

Framtidens statistikproduktion— möjligheter och utmaningar

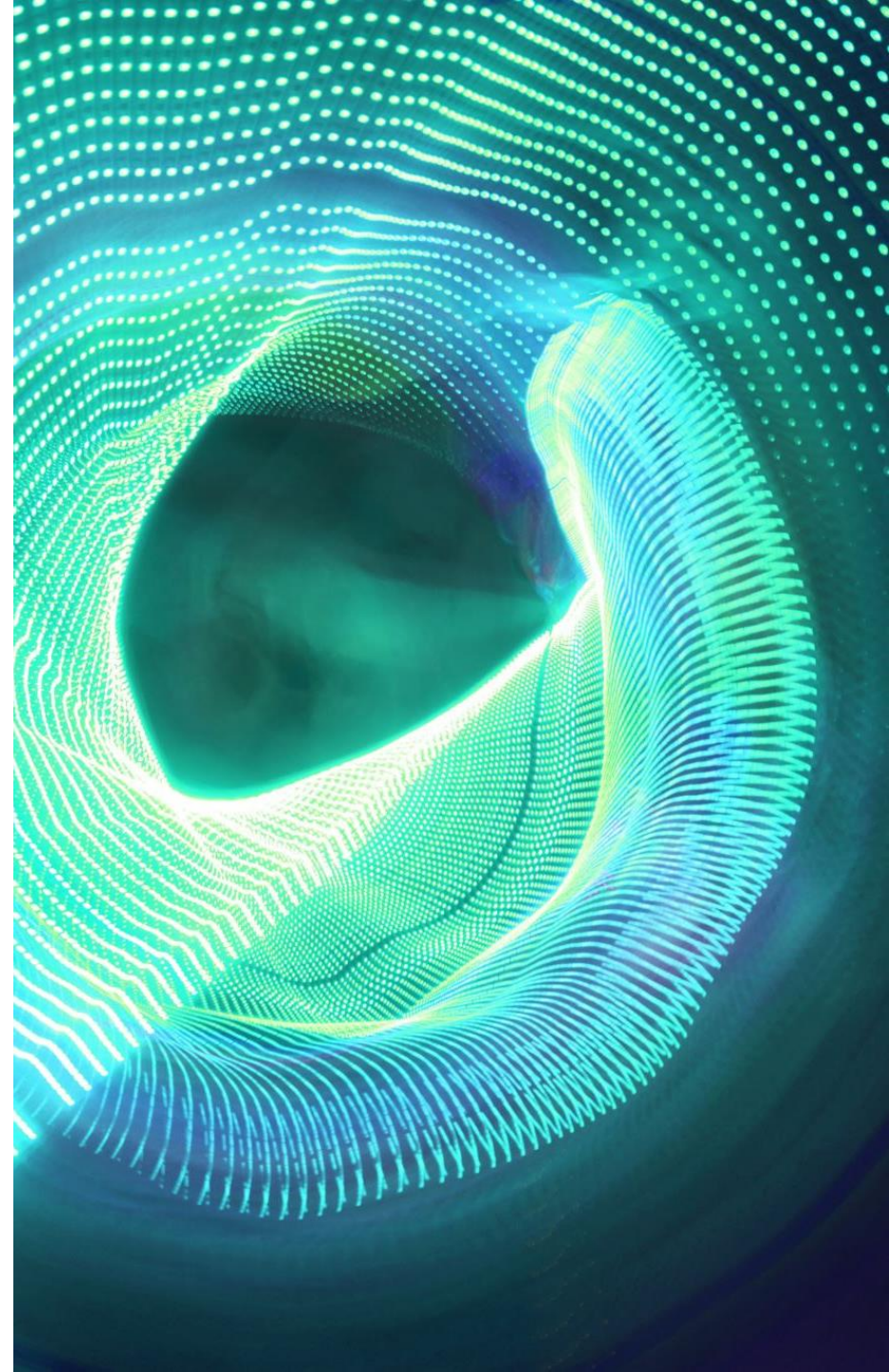
Lilli Japac, SCB

Surveyföreningens jubileumskonferens,
16 oktober 2024



Inledning

- Vad kommer vara det viktigaste de kommande 25 åren?
- Vad finns det för faror, utmaningar och möjligheter för framtiden?



