**Instructions for Labelling in Excel**

**Annotation Workflow**

1. **Open the spreadsheet** containing one comment per row, with a column for each of these labels:
   * spam\_flag
   * sentiment\_manual
   * irony\_manual
   * trust\_manual
   * distrust\_manual
   * mention\_misinformation
   * believe\_misinformation
   * accuse\_misinformation
   * hostility\_manual
   * cogni\_dis\_manual
2. If the “Enable content” thing pops up, click yes.
3. Read each comment in full.
4. Fill in each label using the rules below.

**Label Definitions & Examples**

**1. spam\_flag**

**What to catch:** Self‑promotion or channel plugging.

* **Yes (1)** if the comment invites you to subscribe/follow the commenter’s channel or contains a YouTube channel/video link or reference as promo.
* **No (0)** otherwise.

**Example:** “Check out my channel!” → 1  
“You can read about it at [https://nytimes.com”](https://nytimes.xn--com-9o0a/) → 0

**2. sentiment\_manual**

**Overall emotional valence** of the comment, ignoring sarcasm’s flip:

* **1** = clearly positive (praise, amusement)
* **0** = neutral or mixed (pure description, mixed praise+critique)
* **–1** = clearly negative (criticism, insults, dread)

**Example:**  
“Haha, that was hilarious!” → 1  
“I bet he missed that free throw” → –1  
“It sounds factual.” → 0

**3. irony\_manual**

**Sarcasm or ironic tone**—the literal wording differs from the intended meaning:

* **1** if the comment uses mock‑sadness (“boo‑hoo”), dark humour, absurd exaggeration (“natural selection at work”), or other forms of irony and sarcasm that are so clear you don’t have to figure out if it’s really sarcasm or not.
* **0** otherwise.

“Nice job! Now everyone’s dead, haha” → 1  
“That was awesome!” (straight) → 0

**4. trust\_manual**

**Expressions of confidence or praise for honesty/transparency**:

* **1** if the commenter says “I trust,” “he knows what he’s doing,” “honest,” “genuine,” etc.
* **0** otherwise.

“I trust this reporter.” → 1  
“The reporter said that twice.” → 0

**5. distrust\_manual**

**Open doubt or skepticism** about truthfulness or integrity:

* **1** if they say “I don’t trust,” “they’re lying,” “fake,” “who’s deceiving who,” or question authenticity.
* **0** otherwise.

“This is fake.” → 1  
“What happened here?” → 0

**6. concept\_misinformation**

**References the concept** of misinformation without taking a stance:

* **1** if they use words like “misinformation,” “hoax,” “rumor,” “theory,” “disinformation.” in a Meta-level way.
* **0** otherwise.

“Be careful—there’s so much misinformation out there.” → 1  
“That’s not true.” → 0

**7. believe\_misinformation**

**Endorsement of a widely circulated but unverified/false claim** (conspiracy‑style belief):

* **1** if they say “I believe,” “It’s obvious that…” “It’s true” about a known debunked claim.
* **0** otherwise.

“Everyone knows the moon landing was faked.” → 1  
“Sounds fake.” → 0

**8. accuse\_misinformation**

**Charging the content as false or misleading**:

* **1** if they call it a “hoax,” “fake news,” “they’re lying to you,” “propaganda.”
* **0** otherwise.

“This segment is pure propaganda.” → 1  
“Interesting footage.” → 0

**9. hostility\_manual**

**Aggression, hostility, malice or threats**—language advocating harm or using severe slurs/hate speech:

* **1** for threats (“I’ll beat your ass”), calls for violence, aggressive, and/or derogatory language.
* **0** for mild insults (just leave those unflagged).

“They should be punched in the face.” → 1  
“You moron!” → 0

**10. cogni\_dis\_manual**

**“Ends justify the means” thinking, “it’s for the best”**—expresses positive sentiment towards harmful behaviours/thinking etc. that will cause others harm, but for a good cause or that it will turn out well in the end (in the person’s mind):

* **1** if they speak positively about something that will be harmful to or will be dangerous to some group or person because they genuinely believe that it’s the right thing to do.
* **0** otherwise.

“This guy is so brave! He knows he is serving people food that will kill them in the end, and he hates it, but it will show the people who need it that they need to stop eating these unhealthy things. That is such strength of characters <3” → 1  
“Helping the poor is such a kind thing to do!” → 0

**Quick‑Reference Cheat‑Sheet**

| **Label When = 1** |  |
| --- | --- |
| spam\_flag | Self‑promo (“subscribe,” channel links) |
| sentiment\_manual | +1 praise/amusement 0 neutral/mixed –1 criticism/insults |
| irony\_manual | Sarcasm or tone flip (“boo‑hoo,” dark humor) |
| trust\_manual | Expresses confidence/praise for honesty (“I trust,” “genuine”) |
| distrust\_manual | Expresses skepticism/doubt (“fake,” “lying,” “don’t trust”) |
| mention\_misinformation | References “misinformation,” “hoax,” “rumor,” etc. |
| believe\_misinformation | Endorses a debunked claim (“I believe this conspiracy”) |
| accuse\_misinformation | Calls content false/propaganda/hoax |
| hostility\_manual | Threats, calls for violence, hateful slurs |
| cogni\_dis\_manual | Praises harmful/unethical acts because they believe it’s a good thing (think cults believing messed up things while still having good people in there). |

**Tips for Consistency**

* **One label at a time:** Read the comment fully, then cycle through each label in the order above.
* **Stick to definitions:** If in doubt, default to 0. Only mark 1 when you’re sure the rule applies. (You can turn the background of the cell yellow if you want me to check it further).
* **If you are uncertain,** look through the already labelled comments and see if you can find similar ones where you can see how I have labelled them.
* **Keep notes:** If a comment is borderline or difficult, and you want me to explain it (if you want to help at more than one occasion), jot down a comment and I will reply to it when I see it.