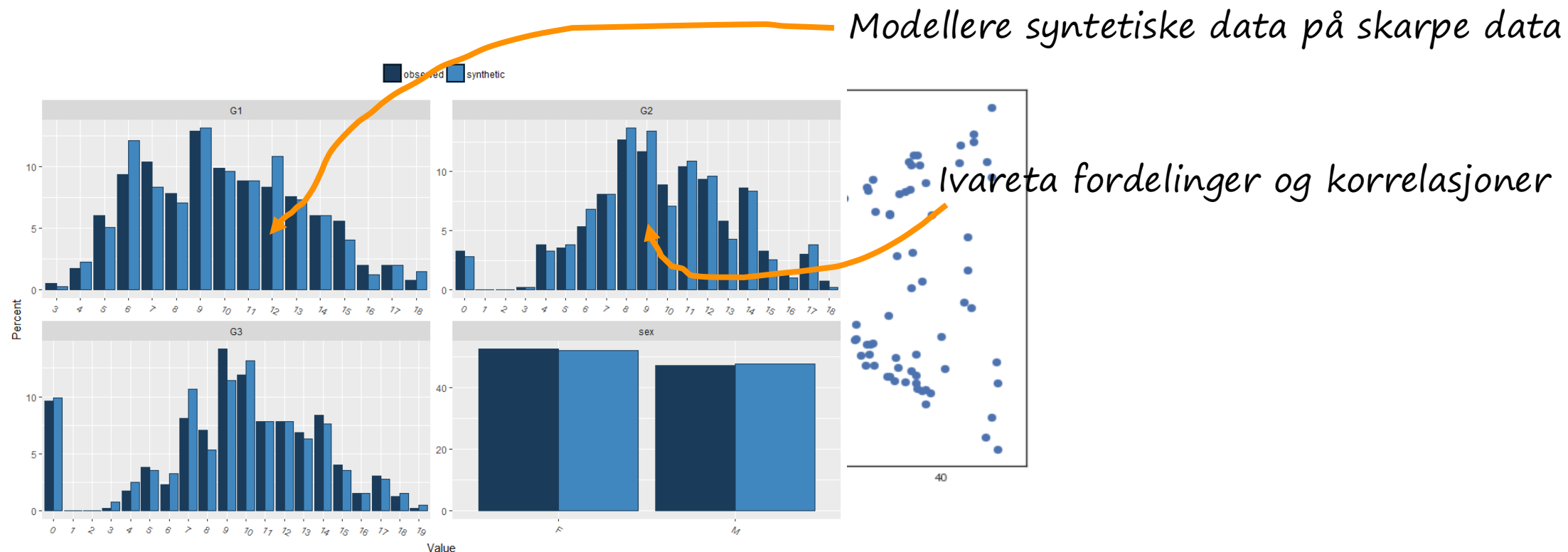


Du må tåle at svarene ikke er helt presise
og identiske fra gang til gang.

