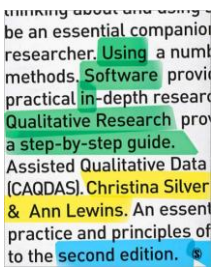
Generative AI for Qualitative Analysis

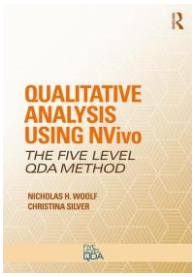
CARMA Online Short Course
2nd – 5th June 2025

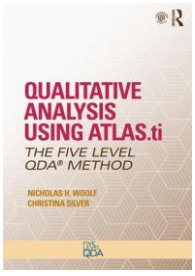
Christina Silver, PhD. FAcSS. SHFEA
Associate Professor (Teaching), Dept. Sociology, University of Surrey
Director of the CAQDAS Networking Project
Director of QDA Services

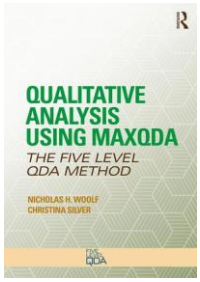
Linktree: https://linktr.ee/Christina_QDAS
Bluesky: @christinaqdas.bsky.social
Mastodon: @ChristinaSilver@sciences.social
Twitter: @Christina_QDAS
LinkedIn: linkedin.com/in/christina-qdas/
Instagram: @christina_qdas
Facebook: facebook.com/QDAservices
YouTube: youtube.com/@ChristinaSilverPhD
Podcast: anchor.fm/christina-qdas
Website: qdaservices.co.uk
Email: christina@qdas.co.uk

1









QDAS
qualitative data analysis services

qdas.co.uk

linktr.ee/Christina_QDAS

CAQDAS networking project
Computer Assisted Qualitative Data AnalysiS

Department of Sociology

University of Surrey

https://www.surrey.ac.uk/computer-assisted-qualitative-data-analysis

2

Course Overview

Day 1: Orientation to AI in the Qualitative Research Workflow

Day 2: AI for Research Design, Reviewing Literature & Data Collection

Day 3: Using AI for qualitative data analysis part #1

Day 4: Using AI for qualitative data analysis part #2
and Working qualitatively in the world of AI

3

Orientation to AI in the Qualitative Research Workflow

Qualitative Data Analysis background
The history of Computer-Assisted Qualitative Data AnalysiS (CAQDAS)
Genres of computer-assistance in the context of qualitative methodologies
Types of AI tool for qualitative research: what do they *actually* do?
The Generative-AI explosion and its impact on qualitative research practice
Balancing human interpretation with computer-assistance
The ethics of using AI for qualitative research

4

Starting points

Are you a skeptic?

worried about the potential dangers of using AI for QUAL?

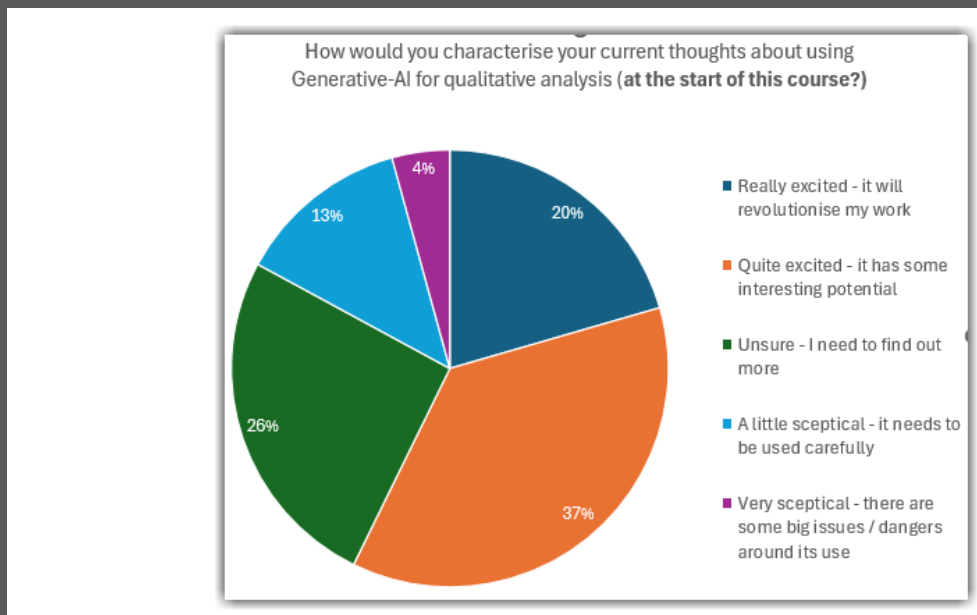
Are you unsure?

Need to know more before taking an informed position?

Are you an advocate?

excited by the possibilities for the practice of QDA?

5



6

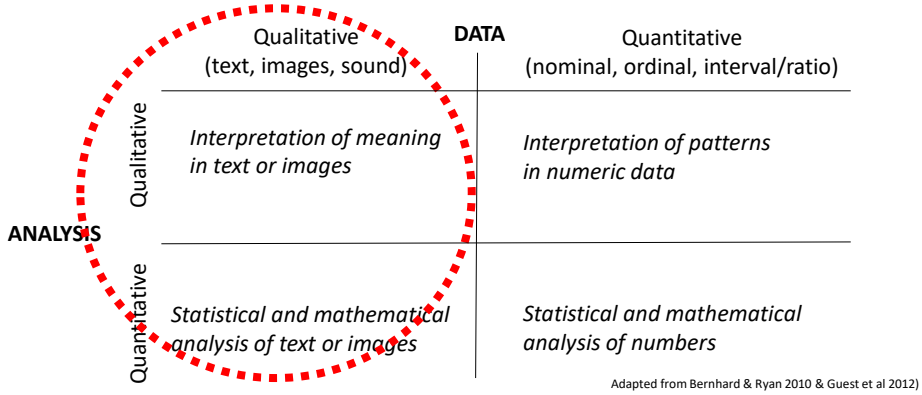
What is qualitative data and what do we do when we analyse it?

7

What is Qualitative Data Analysis?

