Developing a Mental Health Toolkit for Social Media

Lewis Buttle, Matthew Jameison, Daisy Kennedy, Patrick Michaels

Contents

Introduction

Social media has had a transformative impact on how people communicate and express themselves. Facebook, YouTube, Whatsapp, Instagram, and TikTok all report over one billion active users, with a global average daily time spent on social media at almost two and a half hours [1]. The ability for content on these networks to proliferate globally and influence discussion, both online and in the real world, ensures that social media networks have far reaching impacts in the daily lives of billions of people.

One example, which demonstrates the huge effect that social media has, is the recent political impact of Facebook. During the 2010 US congressional elections, Facebook messages directly increased turnout by 60,000 and indirectly, through social contagion, by 280,000. In 2000 Bush beat Al Gore in Florida by fewer than 600 votes, so an additional 340,000 votes, thanks to Facebook, is more than sufficient to swing an election. More recently, the platform was the center of large misinformation campaigns throughout the 2016 US presidential election [2]. Undoubtedly, the influence that social media has in society, and will continue to have, is tremendous.

Moreover, Facebook also has more direct social impacts than affecting elections. In 2012, Facebook modified users' feeds to deliver content with either a more positive or a more negative sentiment. Users who saw fewer positive updates produced more negative posts and vice versa. "Emotional states can be transferred to others …, leading people to experience the same emotions without their awareness [3, p. 8788]". Social media can manipulate human emotion with ease, but what effect is this having on our mental health? This project aims to explore the ways in which social media can negatively affect mental health and, from this, create a toolkit of ways to better manage, or even avoid, these effects.

Background Research

There is a large body of existing research that has started to explore the links between social media activity and issues with mental health, especially among younger demographics. There exists a correlation between time spent on digital activities and mental health problems; adolescents who spend more time on social media sites are significantly more likely to be at risk of suicide or other mental health issues [4]. The rate of mental health issues among adolescents also correlates with smartphone ownership in the US. The rate of mental health issues being diagnosed started increasing around 2011, at which point around half of Americans had a smartphone. By 2015, 92% of teens and young adults owned a smartphone. Additionally, mental health issues are reported more frequently by Gen Z (born 1995-2012) than preceding generations at the same age, and there are more attempted suicides too. Moreover, there appears to be a correlation between lower self-esteem and being more active online [5]. Taken together, this all suggests at least some causal link between social media use and mental health issues in adolescents [6].

Then again, not all social media impacts are negative. Whilst there is a correlation between self-esteem and certain online activities, Gonzales and Hancock [7] argue that this is due to a disconnect between

the perceived self and the ideal self. However, they additionally argue that this phenomenon is not unique to digital spaces. Self-reflection and other unedited views of the self may also occur in physical spaces; online media merely increases the frequency of self-observation. They continue to argue that online activity allows users to self-present a curated representation in a mediated space which, in contrast to unedited views of the self like a mirror, can increase self-esteem.

Social media sites also offer positive value for marginalized groups of people to build communities and understand their identity more, in particular LGBTQ+ identities. One example is AVEN (the Asexuality Visibility & Education Network) which is a forum that is for many people their first and only source of information about asexuality when trying to figure out their identity or to have discussion about it [8]. Other LGBTQ+ identities also form communities and safe spaces online. As offline contexts are often harmful towards LGBTQ+ youth, virtual communities across a range of online spaces provide a sense of community, as well as access to other resources. LGBTQ+ youth who are active in online spaces are better able to explore their identity and benefit from it [9, 10].

Finding camaraderie online is not limited to LGBTQ+ individuals; on Instagram users of '#Depression' have benefited from sharing negative experiences. Andalibi et al. [11] argue that by sharing personal experiences users are working towards understanding their struggles better — a case of 'empowered exhibitionism'. That is by sharing their experiences with others who have similar experiences, they are better able to resolve and understand their struggles.

A commonality here is that social media activity is observed by others and is therefore a type of performance to an audience. The idea that when we communicate we imagine an audience is not new. However, the online nature and scale of social media sites means that users' imagined audiences may not be as close to their actual audiences as they think. When the imagined audience and the real audience are conflicting, there can be negative consequences, not just online but in the real world too. Consider sending a message that is aimed at friends but is also read by colleagues and bosses — this could cause psychological, economic and emotional issues with potentially severe impact [12].

The intersection of many audiences can cause *context collapse*: the spatial, temporal, and social boundaries — which separate different audiences — are no longer present. Context collapse can lead to a variety of responses, some users re-establish context or utilize different social media for different contexts. For example, using Facebook as a 'front page' whilst discussing sexuality on tumblr. A common response is self-censorship. This is to assume that any content is universally visible to the widest imagined audience, and therefore avoid discussing it. Self-censorship can also be defined as restricting yourself to a small, limited set of content that keeps within a constant context. Of course, this still imposes restrictions against the desire for pure self-expression [13].

Self-censorship in online spaces is dictated by context, but that context is composed of more than just the imagined audience. Moderation is ubiquitous in online spaces — it is how harmful or off-topic conversations are pruned, and it ensures that safe spaces remain safe. However, the mechanisms and strictness of moderation policies vary greatly. Differences in how communities are moderated establishes different social norms and therefore different contexts for those communities — an example is how different social media sites are used for different purposes [14].

The effect of moderation is not constrained to self-censorship but may also affect language use. Gibson [15] found that discourse in two LGBTQ+ spaces on Reddit had extreme differences in tone and topic due to having different levels of moderation. The more moderated space had less discussion on money and politics than the other. Increased moderation created a more positive discourse, but this was at the cost of user expression.

Another case study for moderation, and its effects, is digital newspaper comments. Weber [16, p. 954] notes that 'the potential for quality discourse is also determined by the features of the websites, in particular by a website's comment management strategy'. Less restricted comment sections often had higher quality discourse. Ksiazek et al. [17] posit that weaker moderation may lead to greater hostility in the discussion of user content. Though it must be bore in mind that newspaper content is professionally made, unlike social media posts, and so perhaps the two would not receive the same treatment. That is, what is true for newspaper comment sections may not also be true for social media spaces. But the concept

of moderation improving discourse is certainly an interesting point.

Not all moderation policies are made equal. Ksiazek [18] notes only two policies are significant indicators for hostility and civility: prohibiting anonymity and pre-moderation (moderating before the content is public). Removing anonymity leads to lower hostility and civility in discourse, whilst pre-moderation lowers hostility and increases civility. Anonymity is valuable for certain groups of social media users — it makes marginalized groups more likely to be open and active. Furthermore, removing anonymity may have negative consequences on a platform. Pre-moderation on the scale of a digital newspaper is a different problem than at the scale of most popular social media sites; the rate of user-generated content is significantly higher, and the range of conversation types that are permissible is far greater. Moreover, unlike with digital newspapers, user content on social media sites is not restricted to text; identifying problematic content in videos and images is significantly more challenging.

There are several identified factors that link social media use and mental health. The link between daily use and increased risk for mental health issues suggests that reducing social media time may be effective. Similarly the context collapse problem and its negative effects may be rectified with easier to use permission and privacy settings and better general technical literacy — this may also help bridge the difference in the imagined and real audience. Finally issues of moderation can greatly influence the use patterns and effects of social media. It is clear that social media can be beneficial to groups of users but that others may be negatively affected — improving upon these foundational issues may have positive effects for a wide range of users.