**Hur såg det ut för
25 år sedan?**

1999

- Inga smartphones
- Inga sociala media, Poddar...
- CD-skivor och videofilmer--inga streaming tjänster t.e.x. Spotify och Netflix
- gps i bil var nytt
-



Statistikvärlden

- Postenkäter och telefonundersökningar vanliga
- Webbundersökningar var nytt
- Besöksintervjuer ovanliga
- Prismätningar i fält gjordes fortfarande
- Ökande borfall och kostnader i undersökningar
- Hur göra "Mixed-mode"-undersökningar?
-



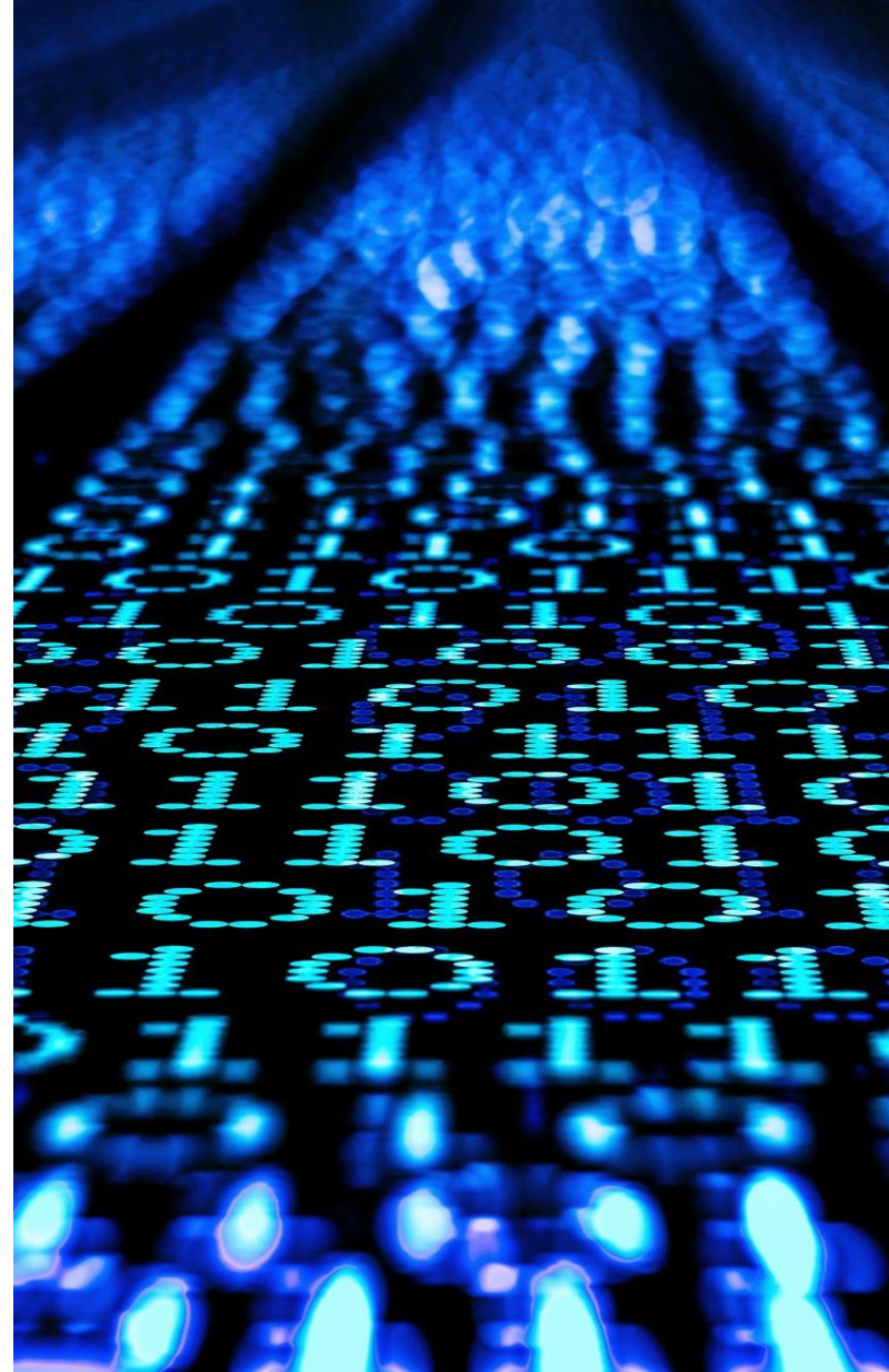
**Hur ser det ut de
kommande 25 åren?**

Omvärlden

- Joint Research Center, 14 Megatrends [Megatrends cards.pdf \(europa.eu\)](#) t.ex.:
- Accelererande teknologikutveckling och “hyperconnectivity”
- Fortsatt urbanisering
- Migrationens ökade betydelse
- Klimatförändringar
- Förändrad säkerhetsparadigm
- Ökad ojämlikhet
-