Qualitative Data Analysis or Qualitative ***Data Analysis***

*"because of a quirk in the English language, the phrase "qualitative data analysis" is mischievously ambiguous". It can mean "**the analysis of qualitative data**" or it can mean "**the qualitative analysis of data**"* (Bernhard & Ryan 2010:4)



Adapted from Bernhard & Ryan 2010 & Guest et al 2012)

8

Qualitative research (Flick 2014)

Focus

- Seeks to understand concepts, experiences, behaviours, beliefs using non-numeric data
- Seeks to understand meanings

Strengths

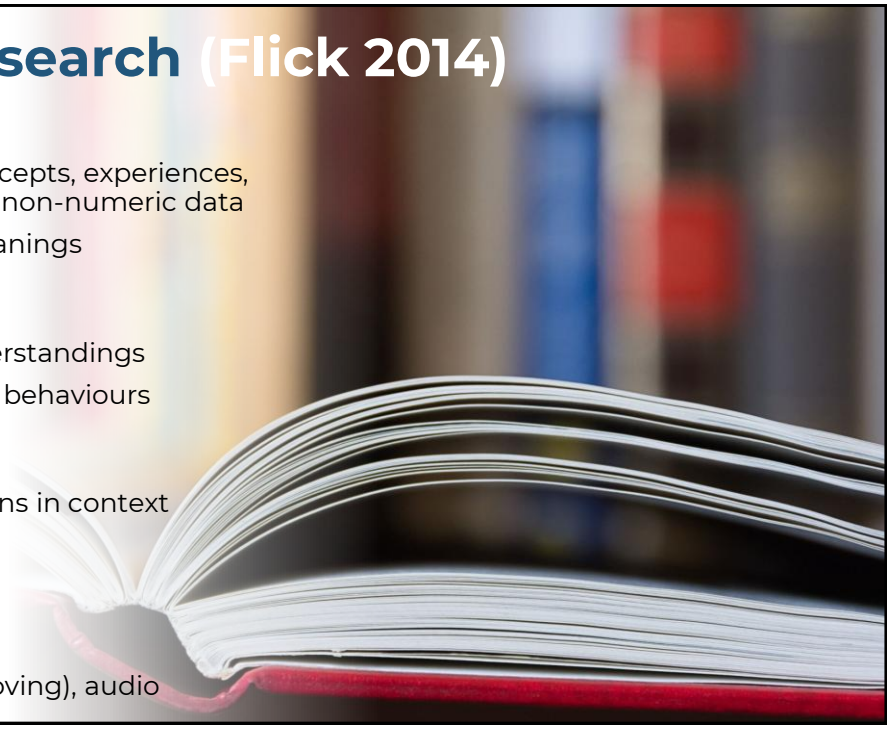
- Generates in-depth understandings
- Uncovers the reasons for behaviours

Uses

- Understand social relations in context
- Generate theories
- Understand complexities

Data

- Text, images (still and moving), audio



9

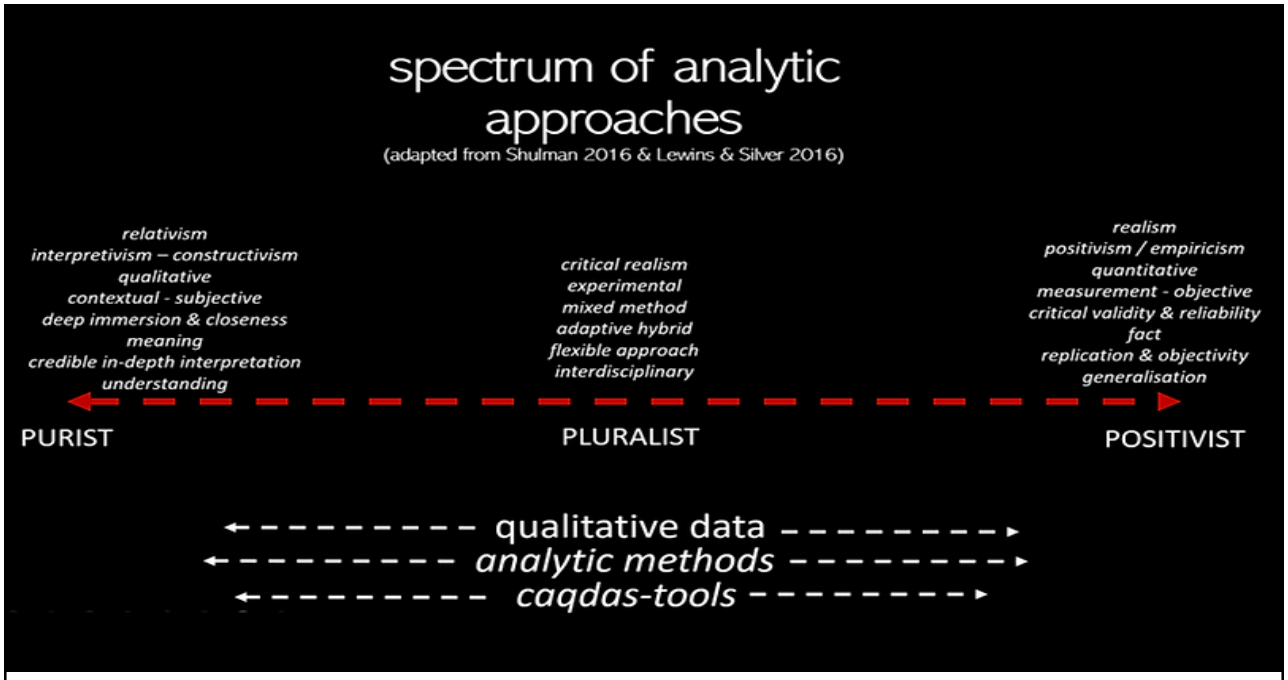
The Analysis of Qualitative Data

Historically the preserve of the human brain to "do" the analysis of qualitative data

Wolcott (1994) proposes 3 ways of "transforming data" into findings at progressively further "distances" from the data (aka "data reduction" – see also Miles & Huberman, 1994).

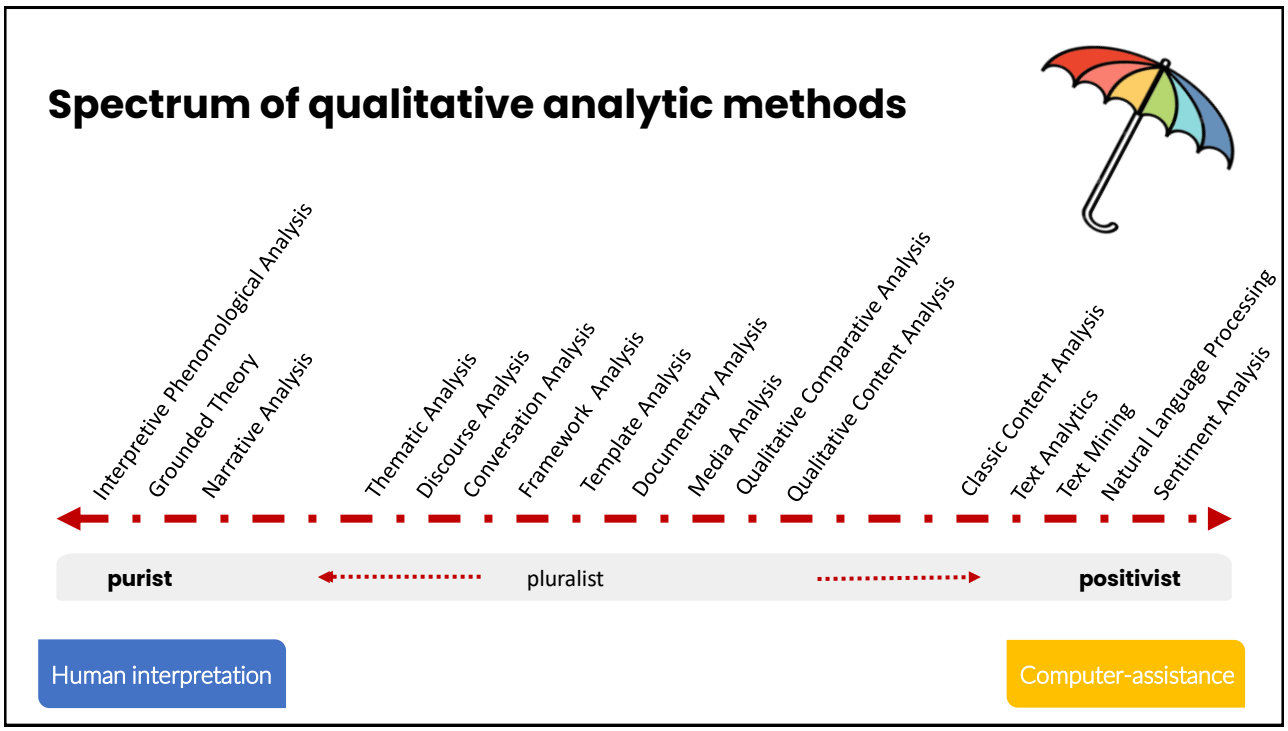
- description as "what is going on here"
- analysis as "how things work"
- interpretation as "what does it mean"