Objectives

A key objective is the research element of the project. It is essential that the problems with social media and their scope is understood. This allows the problem to be looked at and an efficient solution to be found. It is also key to look for existing solutions for these problems and why they may not be being used by large social media companies.

Solutions to the problems of social media should be robust enough that they can deal with the ever changing environment that is presented by a social media platform and ensure they do not impact the user in a way that would drive them away from the platform. As a user's experience cannot be improved

by a healthy social media if they would rather use something else. It is important that any work we do towards the toolkit is well reasoned and based on research.

One of the ways this toolkit would make social media healthier would be to implement features that would reduce a user's screen time with the social media to a reasonable level. Social media addiction is often expressed through excess screen time [19], so the toolkit should look at ways to discourage or prevent users from accessing the social media platform in excess to mitigate the impact or development of a social media addiction. This could be done in a hard limit restriction, e.g. preventing a user using a platform for more than two hours a day. Or it could be more subtle, trying to encourage the user to stop by presenting prompts for them to do something else.

Another way of improving a user's experience would be to reduce the amount of negative engagement that a user experiences. One metric that social media uses to recommend new content to their users is based on the engagement a user has with existing content [3]. However the user's engagement may not be enjoyable for the user, for example if they had an argument. Despite it being an unpleasant form of engagement for the user the algorithms used would recommend more content similar to that which caused the negative engagement. As a result the toolkit aims to take a more nuanced approach by only encouraging the further recommendation of content that is positively engaged with, ignoring or punishing content that is engaged with negatively.

Moderation is a major problem with social media platforms that while solutions have been attempted, it has still persisted with illegal or undesired content slipping through moderation techniques. The toolkit will look at moderation solutions that could perform better than the commonly used by social media platforms which haven't been adopted. By looking at a user's reaction to moderation techniques, the efficacy of these techniques, and the scalability the toolkit would aim to implement balanced techniques which perform well in the environment and are able to effectively filter content.

The objectives described above are summarized below. It is important to note that the team will be using an Agile approach to completing this project, which affords the team the ability to change ideas and requirements later in the project, though this would preferably be avoided.

- Research
  * Look into existing solutions, assess their suitability for the context we're using them in and how practical they are to implement, scalability, and how robust they are
- Reduce screen time of users to ensure that social media is not used excessively
  * Prevent or reduce impacts of addiction
  * Aim of social media shifter to not maximize total time on system
- Reduce unhealthy interactions with social media
  * Discourage hate scrolling
  * An understanding of the semantics of a user's interactions with social media allowing for smarter decisions
- Enlist more effective moderation techniques
  * Minimize illegal and illicit content getting passed through the social media, limiting user exposure
  * Something robust and adaptable to keep up with shifting attitudes of meanings, e.g. new slurs or word reclamations.

Expected Deliverables

Whilst many of the actual project deliverables will be finalized in response to research conducted throughout the project sprints, the team has many very initial, basic ideas for features that could be implemented and investigated. These initial ideas are documented here as expected deliverables, but are likely to change and become much better refined to suit knowledge and justification gained through researching social media issues.

- Research
  * Research into how users react to countermeasures to reduce screen time, so that any implemented methods used are done so that they do not spoil a users experience with the social media.
  * Research into how users react to countermeasures to ensure a user is engaging in a positive way, so that any implemented methods used are done so that they do not spoil a users experience with the social media.
  * Research into moderation techniques and users reactions to them, and balance the choice of technique such that the users feel comfortable and confident to engage with the social media.
  * Subsequently research the technical resources and methods required to produce the features

of the project (Coding languages, libraries, pseudo-solutions, etc).
- Screen Time
  * An on-screen prompt that appears in response to a user's excessive screen time.
  * Create 'breakpoints' in a feed to stop the constant stream of content.
  * A scaling response proportional to screen time, The longer you spend on the app the more frequent these prompts become.
  * Attempt to base these prompts depending on the user's location? For example, if the user is outside and moving, there is likely a more logical conclusion to their high usage, meaning prompts should not be as frequent.
- Engagement Type
  * Produce a natural language processing module that attempts to read a user's mood through comments and posts, and subsequently score the type of content they actually engaged with. If the user is determined to enjoy the content, this is measured as positive, if they were made angry or annoyed by certain posts, this is measured as negative.
  * This module could be used to ensure users are able to engage with relevant safe spaces and subsequently avoid recommending said safe spaces to users who may engage in these spaces negatively.
- Moderation
  * Produce a robust moderation system that can keep up with a changing user environment. For example, by learning to recognise potential new slurs.
  * An effective and scalable solution that would work on very large user bases.
  * A moderation technique that does not dissuade a typical user from posting and engaging with the social media platform.

Resources

For this project to succeed, it was deemed important to denote the resources available to the team that could potentially aid in the project's development. By taking an inventory of these potential resources, it pooled together the known resources from each team member and shared this information amongst the team, to aid in making more resourceful development decisions.

All of the identified resources have to be both accessible and affordable to the team, due to the project's innate limitation of money (the cost of

certain resources) and time (there is realistically little time to develop experience with new technologies/ languages).

Some members of the team are experienced with using python for machine learning purposes, as confirmed by completing the University of Warwick Module CS342, Machine Learning. This will be helpful for any tools for the toolkit that rely on predicting behavior or emotions.

Kaggle Data: Kaggle [20] is an online community based around data science, and provides data sets for open use. This includes data sets relevant to social media posts, which could be used within the project.

Twitter APIs: Twitter [21] allows for the use of live tweets, which could be used as a potential test data stream for this project. However, the fair usage of these tweets may need to be researched further, as to their legal availability.

Python Libraries: Scikit Learn [22] is a machine learning focused library that allows statistical models to be easily constructed and used quickly to create machine learning models for a set of data and output any predictions from that.

Flask: Flask [23] is a RESTful API framework that allows an API to be quickly created with minimum overheads allowing work to be done. Extensions can be added to Flask as needed meaning it should be suitable for whatever our needs become during the proof of concept phase.

NLTK: The Natural Language Toolkit (NLTK) [24] is a python platform for dealing with human language data, allowing for analysis to be done quickly and effectively. It works well with Scikit learn library functions allowing for a smooth and fast development of systems using natural language processing combined with machine learning techniques.

UWCS Server Hosting: The University of Warwick Computing Society (UWCS) allows for members to host websites and data on a server for free. This service may be considered if the project eventually requires any sort of web hosting.

