INFO20003 Semester 2, 2025

Assignment 1: ER Modelling

Due: Week 5 - Friday 29th August 2025, 6:00 PM Melbourne Time.

Submission: Via LMS *https://canvas.lms.unimelb.edu.au/*

NEWSPAPER.NET: 'New, Ethical, Web Social Platform And Personalised Network'

Social media sites nowadays are full of 'AI slop' – "dubious, AI-generated material" (Hoffman, 2024)[1] – amongst other low-quality material. People who go on social media for news frequently find there is a lot of misinformation online. Furthermore, celebrities are finding it hard to engage their fan base authentically, given various fake accounts and scams. As such, it is hard for a typical user to engage responsibly with meaningful content and users, especially as the social media landscape is so fragmented: from X (Twitter) to Instagram to TikTok to Facebook.

As such, you and your classmates decide to launch a startup company that provides a new social media ecosystem. It does not replace existing social media platforms, but rather augments them so that reputable news agencies, organisations, and celebrities could communicate directly with their followers, while ensuring users do not fall for the 'AI slop' or harmful misinformation or scams! This platform is called NEWSPAPER.NET, which stands for "**N**ew, **E**thical, **W**eb **S**ocial **P**latform **A**nd **Per**sonalised **Net**work". NEWSPAPER.NET enables these content creators to complement their existing social media presences with authentic, verifiable, posts.

Your classmates/business partners reckon that you have *the best Database skills* amongst all of them, so they selected you to oversee the database infrastructure for NEWSPAPER.NET. 😊
(The other business partners will take care of the AI and infrastructure side!)

As part of this, you will be creating a MySQL database to store this information. The following specifications have been provided to you to assist in your design.

Hub

A hub is an account owned by a personality or organisation (e.g., Chappell Roan, BlackPink, The University of Melbourne). Hubs have their own profile photo (stored in the database), unique username starting with '@', description, email address, verified status (either Verified or Not Verified), and type(s).

[1] *Hoffman, B. (2024). First Came 'Spam.' Now, With A.I., We've Got 'Slop'. The New York Times. [online] 11 Jun. Available at: https://www.nytimes.com/2024/06/11/style/ai-search-slop.html.*

A hub can be one or more of the following types: {Celebrity, Organisation, Product, Internet Influencer, News Agency}, with NEWSPAPER.NET planning to add more types. To elaborate, hubs such as MrBeast can be an Internet Influencer, a Celebrity, as well as a Product.

A hub has a timeline of posts, which are pieces of content that everyday users can read or otherwise engage with. Of course, there won't be a social network if not for the users (like you or me). A hub may have any number of moderators (see below).

You will be creating the technical infrastructure for the hubs, posts, and users, in addition to several other entities or concepts as detailed below.

💌 *Posts*

Each hub will have at least one post, with no co-creation allowed, as even a newly created hub will need to have a 'welcome' post. For each post, the system records a unique ID (using the UUID format, which looks like "cadc6b45-c296-4cb9-9470-b30aaf4a3cb3"), to track the post. Each post has text, a date-and-timestamp, and an optional preview (a binary object) with corresponding type (either png, jpg, or gif).

There are two mechanisms which allow a hub to link a post to other content:

- A post can link to one or more external URLs such as a YouTube video link or a Reddit permalink, along with the URL creation date/time(s).
- A post can also link to one or more posts by any hub on NEWSPAPER.NET, via UUIDs.

👨 *Users*

Now, everyday users can create user accounts on NEWSPAPER.NET. For each user, the system records their details such as a unique username, first name, last name, a unique email address, a unique phone number, and a unique social media handle and its type (one of Facebook, Instagram, X, BlueSky, XiaoHongShu, or TikTok – this is a fixed list). To ensure that the user behaves responsibly, they have a total number of warnings: for most users, this will be zero; but for irresponsible users, these can go up to a maximum of three before they are automatically terminated. Records of terminated users remain in the system so they cannot rejoin using the same email address, phone number, or social media handle. Users also have a reputation score (decimal numbers, e.g., -1.0 or +42.1), which indicate how helpful and responsible they are: this score is auto calculated by the platform.

NEWSPAPER.NET is not meant to be a full-featured social network: hence it only keeps track of the connections between users (without other features such as messaging). Each user can have up to 500 connections with other users. For each connection, the system stores the start date of the connection and two descriptions (one from the perspective of each user: e.g., "met X at a party" and "Y is a trusted project partner"). If a connection is ended by one of the users, it should no longer be recorded in the system. All connections are bi-directional: a connection is only established when one user initiates it and the other user confirms it by adding their own description; otherwise, it is not recorded.

