Under AI Watch: Understanding Online Behaviors under Supposed AI- Surveillance

EZZEDDINE, Yasmine <http://orcid.org/0000-0002-2810-2231> and BAYERL, Petra <http://orcid.org/0000-0001-6113-9688>

ISSN 17590043

Papers from the British Criminology Conference

An Online Journal by the British Society ofCriminology

Volume 22, 2023

BSC British Society of Criminology

www.britsoccrim.org

British Society of Criminology
PO Box 501, The Nexus Building
Broadway
Letchworth Garden City
Hertfordshire
SG6 9BL

Papers from the British Criminology Conference
An Online Journal by the British Society of Criminology

2023 Conference (27 – 30 June 2023)
Sustaining Futures: Remaking Criminology in an age of Global Injustice
Hosted by the University of Central Lancashire.

With grateful thanks to all our anonymous peer reviewers.

Published annually and available free online at www.britsoccrim.org

Vol. 22

ISSN 1759-0043

Papers from the British Criminology Conference

www.britsoccrim.org ISSN 17759-0043; Vol 22 **Postgraduate Paper**

Under AI Watch: Understanding Online Behaviors under Supposed AI-Surveillance[1]

Yasmine Ezzeddine[2] and Petra Saskia Bayerl[3]

Abstract

Artificial Intelligence (AI) systems being capable of mimicking human intelligence to perform tasks have raised legitimate concerns around ethical and societal implications of their implementation. Despite the fast-paced reproduction of ethical principles to ensure safe and accountable deployment, it would be irrational to consider these sufficient. The adoption of these frameworks heavily relies on citizens' acceptance to the content and the approach of AI implementation. This study focuses on evaluating citizens' behaviours in reaction to assumed AI in online spaces, the factors that trigger rejection and potential changes in behaviour, including potential counteractions. Using an online experiment on Facebook, 30 participants were asked to perform eight tasks, accompanied by think-aloud methodology, under the assumption of AI-surveillance. The findings provide a detailed understanding of the types, reasons, and rationales for agreeing or disagreeing to conduct tasks under assumed AI-surveillance within reallife settings.

Keywords: Online surveillance; Artificial Intelligence; Law Enforcement Agencies; Resistance

---

[2] Yasmine Ezzeddine is an experienced Researcher with a demonstrated history of working in higher education and on multi-national projects. Skilled in areas of Criminal Intelligence, Forensic Sciences, Research Management, Security, Policing and Applied Psychology. She is a PhD student researching Artificial Intelligence use in police surveillance in the UK.

[3] P. Saskia Bayerl is Professor of Digital Communications and Security at CENTRIC, Sheffield Hallam University. Her research interests lay at the intersection of human-computer-interaction, privacy, and transparency management. She holds master's degrees in psychology, linguistics and organisational dynamics from Germany and the US, and a PhD from TU Delft, Netherlands.

Introduction

Surveillance involving systematic monitoring and data collection for purposes of influence or security (Lyon, 2007), are often portrayed as a double-edged sword capable of ensuring protection against crimes on the one hand and facilitating devastating attacks on the other. For instance, surveillance tools, being a source of sensitive information, could cause serious damage when compromised. Recent history is marked by notable attacks targeting industrial facilities such as Trojan Black Energy in 2015 (BlackEnergy, 2021), the WannaCry ransomware attack in 2017 (Mohurle and Patil, 2017) and the Conti ransomware attack on the Costa Rican government (Datta and Acton, 2022). Traditionally, the Panopticon theory (Bentham, 1791) motivated numerous discussions around the costs vs. benefits of surveillance (Foucault, 1991; Orwell, 2000), mostly painting a sinister picture of citizens rejecting surveillance, especially when linked to Law Enforcement Agencies (LEA) (Fussey and Sandhu, 2020). This emphasizes the need to understand the psychological consequences of these technologies in security and criminological domains (Chan and Moses, 2016).

Nevertheless, there seems to be scarce research investigating resistance and counterstrategies to police surveillance and AI-use, particularly on online social platforms. This is where our keen interest in assessing complex perspectives around AI-use in police surveillance stems from, coupled with a curiosity to observe the practical implications of different attitudes in a live experience of online interaction under supposed AI monitoring. The novelty of this research lies in its aim to bridge an important gap between attitudes and behaviours exhibited when assuming AI monitoring of social platforms. The specific design of our study expands knowledge by connecting different disciplines and theoretical frameworks such as self-surveillance (Timan and Albrechtslund, 2015), and factors triggering potential resistance to online monitoring, whether by police or private entities.