GitHub Version Control: Version control will be utilized to minimize the potential risk of losing project work, by having online backups of the code. GitHub will be used to provide this service, which is beneficial specifically for this project as each member of the team will be able to have access to the code, making it more accessible.
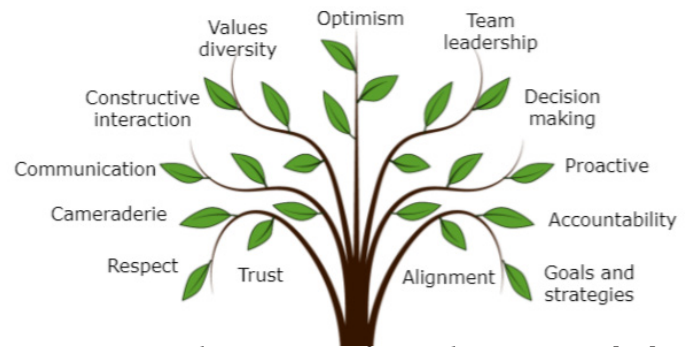
Figure 1.   Characteristics for productive teams [25]

Stakeholders: Whilst there is no specific customer for this project, stakeholders will still be an important resource within this project to inform both technical and management-related decisions throughout the project.

Bi-weekly meetings will be held with the project supervisor (Mike Joy) throughout the project to inform management decisions, and validate work completed every two weeks to ensure the project is succeeding from a less biased view. Mike will be immediately informed of any serious issues that may cause the project to fail, such as serious delays: to serve as an experienced advisor in how to solve any such issues.

Risk Management

A project of this size is inevitably going to encounter some problems along the way. Some problems may be minor and an acceptable risk that can be dealt with when they occur. However, other risks can result in entire project failure. As a result, it is important to anticipate these problems, so that the team can have a resource to look at for guidance on what to do in the event of an issue. Having easy access to a resource like this ensures that problems can be tackled quickly, potentially reducing the severity of the issue. The resource for handling risks is the risk table on page 6, which details: the risk; the area of the project it affects; how it could be handled; and a typology of how we are dealing with the problem.

Project Risk Table

Risk	Risk Type	Risk Management	Risk Response Type
A team member is unable to continue working on the project	Team	Reducing the scope of our project to compensate to prevent burn out in remaining team members	Mitigate
A planned feature is not as feasible as initially expected	Scope	Ensure a feature is fully researched before committing to its development, to avoid time being wasted on trying to implement a solution that would be out of scope	Avoid
Team member does not complete a task by the set deadline	Team	Keep tabs on team members and utilize a pull system to make sure members are not overworked. Be prepared to descope the project	Avoid
Team member keeps missing meetings and not staying in communication	Team	React quickly and contact them via Email, Slack, and Messenger. If still no update, inform the supervisor.	Accept
Dataset is not suitable	Resources	Another dataset would have to be found or Gathered. Existing datasets for Social media content exist, by aggregating several we would lessen the impact of a single dataset being unusable for our project	Mitigate
Team is unproductive/ unmotivated	Team	The project management plan for the team is founded on values proven to be indicative of a productive team	Mitigate
Team members have a disagreement that halts progress	Team	Hold an in-person meeting ASAP with the whole team. Define exactly what the problem is. Allow all parties to express their stance, encouraging honesty and clarity. Brainstorm solutions. Choose a fair solution, expect to compromise in some aspects	Accept
A team member is struggling with a task they have started	Team	First, ask for help. The team members should prioritize their own tasks, and if resources are available, aid with the struggling task. If the struggling task is of low priority and there is not an excess of resources available, consider removing it from the scope	Accept
		Plan to use pair programming for the most difficult tasks, so that the pressure of such a task is not put on a single person	Mitigate
Work is deleted	Resources	A backup should be made of all work that is not made on cloud services, as soon as the work is completed. This should limit the amount of work that could be lost	Mitigate

Risk	Risk Type	Risk Management	Risk Response Type
The scope is completed much sooner than anticipated	Scope	Extra possible scope should be thought of at the start of the project, with the deliverables organized using the MoSCoW method (extra scope is labeled as C - could do if there is enough time)	Mitigate
		Hold a meeting to discuss whether to work more on existing features or to come up with new features	Accept
System requirements are not adequately defined so the team is unsure what features to develop and how	Scope	Use an Agile project management methodology, to allow for changes to requirements during development	Mitigate
Ineffective communication within the team	Team	Assign a project manager and hold them accountable for encouraging and facilitating good team communication	Mitigate
A technology being used is difficult to adjust to and eats into development time	Resources	Clarify the technologies that the team are equipped with, and establish everyone's abilities early on. Only plan for activities which use these resources	Mitigate
A sprint deadline is approaching but the sprint backlog is still full because the team moved slower than expected	Resources	Plan for float time after each sprint, with the first sprint having the longest float time because this is the sprint with the most uncertainty. The sprint review meeting will help to estimate the pace of the team	Mitigate
		Prioritize the sprint backlog at the start of the sprint so that the most important tasks are completed first	
		Consider descoping the feature	

Project Management

Team Values Introduction

The team must ensure the project management strategies in this section result in the basis for a productive team, as this will aid towards the likelihood of the project being successful. Figure 1 outlines the characteristics of productive teams, as formalized by the University of Warwick Module, CS352 — Project Management for Computer Scientists. These characteristics form the team values, helping to establish a common mission for the team to work towards. The objective of this section is to plan a project management strategy that ensures each team value is considered. If all team values are accounted for, then the team can be confident that they are prepared to undertake this project in a productive and successful manner.

Alignment

When making project management decisions in this section, the team members are mindful of the team values, which enables them to be in alignment; working towards establishing a productive team. Also, weekly meetings help to refocus the team on working towards a successful project.

Decision Making

This section designs a process for making transparent, efficient and provably effective decisions throughout the project. Having a clear decision making process aids proactiveness, because it reduces the impact of decision paralysis, as there is a clear strategy for how

to tackle a difficult decision with many options. The decision making process design is influenced by the multi-criteria decision analysis technique described in the Guide to the Project Management Body of Knowledge (*PMBOK*) Section, 8.1.2.4 [26]. The design is enumerated below:

1. **Choose a set of metrics relevant to the decision, e.g. cost, schedule, risks, and alignment with team values.**
2. **Weight the criteria**
3. **List the options for the decision**
4. **Put your name somewhere with the decision for accountability**
5. **Perform the multi-criteria analysis**
6. **Choose the option with the highest score**

This decision is justified on page 9, using the exact process enumerated above - which is to act as an exemplar for the decision making process to be used in this project.

Example of how to make a decision

1. *Choose a set of metrics relevant to the decision*

The criteria chosen are some team values relevant to this decision. These metrics are defined below:

Communication: refers to how much clear and efficient communication the option allows for.

Accountability: refers to the clarity of whose responsibility a decision is.

Values Diversity: refers to how much an approach enables different perspectives and ideas.

Decision Making: refers to how well the option meets the team's value of making transparent, efficient and provably effective decisions.

2. *Weight the criteria*

The weighting of the criteria is shown in the green column headings on page 9, with x2 being worth twice as much as x1.

3. *List the options for the decision*

PMBOK's Section 5.2.2.4 [26] suggests some decision making techniques, which form the options for this decision.

Voting: group discussion of every decision.

Autocratic decision making: allowing one individual to take responsibility for a decision.

Multi-criteria decision analysis: one person making the decision, but before doing so would have to offer any alternatives and justify their choice.

4. *Put your name with the completed table somewhere public for accountability.*

In this example the decision was made by Daisy.

5. *Perform the multi-criteria analysis*

For an example of the table see page 9.