10

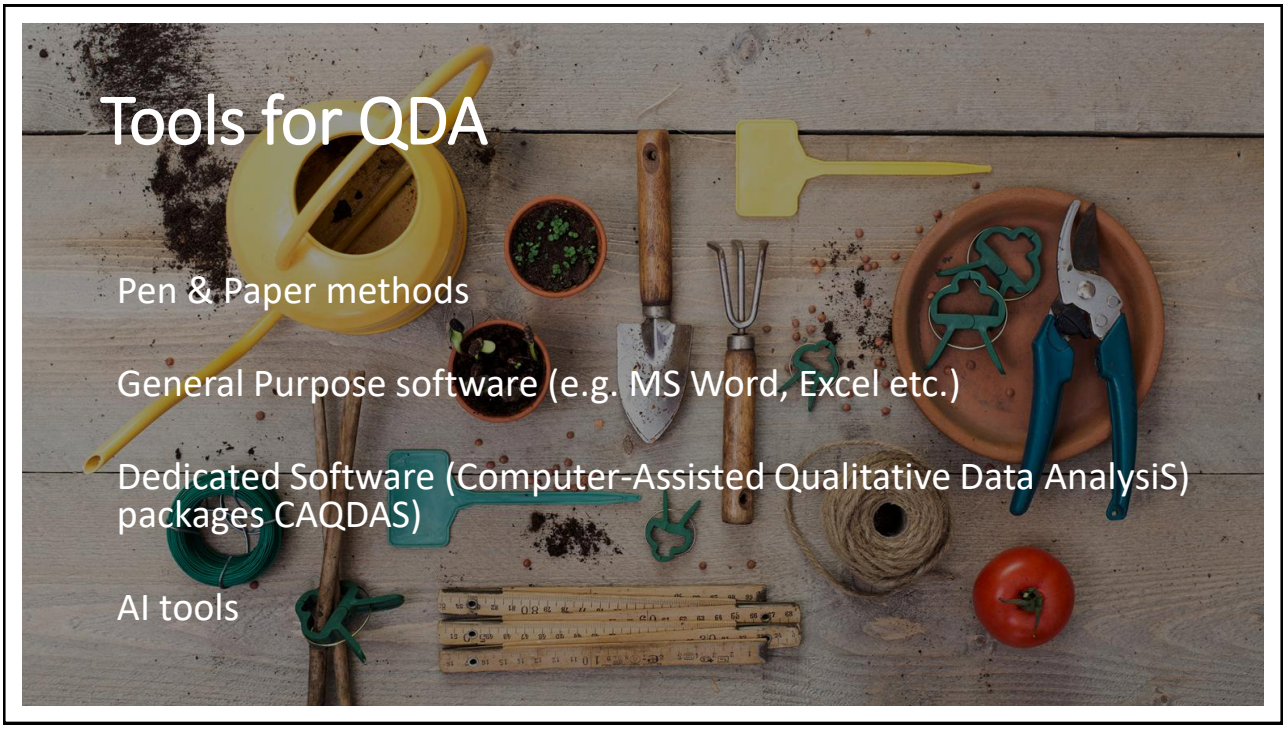


https://www.youtube.com/watch?v=wFabvZJUymI

11



12



13

Computer Assisted Qualitative Data AnalysiS (CAQDAS) packages

Tools designed to facilitate a *qualitative* approach to *qualitative* data.
May also enable handle quantitative (numeric) data and/or tools for quantitative and mixed-methods approaches to the analysis of qualitative data

Include some – but not necessarily all – of the following tools (and maybe others):

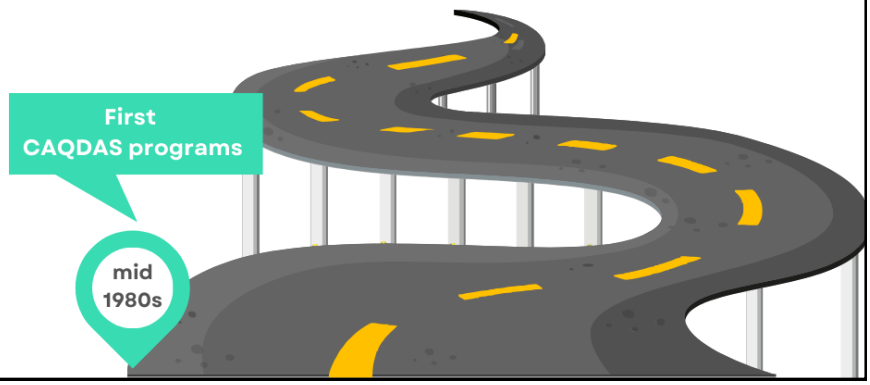
- Content searching tools
- Linking tools
- Coding tools
- Query tools
- Writing and annotation tools
- Mapping or networking tools