Users cannot create posts, but they can engage with posts by commenting at most once on a given post. Users can tag up to 10 other users in each comment. This minimises comment spam and noise. Within each post, comments receive a unique number in sequence, along with the comment's text and the date/time it was made.

👨 *Moderators*

The beauty of NEWSPAPER.NET is that posts can be validated, to reassure users that they are high-quality, informative, and free from misinformation. A User can be assigned by NEWSPAPER.NET to be a Moderator for a Hub for a fixed amount of time. For sensitive or high-traffic hubs, multiple moderators may serve at the same time to help review their activity. In the system, each moderator assignment is given a unique AssignmentCode at the time it is created and is tracked with its start and expiry date–time. All past assignments are retained for use in operational and audit processes. It is not uncommon for a user to be assigned to the same hub multiple times – or even to multiple hubs – if the user is especially trustworthy, but each needs to be tracked separately to maintain transparency/accountability.

Each Moderator can create Reports on any number of Posts (only for the Hubs they oversee), tracked according to the relevant assignment. A Post is not limited to only one Report, such as a Post by a News Agency which may be fact-checked several times. A report needs to track the exact timestamp, the corresponding Moderator, a decision (only one of: Verified Human Content, Verified GenAI Content, Potential Bias, Fake Content, Clickbait, AI Slop, or Uncertain Status), as well as a comment by the Moderator. To recognise the role of AI in helping maintain the ecosystem (e.g., detect fake news), each Moderator can be responsible for at most one AIAgent, which has a name and corresponding source code, although the same AIAgent may be used by multiple moderators. Of course, if a report has been generated with the help of this AIAgent, it should be tracked for transparency.

As for Comments, each Moderator can also create Reports on Comments (again, only for the Hubs they oversee). Similarly, a Comment can have any number of Reports, with each of their timestamps/moderator/comment/ decision tracked. The only difference to the above is that report decisions on Comments can only be one of: (Wholesome Feedback, Informative Content, Rule Violation, or Neutral Content). Because Comments are subjective, Moderators cannot use AIAgents to check Comments. If the comment has been determined to violate rules, the user who authored it will have their warning count increased by one.

Business Requirements

Your database design needs to be able to meet the business's needs to answer the following questions. You do **NOT** need to write SQL queries in your assignment to answer these questions.

1. How many moderators have been assigned to 2 or more different *hubs* (e.g., both BTS and BlackPink)?

2. How many times has the user BreakingNews1234 been warned?

3. Among all hubs that have at least a post containing an external link to a URL in the “twitch.tv” domain, which hub has the highest total number of posts’ reports where the moderator(s)’ decision was “Clickbait”?

4. Which moderators who are connections to EthicalUser123 did not use any AI agents?

5. (For auditing purposes) Which unethical moderators have reviewed their own comments?

6. How many posts are written by hubs that are news agencies, with the text containing ‘breaking news’?

7. How many times did the users Sandy, Farzaneh, and Lillian get tagged in comments, in total?

8. Which hubs have had their posts moderated using at least 3 different AI agents?

9. (For debugging purposes) What is the source code of the AI agent(s) used in moderating post(s) for the hub "MrFeast" on 22 December 2025, 10:15 to 10:20 am (inclusive), where MrFeast's posts have links to posts by AshtonKutcher?

10. (For legal purposes) Which users with reputation less than –50.0 have created comments on posts with higher ratios of 'Fake Content' reports (than any other decision).

11. How many users have more than 400 connections with other users?

12. Which hubs have posted at least 3 posts that were linked to another post?

Instructions

Assignment 1 is worth 10% of your final mark. The assignment will be graded out of 100 marks as described in the table below:

Component	Grading
ER Physical Model with assumptions	80% (8.0 raw final marks)
Conceptual Model in Chen's Notation	20% (2.0 raw final marks)
.MWB Physical Model File submitted	ASSIGNMENT HURDLE

You are to analyse this business case and design a **Conceptual ER Model** in **Chen's notation (MUST be hand drawn using pen/pencil and paper, then photographed or scanned[2])** as taught in class, and a **Physical ER Model** for a MySQL Relational Database in **Crow's foot notation (modelled with MySQL Workbench).**

You may list any assumptions you have made about the model. There is a 200-word limit for assumptions. Assumptions must not be used to simplify the assignment, but only to justify your decision about any ambiguity in the study.