6. *Choose the option with the highest score*

The option with the highest score is the multi-criteria decision analysis (as shown in the example on page 9). This concludes the example of how decisions are to be made in this project. Following this technique ensures that transparent, efficient and provably effective decisions are consistently made throughout the project, aiding the team to be productive and, as such, aiding the project to be successful.

Optimism and Camaraderie

The main project work is performed remotely, with in-person meetings taking place to discuss progress. The regular in-person meetings celebrate accomplishments, which helps the team to maintain enthusiasm for the project, subsequently leading to a sense of optimism within the team.

However, the somewhat formal interactions about successes and goals can lead to a lack of playfulness and empathy within the team, which is known to result in a less productive team. To encourage camaraderie in a mostly online project, the team has made a channel in the project's Slack specifically for team jokes. This channel allows team members to laugh about mistakes and vent their frustrations about difficult parts of the project. This creates a stronger sense of belonging to the team, in turn leading to a more productive team.

Communication

Along with weekly meetings as a team, the team also meets bi-weekly with the project supervisor. The supervisor meetings encourage the team to be clear and direct with their communication, as gossiping or complaining would be unprofessional. This helps the team members to remember to maintain a level of maturity in their discourse.

The team Slack is separated into channels for relevant conversation topics, such as: #resources, #meetings and #proj-specification. This helps to keep conversations clear and organized, which is helpful for anyone who was not part of that conversation to keep informed.

Multi-criteria analysis for decision making

Decision Making Process Option	Communication x2	Accountability x2	Values Diversity x1	Decision Making x2	Score (of 21)
Voting	3/3 Whole group discussion.	1/3 Unclear who's fault a bad decision is.	3/3 Whole group discussion allows multiple views to be considered.	1/3 Time-consuming activity, which means that it is not an efficient technique.	13/21
Autocratic decision making	1/3 Have to trust other people's judgment and cannot share input.	3/3 People know who has made the decision.	1/3 Only one person decides.	3/3 Good accountability leads to better standards and quality; nobody wants to be linked to a bad decision. So the method is effective.	15/21
Multi-criteria decision analysis	3/3 Clearly justified decisions explain exactly why a decision was made - avoiding disputes and gossiping.	3/3 People know who has made the decision.	2/3 Considering alternatives aids open-mindedness. But still the decision is only made by one person	3/3 Transparent and proven to be effective.	**20/21**

Constructive Interaction, Respect and Trust

The sprint review meetings aid constructive interaction as they are designed specifically to be a safe space to give and receive honest feedback, in order to grow and improve as a team. Being able to speak so freely and openly encourages trust and respect between team members, because such honesty empowers everyone to contribute their best during a sprint, so as not to let the team down and be criticized during sprint review.

Methodology Choice Justification

There are two umbrella terms for project management methodologies: Waterfall and Agile. Generally, Waterfall is best for a project with a fixed scope to be completed. However, the team is taking the Agile approach, as this project has a fixed amount of time and resources; the scope is what is to be estimated. This incremental style means the team can begin with a simple social media skeleton, and onto that add features i.e. scope. This plan means that each iteration ends with a working system.

As stated before, Agile is an umbrella term of project management methodologies. This section considers and compares a few of the approaches contained under the umbrella of Agile, in aim of finding a suitable approach for this project. The metrics, by which the approaches are compared, are defined below:

Reliability of Workflow: refers to how the team has other commitments (like coursework) throughout the duration of the project. This metric considers the approach's suitability to a team with an inconsistent amount of time to work on the project.

Meeting a Fixed Deadline: refers to how well the approaches enable the team to meet the strict deadlines involved in this project. One measure of success for the team is reaching schedule targets, so it is important that the approach can satisfy this.

Sureness of Scope: refers to how the team initially has a vague idea of scope. This metric considers the suitability of each approach with regards to handling an unclear set of features.

Measuring Success: refers to how easily the approaches allow the team to validate scope.

Bureaucracy: refers to the amount of meetings and organization involved with each approach. The team does not want excessive bureaucracy in order to maximize the flow of work.

The three Agile approaches being considered for this project are:

Scrum: work is organized into sprints. Items from the product backlog fuel each sprint backlog, defining how much work should be done in a particular sprint. All work is planned at the start of the sprint.

Kanban: work is organized into a pull system, where work is done when resources are available.
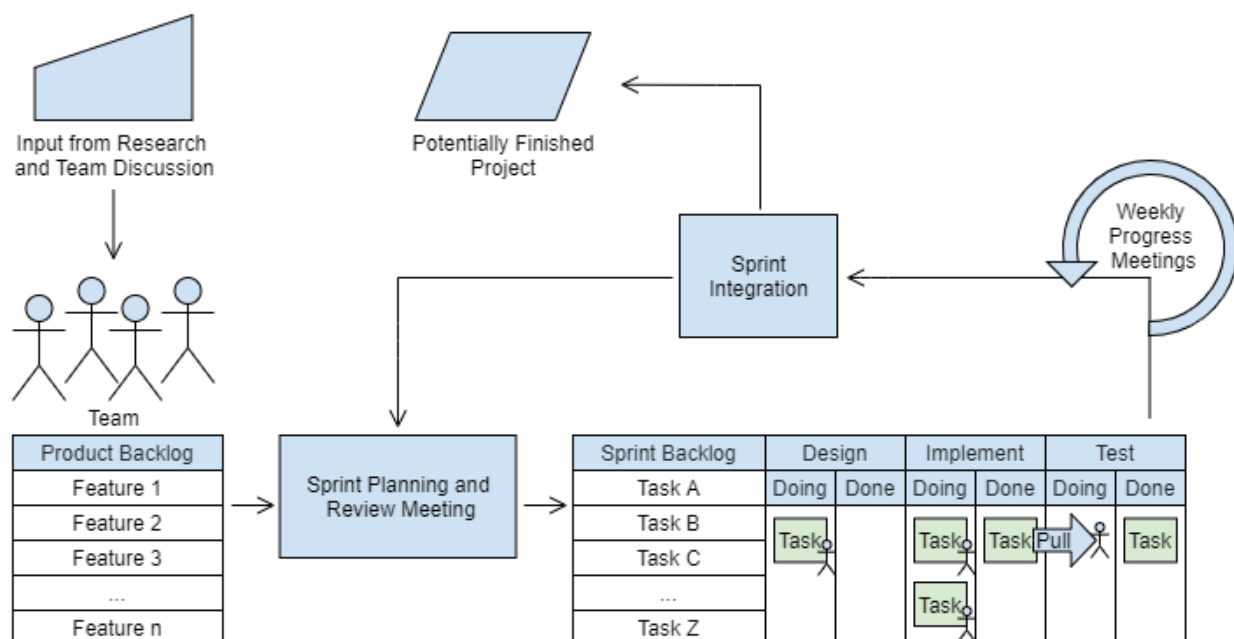
Figure 2. The Scrumban project management methodology to be followed in this project. This defined approach is an example of applying the Goals and Strategies team value, as it shows how the team plans to achieve their development goals.

