Scribe: Survation at scale

Autonomy Institute

EDICa, 13 May 2024

Autonomy’s DNA

The principles that define us as an organisation

Not for profit and independent

We work with progressive organisations and political parties of all kinds.

Funded by a mix of foundations and progressive consultancy work

Progressive

Focusing: freedom, equality, sustainability

Facing: emergent far-right, climate change and climate denial

Walking the walk

Forward-thinking workplace policies

Autonomy’s work

Autonomy develops research, policy and tools for our collaborators to pursue their mission

Research

Autonomy produces data-led research and tools to challenge economic inequality and navigate the climate crisis

Strategy

Consulting for different branches and scales of govt, non-profits, investigative journalists and trade unions to reshape policy and practice

Intelligence

Autonomy’s unique position in between policy and data science allows it to provide bespoke insights and tools for clients

Our approach to AI

Exploiting the novel emergent capabilities of language models whilst remaining cautious of the AI hype bubble

Lossless processing

Organise, categorise and hierarchically summarise without losing the original source data

Enhancing the human researcher

AI should augment how we research rather than do our research for us.

Leveraging ‘mediocre’ computing

Leave complex judgement to humans and leverage the ability of AI to achieve ‘mediocre’ human capability at scale

Large-scale processing of survey data using *Scribe*

What is Scribe?

Scribe is an in-house tool Autonomy has used for scalable qualitative analysis of survey data.

Using Scribe we:

- **Organize and cluster** large volumes of unstructured text (surveys, interviews, consultations
- **Surface emergent themes** from diverse voices, even in large datasets
- **Quantify qualitative patterns** while preserving access to the original text

Rather than replacing researchers, Scribe augments their interpretive work—expanding what's possible in inclusive, rigorous qualitative inquiry**.**

We are now further developing the methodology to support other organisations.

Research using Scribe

The technical steps of Scribe

- **Step 1:** Split up long responses into several smaller responses
- **Step 2:** *Cluster* similar responses by semantic (meaning) proximity
  - We do this by converting the text to an **embedding**
- **Step 3:** Summarise each cluster using an **LLM**

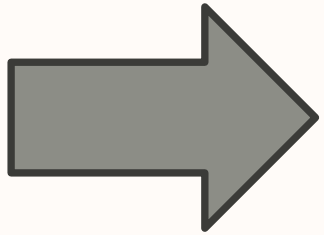
Clustering survey responses

Survey question:

What has the shorter working week changed in your work life?

Original response:

I am much more focused when at work, in order to ensure I finish my work during my working hours, as I know that if I don't, I'll be logging in on my day off to catch up, which totally defeats the point of it all. I love the shorter working week and feel that I get a lot more done in the time I'm in the business now vs before the trial/when I used to still work 5 days a week.



