

Research Design in Practice

Walking Through a Real Study

Beyond Sentiment: Anti-Americanism in Turkish Social Media

Emre Tapan

PhD Candidate, Department of Political Science
Northeastern University

Spring 2026

The Puzzle

NATO ally since 1952.

Yet only **18%** favorable view of the U.S.

Lower than many non-allied states.

How much $\neq$ What
kind

The level of hostility is known.

The *nature* of it is not.

The Puzzle

NATO ally since 1952.

Yet only **18%** favorable view of the U.S.

Lower than many non-allied states.

How much $\neq$ What
kind

The level of hostility is known.

The *nature* of it is not.

Module Connection: Research begins with a puzzle (Van Evera, 1997) — Module I

Opinion vs. Bias

Policy criticism	Entrenched prejudice
Addressable through diplomacy	Resistant to any policy change
“Opinion” zone	“Bias” zone

What kind of hostility underlies anti-Americanism in Turkish public discourse?

Opinion vs. Bias

Policy criticism	Entrenched prejudice
Addressable through diplomacy	Resistant to any policy change
“Opinion” zone	“Bias” zone

What kind of hostility underlies anti-Americanism in Turkish public discourse?

Module Connection: From puzzle to research question — Module I

What Would You Do?

If you wanted to study what *kind* of anti-Americanism exists in Turkey, how would you design the study?

What the Literature Says

Key insight: Anti-Americanism is not one thing. It falls on an *opinion–bias spectrum* and people are selective about what they dislike (Katzenstein & Keohane, 2007; Chiozza, 2007).

The gap: This decomposition has only been done with surveys—never in *text data*, where people express hostility in their own words.

What the Literature Says

Key insight: Anti-Americanism is not one thing. It falls on an *opinion–bias spectrum* and people are selective about what they dislike (Katzenstein & Keohane, 2007; Chiozza, 2007).

The gap: This decomposition has only been done with surveys—never in *text data*, where people express hostility in their own words.

Module Connection: Theory building & identifying gaps (Van Evera; Schedler, 2010)
— Module I

Defining the Concept: Six Dimensions

I separated **what** is criticized from **how** criticism is framed:

Substantive (what):

Political – sovereignty grievances

Cultural – rejection of American values

Religious – U.S. as enemy of Islam

Socioeconomic – critique of capitalism

Discursive (how):

Conspiratorial – hidden orchestration

Dispositional – inherent American character

Defining the Concept: Six Dimensions

I separated **what** is criticized from **how** criticism is framed:

Substantive (what):

Political – sovereignty grievances

Cultural – rejection of American values

Religious – U.S. as enemy of Islam

Socioeconomic – critique of capitalism

Discursive (how):

Conspiratorial – hidden orchestration

Dispositional – inherent American character

Module Connection: Concept formation (Schedler, 2010) — Module I; Operationalization — Module II

What Would You Do?

You have 4.5 million Turkish tweets mentioning U.S. figures.

How would you *measure* which of the six dimensions each tweet expresses?

My Approach: Theory-Grounded LLM Coding

Step 1: Built a detailed **codebook** from theory—definitions, examples, edge-case rules for each dimension.

Step 2: Used an LLM (Gemini 2.0 Flash) with the codebook embedded in the prompt to classify each tweet.

Step 3: Validated LLM coding against human coding (Cohen's kappa, F1 scores).

Why not sentiment analysis? Knowing a tweet is “negative” doesn't tell you *why* or *what kind*.

My Approach: Theory-Grounded LLM Coding

Step 1: Built a detailed **codebook** from theory—definitions, examples, edge-case rules for each dimension.

Step 2: Used an LLM (Gemini 2.0 Flash) with the codebook embedded in the prompt to classify each tweet.

Step 3: Validated LLM coding against human coding (Cohen's kappa, F1 scores).

Why not sentiment analysis? Knowing a tweet is “negative” doesn't tell you *why* or *what kind*.

Module Connection: Measurement validity, inter-rater reliability, content analysis —
Modules II–III

What Would You Do?

You have 4.5 million tweets but can only classify 100,000.

How do you choose which 100,000?