Multi-criteria analysis for development methodology

PM approach Option	Reliability of Workflow x2	Meeting a Fixed Deadline x2	Sureness of Scope x1	Measuring Success x2	Bureaucracy x1	Score (of 24)
Scrum	1/3 May not be able to reliably commit to do an amount of work in the sprint, as the team has other commitments (i.e. varying amounts of coursework)	3/3 You plan towards long term goals.	2/3 Scrum needs an initial vague list of features (product backlog), which we have. But, we don't know the sub-scope of each feature, as it will be planned at the start of each sprint, so if a feature turns out to be more interesting or complex during development, then we won't be able to deviate from that specific feature's plan	3/3 Sprint review meetings allow the team to incrementally measure success. Scope is continually validated throughout development	1/3 Daily meetings to discuss progress would be time-consuming and possibly futile as the team may not be able to work on it everyday. Sprint review meetings would help the team to continually improve and learn how to function more efficiently as a team	17/24
Kanban	3/3 Allows the team to work on the project whenever they are able to. This ties with the Lean Muri principle, that people cannot work at an unsustainable pace e.g. avoiding the overburdening of coursework and the project	1/3 There's no overall long-term goal to reach with Kanban, so meeting a fixed deadline will not be guaranteed	1/3 Hard to estimate how much scope will be completed by the end of the project	1/3 There is no end point/ goal in Kanban, so measuring the success of the project will be difficult, as there is no plan to compare progress to	3/3 No excessive planning or paperwork. Team members just do work when they get time to, rather than wasting time in meetings	14/24

PM approach Option	Reliability of Workflow x2	Meeting a Fixed Deadline x2	Sureness of Scope x1	Measuring Success x2	Bureaucracy x1	Score (of 24)
Srumban	2/3 Without the sprint being planned, it may be difficult to motivate the team to start work on the sprint goal, when they have other commitments. But the flexibility of Scrumban means that the team members do not have to commit to an amount of work, just do as much as they can in the time-scale, with other members being able to take on extra tasks	3/3 With Kanban, the aim is to streamline the delivery process, so the amount of work in a sprint will be more when compared to the more bureaucratic style of a normal Scrum sprint. Scrumban is a leaner version of Scrum, therefore work can be completed faster than in Scrum - at worst at the same speed	3/3 Scrum is ideal when work is in large vague chunks, like in this project. But the Kanban element gives us flexibility during a sprint to take a feature further (or restrict it back more) than perhaps we initially thought we should	2/3 The lack of planning in each spring means that measuring the sprint's success could be difficult, as the work may change throughout the sprint. However, sprint review meetings mean that the team will frequently discuss progress of the product backlog	2/3 Don't necessarily need daily meetings, as work is organized in Kanban. Just need a meeting at the start and end of a sprint for planning and review respectively	19/24

Scrumban: work is organized into sprints. Work is pulled from the sprint backlog during the sprint, like a Kanban board.

The table on page 11 performs the multi-criteria analysis for the decision of which project management approach is the most suitable for this project, this decision was made by Daisy.

Performing multi-criteria analysis on a selection of Agile project management methodologies allowed the team to confidently choose a suitable methodology for this project: Scrumban. Naturally, the approach has been tailored to suit the requirements of the team. The details of this approach are detailed in the following subsection.

Methodology Design & Workflow

The modified Scrumban approach for this project is illustrated in Figure 2. This section explains the diagram and justifies any changes made to Scrumban.

Input and Product Backlog

The background research performed in this specification feeds the product backlog, which will contain a list of vague features for this project to develop. Once the product backlog is established, the team can start the first sprint.

Sprint Planning Meeting and Sprint Backlog

Each sprint begins with a sprint planning meeting. After the first sprint, this meeting will also review the previous sprint. The review stage of the sprint meeting involves reviewing the product, suggesting possible improvements where necessary. These improvements will be listed, with the possibility to be worked on if the team finishes the project earlier than anticipated (this opportunity risk is noted in the risks section of this document). The review will also celebrate the achievements of the team during the sprint, so as to motivate the team to continue their efforts.

The planning stage of the sprint involves choosing a feature to work on, as well as breaking that feature down into more precise, yet still generally vague, tasks to be completed during the sprint. This sprint backlog of tasks can be added to or removed from during the sprint if required, though this is discouraged in order to inspire a more thorough planning phase at the start of the sprint.

Design

The sprint backlog tasks are pulled into action on the sprint's Kanban board. The phases of the Kanban are to first perform any research, design or planning needed for the task.

Implementation

Once this is done, the task is pulled into implementation, either by the same or a different team member.

Testing

After implementation, each task is tested. For some tasks this will involve technical testing of some code implementation. For other tasks, this will simply be a case of measuring success of the activity. For example, a task of finding and preparing a suitable dataset would not involve any functional testing. The success could instead be measured on how similar the dataset is to the design for the required dataset.

After testing, the team must integrate the work that has been done with the skeleton. Each sprint must produce a potentially finished project, as this is what makes our approach incremental. Over the sprints, the team iteratively builds this toolkit for mental health within social media.

Work in Progress Limits

The Kanban board design, in Figure 2, limits the amount of work in progress to two tasks. This is to prevent all team members from, for example, pulling all of the tasks from the sprint backlog and designing them, but never implementing or testing them. This would halt progress, and the sprint could end with lots of design work but very few implemented parts of the product backlog feature. The work in progress limit means that no more than two tasks can be waiting to be implemented or tested at a time. This will keep the workflow moving, as the team would have to pull a task that needs to be implemented before designing any more tasks.

Weekly Progress Meetings

One of the modifications made to Scrumban for this project is that there are no daily stand-up meetings during the sprint. The reason for this is that the team has an unreliable workflow, with other commitments sometimes taking precedence over the project. This means that the team will not be working on the sprint backlog everyday. As a result, the meetings would be a waste of resources, as there would be no progress to discuss on most days. However, the team does plan to have weekly meetings to talk about progress and motivate ourselves for the following week of the sprint.

Sprint Integration

The work performed during the current sprint must be integrated with the rest of the project. This involves not only the integration of code with any relevant

interfaces, but also concluding the feature with regards to updating any paperwork with information about results.

Potentially Finished Project

Each sprint should end with a polished set of work: code should be integrated, commented and formatted to a professional standard; results and evaluations should not be left in note form; and the team should be in a position to start the next sprint with no overlap of work from the previous sprint.

Roles

The team for this project consists of only four members: Lewis Buttle, Matthew Jameison, Daisy Kennedy, and Patrick Michaels. For this reason, the team has chosen to not be overly bureaucratic with the assignment of roles within the team, choosing to avoid assigning roles for activities in which all team members will take part. For example, establishing that all members are on the development team is unnecessary; the responsibility for each team member to work on the toolkit is to be assumed.

One role that the team decided is worth assigning is that of the project manager. This was assigned to Daisy. Having good team leadership allows the team to know who is accountable for clear team communication. For example, the project manager will account for communication failings like a team member not knowing what time a meeting is and therefore missing a session. That is not to say that the project manager is responsible for organizing meetings; they are just accountable for any problems with communication surrounding that. Other areas that the project manager is accountable for are:

- Ensuring fairness during the handling of inter-team disputes
- Making sure that if a risk materializes, then fast action is taken to minimize the damage
- Ensuring multiple team members are not solving the same task
- Creating a space in which all team members feel comfortable to contribute
- Checking in with team members to see if anyone is struggling, and trying to find ways of helping them if they are

Another role that is necessary for this project is a scrum master. The scrum master is responsible for:

- Organizing group meetings during the sprint
- Maintaining contact with the project supervisor during the sprint

The scrum master role is a more active role than project manager; the scrum master is responsible for actually doing the tasks above, not simply to blame if they are not done. For the sake of fairness, the person with this role will be different for each sprint. The allocation of this role is to be determined during the sprint planning meeting.