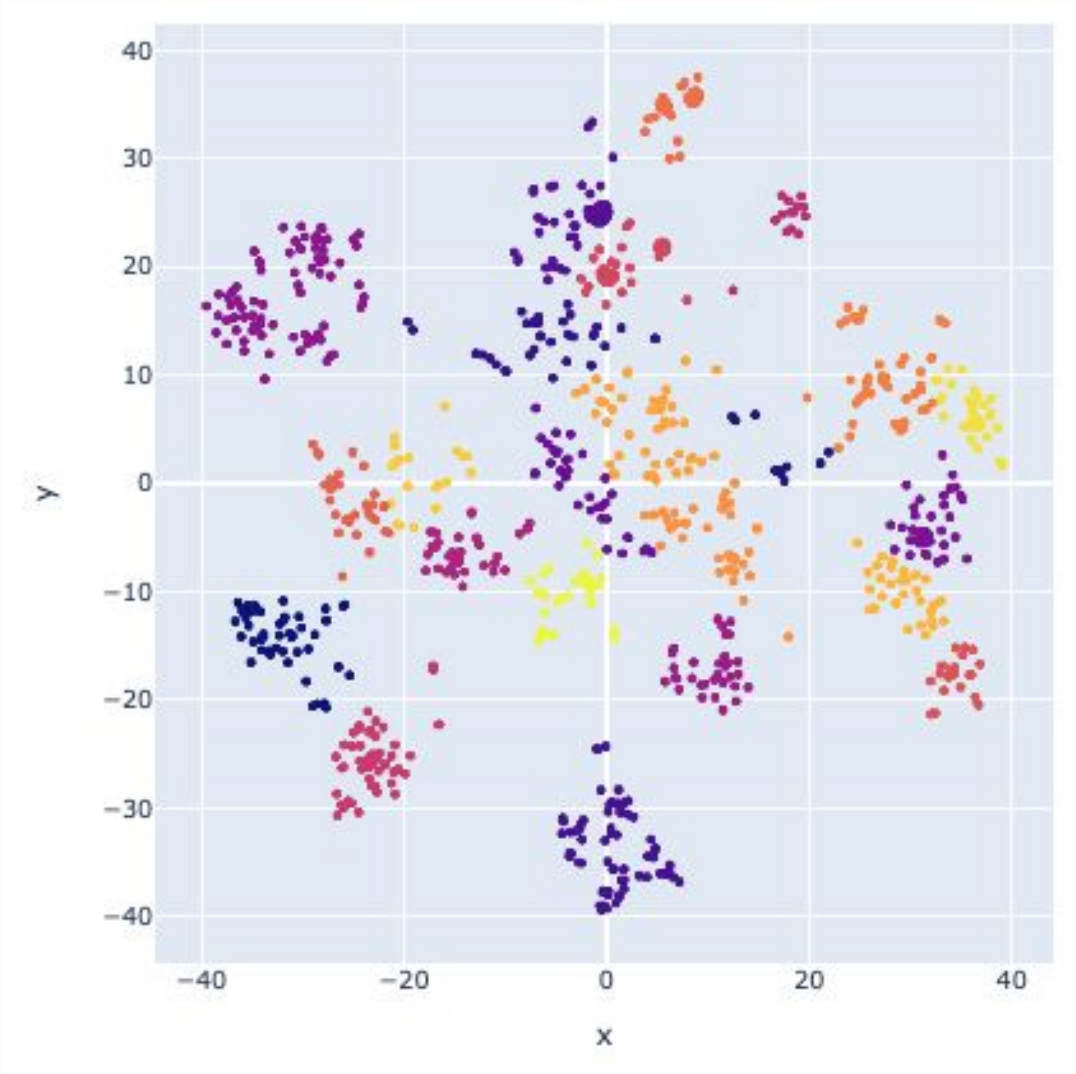
LLM pre-processing

Split response:

- I am much more focused during my work hours due to the decreased working week.
- I have to be more diligent with my time management to make sure I finish my work within the allotted time frame; otherwise, I'd have to catch up on my day off, which defeats the purpose.
- I strongly appreciate the shorter working week.
- I find that I am more productive now, finishing more tasks in the shorter working week as compared to when I was working 5 days a week prior to the trial.

Clustering responses

- By converting each textual response into an ***embedding*** we can project them into ***semantic space***.
- Simply put: This allows us to ***cluster*** responses based on their semantic similarity.



Semantic similarity - Brief example

Semantically identical:

"She gave the book to her friend."

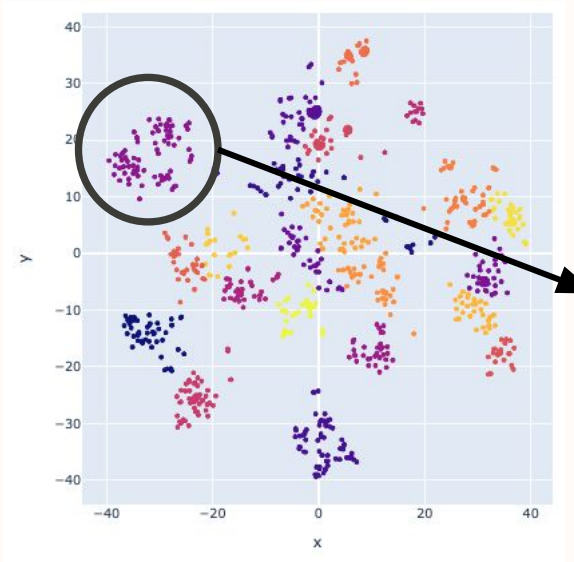
≈ "Her friend received the book from her."

Semantically similar:

"She handed over a book as a gift to her friend."

≈ "Her friend ended up with the book she gave."

