Qualitative methods after generative AI

How to get value out of unstructured datasets

Anders Kristian Munk & Anders Koed Madsen

@TANTlab
www.tantlab.aau.dk

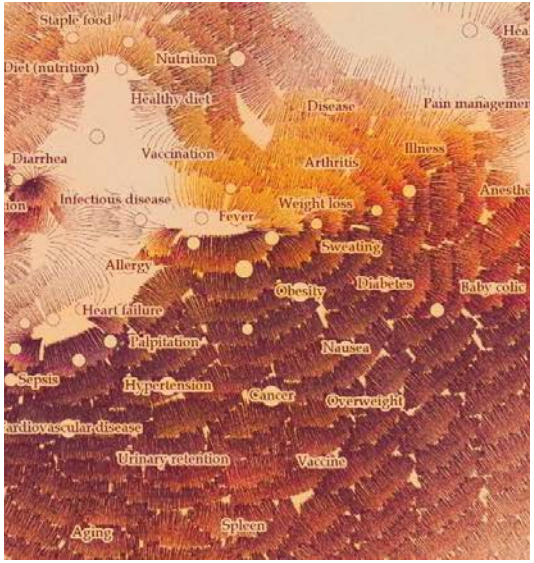
tantlab

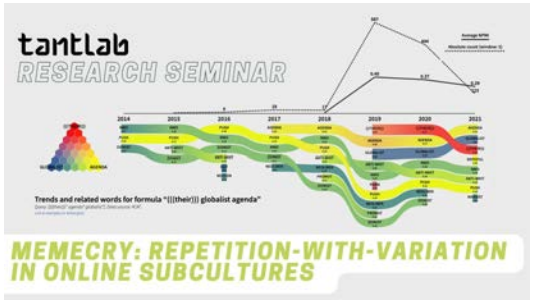
QUANTITATIVE ANALYSIS
LARGE CO-WORD NETWORK
ADD
tanta.lab

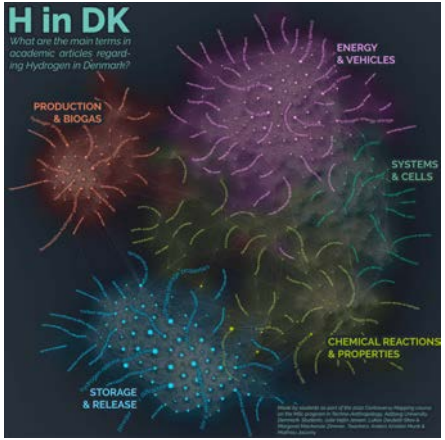
We combine anthropological and computational methods

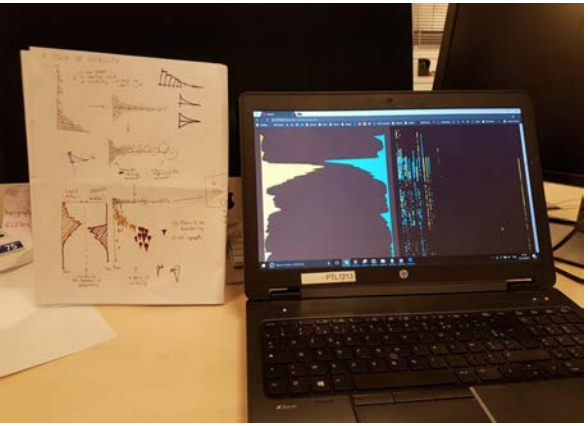
For instance:

- Ethnographic fieldwork
- Interviews
- Big social data (unstructured)
- Machine Learning
- NLP
- Explorative data viz
- Generative AI









In collaboration with partners

As Participatory Data Design

In the format of ‘datasprints’.

To take local conditions and perspective serious

To build situated analytical capacity

Program for the next hour

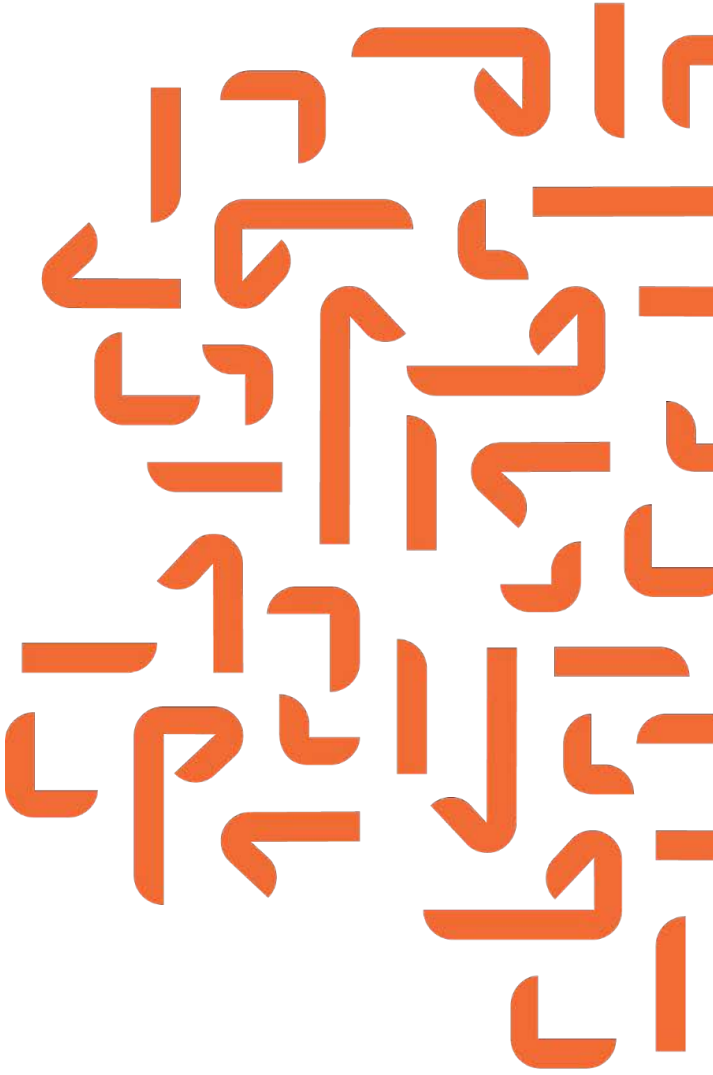
A new data situation: big qual & generative AI
The qualitative world datafied
The rise of Generative AI

Let's turn to the humanities!
New (old) quality criteria for data projects
Searching questions and iterative inquiry

A few cases
PUBLIKUM
UB
ADD controversy map

A new data-situation?
3 organisational attention points

Join us...

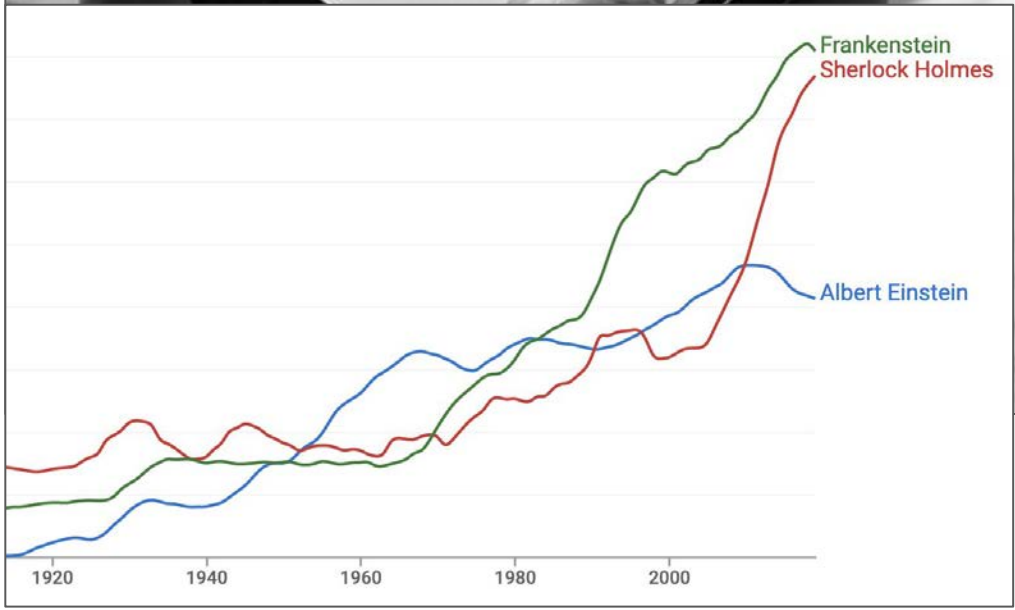


A new data situation

Big qual & generative AI

The qualitative world datafied

- Digitization is not the same as datafication. Qualitative data often get digitized but not datafied.
- Google realized that this in the 1990's and have not since looked back when developings services.
- Many organizations still digitize qualitative content without thinking strategically about data potentials.
- But the mindset is beginning to change....



DATAFIED ARCHIEVES

senseable city lab.

50 ROME

BARCELONA 14

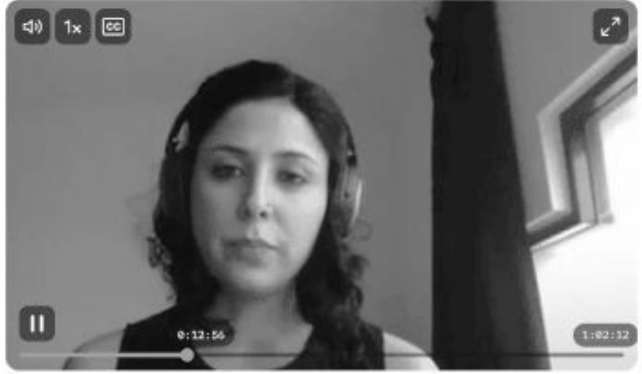
PARIS 14

PRAGUE 11

VIENNA 10

BERLIN 7

DATAFIED AESTHETICS



Summary Timeline Highlights 5

Zuzana Serakova · Quest.io 52% · 32 min

Evan Litvak · Grain 45% · 26 min

Highlights 0 >

Search transcript

Zuzana Serakova

This is who's in it. Okay. It was a meeting about an hour, 10 minutes and it was on the all I need.