Project Management Evaluation

The team has founded a clear and effective methodology for making decisions; multi-criteria decision analysis. This method allows the team members to respect one another's judgment, as the basis for their opinion is well structured and based on clear metrics.

The development methodology chosen is Scrumban. This choice was made using multi-criteria decision analysis, and so is well justified. The Scrumban used in this project has been tweaked to best suit the needs of the project and the team. The methodology has been explained in detail, ensuring the team is in agreement of how the project will be completed.

The aim of this section was to form and justify a project management approach that would enable the team to be both productive and successful. This section has designed the approach to enable the team to possess all of the characteristics that are needed for a productive team. This is indicative of the team's project management approach being strong and reliable, and of this section being successful in its aim.

Schedule

A schedule was constructed to demonstrate how the Scrumban methodology will be effectively utilized throughout the project, as well as to provide a logical justification as to how the project will succeed within the project's limited time frame. The schedule is able to justify this success by illustrating the self-defined deadlines of the team for each specific sprint. This allows the team to subsequently plan the timeline, which will be followed to: produce each feature of the project, and demonstrate that a feasible amount of time has been designated for each sprint.
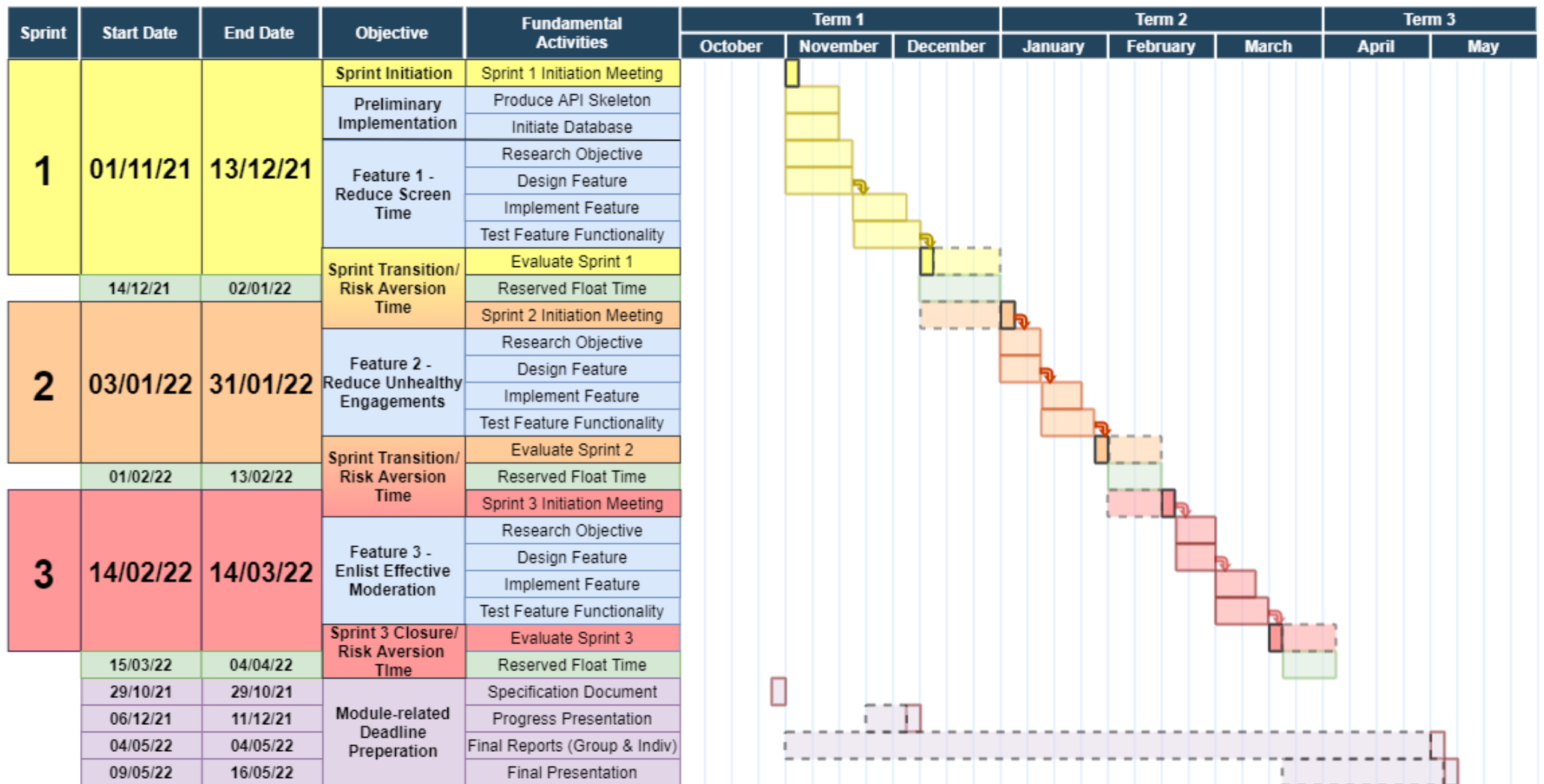
Sprint	Start Date	End Date	Objective	Fundamental Activities	Term 1			Term 2			Term 3	
					October	November	December	January	February	March	April	May
1	**01/11/21**	**13/12/21**	Sprint Initiation	Sprint 1 Initiation Meeting								
			Preliminary Implementation	Produce API Skeleton								
				Initiate Database								
			Feature 1 - Reduce Screen Time	Research Objective								
				Design Feature								
				Implement Feature								
				Test Feature Functionality								
	14/12/21	02/01/22	Sprint Transition/ Risk Aversion Time	Evaluate Sprint 1								
				Reserved Float Time								
2	**03/01/22**	**31/01/22**	Feature 2 - Reduce Unhealthy Engagements	Sprint 2 Initiation Meeting								
				Research Objective								
				Design Feature								
				Implement Feature								
				Test Feature Functionality								
	01/02/22	13/02/22	Sprint Transition/ Risk Aversion Time	Evaluate Sprint 2								
				Reserved Float Time								
3	**14/02/22**	**14/03/22**	Feature 3 - Enlist Effective Moderation	Sprint 3 Initiation Meeting								
				Research Objective								
				Design Feature								
				Implement Feature								
				Test Feature Functionality								
	15/03/22	04/04/22	Sprint 3 Closure/ Risk Aversion Time	Evaluate Sprint 3								
				Reserved Float Time								
	29/10/21	29/10/21	Module-related Deadline Preperation	Specification Document								
	06/12/21	11/12/21		Progress Presentation								
	04/05/22	04/05/22		Final Reports (Group & Indiv)								
	09/05/22	16/05/22		Final Presentation								

Figure 3. A Gantt chart showing the current schedule for the project

It is important to note that the schedule is written at a high level of abstraction, where the activities of the project are written vaguely (for example, 'Implement Feature' being a main activity of each sprint). This vagueness is necessary in order to follow the project's methodology of breaking down the proposed features into sprint backlog tasks during each sprint planning meeting. The team decided it would be unreliable to schedule these activities before they are researched and justified in the sprint meeting. Instead, the proposed solution is broken down to a low enough level as to schedule only the activities that are certain to occur. For example, each sprint will certainly have a pipeline of researching, designing, implementing and subsequently testing the feature. A more specific schedule can be constructed during each sprint's planning meeting.

