

Topic 3 · Part II: Policy Documents

Reading Documents at Scale

Summarize, extract, compare, and classify — with the AI kept honest.

Summarize

Extract

Compare

Classify

Roumen Vesselinov, PhD **Roumen@jhu.edu**

Pick a team · upload the documents · ask the AI · verify every quote

WHAT THIS TOPIC COVERS

From a stack of text to usable facts

Four moves, one discipline: keep the model inside the source text.

- **Faithful summaries** — shrink a long report without inventing anything.
- **Structured extraction** — pull organizations, dates, and figures into a clean table.
- **Classification** — label a document's type, outlook, and policy direction.
- **Comparison** — put several documents side by side and find agreement and conflict.
- **The grounding rule** — every claim carries a short quote, or it is 'not stated'.

WHY THIS IS HARD

Language models are fluent — and that is the risk

The trap

- AI text tools **hallucinate**: they produce plausible sentences that the document never said.
- Fluency reads as authority. A confident, well-written summary can still be wrong.
- The danger grows with long documents you have not read yourself.

The fix

- **Ground every claim** in a short exact quote and name the source document.
- Allow the answer '**not stated**' instead of a guess.
- **Spot-check** two or three quotes by finding them yourself.

THE ONE RULE TO KEEP ON ALL SESSION

The grounding rule

“Support each point with a short exact quote and name which document it came from”

— If the document does not say it, reply 'not stated' rather than guessing.

Team packages of real policy documents

In every team	What it is
Doc1, Doc2, Doc3	Real reports & statements — IMF, World Bank, central banks, rating actions
Same file names	Doc1/Doc2/Doc3 across all teams, so one prompt works for any team
One team each	Each student downloads a single team package and uploads it to Claude
Grounding target	Everything you ask the AI must be traceable to these files

From the Policy Documents page — 30 team packages spanning several regions.

Shrink a document without distorting it

- Ask for **5–7 bullet points using only the uploaded text** — no outside knowledge.
- Require a **short exact quote (under 20 words)** after each bullet.
- Allow **'not stated'** for anything unclear, instead of filling the gap.
- Finish with a **one-sentence takeaway** of the document's overall message.
- Why it works: the quote requirement forces the model back into the source on every line.

Turn prose into a table you can use

Item	Value	Exact quote
Issuing organization	IMF / World Bank / central bank	“ ... ”
Country or region	e.g., Mexico, euro area	“ ... ”
Date / period	e.g., 2025 Article IV	“ ... ”
Headline figures	GDP growth, inflation, policy rate	“ ... ” (one row each)
Main recommendation	the decision or advice, if any	“ ... ”

If a field is not present, the cell reads 'not stated'. Never invent numbers.

Label the document on three dimensions

Ask for a label + a quote

- **Document type:** central-bank statement / IMF assessment / World Bank report / rating action / communiqué / other.
- **Economic outlook:** positive / mixed / negative / neutral.
- **Policy direction:** tighten / ease / hold / not applicable.

Keep it honest

- Each label needs a **one-line justification with a short quote**.
- Allow '**cannot tell from this document**' when the basis is thin.
- Labels without quotes are opinions — send them back.

One table, one column per document

Row	Doc1	Doc2	Doc3
Issuing org	...	...	...
Country / period	...	...	...
Growth figure	...	...	...
Inflation figure	...	...	...
Overall outlook	...	...	...
Main message (≤12 words)	...	...	...

Then: the biggest point of agreement and of disagreement, each with a quote.

Pull every explicit figure into one place

- Extract every stated figure — **growth, inflation, policy rate, unemployment, debt/deficit**.
- Columns: [Document | Indicator | Value | Period | Exact quote].
- **Do not compute or estimate** any figure not written in a document.
- If two documents disagree on the same indicator, **keep both rows** to show the conflict.

PUTTING IT TOGETHER

A short, grounded brief

- Half a page a busy reader could skim, **using only the team's documents**.
- Cover **who issued them and about where/when**, what they say on **growth and inflation**, and **where they agree or diverge**.
- **Every factual claim** carries a short quote and the document name.
- End with one honest line: **“What these documents do NOT tell us: ____”**

How to catch a hallucination

- **No quote → unverified.** Treat any claim without a quote as not yet true.
- **Watch invented precision** — a rate with decimals the source never gave. Ask which sentence it came from.
- If a summary contains something the document does not, **call it out** and re-prompt: 'use only the uploaded text.'
- **Find two or three quotes yourself** in the PDF to confirm they are real.

TAKE HOME

The AI reads fast; you decide what is true

“Reading at scale is not trusting at scale”

— Every number, every claim, traces back to a quote in a document — or it does not count.

Your workflow for Topic 3

1

Pick **one team** on the Policy Documents page and download Doc1, Doc2, Doc3.

2

Open **Claude** in a new tab and upload your team's documents.

3

Copy a prompt from the Topic 3 file — the prompts work for any team.

4

For every fact, demand a **short exact quote** and the document name.

5

Spot-check quotes against the source; mark anything unsupported.