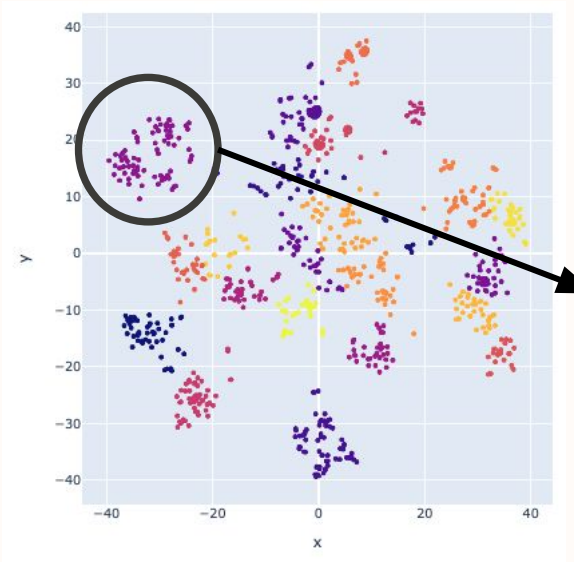
Clustering based on semantic similarity



Q: Thinking about your local community, what is the biggest change you'd like to see in the next year or two?

- "Improved road safety measures (crossings etc)"
- "More spent on road improvement"
- "improved infrastructure
- "Improved infrastructure"
- "Better roads"
- "Repair and renewal of infrastructure such as potholes and sewage works"
- "Improved road surfaces"
- "Road pot holes"
- "Potholes being repaired"
- "cleaner roads"x
- "Better improved roads"
- "Broadband"
- "Better roads and cared for public land"

Clustering based on semantic similarity



Q: Thinking about your local community, what is the biggest change you'd like to see in the next year or two?

- "Improved road safety measures (crossings etc)"
- "More spent on road improvement"
- "improved infrastructure
- "Improved infrastructure"
- "Better roads"
- "Repair and renewal of infrastructure such as potholes and sewage works"
- "Improved road surfaces"
- "Road pot holes"
- "Potholes being repaired"
- "cleaner roads"x
- "Better improved roads"
- "Broadband"
- "Better roads and cared for public land"

Summary:
The common theme among this subset of responses is the desire for improved road infrastructure, including repairing potholes, improving road surfaces, and implementing road safety measures. There is also a mention of better cycling infrastructure and reduced traffic congestion. Some respondents also mentioned the need for improved public transportation and broadband services.

Summarising clusters

The common theme among this subset of responses is the desire for improved road infrastructure, including repairing potholes, improving road surfaces, and implementing road safety measures. There is also a mention of better cycling infrastructure and reduced traffic congestion. Some respondents also mentioned the need for improved public transportation and broadband services.

The main theme of these responses is the desire for greater equality, specifically in terms of reducing poverty, homelessness, and food insecurity. Additionally, there is a suggestion to increase the availability of food banks and community dining options to support those in need.

The group of responses suggests that some people would like to see more elections, increased community events, improved safety, better racial equality, reduced immigration, increased funding for climate-related projects, and addressing issues with housing affordability.

The majority of respondents in this subset expressed uncertainty or lack of knowledge about the biggest change they would like to see in their local community in the next year or two.

The largest change desired by the local community is an improvement in healthcare services, particularly in terms of accessibility to doctors, dentists, and mental health services. There is also a desire for better integration between hospitals and care facilities, shorter waiting times, and increased availability of NHS dentists.

The majority of respondents in this subset of survey responses express a desire for better access to social services, more investment in local services and facilities, improved accountability and leadership from local councils and government, and increased support for the community and young people. There is also a desire for better public services, maintenance of the local environment, and addressing poverty and mental health issues.

Scribe for EDI - Potential use cases

Intersectional insights

Free-form text is richer and will allow for “unknown unknowns” in survey responses, as opposed to categorical prescribed surveys. However, traditional methods scale poorly with large volumes of free-form text. Scribe makes free-form text scalable.

Surface marginalized voices

Scribe helps ensure that underrepresented or minority perspectives are surfaced in large datasets, rather than being drowned out by majority trends. This can be combined with demographic data in addition.

Faster feedback loops for engagement

By reducing analysis time from weeks to hours, Scribe enables shorter cycles between data collection and community reflection, empowering more inclusive, iterative research practices.

Demo

Demo at:
https://static.autonomy.work/adu/public/scribe-app/