surrey.ac.uk/computer-assisted-qualitative-data-analysis/about

14

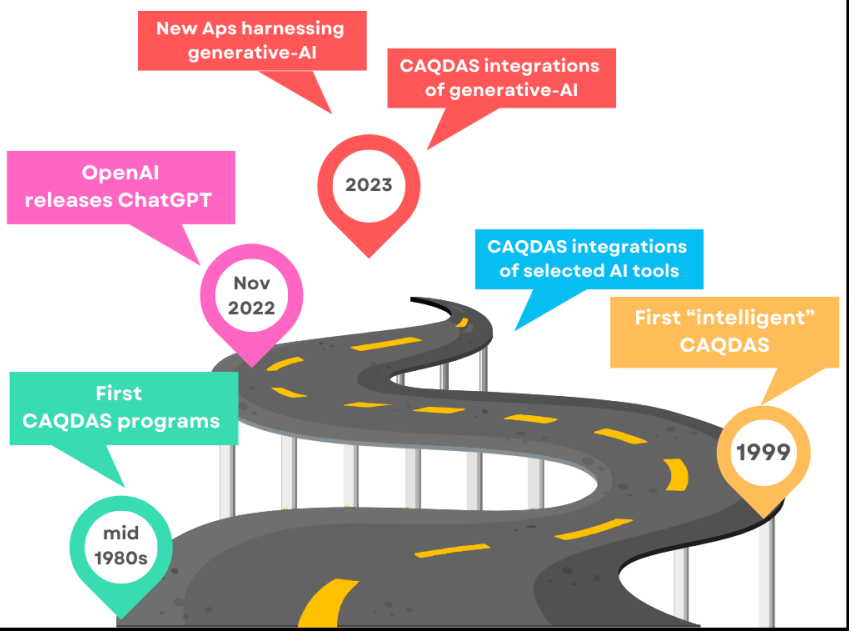
A little bit of history

CAQDAS = Computer Assisted Qualitative Data AnalysiS



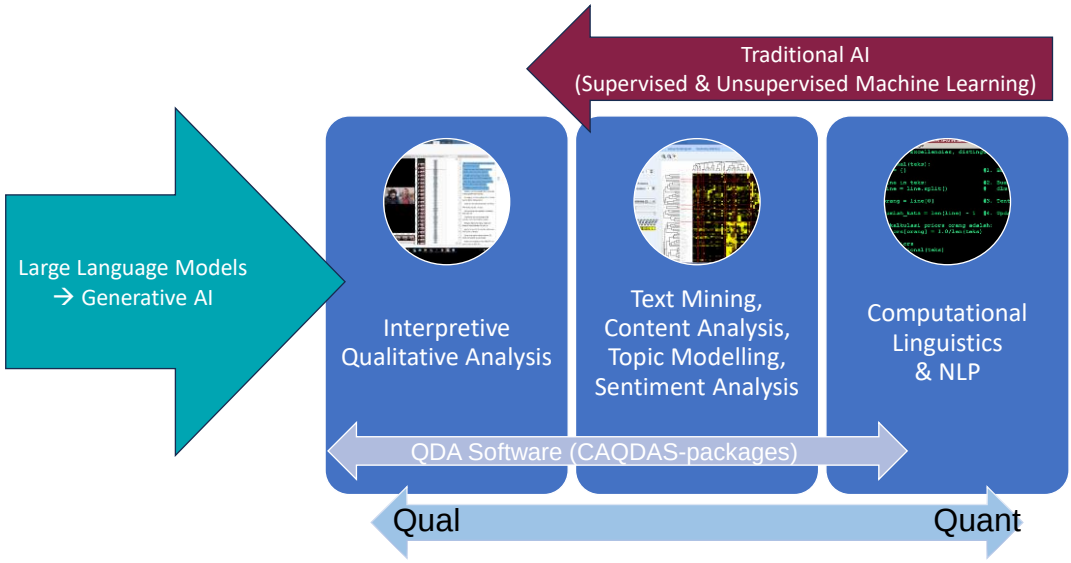
15

A little bit of history



16

Traditions of tools for text analysis



17

Technical Definitions of AI

- *At its simplest form, artificial intelligence is a field, which combines computer science and robust datasets, to **enable problem-solving**. It also encompasses sub-fields of machine learning and deep learning […]*

- *These disciplines are comprised of AI algorithms which seek to create expert systems which **make predictions or classifications based on input data.***

https://www.ibm.com/topics/artificial-intelligence

18

Generative-AI

- *"…refers to **deep-learning models** that can take raw data […] and "learn" to generate **statistically probable outputs when prompted**. At a high level, generative models encode a simplified representation of their training data and draw from it to create a new work that's similar, but not identical, to the original data.*

- *Generative models have been used for years in statistics to analyze numerical data. The rise of **deep learning**, however, made it possible to extend them **to images, speech, and other complex data types**".*

https://research.ibm.com/blog/what-is-generative-AI

19

Large Language Models

- *"Over the past few years, we have taken a **gigantic leap** forward in our decades-long quest to build intelligent machines: the advent of the **large language model, or LLM.***

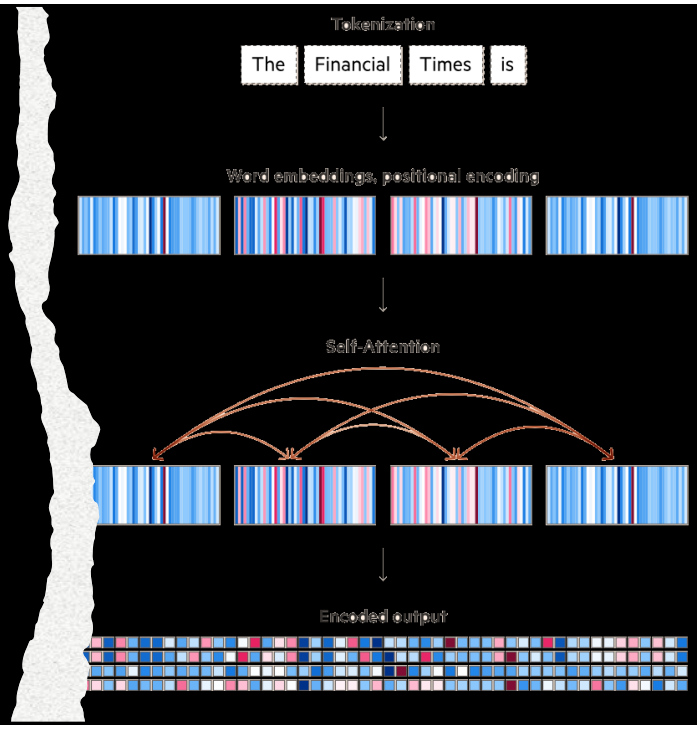
- *This technology, based on research that tries to model the human brain, has led to a new field known as generative-AI – software that can **create plausible and sophisticated text, images** and computer code at a level that **mimics human ability"***

https://ig.ft.com/generative-ai/

20

Generative-AI exists because of the transformer

https://ig.ft.com/generative-ai/



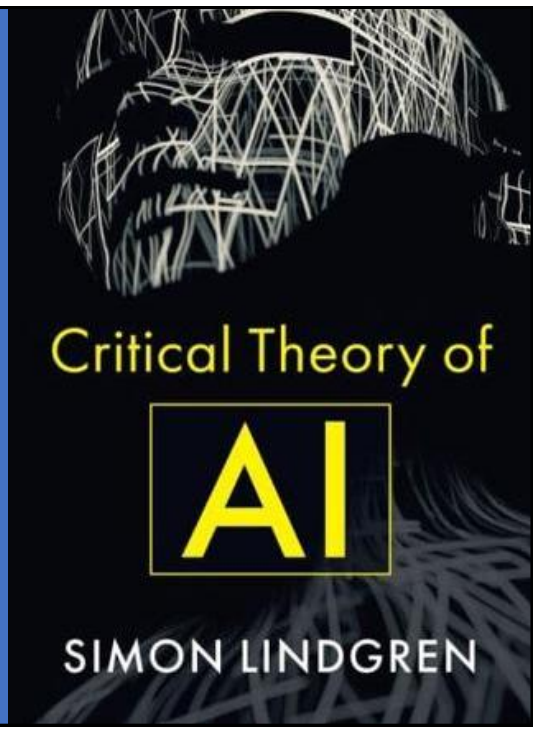
21

Critical Definitions of AI
Simon Lindgren

"AI as a socio-political reality....currently existing technology and its ramifications for people, politics, and power" (Lindgren 2024:4)