Syntetiske data

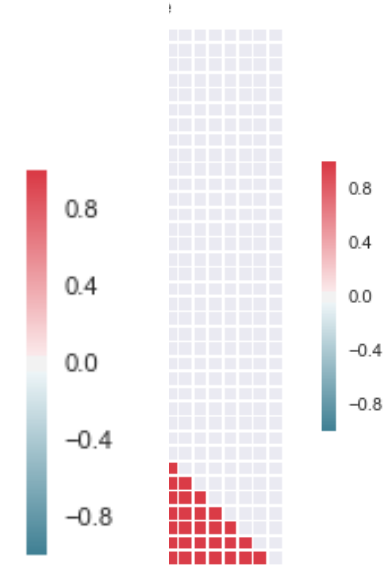
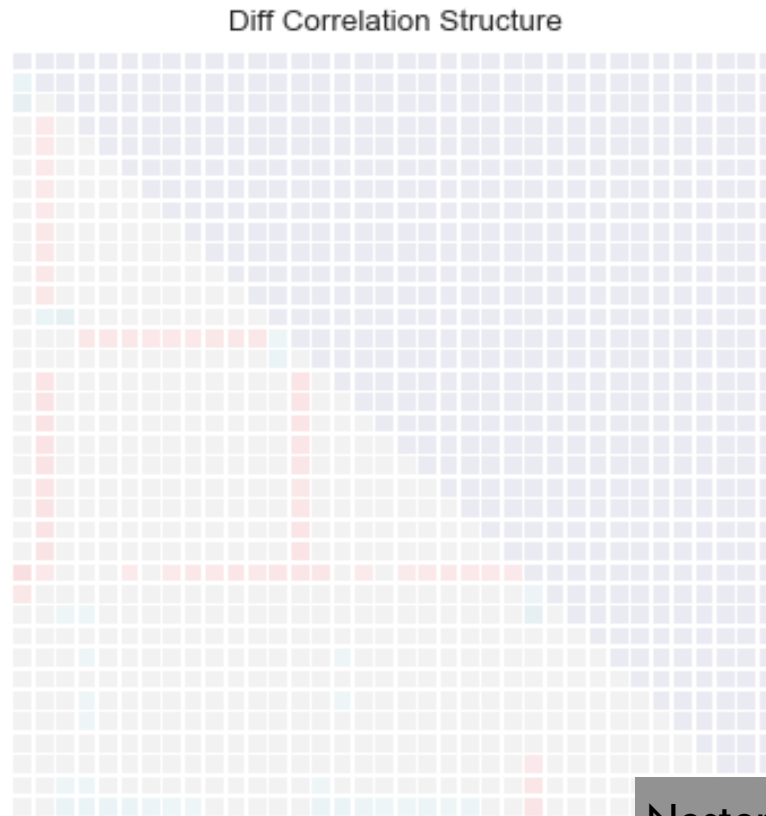


Lage et nytt *kunstig* dataset basert på en *modell* av et “skarpt” dataset



Bevarer **statistiske** sammenhenger!

Skarpe - syntetiske



Nesten ikke noe informasjonstap!

Pros



Alle datapunkter er fiktive!



Du bevarer datastrukturen!



En god syntetisering bevarer viktige sammenhenger og statistiske egenskaper.

Supert for analyser og maskinlæring!

Cons



Alle er fiktive, men du kan fremdeles lære masse om en populasjon og typiske egenskaper til en subpopulasjon.



Usikkert om syntetisk = “anonymt”.

Kanskje bør man teste i tråd med Art 29 WP anbefalingene?*



Er det virkelig lovgivers intensjon å la *informasjonsrike syntetiske data* kunne brukes til hvilket som helst formål?

Velkommen til Dolly

Dolly er NAVs selvbetjeningsløsning for å opprette syntetiske data. I Dolly kan du opprette syntetiske personer med forskjellige egenskaper, og tilgjengeliggjøre dataene i valgte testmiljøer.

[Logg inn med NAV epost](#)[Logg inn med BankId](#)

	Ident	Navn	Kjønn	Alder	Gruppe	
	24510199266 	Usymmetrisk Juice	Kvinne	23	→ Vis i gruppe	
	13849297648 	KJEMPENDE SJESELONG	Kvinne	33	→ Vis i gruppe	
	19448318888 	MAGNÍFICO AKKORD	Kvinne	42	→ Vis i gruppe	
	11410088381 	Sigen Grevling	Mann	25	→ Vis i gruppe	
	11487224938 	Rik Halvmeter	Mann	53	→ Vis i gruppe	
	67448239168 	INKONSEKVENT GALLUPMÅLING	Mann	43	→ Vis i gruppe	
	06527607793 	Emosjonell Mais	Mann	48	→ Vis i gruppe	
	25497016870 	Oksydert Båndsalat	Kvinne	54	→ Vis i gruppe	
	14108921111 	Ru Fasade	Mann	35	→ Vis i gruppe	

Hva med tekst?

