



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY

Digital Content Direction

Prepared by Dr. Sherihan Mohamed

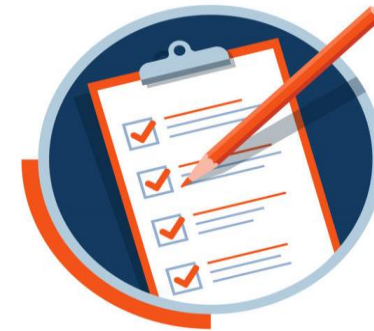
DGM 733

Lecture 10

Remember the last Lecture we talk about:

(Online Community Management)

- * What Are Online Communities?
- * Online Community Managers.
- * Who's Who in the Online Community?
- * Conclusion.
- * References.



Online Community Management and Content Moderation



Lecture 10

Lecture Outlines:

(Content Moderation)

- * Content Moderators.
- * Content Producers.
- * Responding to Hostility, Misinformation and Disinformation.
- * Resources for Community Managers and Content Moderators.
- * Conclusion.
- * References.



Lecture 10

Digital Content Direction

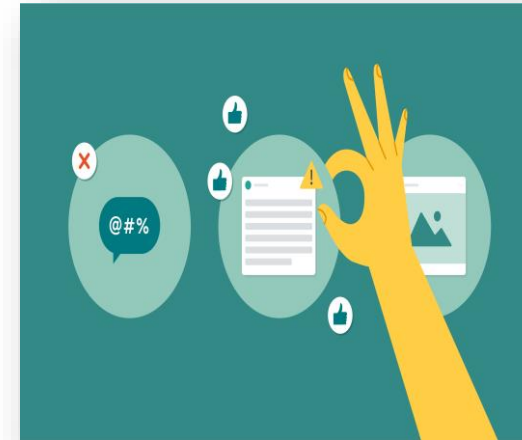


Content Moderators



The terms 'online community management' and 'content moderation' are sometimes used interchangeably; in fact, they have subtly Different meanings.

We have provided a definition of the online community manager before, and a snapshot of the tasks they are required to perform. Content moderators have a somewhat more specific role:





They decide which content can be posted to these spaces and which content must either be deleted or banned outright.

Content moderation can be undertaken by community managers, or by a designated moderator. According to the ACM, content moderation can also be undertaken by:

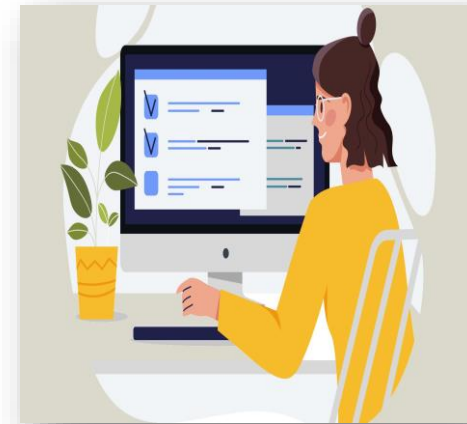
- Social media platform operators
(note: this is typically performed by outsourced contractors)
- Community platform operators
(often performed as a service extension of platform licensing or purchase)





- Organisational representatives who have administrative access and privileges (e.g., media professionals, marketing professionals, customer service agents, product managers)
- Online community members and users.
- Automated machine tools (such as keyword filters programmable by humans).

Moderation can be undertaken prior to material being published (pre-moderation); after published material has been flagged as offensive, or harmful, by community members (reactive moderation); or after publication, when a moderator surveys the material and finds something that goes against community guidelines. Of the three, pre-moderation is potentially the riskiest.



For example, as the ACM point out,
with material that is published after review,
but that is found to contain offensive or
harmful content, 'it is harder to argue that
content was not endorsed by the
administrator, community or host
organisation'.

Moderation can be undertaken manually, e.g., the moderator scans through posts to ensure that all content adheres to community standards and deletes that which does not.

In recent times, automated software has increasingly been deployed for these purposes. That software includes Perspective API, which has been described as ‘a system for the automated detection of “toxicity”’.



Content moderation itself is Not New. ‘What is new’, writes Sarah Roberts, whose book Behind the Screen (2019) examines this line of work, ‘is the industrial-scale organized content moderation activities of professionals who are paid for their evaluative gatekeeping services, and who undertake the work they do on behalf of largescale commercial entities’. The proliferation of platforms has meant a proliferation of online communities to manage and content to moderate.

