Human-Computer Interaction

Qualitative Analysis

Grounded Theory + AI-Assisted

Professor Bilge Mutlu

Today's Agenda

- Topic overview: *qualitative analysis, grounded theory & open coding*
- AI-assisted coding and how to audit it
- Hands-on: *open-code a transcript*
- Assignment + project status

Qualitative data analysis methods

- Content analysis
- Discourse analysis
- Narrative analysis
- **Thematic analysis**
- **Grounded Theory**

What is Grounded Theory?[1]

- An *approach* for areas where little is known, or for a fresh take on the known
- A *method* to systematically build integrated sets of concepts from data
- A *process* of interplay between the researcher and the data
- A *theory* shaped by the data, as the data are shaped by the researcher

[1] Strauss, A. & Corbin, J. (1998). *Basics of Qualitative Research* (2nd ed.). Sage, pp. 11–15, 42.

The analytic stance[2]

Muller: *curiosity, creativity, and surprise* as analytic tools

- **Abduction:** find a surprise, then try to explain it
- **Constant comparison:** data with data, *and* data with the emerging theory
- **Open coding:** codes not governed by prior theory
- **Surprise** says where to look next; none says you're done

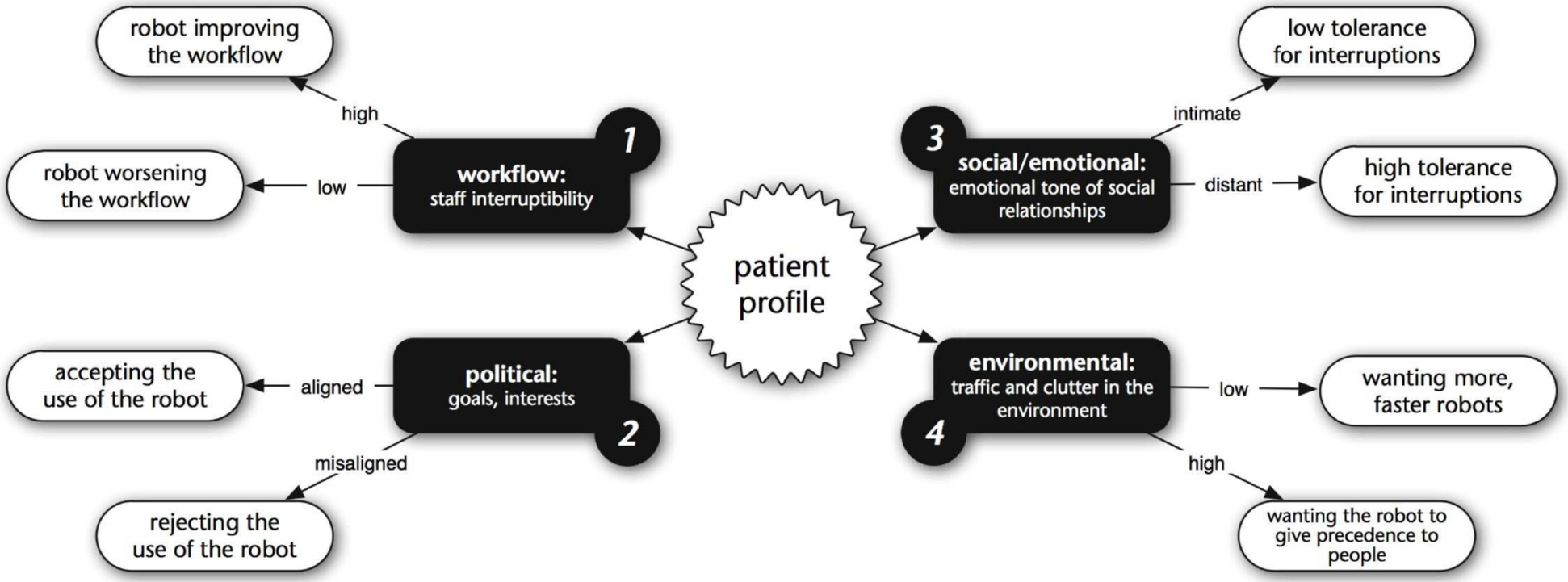
[2] Muller, M. (2014). Curiosity, Creativity, and Surprise as Analytic Tools: Grounded Theory Method. In Olson & Kellogg (eds.), *Ways of Knowing in HCI*, Springer, 25–48.

Grounded theory in practice[3]

- 15-month ethnography of a delivery robot in a hospital
- Applied **grounded theory analysis**: open coding, axial coding, model building
- A model of four factors that shaped acceptance
- Where the robot didn't fit a unit's work or goals, staff rejected it

[3] Mutlu, B. & Forlizzi, J. (2008). Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction. *HRI '08*, 287–294.

From codes to a model[4]



[4] Mutlu & Forlizzi (2008), Fig. 2.

Open coding

Naming concepts that are significant in the data: abstract representations of events, objects, relationships, interactions.

