Sustainable Data Science and Artificial Intelligence

Tutorial #2 – Part 1

|  |  |
| --- | --- |
| **My Name:** |  |

**Instructions**

This activity is focused on content moderation. It is the process of reviewing, monitoring, and managing user-generated content (UGC) on online platforms to ensure it follows community guidelines and legal standards. This includes text, images, videos, and links shared on social media, forums, and websites.

Some of the reasons why content could be removed online:

1. **Hate Speech & Discrimination:** Content that attacks or promotes violence against people based on race, ethnicity, nationality, religion, gender, sexual orientation, disability, or other protected characteristics.
2. **Harassment & Cyberbullying:** Direct insults, threats, or persistent attacks on individuals. Doxxing (sharing personal/private information).
3. **Violent or Graphic Content:** Real-world violence (shootings, assaults, animal cruelty). Gore or explicit injuries (depends on platform policies).
4. **Misinformation & Fake News:** False claims about public health, elections, or scientific facts.
5. **Inciting Violence, Terrorism, or Illegal Activity:** Calls for violence, riots, or terrorist attacks. Selling or promoting illegal drugs, weapons, or human trafficking.
6. **Self-Harm, Suicide, or Eating Disorders:** Encouraging suicide, self-harm, or eating disorders (e.g., pro-anorexia content). Some platforms offer support resources instead of removing the post.
7. **Child Exploitation & Nudity:** Child sexual abuse material (CSAM) – zero tolerance. Unwanted sexual advances or explicit content.
8. **Copyright & Intellectual Property Violations:** Using copyrighted music, videos, or images without permission. Fake giveaways pretending to be official brands.
9. **Spam, Scams, & Phishing:** Fake links pretending to be official sites to steal passwords. "Get rich quick" schemes or fake giveaways.

When reviewing content, each post can be classified into one of three categories:

* **NOT REMOVED**
* **FLAGGED FOR REVIEW**
* **REMOVED**

For each of the following posts, review and decide a category for the post, and also give a 25-50 word rationale for your decision (I’ve done the first one for you).

Review 1:

|  |  |
| --- | --- |
| **Posting:** | *"I hate all [ethnic group]. They should be wiped off the earth!"* |
| **Decision:** | **REMOVED** |
| **Rationale:** | I would remove this post as it appears to be an example of hate speech, and I think it is also inciting violence against a particular group. |

Review 2:

|  |  |
| --- | --- |
| **Posting:** | *"COVID-19 vaccines contain microchips to control your brain. Wake up, people!"* |
| **Decision:** |  |
| **Rationale:** |  |

Review 3:

|  |  |
| --- | --- |
| **Posting:** | *"Ugh, I failed my exam. I should just disappear forever..."* |
| **Decision:** |  |
| **Rationale:** |  |

Review 4:

|  |  |
| --- | --- |
| **Posting:** | *"Just broke my leg 😭 Here’s a pic of the X-ray!"* |
| **Decision:** |  |
| **Rationale:** |  |

Review 5:

|  |  |
| --- | --- |
| **Posting:** | *"Check out this awesome new AI tool! [link]"* |
| **Decision:** |  |
| **Rationale:** |  |

Sustainable Data Science and Artificial Intelligence

Tutorial #2 – Part 2

|  |  |
| --- | --- |
| **Group****Members:** |  |

The following five scenarios are similar to the previous ones, except I want you to do these five in a group, so you have to figure out some way of getting consent within the group.

Review 6:

|  |  |
| --- | --- |
| **Posting:** | *"Let’s bomb this government building! Meet at 5 PM."* |
| **Decision:** |  |
| **Rationale:** |  |

Review 7:

|  |  |
| --- | --- |
| **Posting:** | *"He’s so ugly, I bet no one loves him. LOL!"* |
| **Decision:** |  |
| **Rationale:** |  |

Review 8:

|  |  |
| --- | --- |
| **Posting:** | *"This politician is a liar! They never keep their promises!"* |
| **Decision:** |  |
| **Rationale:** |  |

Review 9:

|  |  |
| --- | --- |
| **Posting:** | *"This new diet pill burns fat overnight! DM me to buy! "* |
| **Decision:** |  |
| **Rationale:** |  |

Review 10:

|  |  |
| --- | --- |
| **Posting:** | *"Here’s a funny meme of a cat with a knife! 😂🔪"* |
| **Decision:** |  |
| **Rationale:** |  |

**Submission Instructions**

Please submit in the following two ways

|  |
| --- |
| Please e-mail me your completed solution, and with filename:*Familyname\_FirstName\_TU850\_1\_SDSaAI\_Tutorial2.docx*e.g.*Gordon\_Damian\_TU850\_1\_SDSaAI\_ Tutorial2.docx*e-mail to Damian.X.Gordon@TUDublin.ie with subject heading as follows: [DT850/1] SDSaAI Tutorial2 |
| Please submit into Brightspace in:* Assessment
	+ Assignments
		- Tutorial #2
 |

For the email, please include the following message:

|  |
| --- |
| **Damian,****I am a student in your Sustainable DS & AI (CMPU1040) class in the BSc in DS & AI (DT850/1).****Please find attached Tutorial #2****Regards,****Your Name****Student Number.****DT580, BSc in DS & AI** |