**Ingen ryddig, fin tabell
med kolonnenavn...**

**Mange måter å uttrykke
det samme på...**

Forskeren Marie Curie, kjent for sine banebrytende oppdagelser innen radioaktivitet, ble alvorlig syk. Hun led av en mystisk blodsykdom som legene kalte "aplastisk anemi". Sykdommen var forårsaket av hennes arbeid med radioaktive stoffer. Curie ble internasjonalt kjent for sitt arbeid med to nye grunnstoffer, og hun er den eneste personen som har mottatt Nobelpriser i to ulike disipliner. Datteren, Irène Joliot-Curie, som fulgte i morens fotspor, beskrev henne som en utrettelig forsker: "Hun visste at arbeidet kunne koste henne livet, men mente kunnskapen var verdt prisen." Irène ble selv tildelt Nobelprisen for sitt arbeid med radioaktivitet i 1935.

**Opplysningene kan opptre i én setning....
...og så komme tilbake flere paragrafer senere**

**Det samme ordet betyr ikke
alltid det samme...**

Forskeren Marie Curie, kjent for sine banebrytende oppdagelser innen radioaktivitet, ble alvorlig syk. Hun led av en mystisk blodsykdom som legene kalte “aplastisk anemi”. Sykdommen var forårsaket av hennes arbeid med radioaktive stoffer. Curie ble internasjonalt kjent for sitt arbeid med to nye grunnstoffer, og hun er den eneste personen som har mottatt Nobelpriser i to ulike disipliner. Datteren, Irène Joliot-Curie, som fulgte i morens fotspor, beskrev henne som en utrettelig forsker: *“Hun visste at arbeidet kunne koste henne livet, men mente kunnskapen var verdt prisen.”* Irène ble selv tildelt Nobelprisen for sitt arbeid med radioaktivitet i 1935.

Forskeren Marie Curie, kjent for sine banebrytende oppdagelser innen radioaktivitet, ble alvorlig syk. Hun led av en mystisk blodsykdom som legene kalte “aplastisk anemi”. Sykdommen var forårsaket av hennes arbeid med radioaktive stoffer. Curie ble internasjonalt kjent for sitt arbeid med to nye grunnstoffer, og hun er den eneste personen som har mottatt Nobelpriser i to ulike disipliner. Datteren, Irène Joliot-Curie, som fulgte i morens fotspor, beskrev henne som en utrettelig forsker: “Hun visste at arbeidet kunne koste henne livet, men mente kunnskapen var verdt prisen.” Irène ble selv tildelt Nobelprisen for sitt arbeid med radioaktivitet i 1935.

Direkte identifiserende

Indirekte identifiserende

Konfidensielt

Forskeren *******, kjent for sine *******, ble alvorlig syk. Hun led av en mystisk blodsykdom som legene kalte *******. Sykdommen var forårsaket av hennes arbeid med radioaktive stoffer. ******* ble internasjonalt kjent for sitt arbeid med to nye grunnstoffer, og hun er *******. Datteren, *******, *******, beskrev henne som en utrettelig forsker: «*Hun visste at arbeidet kunne koste henne livet, men mente kunnskapen var verdt prisen*». ******* ble ******* for sitt arbeid med radioaktivitet i *******.

Direkte identifiserende

Indirekte identifiserende

Konfidensielt

Kan vi si noe om hvor “godt” dette egentlig er?

Når er det egentlig “godt nok”...?



Juristen

personopplysning

ikke personopplysning

Når er det egentlig bra nok?

1010
1010

Teknikeren

Vasket tekst



OK?

OK?

Finner opplysninger i tekst
– kan de også finne personopplysninger?

Forstår kontekst
– bedre kontekstuell forståelse av hva som kan være en personopplysning i tekst?

Store språkmodeller

Genere forslag
– kan de generere gode forslag som erstatter personopplysninger i tekst uten at vi mister meningsinnhold?

“Vet mye” (stort treningsgrunnlag)
– kan de brukes for å teste/knekke anonymiseringen?