Statistikproduktionen i framtiden

- Kombinera flera datakällor—undersökningar och digitala data
- Statistikbyråer inte längre de som samlar in mest data
- Nytt lagförslag (Regulation (EC) No 223/2009) tillgång till privata (digitala) data
- Användning av digitala data—nya insikter, förbättra eller ersätta befintliga undersökningar
- Nya metoder för att t.ex. samla in, integrera, bearbeta och analysera data



Vi kommer fortfarande gör undersökningar

- Många digitala datakällor har få variabler—samband mellan variabler
- Samplingteori och undersökningar för att (Heeringa and Lyberg 2021):
 - skatta parametrar i modeller,
 - utvärdera modeller,
 - justera för brister i data och
 - som benchmark.
- Nya metoder för populationer som är svåra att nå, få med och intervjua
- “Smart surveys”

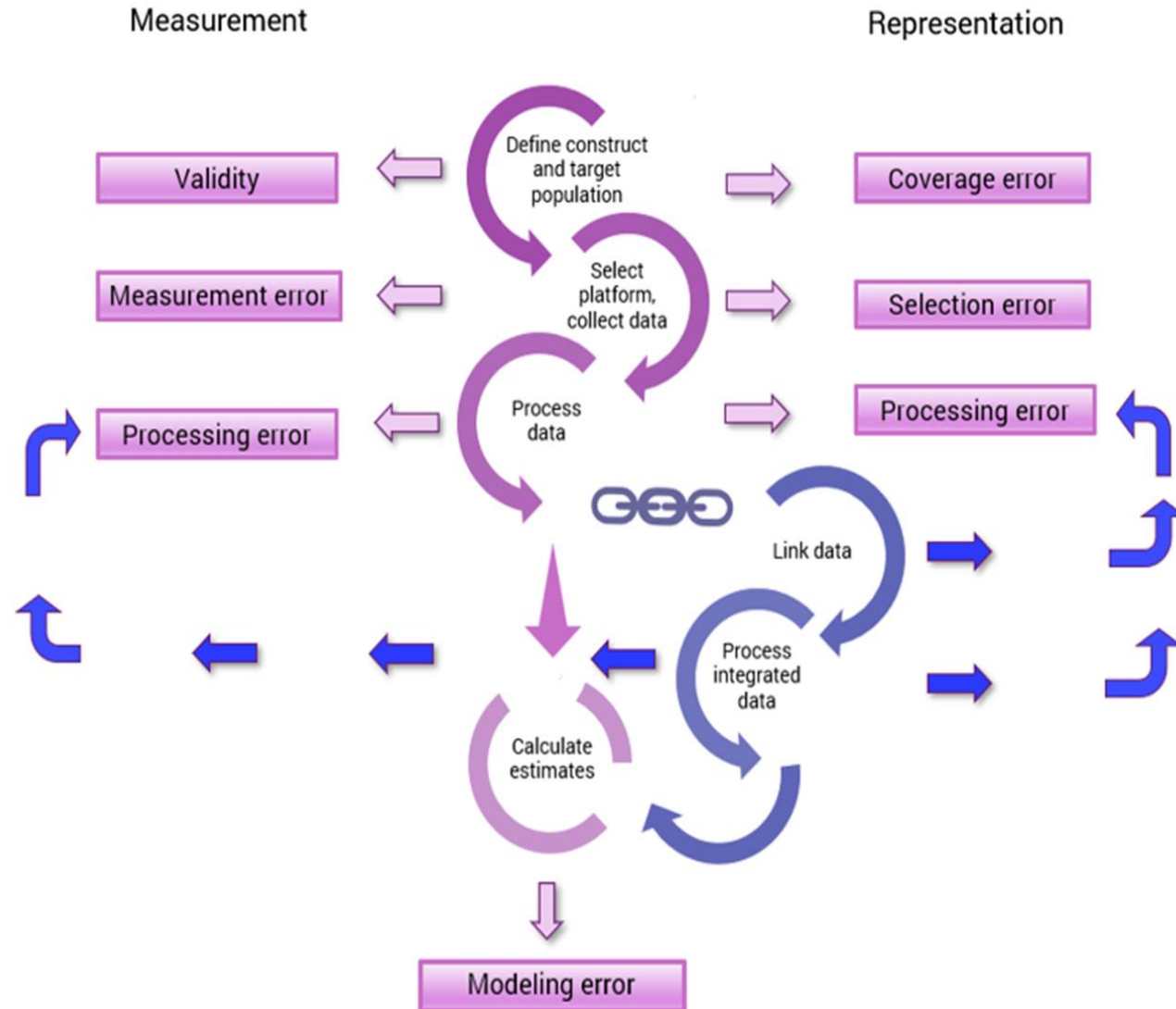
“Smart surveys”