This for sure. Yeah. And I love the search too, because like you said, you said it per dense. Like it's almost like we gave so muc into one tool, and now it does the thing it do. It's just like, here's everything. Todd single thing with Todd and it's just like, understand everywhere this is occuring.

Evan Litvak

And again, I'm like, maybe like, you know, unnecessarily grateful, but like whatever I Simplicity. I want to add a tag. I guess I thing, then I have to like type in here. I' finally I understand.

I don't know what this meeting is. This is internal working session. Okay. Maybe. And the fixed, this one where like the, you kno will work, but okay. Like for, for like fiv up until right now, you typed one tag in. A tag was like, it was like, Flashing there,

DATAFIED MEETINGS

The rise of generative AI

slido

How do you think generative AI will help us work with qualitative data?

ⓘ Click **Present with Slido** or install our Chrome extension to activate this poll while presenting.

Making life *easier* for qualitative researchers

- Transcribe my interviews
- Annotate my documents
- Summarize my fieldnotes
- Caption my videos
- Visualize the workshop ideas I developed with my informants
- And do it at scale!

Urbanist AI

Making life *different* for qualitative researchers

- Let me be the curator
  - ...you suggest analysis, I choose and synthesize
- Simulate my informants
  - ...what did I forget to ask?
  - ...what would they have thought about x?
- Or simulate me!
  - ...semistructured interviews with a chatbot?
- Let my readers interact "directly" with my data
  - ...without giving them access to the data

And making life *difficult* for qualitative researchers

- Who is this new collaborator?
- Does it have a voice? Does it have values of its own?
- How do I know, what they are?
- Can I affect them?
- How do I know if I am speaking to the foundation model or to my own data?

CREDIT: STABLE DIFFUSION, FEBRUARY 2023

"Wedding ceremony for my daughter"

Munk, A.K. (2023). **Coming of Age in Stable Diffusion.** Anthropology News

Let’s turn to the humanities

New (old) quality criteria for data projects

Quality criteria in a positivist paradigm

Reliability
Can this be reproduced?

Validity
Are we measuring what we claim to measure?

Representativeness
Do we have the right sample?

Generalisability
Do our findings apply to other situations?

Quality criteria in a hermeneutic paradigm

Recursivity
Are we adapting our research design to the field?

Credibility
Will our findings be accepted by the field?
Do they matter to our stakeholders?
A.k.a. ecological validation

Reflexivity
Have we been transparent about our process?
And have we reached a point of saturation?

Situationality
Do we have a uniquely adequate understanding of the situation in question, on its own terms?

slido

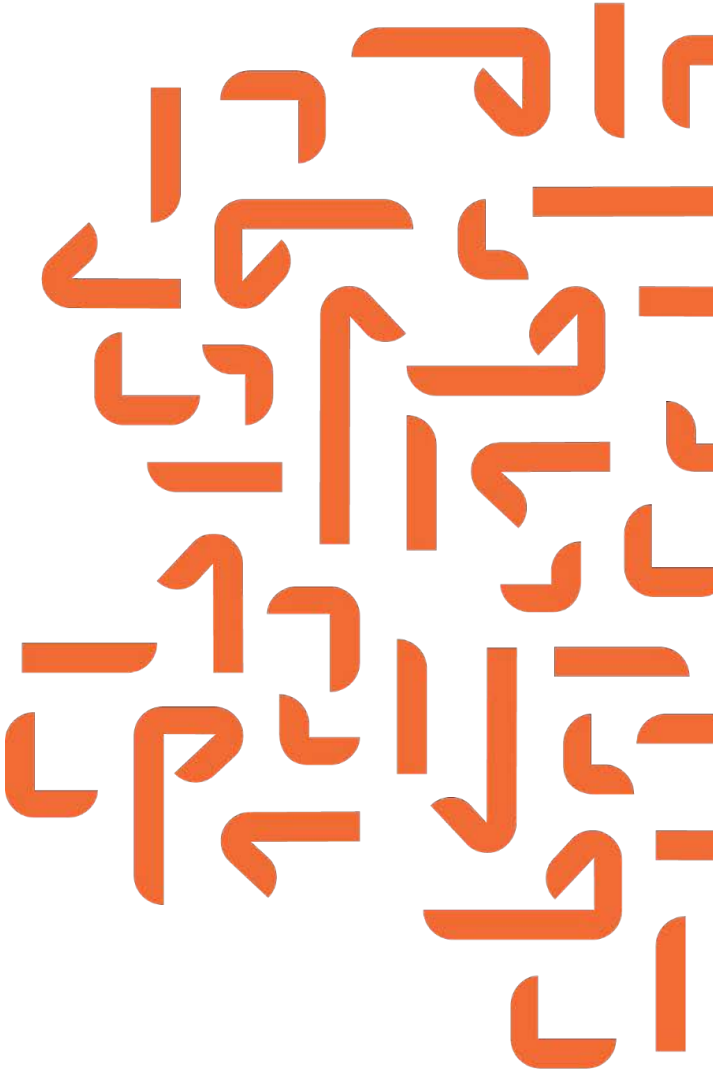
What is the biggest ethical challenge with large language models and the way we treat them as intelligent beings?

ⓘ Click **Present with Slido** or install our Chrome extension to activate this poll while presenting.

What is the biggest ethical problem with that question?

- Assumes we all agree what “ethical” is
- Assumes we all know about LLMs
- Assumes you have also thought about the way we anthropomorphize LLMs and that you find it problematic
- Therefore high risk of ignoring what ethical concerns you may actually have with generative AI
- If you even have concerns at all

Perhaps we should have first spent time exploring how problems around generative AI get framed differently by different groups, e.g. online or in the news media?



Searching questions

(how to formulate good problems?)

CENTER FOR ORGANIZATIONAL INNOVATION PÅ COLUMBIA UNIVERSITY

David Stark

Searching questions

“The first temptation for leaders [...] is to immediately reframe ambiguous situations as cases of **analytic problem solving** to clearly identify the problem, break it down into independent components, and organize a series of decisions about how best to solve them”.

David Stark

Searching questions

“The fundamental challenge is the kind of search when you do not know what you are looking for but will recognize it when you find it [...] perplexing situations provoke **innovative inquiry** [...] why not build organizations that generate such situations?”

....and how can changed data practices help generate those situations

Participatory Data Design

“Local trust in improved by involving those affected by datafication in the iterative loop described here.”

Problem Identification

There is something we want to know more about. Perhaps the problem is ‘wicked’ in the sense that it is ill defined. Existing concepts loose power.

Re-formulate Problem

Adjust the framing of the problem [illegible] observed. Accept that problem [illegible] beta’.

Data Detective Work

Go on a treasure hunt for relevant (digital) traces left by those engaged with the problem. Or find ways of producing those [illegible] that traces will exist in different places [illegible].

Analysis

Use algorithmic techniques to find patterns in relevant traces. Get new ideas as to how the problem could be framed and understood by different groups.

A few Cases

PUBLIKUM, UB, ADD

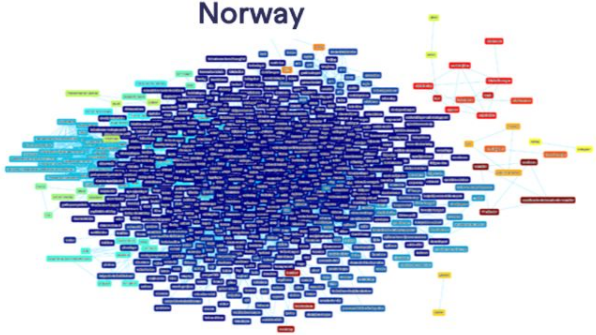
Norsk filminstitutt
FESTIVAL DE CANNES
UN CERTAIN REGARD
2022 OFFICIAL SELECTION
OSLO
PICTURES
How is narcissism different
in Scandinavia?
SICK OF MYSELF
A film by Kristoffer Borgli
Co-funded by the
Creative Europe MEDIA Programme
of the European Union
PUBLIKUM / will&agency

Norsk filminstitutt

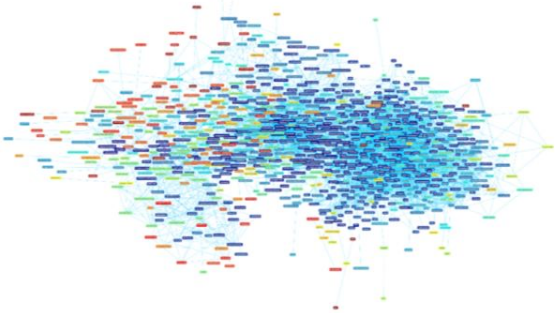
OSLO PICTURES

Narcissism in Scandinavia

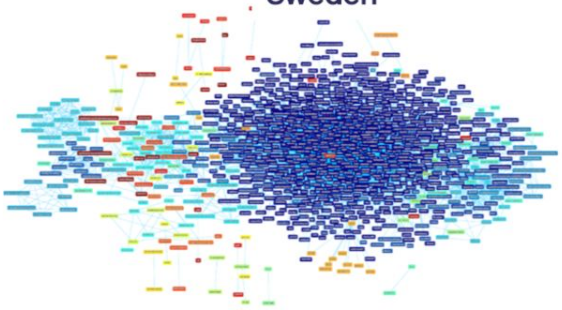
Norway



Denmark



Sweden



Co-funded by the Creative Europe MEDIA Programme of the European Union

PUBLIKUM / will&agency

Nya perspektiv på psykisk ohäls
behövs

Debatt och insändare • Psykisk ohälsa är ett mångfacetterat begrepp och omfattar ett brett spektrum av symtom. Beroende på vem som definierar psykisk ohälsa, och i vilket sammanhang detta sker, kan innebörden av begreppet se väldigt olika ut. Vilken syn vi utgår ifrån har en avgörande betydelse för utgången av våra ansträngningar för att både förebygga och behandla de symtom som vi idag kallar för psykisk ohälsa.