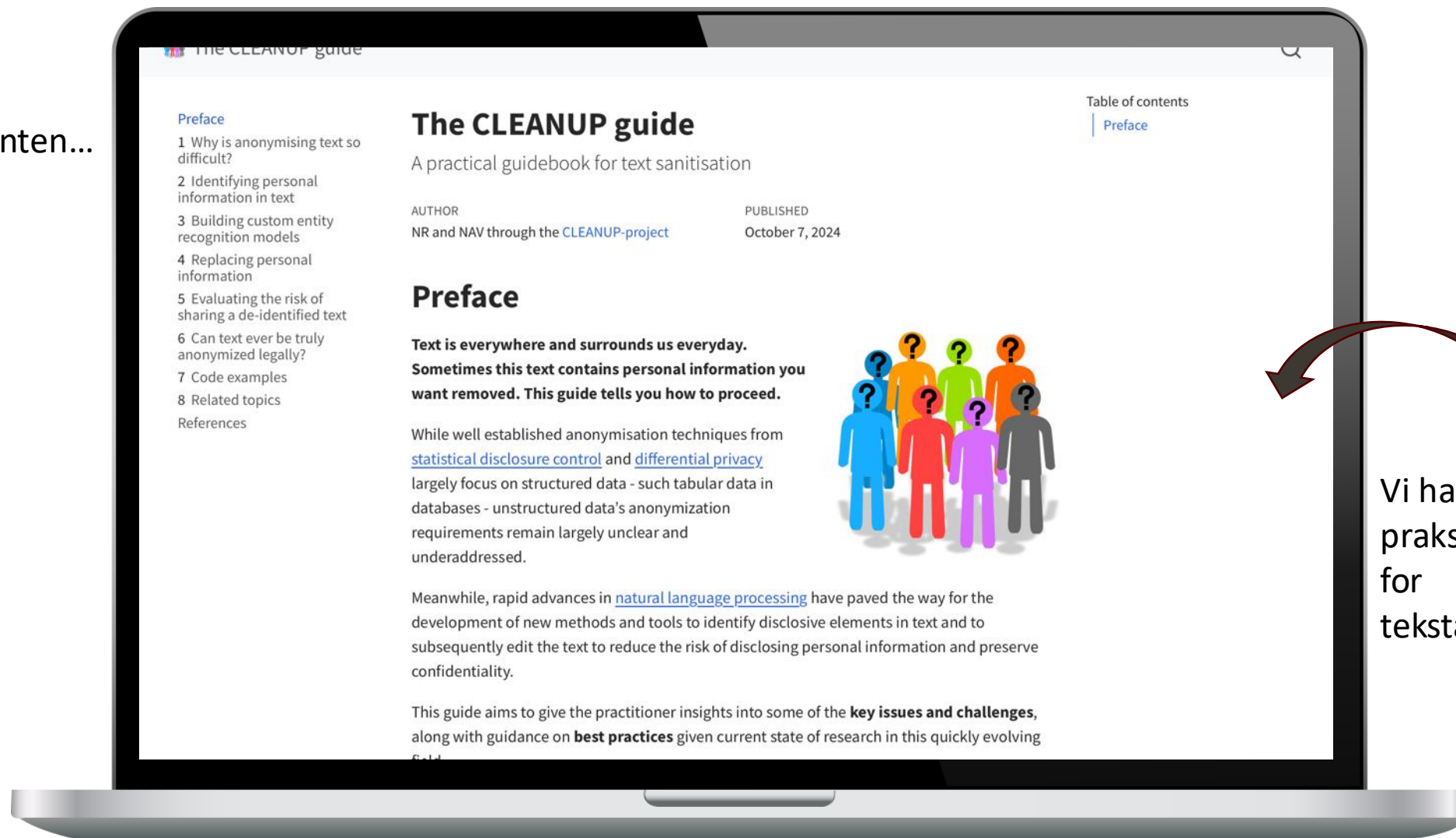
Copilot

Hvem tror du dette er? Forskeren ***, kjent for si...

Hvem tror du dette er?

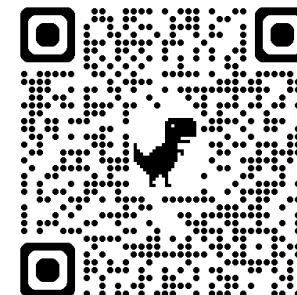
Forskeren ***, kjent for sine ***, ble alvorlig syk. Hun led av en mystisk blodsykdom som legene kalte ***. Sykdommen var forårsaket av hennes arbeid med radioaktive stoffer. *** ble internasjonalt kjent for sitt arbeid med to nye grunnstoffer, og hun er ***. Datteren, ***, ***, beskrev henne som en utrettelig forsker: «Hun visste at arbeidet kunne koste henne livet, men mente kunnskapen var verdt prisen». *** ble *** for sitt arbeid med radioaktivitet i ***.