Sampling Design: Three Decisions

Decision 1: Stratify by time, source, and engagement.

Ensures every year, every U.S. target, and both viral and quiet tweets are represented.

Decision 2: Oversample zero-engagement tweets (70% of sample vs. 49% of frame).

Viral tweets may reflect algorithms, not typical attitudes.

Decision 3: Damp author concentration.

Top 1% of authors produced 49% of tweets. Two-stage PPS sampling reduced their share to 12%.

Sampling Design: Three Decisions

Decision 1: Stratify by time, source, and engagement.

Ensures every year, every U.S. target, and both viral and quiet tweets are represented.

Decision 2: Oversample zero-engagement tweets (70% of sample vs. 49% of frame).

Viral tweets may reflect algorithms, not typical attitudes.

Decision 3: Damp author concentration.

Top 1% of authors produced 49% of tweets. Two-stage PPS sampling reduced their share to 12%.

Module Connection: Stratified sampling (Lec 7); Selection bias (Lec 6) — Module II

Spot the Tradeoff

By oversampling quiet tweets and damping prolific users, what do we *gain*? What might we *lose*?

The Tradeoff

What we gain:

- Better representation of “ordinary” discourse
- Protection against a few loud voices distorting results
- More credible population estimates

What we risk:

- Underweighting influential voices that shape debate
- Needing statistical weights to adjust back
- Design complexity

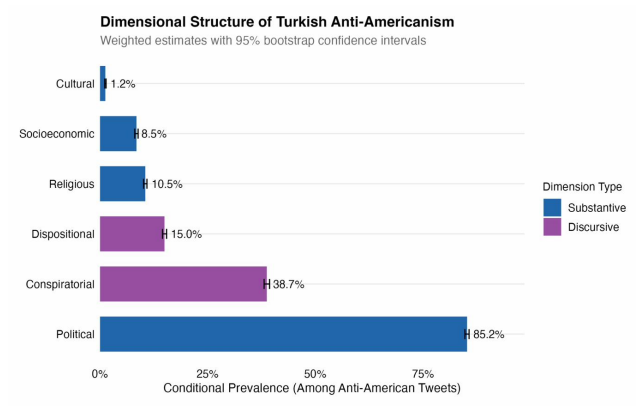
Every sampling choice is a tradeoff. The key is to make it deliberately, not accidentally.

Guess Before I Reveal

If Turkish anti-Americanism has six dimensions...

Which dimension do you think *dominates*?
Which do you think is *least* common?

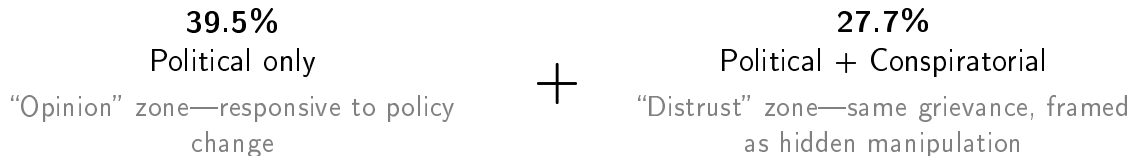
Result: The Dimensional Structure



85.2% political. Only **1.2%** cultural. The near-absence of cultural content rules out “clash of civilizations.”

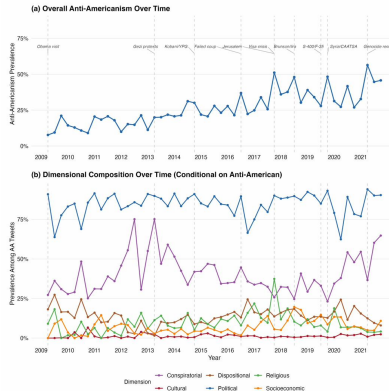
Result: The Political-Conspiratorial Nexus

67% of anti-American content falls into just two patterns:



The conspiratorial overlay makes policy criticism *resistant to correction*.