#3
Archetypes.

#3 Archetypes:

Three versions of a bastard

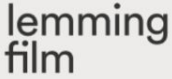
Denmark: **Dangerous** bastards

Sweden: **Poor** bastard

Norway: **Useful** bastard

Co-funded by the Creative Europe MEDIA Programme of the European Union

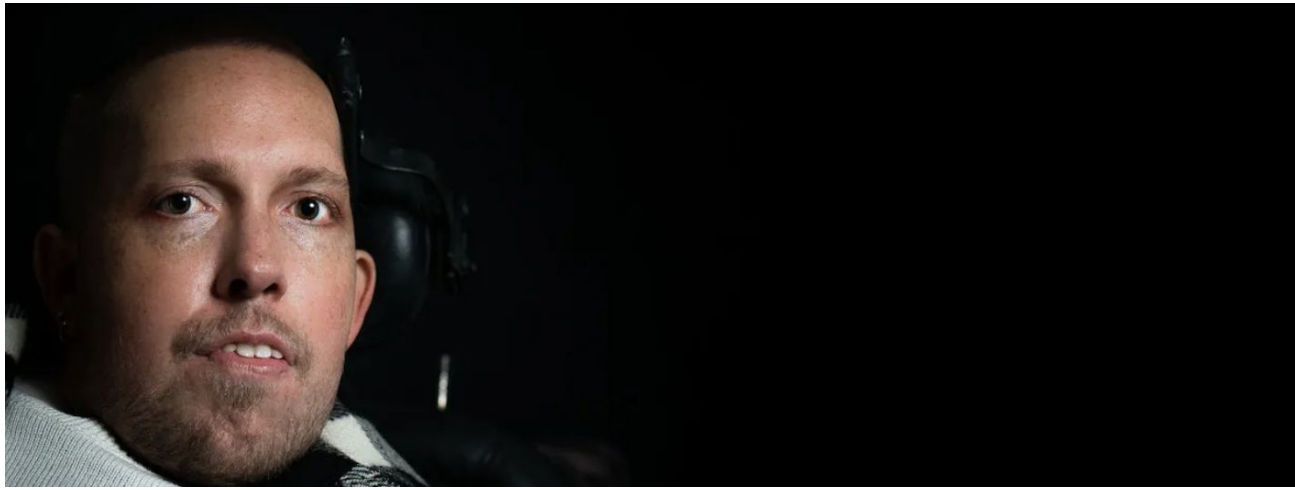
PUBLIKUM / will&agency



Co-funded by the
Creative Europe MEDIA Programme
of the European Union

URBAN BELONGING

EUROPEAN UNION PRIZE FOR
CITIZEN SCIENCE



The Urban Belonging Project

What makes you feel that you belong in the city??

To design a more socially sustainable city, we long for more diverse and nuanced insights about how the city works as a space of belonging - not seen from the professional eyes of those who plan it, but from those who live in it. To explore this, the Urban Belonging project was initiated in 2021 by a collective of planners and scholars in Copenhagen and Amsterdam with the ambition of mapping lived experiences of local communities in the city.

LGBT+ DANMARK

mino danmark

Hugs & Food

SIND

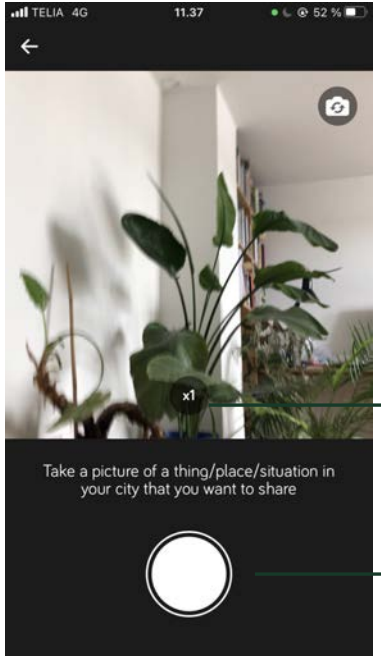
Danske Døves Landsforbund

Dansk Handicap Forbund

Gehl

Participatory photography

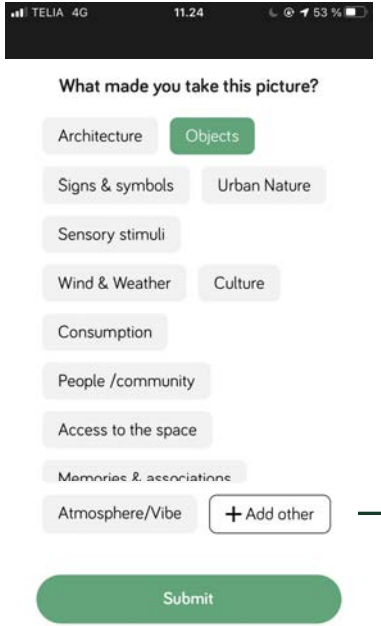
Participatory map-making



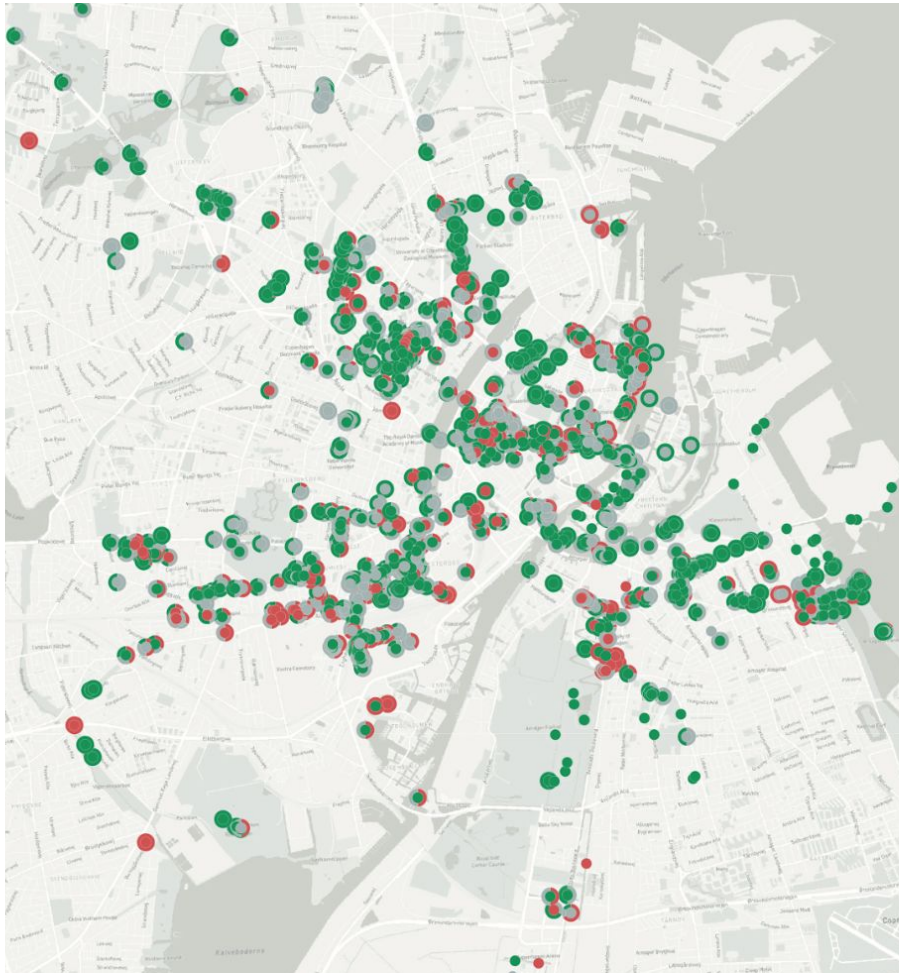
Use this button to zoom

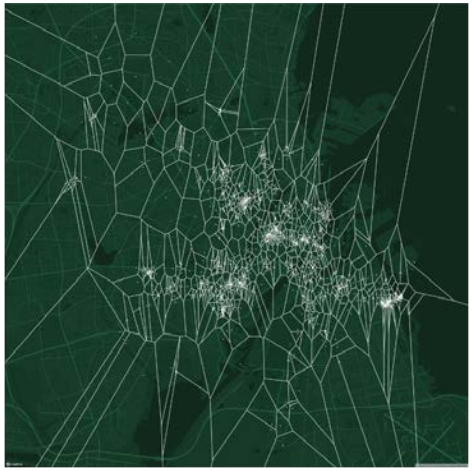
Use this button to take a photo

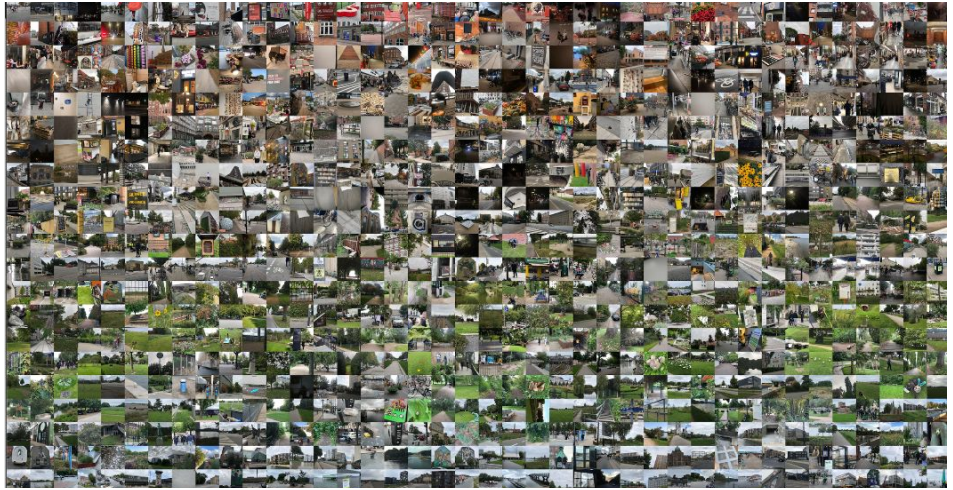
Use the slider to indicate how much or how little the photo is for you