"...it is not an autonomous force" (p9) "it both ***consists*** of things, and ***does*** things" (p24)

"AI assemblage" = combination of technical aspects and contexts that surround and are impacted



22

Critical Definitions of AI: Emily Bender

In fact, this is a **marketing term**. It's a way to make certain kinds of automation sound sophisticated, powerful, or magical and as such it's a way to dodge accountability by making the machines sound like autonomous thinking entities rather than tools that are created and used by people and companies.

It's also the name of a **subfield of computer science** concerned with making machines that "think like humans" but even there it was started as a marketing term in the 1950s to attract research funding to that field.

I think that discussions of this technology become much clearer when we replace the term AI with the word "**automation**".

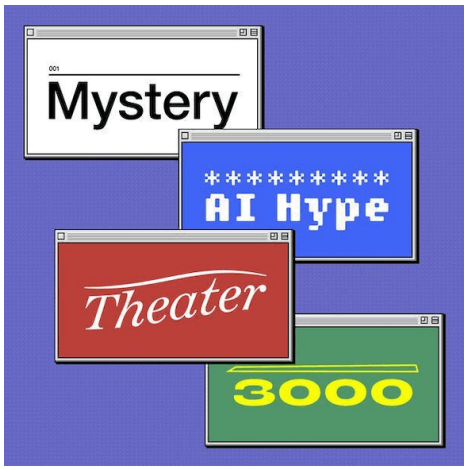
Then we can ask:

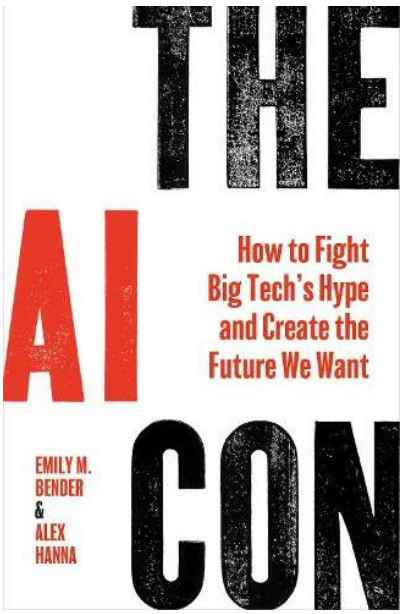
- What is being automated? Who's automating it and why? Who benefits from that automation? How well does the automation work in its use case that we're considering?
- Who's being harmed? Who has accountability for the functioning of the automated system?
- What existing regulations already apply to the activities where the automation is being used?

https://medium.com/@emilymenonbender/opening-remarks-on-ai-in-the-workplace-new-crisis-or-longstanding-challenge-eb81d1bee9f

23

Prof. Emily M. Bender Dr. Alex Hanna





24

Genres of AI in Qual

Traditional ("old skool") and Generative-AI

25

Definitions: Traditional-AI and Generative-AI

(explanation here)

Traditional AI (aka Narrow or Weak AI) are systems designed to respond to a particular set of inputs. They have the capability to learn from data and make decisions or predictions based on that data.

Generative AI [...] can create something new [...] from the piece of information you gave it. Models are trained on a set of data and learn the underlying patterns to generate new data that mirrors the training set.

26

Examples

Traditional-AI

- Exploratory Text Mining

  Application of Natural Language Processing (NLP), statistical and unsupervised machine learning techniques for rapid extraction of topics, themes and patterns in text responses: e.g. topic modeling, sentiment analysis
- Automatic Document Classification

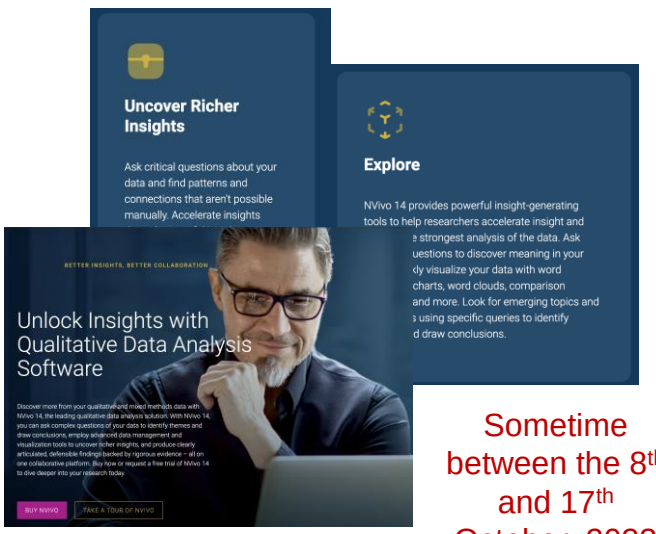
  Application of supervised machine learning on previously categorized text responses for the automatic classification of new uncategorized text responses.

Generative-AI

- General-purpose Chatbots
- New Apps Harnessing Generative-AI for aspects of qualitative research process
- Integrations of Gen-AI capabilities into existing CAQDAS-packages
- Bespoke tools developed in-house

27

WHEN did AI "happen" in Qualitative Analysis? Awareness and Discourses



Sometime between the 8th and 17th October, 2023

28

What's possible?