Result: Dimensions Shift with Events



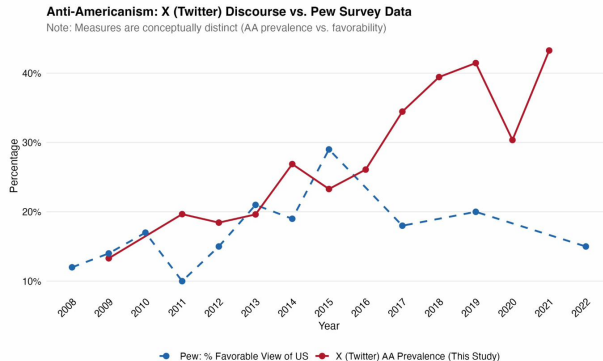
Gezi → conspiratorial spike. Jerusalem → religious spike. Sanctions → economic spike.
If hostility were fixed prejudice, the mix wouldn't shift with events.

Guess Before I Reveal

I compared my Twitter data to Pew's survey data on Turkish attitudes toward the U.S. (2008–2022).

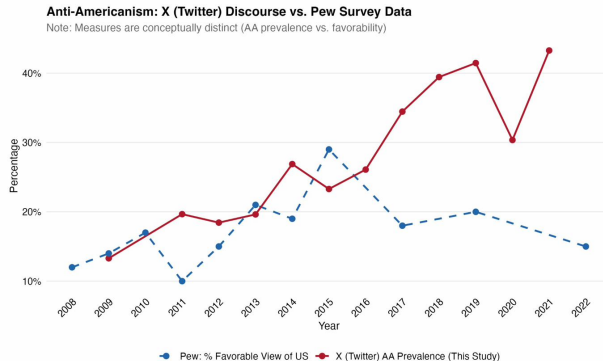
Do you think the two measures *track each other*?
Or do they *diverge*?

Result: Discourse vs. Survey Data



They tracked until ~2015, then diverged. The gap itself is a finding.

Result: Discourse vs. Survey Data



They tracked until ~ 2015 , then diverged. The gap itself is a finding.

Module Connection: Triangulation is a key benefit of mixed methods (Lec 16) — Module IV

Spot the Tradeoff

This study used social media data + LLM coding.

What can this design tell us that surveys *cannot*?
What can surveys tell us that this design *cannot*?

Honest Limitations

External validity: Twitter users $\neq$ the Turkish population. They skew younger, more urban, more politically engaged.

Internal validity: The temporal analysis is descriptive. We can't prove crises *cause* dimensional shifts—only that they co-occur.

Measurement: LLM validated against one human coder so far. The LLM is a “black box”—we see accuracy but not reasoning.

Honest Limitations

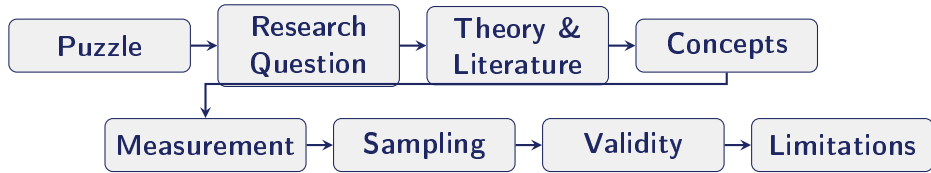
External validity: Twitter users $\neq$ the Turkish population. They skew younger, more urban, more politically engaged.

Internal validity: The temporal analysis is descriptive. We can't prove crises *cause* dimensional shifts—only that they co-occur.

Measurement: LLM validated against one human coder so far. The LLM is a “black box”—we see accuracy but not reasoning.

Module Connection: Internal & external validity, coverage bias (Lecs 5–6) — Module II

The Design Chain



Every choice constrains and enables what comes next.

Design is an argument for why your evidence answers your question.

Your Turn: Pick One Design Choice

Which of these did your team navigate? **Pick one** and be ready to share.

- | | |
|----------------------|------------------------|
| ➊ Research question | ➋ Measurement approach |
| ➌ Conceptualization | ➍ Validity concern |
| ➎ Operationalization | ➏ Key tradeoff |
| ➐ Sampling strategy | ➑ Honest limitation |

What did you choose, and what was the tradeoff?

Thank You!

Emre Tapan
PhD Candidate, Political Science
tapan.y@northeastern.edu

POLS 2399 – Research Methods · Spring 2026