If you took the photo for another reason you can make your own annotation by clicking ‘add other’



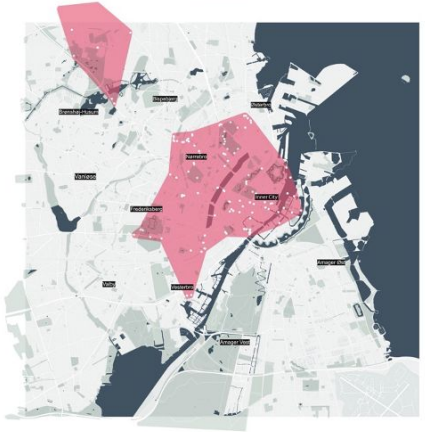




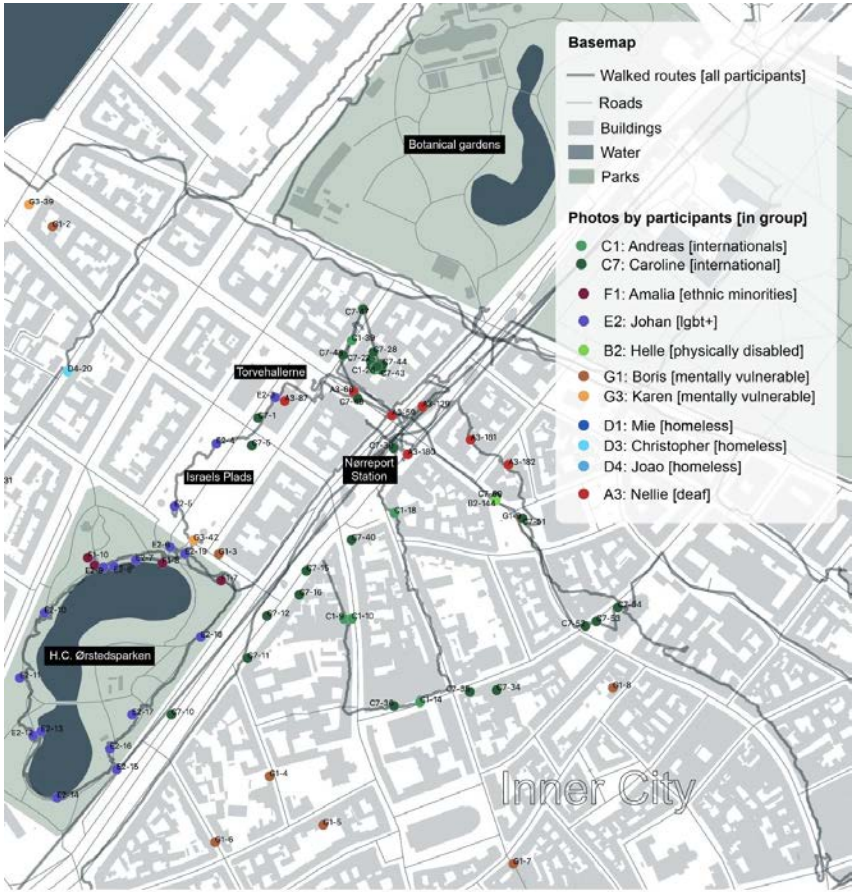
Internationals group

Physically disabled group

LGBT+ group



Mentally vulnerable group



S
T
U
Johan [E2]

V
W
X
Amalia [F1]

(S)

(T)

(U)

(V)

(W)

(X)

tantlab MASSHINE

Filters

Search for a user

Type a username...

Search for some text

Type a search query...

Select a user segment

british DDL DHF ENG HUGS LGBTQ MINO Researcher SIND

Do you feel this is for you?

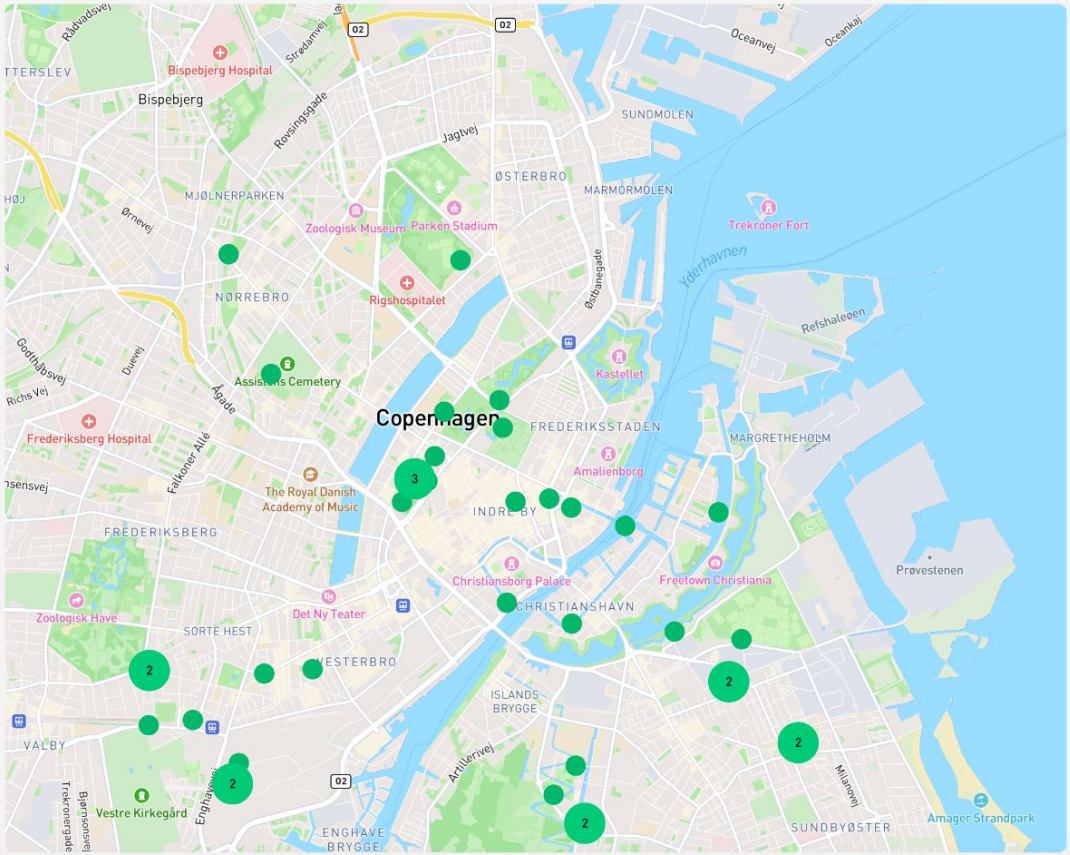
Not for me | Ambiguous | For me

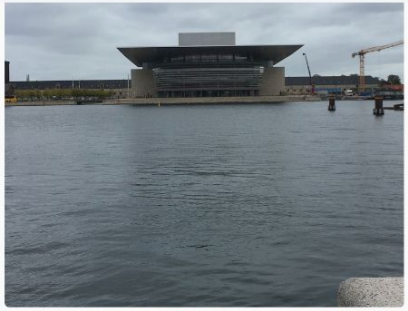
Do people agree on their response?

High agreement | High disagreement

What do you notice in this picture?

Walks Photos Areas User Segments





Author's Annotation DHF

Do you feel this is for you?

Not for me | Ambiguous | For me

What made you take this picture?

Architecture | Urban Environment | Culture | Consumption | People /community | Memories & associations | Atmosphere | God underholdning

User Reactions All user segments

Do you feel this is for you?

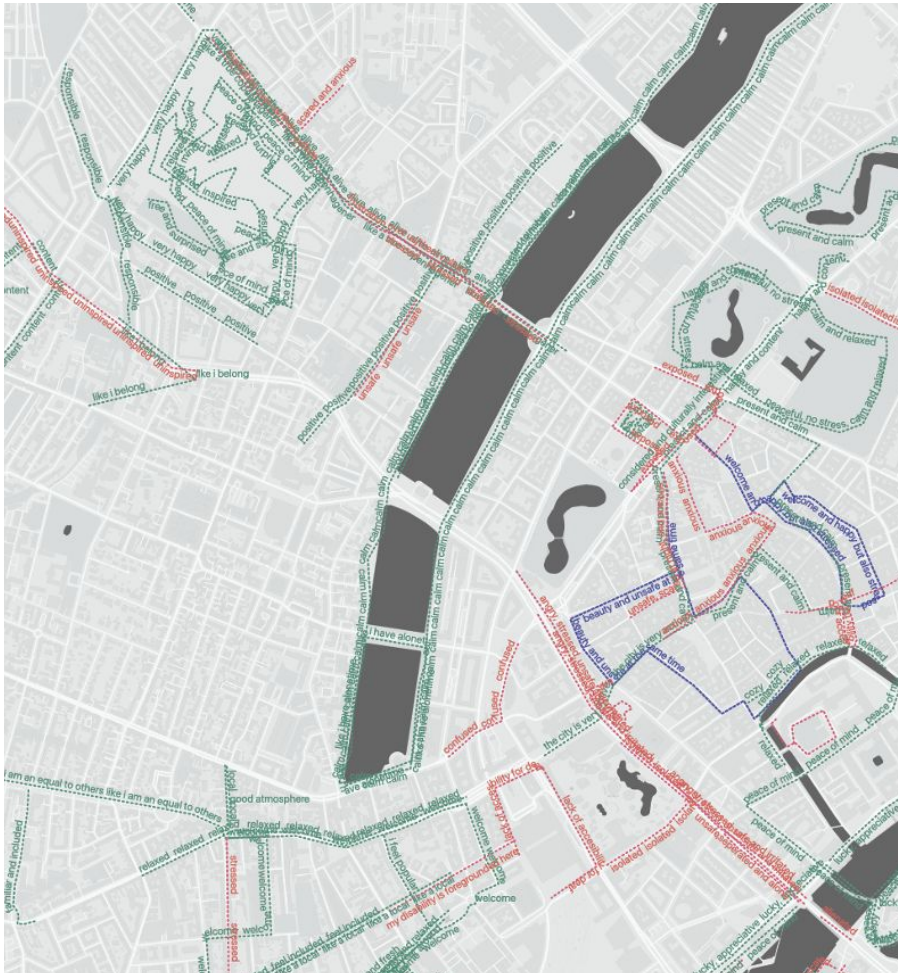
Not for me	2	Ambiguous	4	For me
25% (1)	25% (1)	25% (1)	25% (1)	0% (0)