Each sprint is given a start and end date and their activities are given unspecific blocks of time on the Gantt chart. This helped to realize which activities are parellizable, and which activities are dependent on other activities being completed (denoted via the colored arrows). Identifying these parellizable activities is a benefit to the team, as it shows which activities can be worked on concurrently. This improves the workflow by maximally utilizing the team as a resource by being able to delegate these parallel work packages.

In essence, the schedule gives the team a solid insight into how the project can realistically be completed using the Scrumban methodology. It also gives the team confidence that the project's scope is of a feasible size in terms of the project's strict limitation of time.

Scheduling Risks & Mitigation

A limitation of following a Scrum-based approach, and abstracting the project activities, is that it becomes difficult to describe how much actual work the project requires, as the team does not currently have the specific activities of each sprint. This is a large risk, as each feature may require much more work than expected, which would make it likely that deadlines will not be met. An additional risk is that the team has no experience working together, which could hinder the effectiveness of Scrumban and increase the likelihood of delays if the team is unable to work together effectively. However, these risks have been mitigated in a number of ways.

Firstly, large amounts of float time (shown in green) has been reserved between each sprint in the schedule to combat the potential unreliability

of the scheduling. This float time is a vital resource that will be used to cushion each sprint's runtime, by being available at the end of each sprint if the sprint is going to overrun the proposed deadline. The float time after the first sprint is longer than the subsequent float times. This is because after the first sprint, the team will have a better idea of development pace, and so can more accurately plan the following sprints. Alternatively, the schedule shows that if a sprint does finish on time, then the start of the next sprint can be realistically initiated any time during the leftover float time period, to further mitigate the risk of the next sprint overrunning by starting earlier. In summary, the designated float time can be used to allocate more time to the end of a sprint, or the start of a sprint, for the sake of keeping the project on track to meet the loose deadlines.

Another mitigation is that the fundamental methodology behind Scrumban sprints has been adapted to suit this project. Specifically, the length of the sprints are longer than traditional sprints, ranging from 4-6 weeks depending on the workload expected from the sprint. This was done to combat the unreliability of the team's workflow throughout term, as the team is likely to be busier toward the end of each respective term with other academic obligations. A longer sprint period will put less pressure on the team to output a consistent amount of work every week, and instead output a controlled amount each week to gradually finish a feature by the end of a sprint. This workload will be organized through a team planning meeting at the start of a sprint to outline the sprint goals and schedule. This will be followed by weekly 'stand-up' team meetings to set the workload for the following week, as well as evaluate if the sprint is going at a controlled pace to finish the feature. These weekly meetings will serve as a means of deciding if using float time is necessary. In summary, by utilizing a flexible version of Scrumban, the team has an opportunity to evaluate progress weekly, and ensure that any potential delays or risks can be identified effectively and the necessary actions can be quickly taken to correct them. It will be the entire team's responsibility during these meetings to identify and discuss these delays.

In terms of dependency risk, the schedule illustrated to the team which activities could potentially bottleneck development: a potential risk in which development hinders on a single, potentially drawn out task. The schedule illustrated that there are no single major bottlenecks during development, as each

task has at least one other possible task that could be worked on in parallel. The closest task to a potential bottleneck is the development of the API skeleton, as no features can be implemented without this task being completed. Hence, it was deemed important that this task is focused on to swift completion, to avoid delaying the implementation of any features. This task was not deemed too dangerous of a bottleneck, as the research and design of the first feature can be completed in parallel with this task, as these tasks do not rely on the API skeleton existing. This risk was addressed in the schedule by granting the first sprint an extra 2 weeks of development time, compared to the other sprints. This extra time is thought to be a necessary mitigation for any delays caused by the possibility of this bottleneck occurring.

Measuring Success

To evaluate if the project has been successful in satisfying the set objectives, the team discussed formal testing methods that would be suitable for measuring the success of each feature produced by the project. These testing methods will be constantly employed throughout the project sprints, to consistently validate and justify the work produced at each stage of development to ensure the project is directed toward success.

To test feature functionality, there will be functional test plans written for each feature during the design phase of each feature's respective sprint. These functional tests will be written at the start of a sprint once a feature has been thoroughly researched, and will serve as the 'checklist' for how the team requires a feature to act, to consider it to be functioning as expected: and therefore complete. These tests will act as requirements during each sprint, and will be used during weekly stand up meetings to help measure the progress of the sprint, by checking how many tests a current feature version passes. Additionally, if the work produced can pass a satisfactory amount of these tests, the team can justify closing the sprint and be confident the feature is functioning as expected.

To evaluate if the developed features have succeeded in meeting the original objectives, internal and external objective evaluation will occur. Internal evaluation will consist of the team personally evaluating a completed feature against the original objective, mainly answering the questions 'Does this feature help combat mental health issues?' and 'Is this feature suitable for a social media platform?'. Additionally, external evaluation will aim to answer the same

questions for each developed feature, but based on opinions from outside the team. Opinions that will be formally considered will be the project supervisor, as well as surveying peers and acquaintances of the team that are familiar with social media. The team aims to collect the opinions of at least 25 people, to generate a decent pool of opinions. Ideally, a larger number of people would better support this project, however, it is a limitation of the project that it would not be feasible to find a larger group of willing testers beyond peers and acquaintances, ultimately due to a lack of funding and time to distribute the questionnaire effectively.

The opinions will be collected via a digital survey form that the team will devise, where a person taking the survey will be allowed to interact with each of the developed features, and subsequently answer questions relating to the feature's possible effects on a user's mental health, as well as suitability on a social media platform. The team will then deduce the conclusions from these opinions, to be used to further evaluate each feature. The ethics and legality of performing this survey have been considered, and are discussed in the legal, social and ethical issues section.

Legal, Social, and Ethical Issues

One of the methods of measuring success discussed for this project is to survey acquaintances on the perceived efficacy of our toolkit. The nature of this project means that ethical approval is not required for such a survey [27].

Due to the project's nature of investigating social media features, the project is likely to require the use of existing social media data. The retrieval and usage of this data is subject to legal and ethical dispute due to how it could breach GDPR. Hence, it is necessary to document how the team will handle data with caution to construct the project fairly and legally.

Specifically, the project will require examples of posts made to social media to test any features, and likely train and test any learning models. To keep this data usage legal and ethical, the team has identified Kaggle, which provides large public datasets, many of which relating to social media posts (discussed further in the 'resources' section). With Kaggle, the team can easily check the data licenses of any datasets that we want to use. If the license permits use for this project (the Creative Commons license is generally permissive), then the team can use the dataset and avoid the GDPR concerns of collecting the data ourselves.