Skjer mye på
forskningsfronten...



Vi har samlet beste
praksis i en veileder
for
tekstanonymisering...

<https://data.nav.no/fortelling/cleanup-guide/>



Aino

chat.dnb.no/index.html?seg=pm&la=NO



Privat

Aino er en chatbot som yter kundeservice. Ikke oppg personopplysninger til chatboten. Enkelte svar er generert med kunstig intelligens og er merket med et tag. KI-genererte svar kan inneholde feil. Du kan lese mer om personvern her: [Ditt personvern](#).

17:06:46

God ettermiddag!
Jeg er Aino, DNBs digitale medarbeider 😊

Skriv med få ord hva jeg kan hjelpe deg med.

 Rabatt på bilforsikring

Chat



Ved å chatte med oss samtykker du til vår [personvernerklæring](#). Vi oppfordrer deg til å lese denne. Vennligst ikke oppgi personopplysninger i chatten.

Samtykker du til dette?

Nei

Ja

Skatteetatens chatbot



Hei! Vi chatter anonymt. Ikke skriv ting som navn, adresse eller fødselsnummer.

Om personvern

Hva kan jeg hjelpe deg med?

Frida

Privat - bokmål



For å chatte med oss er det viktig at du ikke skriver opplysninger i chatten som gjør at vi kan identifisere deg eller andre.

Godtar du dette?

Nei

Ja

Skriv her, unngå personlig info

Kan vi **opplyse** om sensitiv informasjon **før** det blir lagret?

Kan vi kanskje bruke **KI** til dette?

TryggTekst

(under utvikling)

Kan KI hjelpe oss med oppdage *særlige kategorier* av personopplysninger i fritekst?

TryggTekst

Språkmodeller finner personopplysninger

Kan KI hjelpe oss med oppdage *særlige kategorier* av personopplysninger i fritekst?

“ Du har innsett at det ikke vil være mulig for deg å fortsette i arbeid som frisør hos nåværende arbeidsgiver. Du sliter med senebetennelse i en arm, og er i gang med behandling hos fysioterapeut. Arbeidsplassen har ikke andre arbeidsoppgaver å tilby deg, og du er derfor på let etter annet arbeid.

Samtalereferat



Et fiktivt samtalereferat

Dato (obligatorisk)

11.6.2025



Møteform (obligatorisk)

Oppmøte



Samtalereferat (obligatorisk)

Du har innsett at det ikke vil være mulig for deg å fortsette i arbeid som frisør hos nåværende arbeidsgiver. Du sliter med senebetennelse i en arm, og er i gang med behandling hos fysioterapeut. Arbeidsplassen har ikke andre arbeidsoppgaver å tilby deg, og du er derfor på let etter annet arbeid.

4701 tegn igjen

☒ Klarspråkhjelpen



Sjekk teksten med TryggTekst



- Helseopplysninger
sliter med senebetennelse

Hva med anonymiserings*prosessen*?



Er prosessen å anse som *selvstendig behandling* av personopplysninger eller kun som en *teknikk* på lik linje med aidentifisering, pseudonymisering eller kryptering?



JA - i følge personvernteorien er *anonymiseringsprosessen en selvstendig behandling av personopplysninger* og må følge personopplysningslovens krav



Personopplysninger



Syntaktisk anonymisering

Differential privacy

Syntetisering

++



Anonyme opplysninger

Rettslige krav til anonymiseringsprosessen

- Rettslig behandlingsgrunnlag
- Formålsangivelse og forenlighetsvurdering?
- Opplysningene som skal være gjenstand for anonymisering må være lovlig innhentet
- Formålsbegrensning
- Dataminimering
- Den registrertes rettigheter (åpenhet, kontroll)
- Risikovurdering av anonymiseringsprosessen



Sikre og dokumentere at dataene er reelt anonymisert!

Takk for oss!



Spørsmål?

robindra.prabhu@nav.no
sara.elise.wollo@nav.no