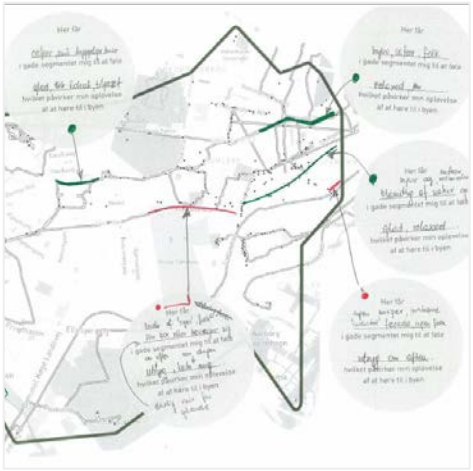
Participants doing the data interpretation

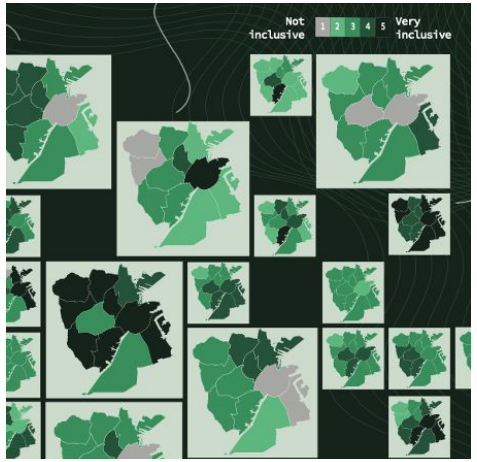
Crafting situated, subjective stories

Title:
More wheels than human

Description
This photo shows how you as a wheelchair user often gets send on detours in the city. As the sign shows, I have been thought of, but I have been thought of as a vehicle and not as a social being. I must move forward on the bike lane, a potentially dangerous situation. At the same time, I need to split up with my partner, friends, or others that I am out with. It feels unsafe and somewhat dehumanizing.


like i belong
good atmosphere
welcome
present and calm
peace of mind
relaxed




Not inclusive
1 2 3 4 5
Very inclusive

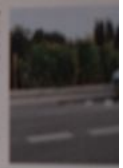
1
2
3
4
5
6
7

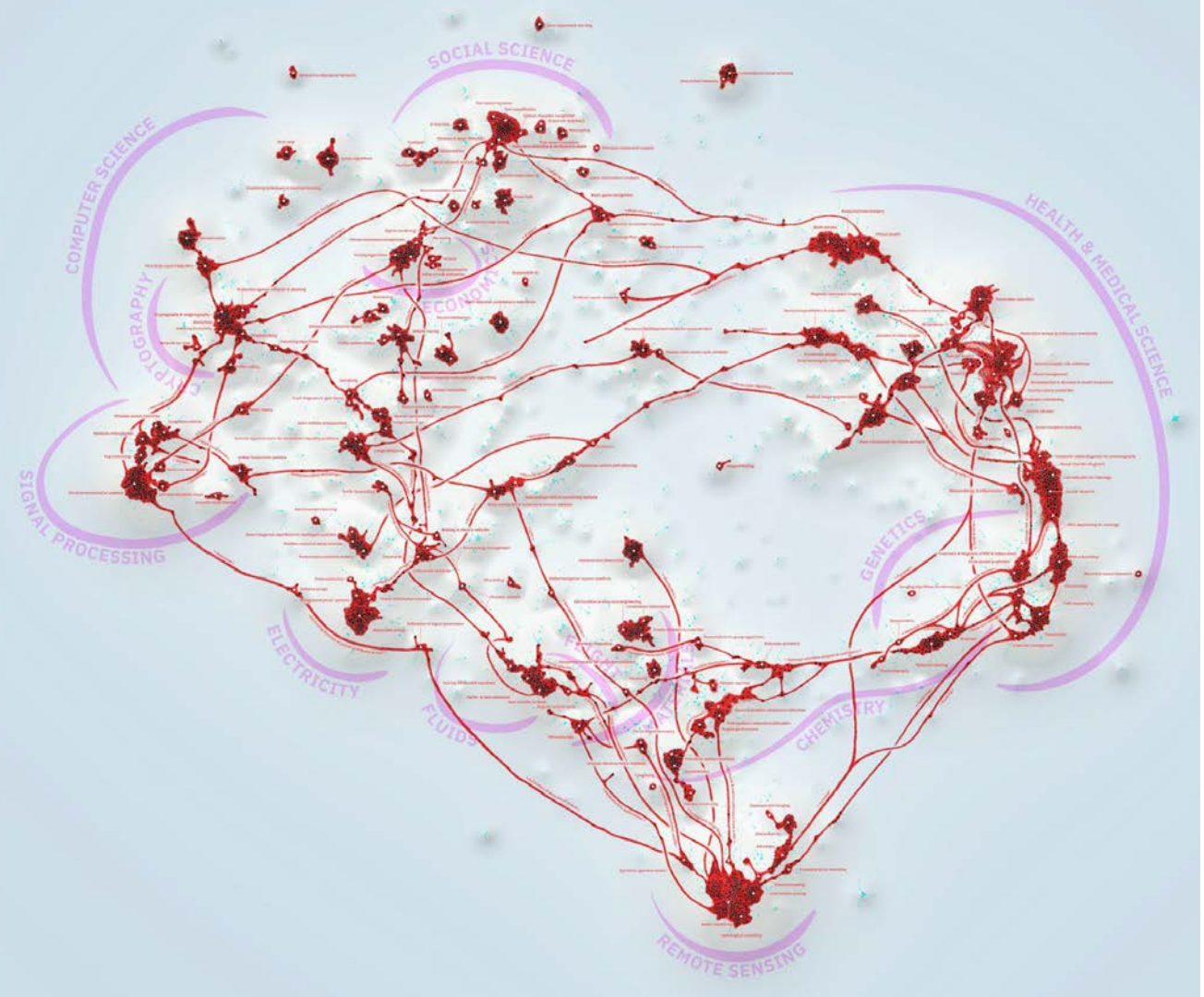
Seeing the city with the eyes of those who live in it

Over ten days, participants have captured 1459 photos of Copenhagen that show how the city is experienced from their point of view. Based on what matters most to them, participants have selected their favourite photos, and inscripted them with a personal story - narrated in their own voice.

The selected photos - and the stories that go with them - are presented here, divided in themes that center around: nature, place identity, sancturies of