Current Developments in AI across the Qualitative Research Cycle

29

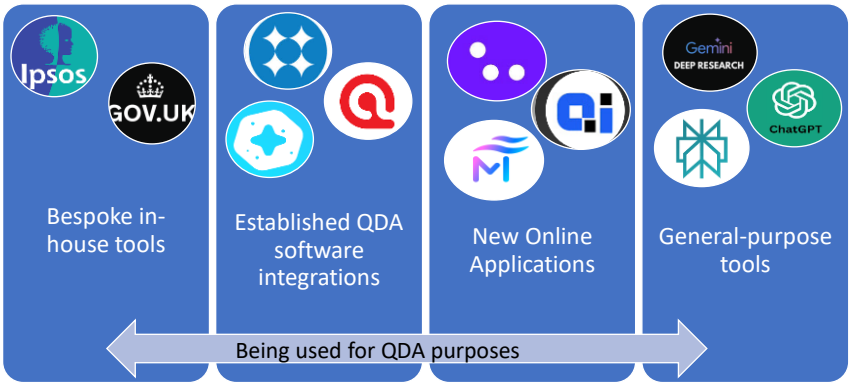
GenAI Capabilities afforded by LLMs in Qualitative Research

Generate	get ideas for projects & create data
Converse	ask questions via chat interface
Summarise	reduce data by condensing / synthesising
Convert	turning one thing into another → transcription of audio/video
Label	tag and categorise content

30

Christina Silver, PhD
christina@qdas.co.uk

Genres of GenAI for Qualitative Analysis



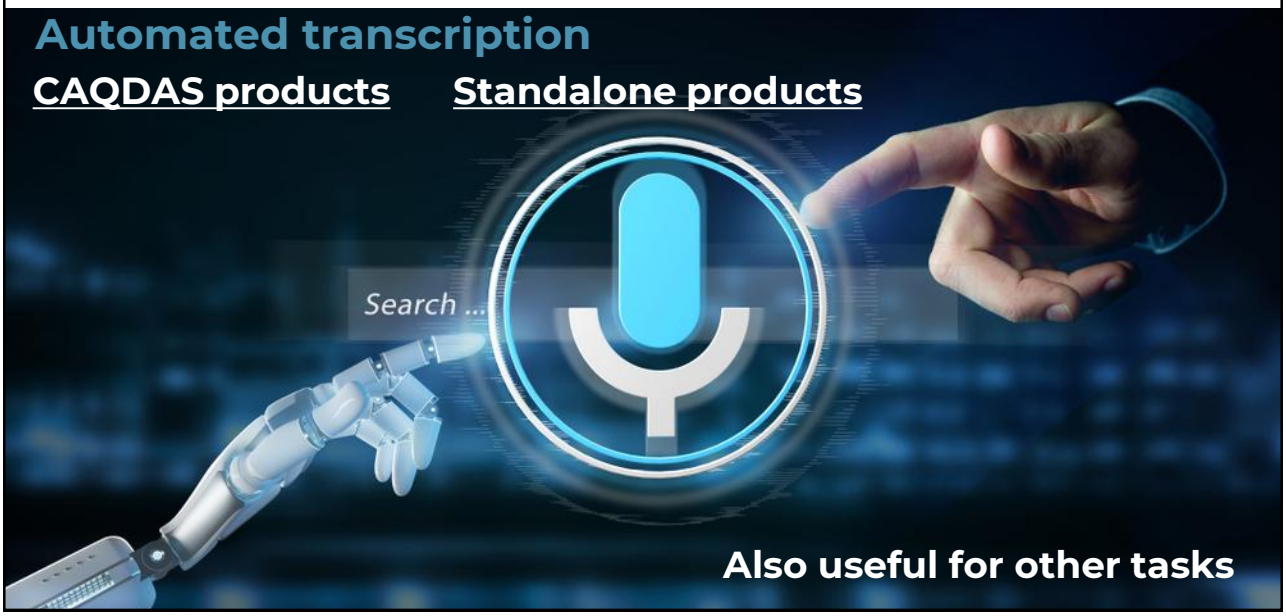
31

Generate (creating something new)

- **Generate ideas**
- **Generate data**
- **Generate "interpretations"**

32

Convert



33

Convert *speech-to-text* Automated transcription

Standalone products

- Numerous options: can be formatted for CAQDAS import

CAQDAS products

- NVivo transcribe (separate)

https://lumivero.com/solutions/aggregate/transcription-native-language-processing/

- Quirkos transcribe (integrated)

https://www.quirkos.com/learn-qualitative/qualitative-automated-transcription.html

- Transana (integrated)

https://www.transana.com/blog/2023/12/08/automated-transcription-comparing-models/

- MAXQDA (integrated & separate)

https://www.maxqda.com/automatic-transcription

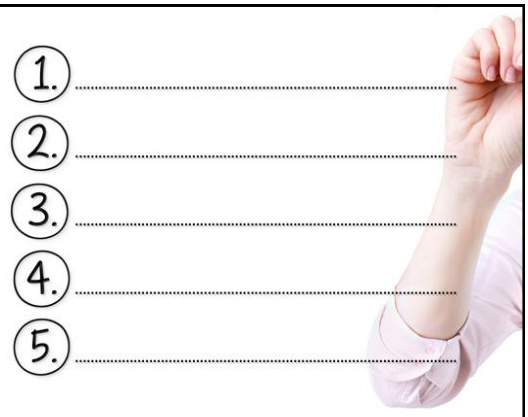
- ATLAS.ti (integrated)

https://atlasti.com/auto-transcription

34

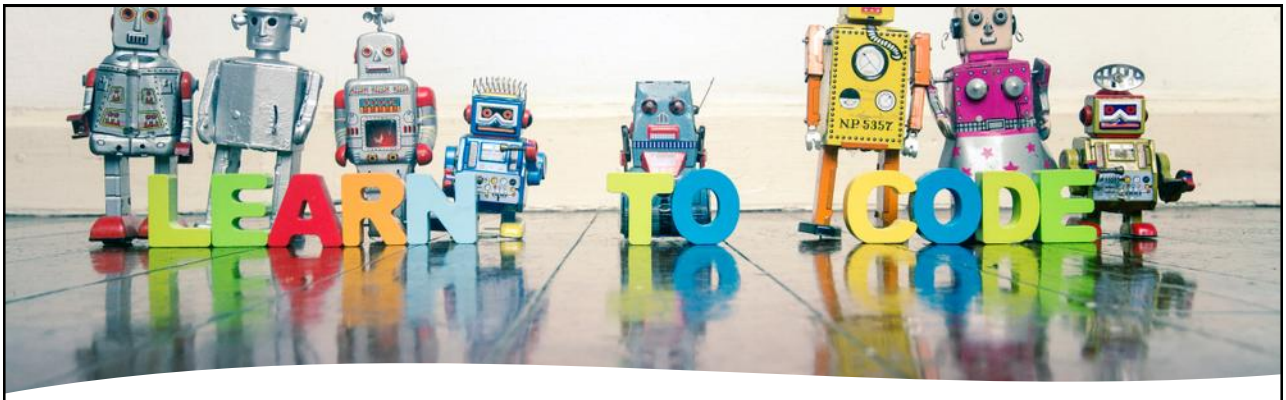
Summarise

- Very quickly will automatically summarise text
- Form of data reduction
- Different levels of summary can be generated
  - Amount of data
  - Length of summary
  - Format of summary



Enabled in standalone Chatbots (ChatGPT etc.) and some established CAQDAS packages (AILYZE, ATLAS.ti, CoLoop, MAXQDA, NVivo, Reveal)

35



Label

Categorise data in terms of what's "going on"