- Smarta enheter (t.ex. Mobiltelefon) används för att samla in information=>kombination av **aktiva och passiva data**
- Forskningen visar t.ex. att:
 - **Mätfel** i undersökningar kan ibland minskas genom passiva data (Luiten et al 2021, **fysisk aktivitet**)
 - Smart surveys är **komplexa** och innehåller många steg
 - Låg svarsfrekvens och viljan att dela passiva data är **låg** t.ex.
 - > Holländsk studie (Struminskaya et al 2021) dela sin GPS-plats vid en tidpunkt 46%
 - > Holländsk TUS (Scherpenzeel 2017), samla in GPS-plats i 2 dygn, 17%
- Mobiltelefoner, **täckning** lägre i vissa grupper t.ex. äldre
- Låg **svarsfrekvens** i vissa grupper t.ex. unga, låg utbildningsnivå, utländska medborgare (Keusch et al 2021)
- Kan påverka **undersökningsvariabler** t.e.x. överskattning av användning av social media (Keusch et al 2021)

Kvalitet

- **Tillförlitlighet** kommer fortfarande vara viktigt för användare d.v.s.:
 - statistiken är **representativ** för den grupp som användaren är intresserad av och
 - statistiken **mäter** det koncept som användaren är intresserad av.
- **Digitala data** och tillförlitlighet
- Visa på bra kvalitet, kräver **metodstudier**
- Transparens och kommunicera eventuella kvalitetsbrister

Vad är så speciellt med digital data?



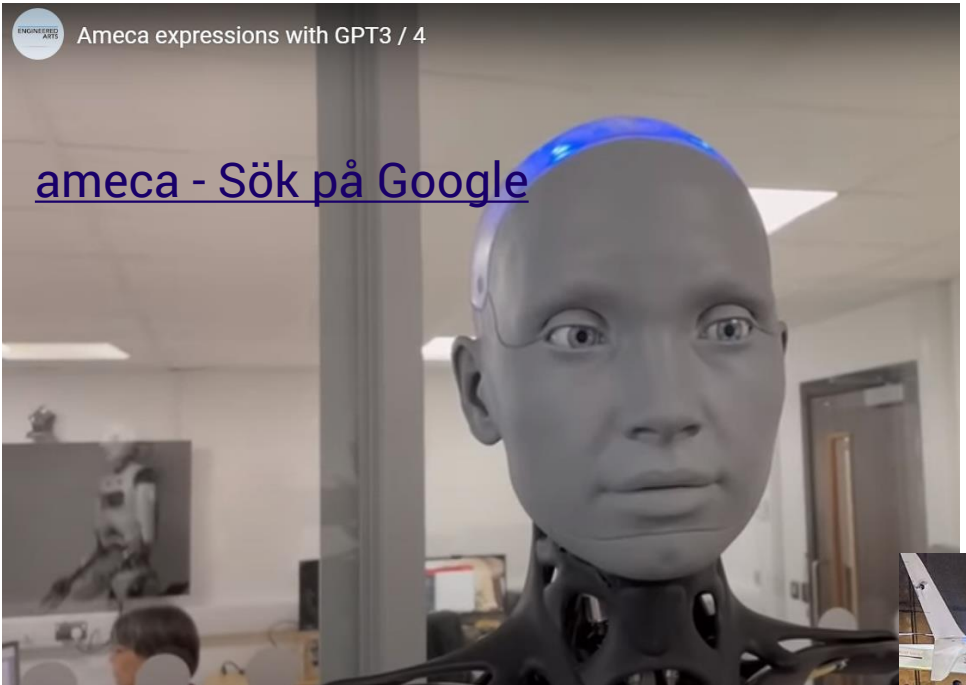
- Vi saknar ofta **detalj**kunskap om datakällan och hur den genererats
- **Konceptet** kommer först
- **Plattforms**populationen skiljer sig ofta från målpopulationen
- Ofta behöver digitala data **integreras** med andra datakällor
- **Nya metoder** och modeller används i flera steg
- När vi integrerar data så bidrar varje datakälla till **det totala felet** i skattningen

AI

SCB

Technology and AI

[Atlas Gets a Grip | Boston Dynamics - YouTube](#)



ENGINEERED ARTS Ameca expressions with GPT3 / 4

[ameca - Sök på Google](#)



TAKING NOTES AT MEETINGS

An AI-powered tool, **Reflect**, can help summarize key points of our meetings and transcribe voice accurately.