Sancturies of nature

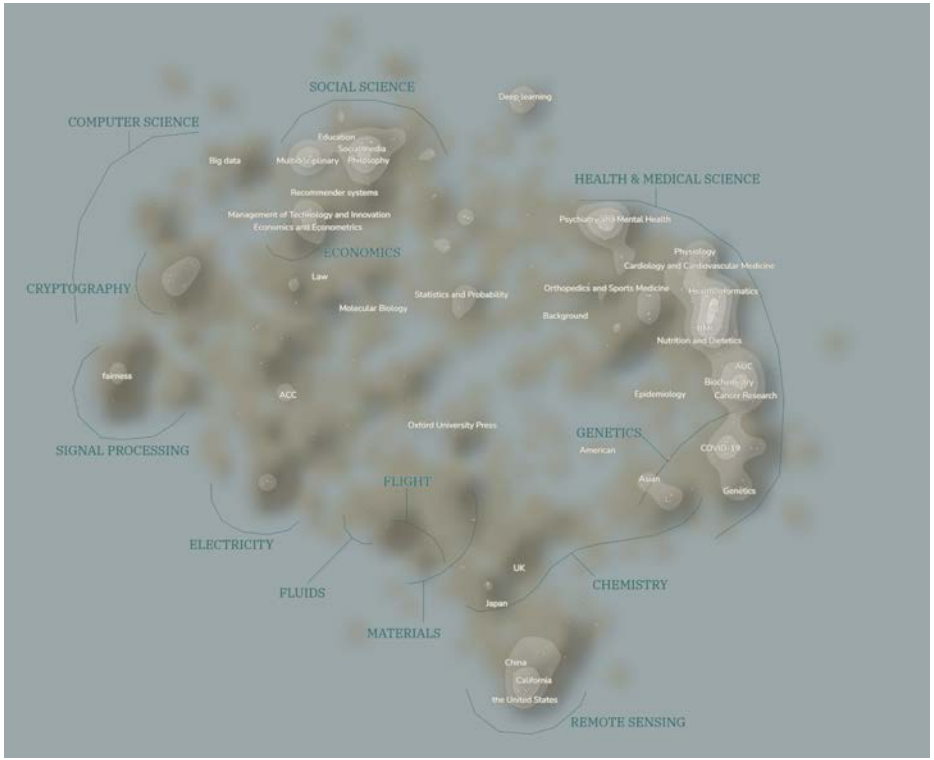


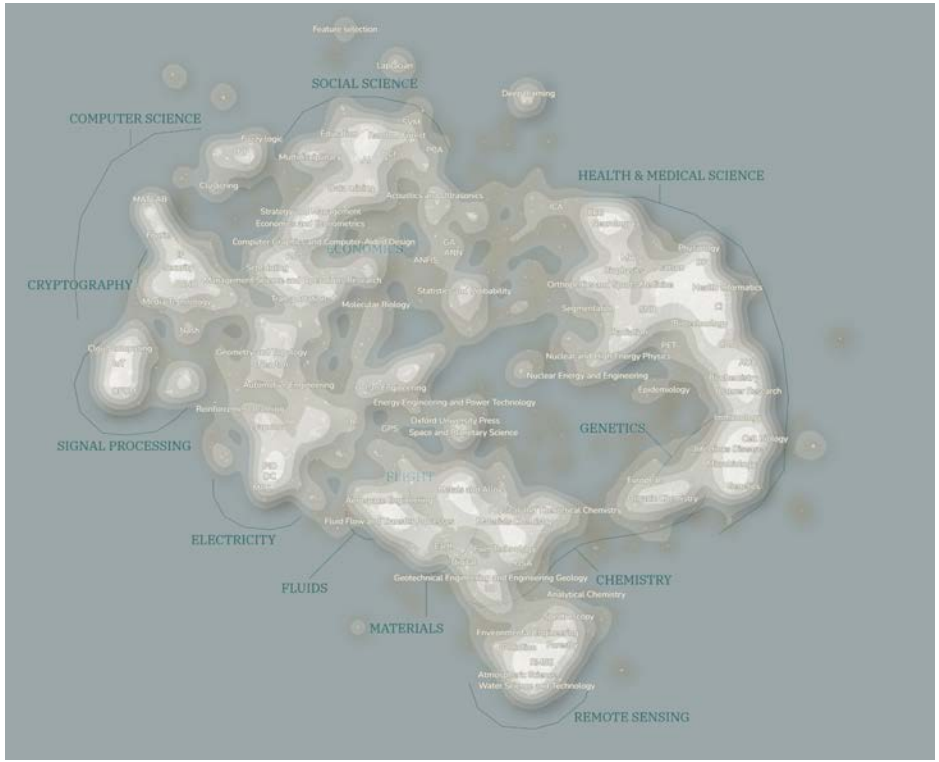


How does AI have agency in science & technology?

Corpus of 1M scientific abstracts about algorithms, machine learning, or artificial intelligence.

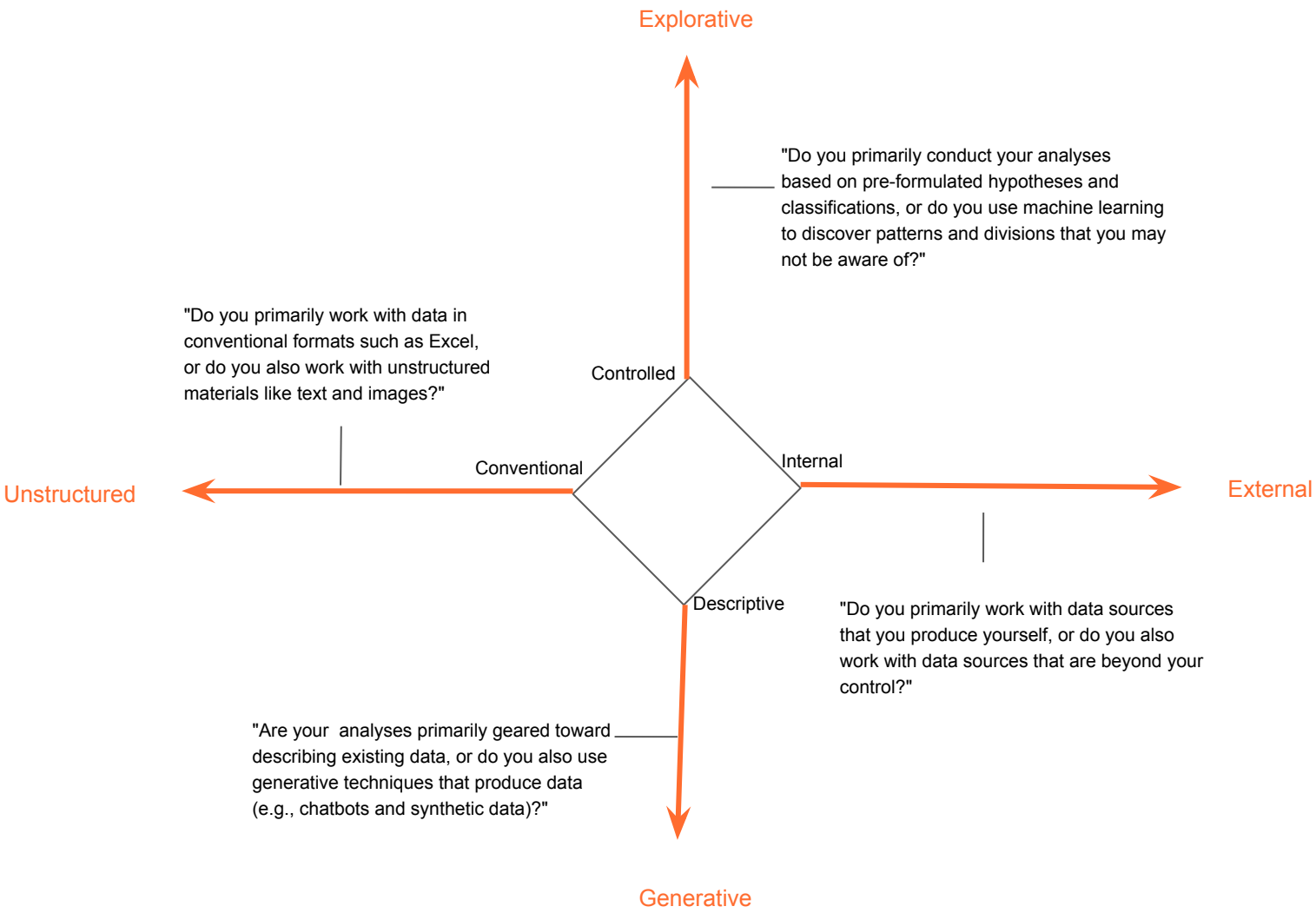
Mapped as a semantic network and qualitatively annotated.

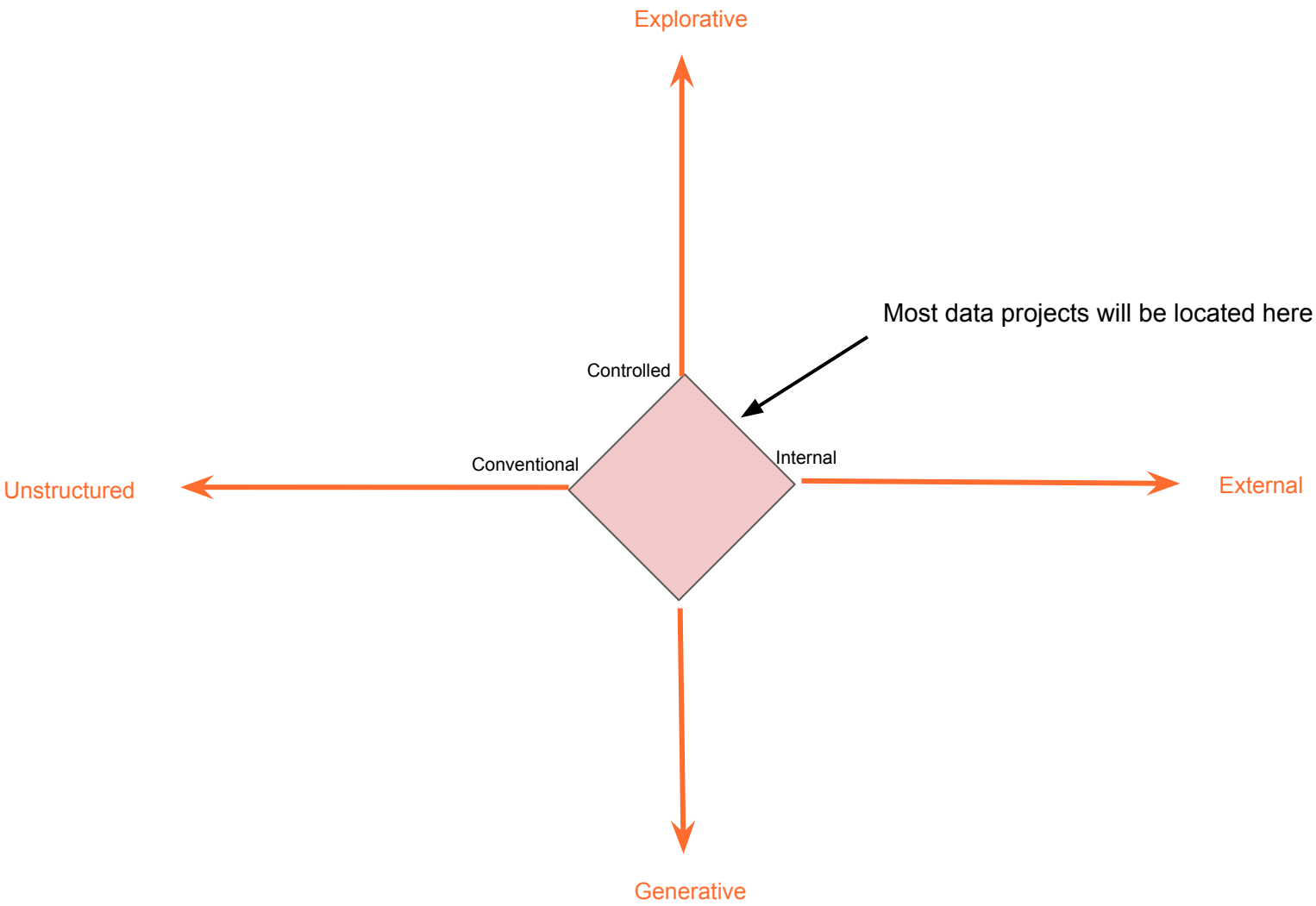




Distribution of articles about racial bias and discrimination (left) vs all articles (right) in the semantic network

A new data situation?