Lecture 10

Digital Content Direction



Content Producers

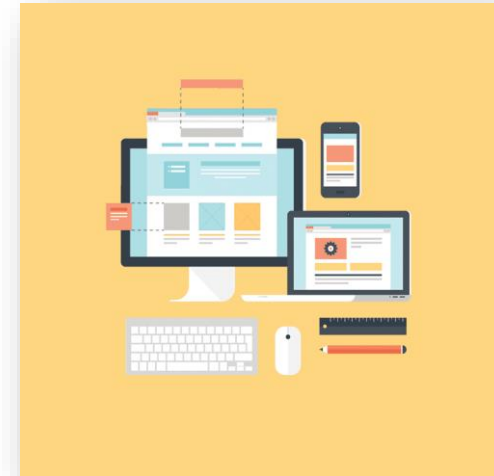




So, are online community managers and content moderators content producers?

The answer would seem to be 'no', at least at first glance. These media professionals are not (usually) charged with producing vlogs and podcasts; they're not the ones who produce op eds or promotional material.

A more careful assessment reveals, however, that community managers and content moderators help produce something that is just as important as those artefacts and that is the user's experience of a particular online community.





Through deciding what content is acceptable for a particular community and what is not, these media professionals play no small role in determining whether somebody will or will not return to—or, more accurately, be permitted to re-enter—the community they oversee.



In an incisive 2018 study, the US scholar Tarleton Gillespie writes that content moderation is hugely important from the perspectives of profit and public health:

“Platforms must, in some form or another, moderate: both to protect one user from another, or one group from its antagonists, and to remove the offensive, vile, or illegal as well as to present their best face to new users, to their advertisers and partners, and to the public at large.”





In doing this, content moderators are required to adhere to guidelines enforced by the community manager.

Indeed, as we noted before, community managers are sometimes tasked with content moderation activities.

In the next section, we examine how these media professionals respond to digital hostility.

This encompasses the 'offensive, vile, or illegal' content that Gillespie describes above; it also encompasses content that may appear innocuous, but might in fact promote harmful views.



Lecture 10

Digital Content Direction



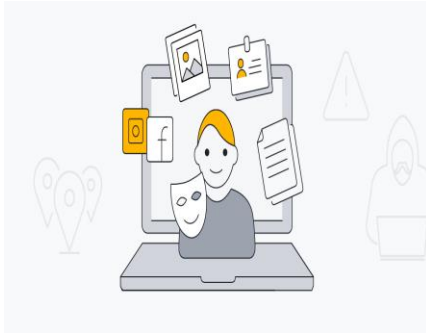
Responding to Hostility,
Misinformation and Disinformation



We use the term digital hostility to describe behaviour that primarily or exclusively seeks to anger, offend or humiliate others.

Digital hostility can take multiple forms, including:





- **Trolling:** the deliberate posting of content that is designed to generate a heightened (and usually adverse) response from the recipient.
- **Doxxing:** the unauthorised public sharing of another individual's personal details (e.g., address, phone number, photographs of loved ones).

- **Ad hominem attacks:** attacking a person rather than their ideas or actions.

Additionally, there is the abuse that moderators are subject to for just doing their jobs. This abuse can stem from the decisions made by community managers and content moderators in enforcing community standards.



Online communities can also be plagued by **misinformation** and **disinformation**.

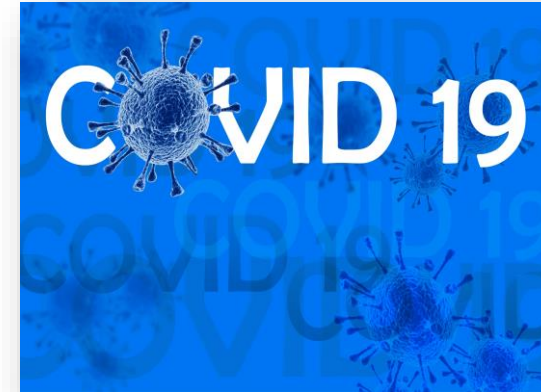
Misinformation refers to factually inaccurate material. This inaccuracy can be attributed to a number of factors, including poor fact checking; it's not necessarily intended to harm.

Disinformation refers to factually inaccurate material that is spread with the specific aim of causing unrest, be this political, physical or psychological.



The riots at the Capitol building in Washington in January 2021 is perhaps the best-known example of the unrest that disinformation can cause. Neither misinformation nor disinformation are new, but both have proliferated during the COVID-19 pandemic.

The misinformation and disinformation that we are referring to concerns the novel coronavirus: how the virus is being used by powerful forces to manipulate the broader population, the efficacy (or lack thereof) of vaccines, and whether COVID-19 even exists.





Sometimes, online misinformation and disinformation can take overtly hostile forms. These include YouTube and Instagram videos in which self-styled 'experts' (many of whom do not have medical qualifications) decry the novel coronavirus as a government fabrication and vaccines as dangerous, even lethal.