A part of a versatile service, **Notion AI** can collect, summarize, and highlight important ideas.

AI meeting note-taking service, **Otter** can capture audio and transcribe it into text.

Fireflies specializes in recording, transcribing, summarizing meetings.

The AI note taker that transcribes, summarizes and generates insights from your calls with customers, prospects, and your team.

Notion

Reflect

Otter

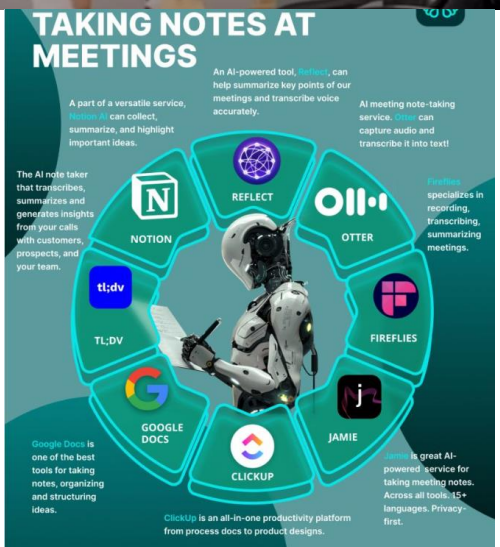
Fireflies

Jamie

ClickUp is an all-in-one productivity platform from process docs to product designs.

Google Docs is one of the best tools for taking notes, organizing meeting notes, and structuring ideas.

Jamie is great AI-powered service for taking meeting notes. Across all tools, 15+ languages. Privacy-first.



[EVTOL: Geely Aerofugia Unveils Self-developed Flying Car - Asian Sky Group](#)

Från el till ai på 100+år



Elektrisk belysning när som helst vilket förlängde de produktiva timmarna.



Hushållsapparater: kylskåp, tvättmaskiner,...
Smarta apparater (lär sig användares preferenser), sömlös anslutning mellan enheter.



Kommunikation: telegraf, telefon främjade bättre anslutningsmöjligheter och affärsinteraktioner. **Realtidsöversättning tar bort språkbarriär, bättre kundservice.**



Underhållning: radio och tv förde in underhållning i människors hem,. **AI kan generera musik, konst och annat kreativt innehåll.**



Industriell revolution: eldrivna fabriker och tillverkningsprocesser. **AI-drivna robotar och automatisering**



Transport: Elektriska spårvagnar, tåg och kollektivtrafiksystem . **Optimera logistik och leveranskedjor. Autonoma fordon, och förbättrade trafikflöden.**



Medicin: röntgen, elektriskt driven medicinsk utrustning...**Avancerad bildanalys bättre och snabbare diagnoser, AI-drivna robotar i patientvården. Förbättrad logistik.**



AI i statistikproduktionen

- AI, ML, Generative AI, open-source large language models
- Några exempel:
 - -SAS to R
 - Förbättra sökmotor Office of National Statistics (UK) and International Monetary Fund (IMF)
 - Kodning t.ex. yrke
 - Ersätta intervjuare
- UNECE-rapport med exempel



Problem statement

- Current website search looks at prevalence of keywords in the title and subtitle, rather than meaning
- E.g. Results differ wildly for 'how many people died' and 'registered deaths'.
- Even if the right bulletin is returned, the user needs to search through the page to return the right information