Additionally, the team has acknowledged the possibility of producing a 'dummy' dataset from fictitious, generated data. This could be used as a backup data option to Kaggle's public datasets, as fictitious data would not be a concern for legality as no data is collected. However, the data being fictitious would provide weaker, less realistic models and is therefore the less desirable option compared to real, collected data. This mitigation will only be employed if the team struggles at a later date with gaining consent for use of Kaggle datasets.

Conclusion

To summarize, social media has a big impact on the mental health of its users. This project aims to design and implement some tools to help in varying areas of this problem. The areas to be addressed are: a reduction in screen time; a reduction in unhealthy interactions; and an improvement of moderation techniques. A set of expected deliverables have been devised, though these will be refined more during the sprint planning meetings for their respective objectives.

An non-exhaustive list of resources, which the team has at their disposal, has been collated. The resources show the team's limitations, in terms of our abilities with languages and tools. This will be helpful when planning the sprint backlog, as the team will be able to identify the nature of tasks that are achievable by the team.

The likelihood of risks occurring has also been addressed. A plan of how to mitigate the impact of these risks has been devised. This will act as a first point of call when problems inevitably arise. Having this list enables the team to react quickly, hopefully reducing the impact of the problem.

The project management plan for the project is based upon a set of team values that are proven to be characteristics of successful teams. This means that the team is confident that the project management approach, described in this document, is an effective one. The team has a clear process for making decisions (multi-criteria decision analysis) which will help to avoid disputes and conflicts throughout the project. The development methodology for this project is Scrumban. This choice has been thoroughly justified and designed, in order to ensure the team are all aware and on the same page.

Important roles and their responsibilities have been defined. A schedule has been made, which organizes the project into three main sprints, with a float time between each one to mitigate the uncertainty of the sprint activities. Finally, the team has outlined what success would look like to the project, allowing the project to reach a definitive end point.

References

[1] WeAreSocial, "Digital 2021 October Global Statshot Report," Oct. 21, 2021. [Online]. Available: https://wearesocial.com/uk/blog/2021/10/social-media-users-pass-the-4-5-billion-mark/ [Accessed Oct. 26 2021].

[2] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, J. H. Fowler, "A 61-million-person experiment in social influence and political mobilization", *Nature*, vol. 489, pp. 295–298, Sep 2012.

[3] A. Kramer, J. E. Guillory, J. Hancock, "Experimental evidence of massive–scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 24, pp. 8788–8790, 2014.

[4] K. E. Riehm, K. A. Feder, K. N. Tormohlen, R. M. Crum, A. S. Young, K. M. Green, L. R. Pacek, L. N. La Flair, R. Mojtabai, "Associations Between Time Spent Using Social Media and Internalizing and Externalizing Problems Among US Youth," *JAMA Psychiatry*, vol. 76, no. 12, pp. 1266–1273, 2019.

[5] I. Pantic, "Online Social Networking and Mental Health," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 10, pp. 652–657, Oct 2014.

[6] J. M. Twenge, T. E. Joiner, M. L. Rogers, G. N. Martin, "Increases in Depressive Symptoms, Suicide–Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time," *Clinical Psychological Science*, vol. 6, no. 1, pp. 3–17, 2018.

[7] A. L. Gonzales, J. T. Hancock, "Mirror, Mirror on my Facebook Wall: Effects of Exposure to Facebook on Self–Esteem," *Cyberpsychology, Behavior, and Social Networking*, vol. 14, no. 1–2, pp. 79–83, 2011.

[8] C. D. Chasin, "Making Sense in and of the Asexual Community: Navigating Relationships and Identities in a Context of Resistance," *Journal of Community and Applied Social Psychology*, vol. 25, no. 2, pp. 167–180, 2015.

[9] L. Lucero, "Safe spaces in online places: social media and LGBTQ youth," *Multicultural Education Review*, vol. 9, no. 2, pp. 117–128, 2017.

[10] S. L. Craig, L. McInroy, "You Can Form a Part of Yourself Online: The Influence of New Media on Identity Development and Coming Out for LGBTQ Youth," *Journal of Gay and Lesbian Mental Health*, vol. 18, no. 1, pp. 95–109, 2014.

[11] N. Andalibi, P. Ozturk, A. Forte, "Sensitive Self–disclosures, Responses, and Social Support on Instagram: The Case of #Depression," In Proc. 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 2017, pp. 1486–1500.

[12] E. Litt, "Knock, Knock. Who's There? The Imagined Audience," *Journal of Broadcasting and Electronic Media*, vol. 56, no. 3, pp. 330–345, 2012.

[13] A. E. Marwick, D. Boyd, "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience," *New Media and Society*, vol 13, no. 1, pp. 114–133, 2011.

[14] S. Duguay, "'He has a way gayer Facebook than I do': Investigating sexual identity disclosure and context collapse on a social networking site," *New Media and Society*, vol. 18, no. 6, pp. 897–907, 2016.

[15] A. Gibson, "Free Speech and Safe Spaces: How Moderation Policies Shape Online Discussion Spaces," *Social Media and Society*, vol. 5, no. 1, 2019.

[16] P. Weber, "Discussions in the comments section: Factors influencing participation and interactivity in online newspapers' reader comments," *New Media and Society*, vol. 16, no. 6, pp. 941–957, 2013.

[17] T. B. Ksiazek, L. Peer, A. Zivic, "Discussing the News: Civility and hostility in user comments," *Digital Journalism*, vol. 3, no. 6, pp.850–870, 2015.

[18] T. B. Ksiazek, "Civil Interactivity: How News Organizations' Commenting Policies Explain Civility and Hostility in User Comments," *Journal of Broadcasting and Electronic Media*, vol. 59, no. 4, pp. 556–573, 2015.

[19] D. Hartas, "The social context of adolescent mental health and wellbeing: parents, friends and social media," *Research Papers in Education*, vol. 36, no. 5, pp. 542–560, 2021.

[20] Kaggle, "Kaggle Home Page," 2021. [Online]. Available: https://www.kaggle.com/ [Accessed Oct. 27 2021].

[21] Twitter, "Twitter Developer Platform," 2021. [Online]. Available: https://developer.twitter.com/ [Accessed Oct. 27 2021].

[22] Scikit, "Scikit-learn Homepage," 2021. [Online]. Available: https://scikit-learn.org/stable/ [Accessed Oct. 27 2021].

[23] A. Ronacher, "Flask," 2010. [Online]. Available: https://palletsprojects.com/p/flask/ [Accessed Oct. 27 2021].

[24] S. Bird, E. Loper, E. Klein, *Natural Language Processing with Python*. O'Reilly Media Inc, 2009.

[25] University of Warwick. (2020). CS352 Project Management for Computer Scientists.

[26] Project Management Institute, 2017. *A guide to the project management body of knowledge*. 6th ed. Newtown Square, Pennsylvania, USA: Project Management Institute.

[27] M. Joy, "Ethical Consent" University of Warwick Department of Computer Science Webpage, 2020. [Online]. Available: https://warwick.ac.uk/fac/sci/dcs/teaching/ethics/ [Accessed Oct. 27 2021].