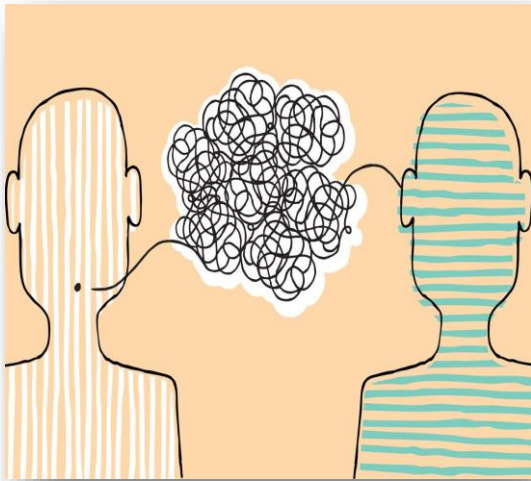
On other occasions, online misinformation and disinformation can appear more inoffensive; it can only be understood by those who are familiar with the codes used by proponents.

We refer here to the 'dog whistle' or a 'phrase that may sound innocuous to some people, but which also communicates something more insidious either to a subset of the audience or outside of the audience's conscious awareness — a covert appeal to some noxious set of views'.



A key example of the dog whistle is the slogan ‘do your own research’, which has been adopted by exponents of COVID-19 conspiracy theories.

‘Do your own research’ sounds like eminently sensible advice—surely one should become well-versed in the topics they’re discussing—until you consider that ‘research’ can mean relying on the scientifically unproven ideas published by members of the public on platforms such as Telegram or 4Chan, and not rigorous, peer-reviewed research.



To detect inoffensive-seeming content, including dog whistles, **community managers** and **content moderators** must possess a considerable deal of media literacy.

They must be able to detect dog whistles and coded, covert endorsements of conspiracy theories and forms of bigotry (racism, anti-Semitism, misogyny, ableism, homophobia, transphobia).

Also, community managers and moderators are required to develop a strong sense of resilience. The term 'resilience' refers to the ability to emotionally withstand sometimes profound challenges and difficulties.

Resilience is Not an innate quality, it's not something that one is born with. Resilience is rather a quality that needs to be cultivated and maintained by the individual. This can be achieved through measures such as the following:





- Access to counselling and psychological services.
- Regular breaks during a work shift to allow the worker to take time away from the challenging content and activity they have been dealing with.

- Support groups in which managers and moderators can discuss the challenges, they have faced in their group, in a non-judgmental environment composed of individuals who have faced and successfully addressed these challenges.
- Reminding oneself that attacks and disinformation say more about their proponents than they do those who are charged with removing them from platforms.

There are no easy answers when it comes to cultivating resilience, **but it is important to do so**; as one 2019 article points out: ‘Moderators battle depression, addiction, and even post-traumatic stress disorder from the endless parade of horrors they consume’.





These are workplace injuries, and they have not always been fully appreciated by media outlets. For example, in a recording leaked in 2019, Mark Zuckerberg (chief executive of Facebook) was heard describing reports by moderators of the distress they experienced in doing their jobs as being ‘a little over-dramatic’.

Lecture 10

Digital Content Direction



Resources for Community Managers
and Content Moderators



Online community management and
content moderation can be challenging
professions. Fortunately, there are a number
of resources and avenues of support for these
professionals.

In Australia, where the authors of this
book are based, there are Australian
Community Managers (ACM).



AUSTRALIAN COMMUNITY MANAGERS

ACM is an online community that seeks to 'grow and enhance community management practice through training, resources, mentoring and meaning-rich connection' (ACM Website).

They also hold the annual Swarm conference, in which community managers and researchers come together to share knowledge and advice.

There is software devoted to building online communities.

This software includes Higher Logic Online Community, which has been described as a 'human-focused engagement platform' (Higher Logic Online Community Reviews undated).



It also includes Mobilise, which purports to provide companies with a new place to engage prospects, partners and customers by providing a private place to build a professional network (Mobilise Reviews undated)



Lecture Outlines:

(Content Moderation)

- * Content Moderators.
- * Content Producers.
- * Responding to Hostility, Misinformation and Disinformation.
- * Resources for Community Managers and Content Moderators.
- * Conclusion.
- * References.



Lecture References:

* Jenkins, H., Ford, S., & Green, J. (2013),
Spreadable media: Creating value and meaning in a networked culture, New York University Press.

* Jay Daniel Thompson & John Weldon (2022),
Content Production for Digital Media: An Introduction, Springer Nature Singapore.





مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY



Thank You





مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY

Digital Content Direction

Lecture 9

DGM 733

Prepared by:
Dr. Sherihan Mohamed

www.esu.ac.ae