StatsChat

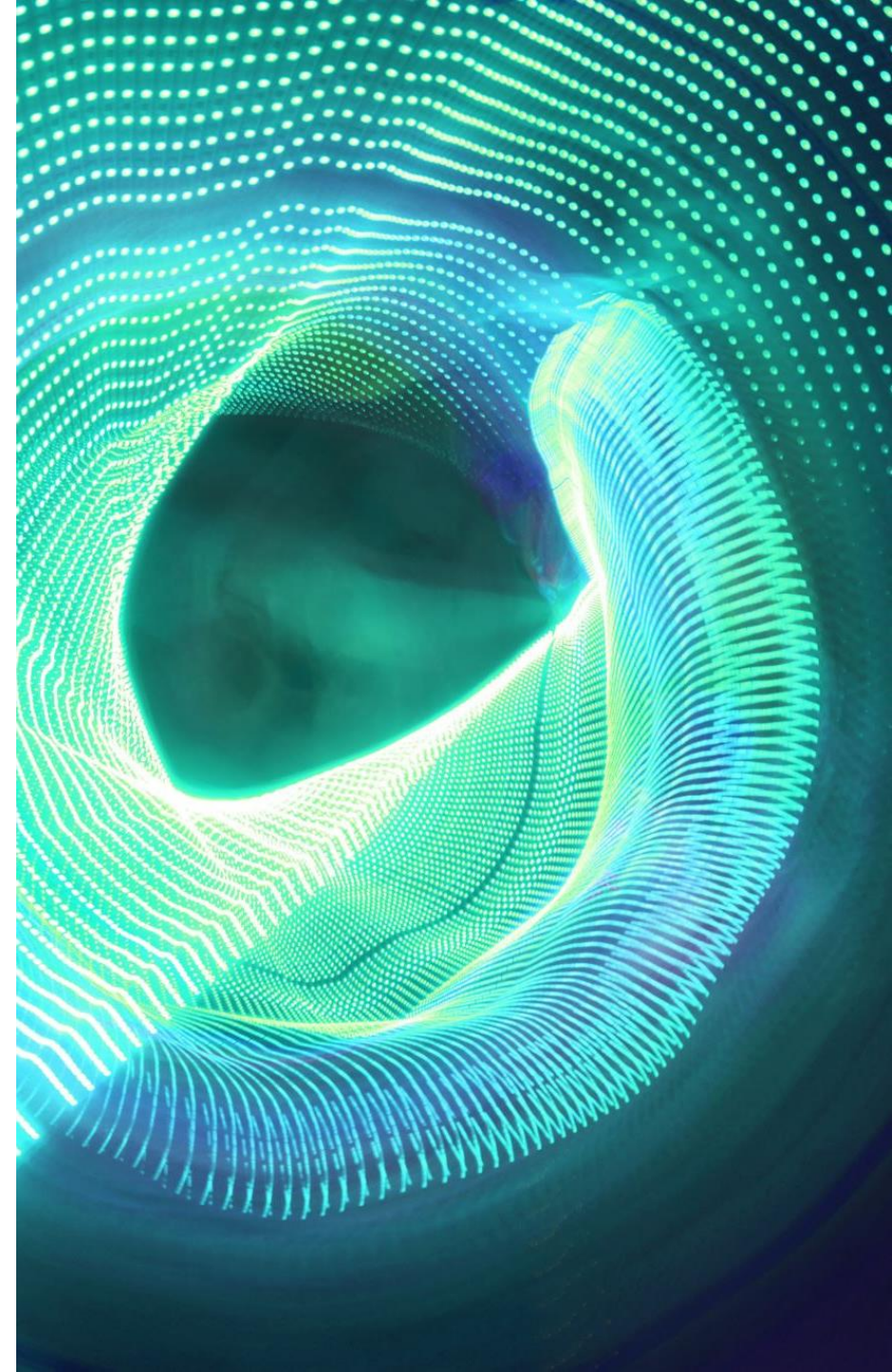
- **Experimental tool understands the meaning behind user queries and 'reads' relevant documents before generating a summary answer.**
 - Open tools / transparent to scrutiny
 - Reusable / flexible – easy to change
 - Secure design - data never leave ONS infrastructure
 - Control over answer – user doesn't directly engage with LLM

The screenshot shows the search results for 'how many people live alone?' on the Office for National Statistics website. The search bar contains the query, and the results show 24 results. The top result is 'Internet access - households and individuals, Great Britain', released on 12 August 2019. The snippet mentions that internet access in Great Britain includes 'how many people' who have the internet and how they use it.

The screenshot shows the StatsChat experimental tool interface. The user asks the question: 'how many people live alone in the UK?'. The tool provides a 'Most likely answer': 'The number of people living alone in the UK in 2022 was 8.3 million, this represents 13% of the household population and 30% of all households; the majority (53%) of these households were women living alone.' The answer is based on the publication 'Families and households in the UK: 2022', released on 18 May 2023.

Avslutningsvis

- Vad kommer vara det viktigaste de kommande 25 åren?
 - AI-utvecklingen
- Vad finns det för faror, utmaningar och möjligheter för framtiden?
 - Takten i utvecklingen
 - Implementeringstakten
 - Ökad effektivitet
 - Kvalitet?



Tack!

lilli.japec@scb.se