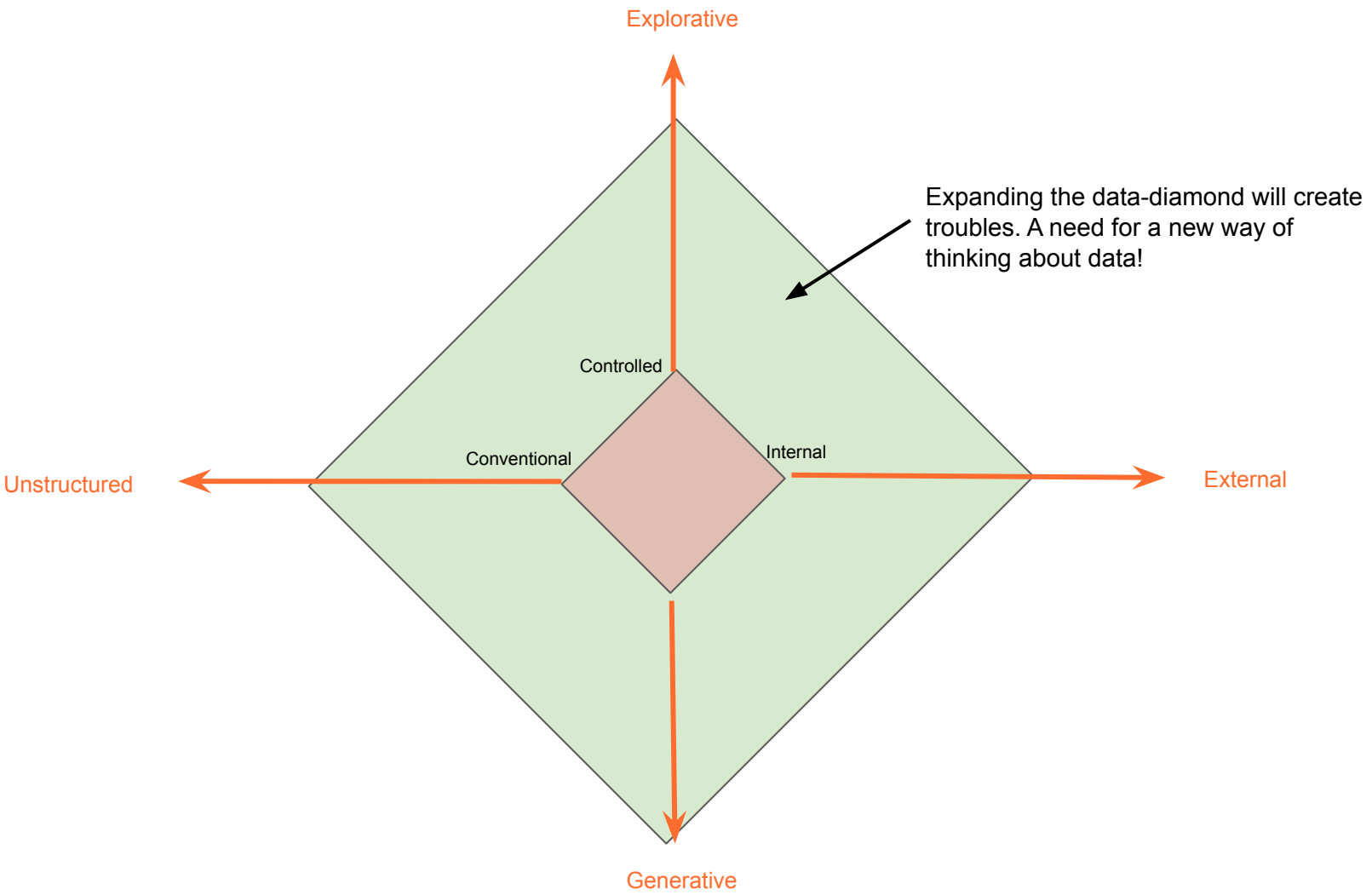

Explorative
"Do you primarily conduct your analyses based on pre-formulated hypotheses and classifications, or do you use machine learning to discover patterns and divisions that you may not be aware of?"
"Do you primarily work with data in conventional formats such as Excel, or do you also work with unstructured materials like text and images?"
Controlled
Conventional
Internal
Unstructured
External
Descriptive
"Do you primarily work with data sources that you produce yourself, or do you also work with data sources that are beyond your control?"
"Are your analyses primarily geared toward describing existing data, or do you also use generative techniques that produce data (e.g., chatbots and synthetic data)?"
Generative


Explorative
Most data projects will be located here
Controlled
Conventional
Internal
Unstructured
External
Generative

slido

Are your projects adventurous on the data diamond?

ⓘ Click **Present with Slido** or install our Chrome extension to activate this poll while presenting.

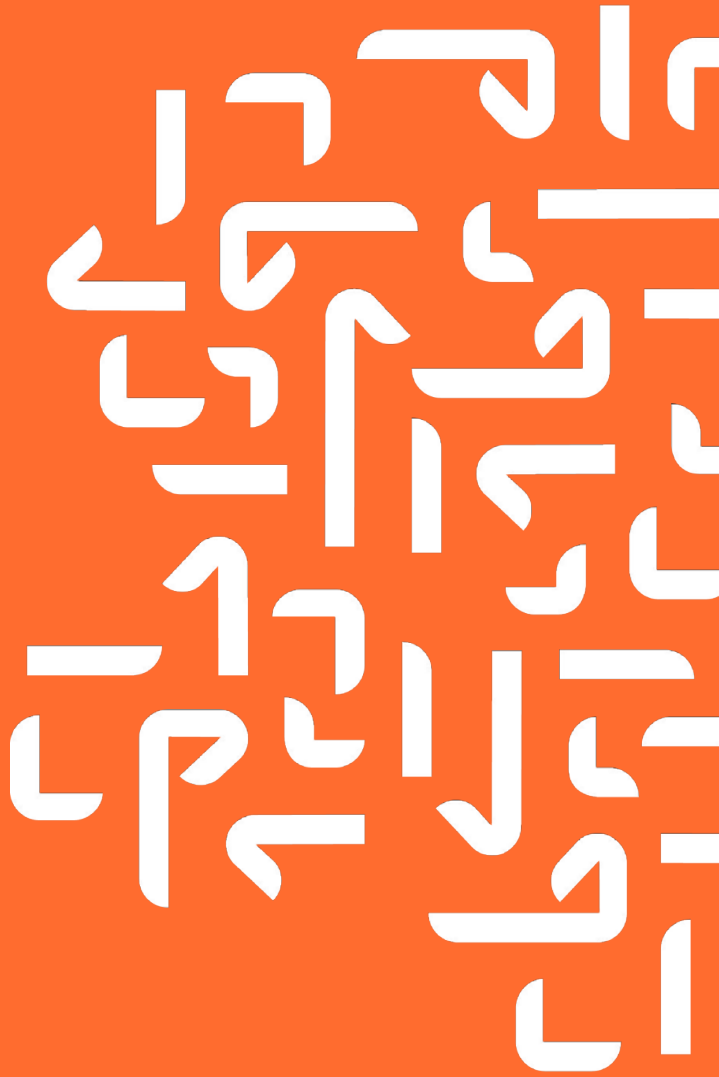

Explorative
Expanding the data-diamond will create troubles. A need for a new way of thinking about data!
Controlled
Conventional
Internal
Unstructured
External
Generative

3 organisational attention points

1. Liberate Data from NPM

General observation: We often use data to legitimize upward within a system of KPIs created for the purpose of comparing organizations. However, this logic makes data inflexible and detached from what makes sense on the ground in the specific organizational context. We (sometimes) need to liberate data from this frame.

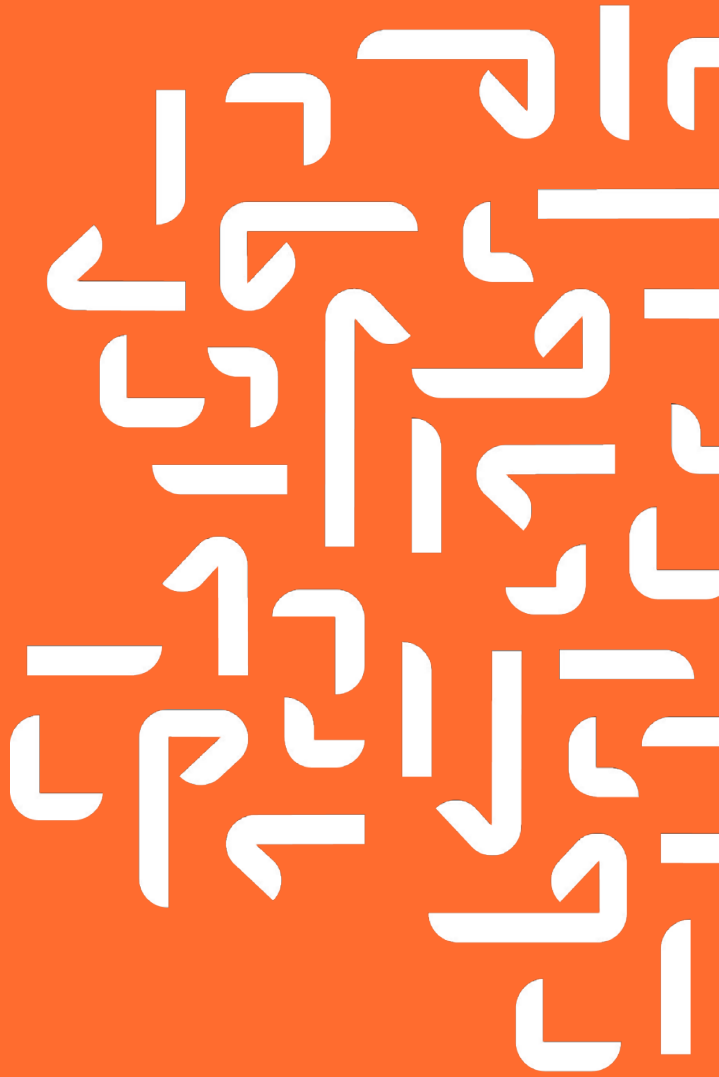
Specific case: The principal in the schooling system in a big Danish municipality felt locked in the standard surveys measuring the well-being of pupils. She used topic modeling on texts to source the pupils concerns and interpreted these with the pupils themselves. She sacrificed comparison but learned something about the unique situation at the school. AI was a sparring partner rather than a control mechanism.