Qualitative coding: conceptual labeling
Factual grouping: organisational labeling

36

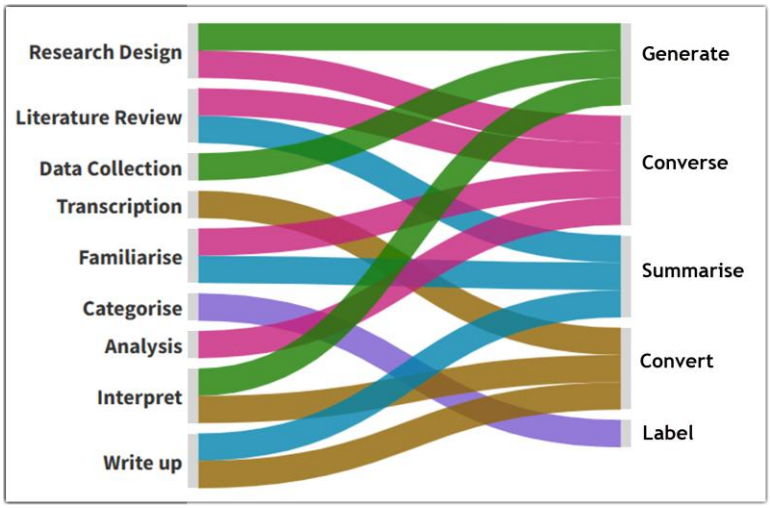
Converse: *Chatbots*

- Receive answers to any question – responses generated from AI large language models (LLMs)
- Conversational interface and experience
- Prompting is important

ChatGPT Claude Gemini Copilot perplexity

37

WHEN can GenAI be used in Qual Projects?



38

GenAI throughout the Qualitative Research Workflow

- ChatBots for ideation
- Deep Research

Research Design

AI interviewer tools
- via WhatsAP (e.g. Yazi)
- via Bots (e.g. AILYZE, Qualia)

Create 'silicone' or 'synthetic' data
- using ChatBots,
- AI Personas (e.g. Ipsos UK)

Literature Review

AI add-ins to existing Reference Managers
AI tools to find and summarise literature
(e.g. Litmaps, Scite, Aithor, etc.)

Data Collection

Transcription

- **General-purpose** transcription (e.g. Otter.ai)
- Auto-transcription as **part of data collection** (e.g. Teams and Zoom transcripts)
- Transcription **via QDA tools** (e.g. NVivo, MAXQDA, Quirkos, Transana)

Familiarise

Categorise

Analysis

Interpret

AI tools for qualitative analysis
- integrations into existing tools (e.g. ATLAS.ti, MAXQDA, NVivo, QualCoder, WordStat, QDA Miner)
- new Aps (e.g. AILYZE, Reveal Qinsights, Tailwind)

Write up

- **AI writing tools** for refining (e.g. ChatBots and bespoke AI writing tools e.g. Jasper, Copy.ai etc)

The risk of "qualitative deepfake"
qdas.co.uk/post/ai-deepfake-qualitative-analysis

39

Note; detail subject to change as tools develop

GenAI tools for QDA

and lots of others

Established CAQDAS-packages (more may appear)						New Online Applications
ATLAS.ti 1989/2023	MAXQDA 1989/2023	QualCoder 2019/2024	NVivo 1987/2024	WordStat 1998/2025	QDA Miner 2004/2025	AILYZE 2023 CoLoop 2023 Reveal 2024 QInsights 2024 Tailwind 2025
Transcription	Transcription		Transcription	Text Extraction Summarisation Topic Extraction Topic Naming Topic Grouping Syntactic phrase classification Named Entity classification Custom scripts	Text Extraction Summarisation Code Similarity Searching Table Filtering Follow-up AI queries	(Transcription)
Summarisation (multiple levels)	Summarisation (multiple levels)		Summarisation (multiple levels)			Summarisation (transcript level)
	Explanations		Explanations			
Code Suggestions	Code Suggestions		Code Suggestions			NO CODING
Coding (automatic & intentional)	Coding (1 doc/code at a time)	Coding (search & code-based)				
Chat (conversational)	Chat (conversational)	Chat (code/topic/ general)				Chat (conversational & grid)

Plus automatic analyses in some new APs, e.g. MyRA

41

Bespoke vs General Purpose? Methodological Considerations

	Bespoke tools (CAQDAS-integrations / new Apps)	General Purpose tools (e.g. ChatGPT, Claude, CoPilot, etc)
Qualitative analysis focus	• Project-level use of research context to guide model (e.g. new Apps) • Prompting to focus specific analytic interrogations (e.g. MAXQDA AI-summaries) • Optimised for qualitative analysis tasks: tailoring AI algorithms and functionalities to specific QDA needs, and data types	• Generalised tool with no focus on qualitative analysis needs
Analytic integration	• Integration with other qualitative analysis features (e.g. MAXQDA) or other research tools (e.g. CoLoop) • Usually good and interactive citations (validation) • Usually, can identify speaker roles and focus accordingly	• Some 3rd party plugins, may require technical know-how • Notoriously bad at accurate citation • No notion of speaker roles and unreliable at distinguishing them
Transcription features	• Lower word error rate (WER) (sometimes accuracy confidence specified/highlighted (e.g. CoLoop, Transana) • Speaker identification / labelling (in some cases influencing retrieval) • Formatted for analysis • Choice of models to balance speed / accuracy / privacy (e.g. Transana) • Keyword / phrase focusing, pronunciations (e.g. MAXQDA)	• Typically, lower accuracy • Not designed for research purposes so usually requires cleaning
Qualitative Coding	• Harnesses GenAI for purposes of qualitative coding	• Cannot code in traditional sense (cannot systematically link for gathering purposes)
Collaboration	• Specific tools for enabling collaboration	• Limited collaboration capabilities