[2] *NOTE: This decision has been made in the interest of fairness. From experience, letting students use technology to draw these ER models has led to potential inequity, as some students have the advantage of e.g., 'autocorrected' images or diagram assistance. If a student doesn't have a camera or scanner, the Library has free access to scanning: see How to Scan in https://studentit.unimelb.edu.au/student-print*

Assignment Submission

Please pay special attention to the penalties listed [⚠].
You are to submit the assignment under the Assignments tab on Canvas LMS. The submission will require you to **submit two (2) files:**

1) **A SINGLE PDF document** containing:
   a) a legible (hand drawn) picture of a conceptual model in Chen's notation, scanned/photographed.
   b) a screenshot/export of your Physical ER Model done in MySQL Workbench. Ensure that tables are fully expanded so all attributes are readable. Please ensure the image is not blurry.
   c) any assumptions you made (limit of 200 words).
   Note: All these components must be within a **SINGLE** PDF document. You can use an online tool like https://smallpdf.com/merge-pdf to merge multiple PDFs together into a single PDF.

2) **A copy of your final .MWB MySQL Workbench file of your Physical ER model.**
   a) This .MWB model will only be used by staff in circumstances where the screenshot/export in the PDF is unreadable (**i.e., if we have to open this .MWB file, a penalty of 10% will occur** ⚠). So, remember to include a clear screenshot/export of your physical model in the PDF!
   b) If the screenshot/export of your physical model is unreadable and the .MWB file was not submitted, you will receive zero marks for your physical model ⚠.
   c) Even if the screenshot/export of your physical model is readable but the .MWB file is missing, your marks will be WITHHELD until you provide it, as it is an ASSIGNMENT HURDLE ⚠.

Remember, if you fail to submit clear and legible models your assignment will be penalised ⚠.
If you submit non-hand-drawn models for Submission (1a) your assignment will be penalised ⚠.

Please make sure that you **submit** your files on Canvas ⚠. After uploading the files, you need to press **'Submit Assignment'** to submit the files. If you submit late because you failed to press the submit button and only noticed this after the deadline, your submission will be considered late just like any other late submission to maintain fairness for all students.

Unless you have an approved extension (see below), **you will be penalised -10% of the maximum number of marks in the assignment per calendar day that your submission is late** ⚠. For instance, if you received a 78% raw score, but submitted 2 days late, you'd receive a 58% for the assignment.

Requesting a Submission Deadline Extension

To request an extension:

1. **Follow the process and policy on Canvas -> Modules -> FEIT Extensions and Special consideration (***https://canvas.lms.unimelb.edu.au/courses/215399/pages/feit-extensions-and-special-consideration?module_item_id=6885297***)**
2. Forward any paperwork received from the previous step to the central subject inbox:

   INFO20003-semester2@unimelb.edu.au

   - from your university email address, supplying your student ID, how many days you'd like to extend, and any extra evidence that can support the number of days you are requesting.

3. **Short extensions applied through online declaration forms must be forwarded before 10:00am, 29 August 2025 in order to receive approval from the subject team.**
4. Please include in the subject [INFO20003 Assignment 1 Extension].
5. If your submission deadline extension is granted you will receive an email reply granting the new submission date. **Do not lose this email!**

Reminder: INFO20003 Hurdle Requirements

To pass INFO20003, you must pass two hurdles:

- **Hurdle 1:** Obtain at least 50% (15/30) or higher for the three assignments (each worth 10%)
- **Hurdle 2:** Obtain at least 50% (35/70) or higher for the combination of quizzes and end of semester exam.

It is our recommendation to students that you attempt every assignment and every question in the exam.

Reminder: Academic Integrity

"You are not permitted to use an AI tool in any way when completing your work. Any use suspected will be reported as potential academic misconduct and subject to appropriate penalties."

⚠ **Students are kindly reminded to always comply with the University's Academic Integrity Policies.**

Please read https://academicintegrity.unimelb.edu.au/plagiarism-and-collusion and please note:
"If a student is found to have deliberately plagiarised or colluded the penalties are severe and can include failure of a subject or exclusion from the University. The University provides extensive resources and educates students about academic integrity so that students are aware of what constitutes plagiarism and collusion, and the consequences of those practices."
(Source: https://academicintegrity.unimelb.edu.au/plagiarism-investigation-and-penalties)

⚠ **Note that you will have ER Modelling questions in the exam as well, which are probably tougher than these ones here: cheating will not help you in the exam. So please, no LLM or other AI assistance within this assignment.**

GOOD LUCK!