- Read line-by-line: what actions, feelings, consequences appear?
- Stay close to participants' words (**in-vivo codes**) *before* abstracting
- Constantly ask: *"What is happening here?"*

In-vivo → concept

Raw quote	In-vivo code	Concept
"I keep hitting refresh because it freezes."	"hitting refresh"	coping strategy
"I just wait and hope it loads."	"wait and hope"	user resignation
"It's so annoying. I quit the app."	"quit the app"	frustration → abandonment

Inductive vs. deductive coding

Inductive

- Codes *emerge* from the data
- Stay close to participants' language
- Build theory, discover patterns

Deductive

- Codes start from theory or prior work
- Apply a predefined codebook
- Test, refine, extend theory

Reliability: Cohen's κ[5]

Independent coders, same segments: how much do they agree *beyond chance*?

$$\kappa = (P(a) - P(e)) / (1 - P(e))$$

κ	Agreement	κ	Agreement
< 0	poor	.41–.60	moderate
0–.20	slight	.61–.80	substantial
.21–.40	fair	.81–1.00	almost perfect

[5] Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.

The retrain loop

1. Choose the measure (Cohen's κ) & a target (κ ≥ .80)[6]
2. Pick a **reliability sample** (~10% of the data)
3. Train a second coder; code the sample independently
4. Compute κ → **reconcile disagreements** → revise codebook → recode
5. Report inter-rater reliability

---

[6] Lombard, M., Snyder-Duch, J. & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4), 587–604.

Does grounded theory need κ?[7]

- "Grounded theory rarely, if ever, requires IRR"
- When codes are the *process*, not the product, κ adds little
- κ fits best where the codes *are* the result: content analysis, a fixed codebook, many coders
- Using it anyway? Say what it's for

[7] McDonald, N., Schoenebeck, S. & Forte, A. (2019). Reliability and inter-rater reliability in qualitative research. *PACM HCI*, 3(CSCW), Art. 72.

From codes to theory

Increasing abstraction, across coding cycles:

1. **Open codes** → "delays," "bugs," "frustration"
2. **Axial codes** → "breakdowns in interaction"
3. **Selective coding** → a model of *"user adaptation to unreliable systems"*

Data → Codes → Themes → Relationships → Theory → Design implications

AI-assisted coding: newer models

- GPT-3 + expert codebook: κ .61 and .38 (experts: .88, .90)[8]
- GPT-4, one code per prompt + rationale: mean κ .68 (range .30–1.00); GPT-3.5: .34[9]
- The rationale alone raised κ from .59 to .68
- Still weak on implicit meaning; small interview sets: hand-code

[8] Xiao, Z., et al. (2023). Combining codebook with GPT-3 for deductive coding. *IUI '23 Companion*, 75–78.

[9] Dunivin, Z. O. (2024). Scalable qualitative coding with LLMs. arXiv:2401.15170; *EPJ Data Science* 14, 28 (2025).

The audit stance

The LLM is a **coder to audit**, not an oracle to trust.

- **Drift:** code meanings shift across a long transcript
- **Over-smoothing:** disconfirming cases flattened into tidy themes
- **Fluent agreement** that looks like reliability you didn't earn
- Keep the **chain of evidence**: quote → code → category → claim

Maintaining transparency

Inter-rater reliability only addresses subjectivity *in coding*. Also:

- Keep an **analytic memo** for each major insight or decision
- Document **code evolution** (merged / renamed / dropped)
- Link evidence: **quote → code → category → claim**
- Maintain a shared **codebook** in teams

In-class activity: *open-code a transcript* (30 min)

In your pod, code a short excerpt (yours, or from Q1/Q2):

1. **(10 min) In-vivo → concept.** Open-code ~15 segments; build a small codebook.
2. **(10 min) Second coder + reconcile.** A pod member codes it independently; compute **κ**, reconcile, revise the codebook.
3. **(10 min) AI-assisted pass + audit.** Give an LLM your codebook; mark where it helped or over-smoothed.

Grab the worksheet

- One copy per pod: File → Make a copy
- Nothing to submit today
- It seeds your *First-Pass Qualitative Analysis*

da.gd/IhUtSS



No qualitative data?

Code a **practice set** as a thought experiment. We grade the method, not the fit.

- **Q1** · Creative professionals on AI
- **Q2** · Cancer-care staff on AI
- Codebook + κ templates for today

da.gd/d8BEE



Assignment: *First-pass qualitative analysis*

Individual write-up, started with your team in class. **Due Fri Oct 30, 9:30 am.** Bring it to Methods 8.

1. **Data & scope:** what you coded (source, amount) + the analytic question
2. **Codebook:** codes with definitions; in-vivo examples for the key ones
3. **Coding approach:** open/axial or grounded-theory move; how codes were derived

Assignment, *continued*

- **4 · Reliability:** two coders on a shared segment + κ (or your inter-rater plan) + a retrain note
- **5 · Emerging themes:** 2–4 candidate themes with coded evidence
- **6 · AI-assist audit:** if you used an LLM, what it did + one validity caution you applied

Deliverable: ≈1–2 pp + codebook, individual, on Canvas.

Graded on completeness, soundness, feasibility (rubric on Canvas).

Project status

Milestone (Wk 8–11): data collection underway

- Each team posts 3 lines on Canvas by **11:59 pm tonight**: **done · blocker · next**
- Hailey reads them; we follow up with teams that name a blocker

What's Next

- **Methods 8: Statistics I, Descriptive + Effect Sizes & CIs** (Fri Oct 30)
  - **Read:** the textbook **Ch. 4** (Statistical analysis) · Cumming (2014), *The New Statistics*
  - Bring your **first-pass qualitative analysis** write-up
- **Monday:** Agents, Delegation & Embodiment