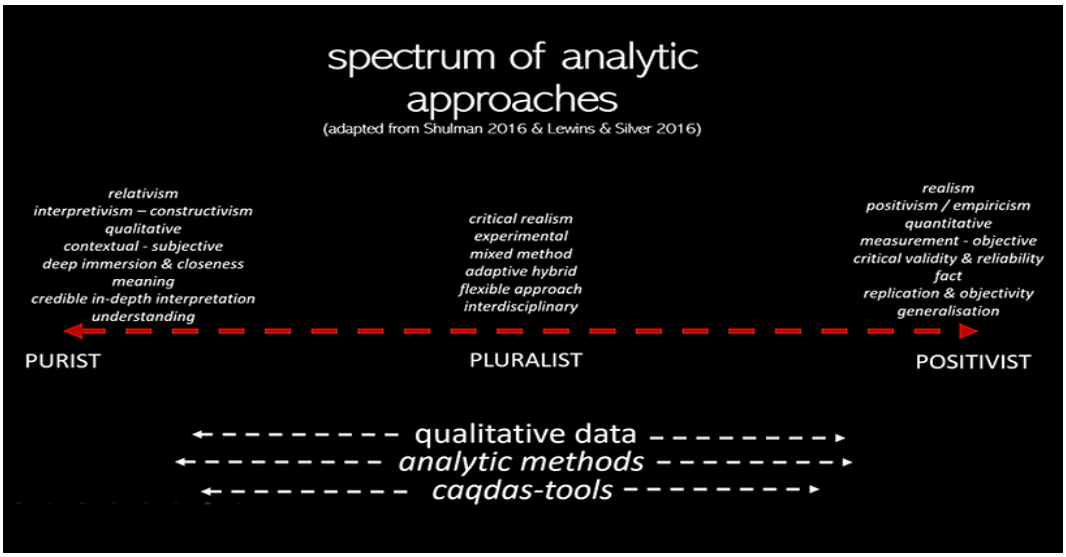
43

Bespoke vs General Purpose? Technical Considerations

	Bespoke tools (integrations / new Apps)	General Purpose tools (e.g. ChatGPT, Claude, CoPilot, etc)
AI Engines & Models	Not necessarily locked-in • Use of multiple available models (e.g. CoLoop, MAXQDA) • User-choice of models (e.g. WordStat)	• Uses models available via specified engine (e.g. ChatGPT uses OpenAI, etc.)
Hallucinations	• Typically built on Retrieval-Augmented Generation (RAG) systems designed to ground responses in data contained in the system • Responses usually directly traceable to source material (enhancing transparency)	• Trained to give an answer – may hallucinate to provide an answer • Often not reliably traceable
Context limits	• Restrictions in place to ensure maximum accuracy (e.g. MAXQDA amount of data that can be summarises/conversed with • Multi agent architecture such that context is handled dynamically (multiple sequential queries to different models to respond to a query (e.g. CoLoop)	• Fills entire context with as much info as possible regardless of relevance (max 100,000 words ~ 10 hours audio). • Good at answering questions about single documents but prone to hallucinate when working across many.
Data security	• Specific assurances aligned with research principles guaranteed. • Aligned with relevant regulations	• Big Tech companies frequently change T+Cs and are subject to competing interest for products
Support	• Support team conversant with research contexts and specifics of qualitative analysis	• No support

44

AI in the context of QDA



45

Signposting: Planning for the use of AI in QDA

Technical
- capabilities
- Tool transparency
- Analytic Integration

Ethical
- Development/functioning of LLMs
- Data privacy/security
- Legitimacy/ authorship

Methodological
- Type & sequence of assistance
- Units of analysis
- Interpretive intelligence

46

Navigating ethical issues in using Generative-AI

- Development of Large Language Models (LLMs) +
- Environmental Impact of using LLMs +
- Ethics Guidelines for QDA +
- Data Security when using AI tools for qualitative analysis +
- New ways of thinking about Informed Consent +
- Bias +
- Authorship and Publication +

[Ethics of AI in QDA] | [CAQDAS]

47

Ethical Considerations

Development of LLMs – stolen data (fair use?), model bias, exploitation during fine tuning, relevance to research-needs, model collapse

Environmental & Societal Costs of use – Energy & Water Distribution of environmental impact (systematic techno-colonialism) https://ai-impact-risk.com/ai_energy_water_impact.html

Litigation – unknown impact
https://www.hinckleyallen.com/publications/industry-today-ai-training-data-the-copyright-controversy/

Fairly Trained Certified

fairlytrained.org/certifications

48

Data Privacy and Security

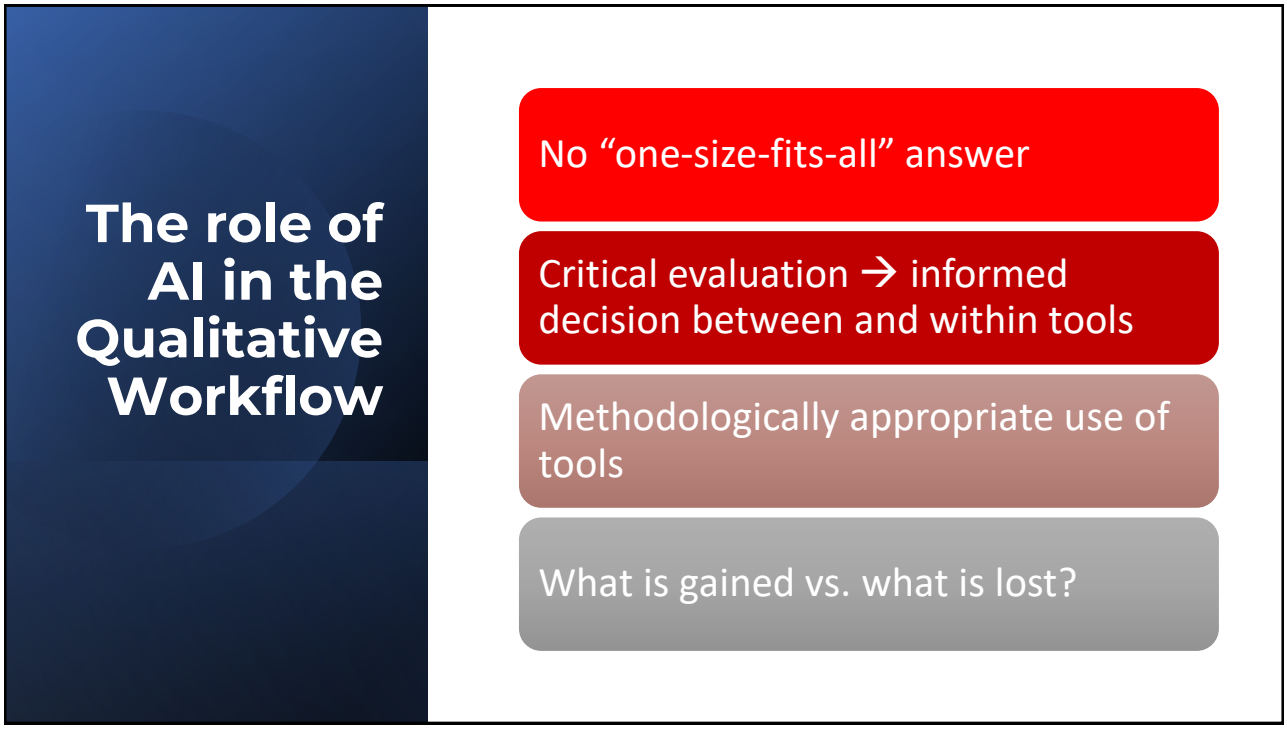
Privacy: (rights) expectations surrounding personally identifiable information (PII)
Security: (protection) from unauthorised access/use/risk

- Data collection, storage, analysis
- Anonymity, consent, explicit uses, encryption

Meanings of "informed consent"

49

50


The role of AI in the Qualitative Workflow
No “one-size-fits-all” answer
Critical evaluation → informed decision between and within tools
Methodologically appropriate use of tools
What is gained vs. what is lost?

51