2. Be aware of ‘shadow datafication’

General observation: If leadership does not develop guidelines for the use of new types of relevant data, "shadow datafication" from the bottom up can quickly emerge. It risks becoming a form of data practice that is difficult to keep track of and often takes on a life of its own within the organization.

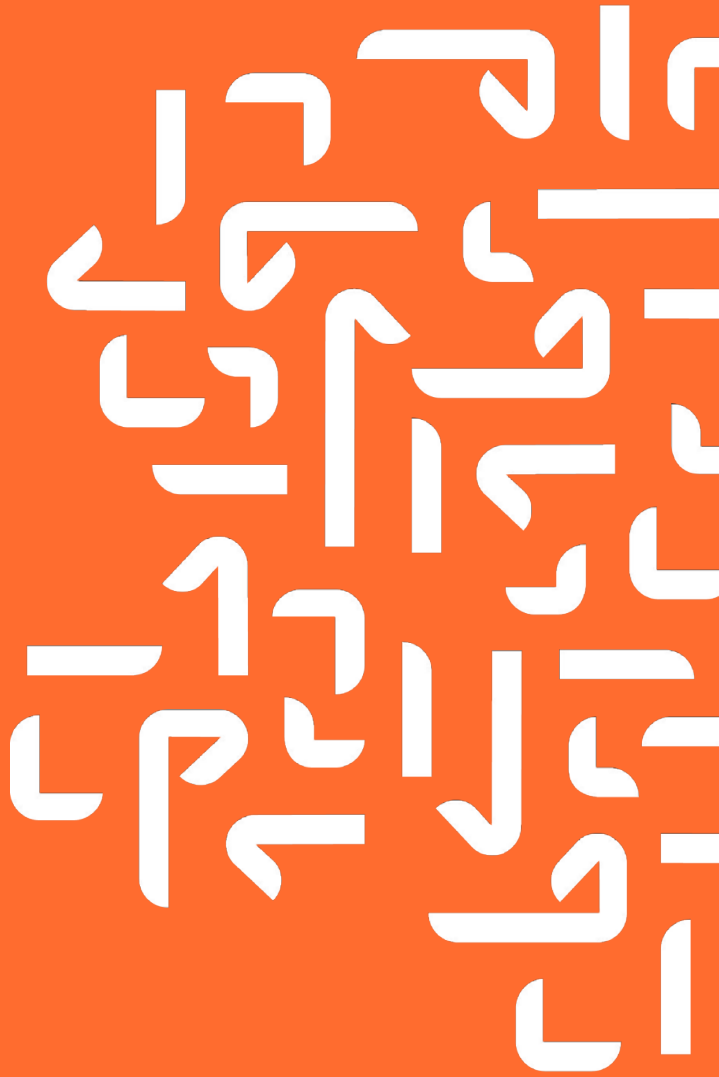
Specific case: In a larger Danish municipality, there was a lack of guidelines for using social media to gain insights into citizens. As a result, employees used openly accessible copies of Facebook as an 'ad hoc' database in their work.



3. Revisit the precautionary principle

General observation: There is a widespread tendency to proceed cautiously, especially when GDPR is involved. This means a general sceptical attitude towards ‘open source’ solutions and a tendency to ignore the consequences of NOT working with AI.

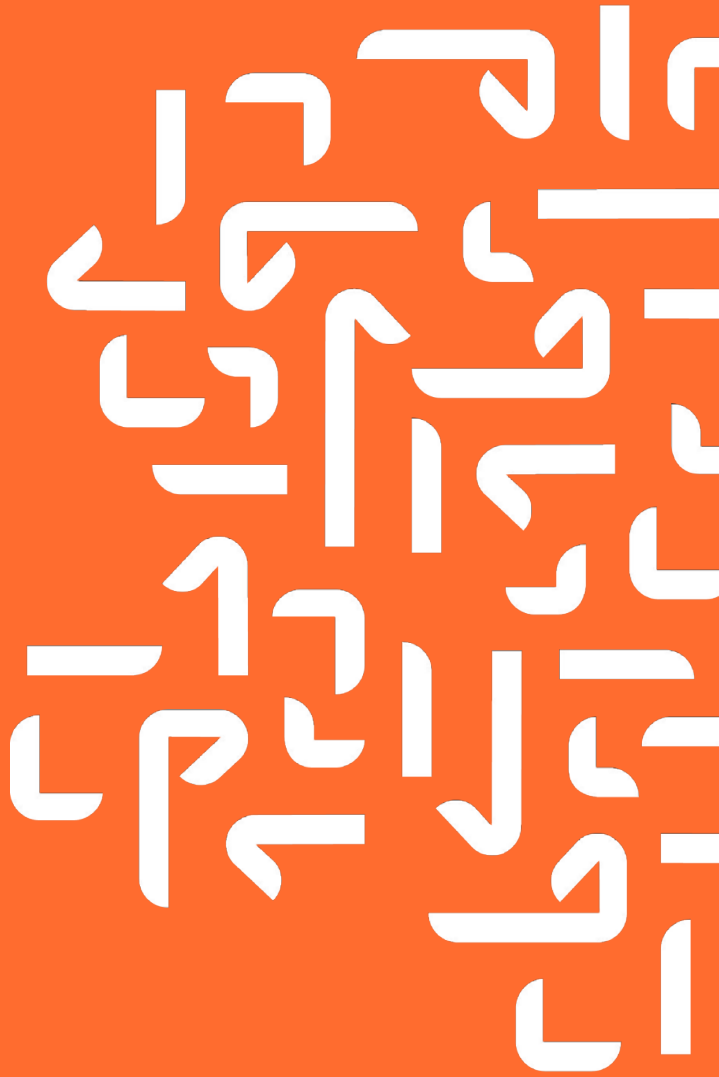
Specific case: Municipal projects working with data-driven methods often get blocked because of an institutional fear of breaking with legal and ethical standards. Some legal departments seem to over implement GDPR and relevant open source tools take a long time to get accepted. Something that makes shadows thrive.



4. Look at inherited Infrastructures

General observation: Organizations' work with data is heavily influenced by the friction between different 'inherited' systems and standards. It is often hard to break free of the infrastructure even though the will is there.

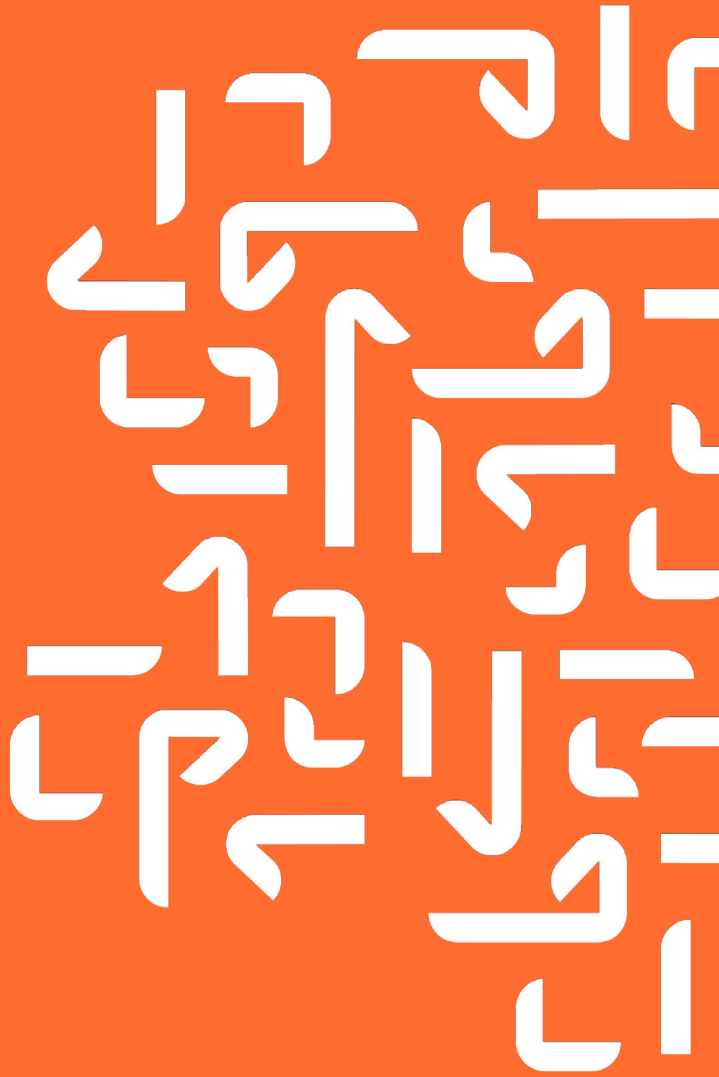
Specific case: In emergency services, alarm calls are reported in a system that primarily handles structured data. However, if an emergency response receives an incorrect code, ambulance personnel must provide feedback in free text and explain why they were dispatched incorrectly. The system cannot capture this learning.



5. Understand divergent ‘data-frames’

General observation: Even though employees may appear to support the idea of optimizing data usage, different groups of employees often interpret the meaning of data differently. Especially in the new data-situation. Divergent understandings of data among staff can hinder the translation of vision into practice.

Specific case: The City of Copenhagen's Environment and Technical Department has a 'smart city' strategy with broad initial support, but it is manifested differently in various sub-departments.



6. Rethink data-competencies

General observation: Data is not what it used to be. AI often relies on unstructured data in the form of texts and images with specific history and context. Programming is not what it used to be either. ChatGPT is good at writing code. Perhaps future skills will be more aligned with 'prompt engineering,' which involves having conversations with bots.

Specific case: When you recruit, consider that future data skills are partly in the realm of the humanities.

Join us...

Course module

15 ECTS for **Master in IT** on **Artificial Intelligence and Qualitative Methods**

Learn how to power your qualitative analyses with computational tools

First iteration Autumn 2024

Conference

What will generative AI mean for research methods in the SSH?

Keynotes by Ted Underwood, Jill Rettberg, Petter Törnberg, Katja de Vries and Jakob Stegelmann.

Registration: ssh.aau.dk/MASSHINE23