In an era dominated by smart technologies that are “profoundly transforming social life, identities and relations” (Smith et al., 2017, p.259), it is crucial to investigate people’s interactions and rationales of merging their physical and virtual existences, which equally contribute to the breadcrumbs constituting their digital footprint (Laufs and Borrion, 2022). The aim is to observe the influence of AI-driven monitoring on citizens’ engagement with different content types on social media. Based on research about the influence of surveillance on behaviour (Ezzeddine et al., 2023), we seek to evaluate when citizens would draw the line for police online monitoring, triggered by which factors, if any, and for what purposes. Briefly said,

we aim to answer the following research question: What triggers resistance in citizens in response to online surveillance by police compared to other entities?

Methodology

Approach

The approach consisted of an online experiment where participants were repeatedly reminded of potential AI-use while performing a series of tasks using their own personal Facebook accounts. Facebook was chosen as it is still one of the most dominantly used platforms (Snelson, 2016), making it “a potentially rich source of qualitative data for researchers” (Franz et al., 2019, p.1). We observed participants' behaviours across three contexts: *Animals World* page, *Debate UK Politics* and *Yorkshire: Crime and Incidents*, where they were reminded of AI online monitoring by police and third parties. This was accompanied by verbalisation to collect concurrent insights into the effect of the contextual manipulations.

Participants

Individuals over 18 years old, who have a Facebook account and were willing to use it to engage with the experimental tasks, were recruited through online advertisements, LinkedIn, and flyers. Participants were also approached through direct contact (in-person or by email) based on referrals (snowball-sampling). The recruitment information contained detailed explanations about what to expect, time needed to complete the experiment (45 minutes to an hour), and the incentive participants received for their time (£20 Love2Shop voucher).

A total of 30 participants agreed to take part (13 women, 17 men) with an average age of 36 years. All participants had (at least) an undergraduate degree, 12 of them identified as members of an ethnic minority and 21 of them worked in areas related to security. Table 1 provides details on the sample.

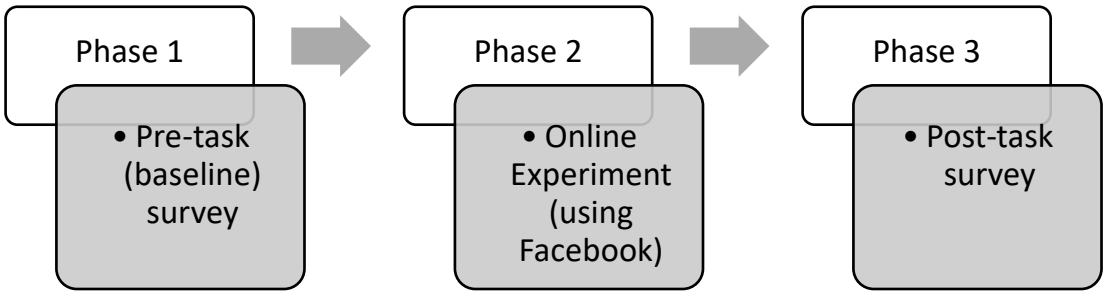
Table 1: Demographic Information of Participants

Participants	Average Age	18-34 years	>35 years	Gender distribution Women / Men	Ethnicity minority/majority	Educational Level Univ. Degree/Master's	Occupation Securityrelated/nonsecurity related
30 participants	36.3	70%	30%	43.3 / 56.7%	40%/53.3% 6.7% prefer not to say	93.3% /6.7%	26.67%/70% 3% prefer not to say

Data collection

The study was conducted remotely in three phases (Figure.1) using MS Teams. Before the experiment, participants received an email with the Information Sheet and Consent Form, while Consent was obtained prior to scheduling the meeting. Phase 1 was a pre-task survey with ten baseline multiple-choice questions on self-rated knowledge of AI and social media activity. In addition, participants were introduced to the *think-aloud methodology* (Ericsson and Simon, 1993) using a short YouTube video, which they were then asked to practice by describing an event that happened to them recently. In Phase 2, the online experiment was conducted.

Figure 1. Phases of the Study



Participants were asked to share their screen and conduct eight tasks on each of three preselected Facebook pages, while verbalizing their thoughts (Güss, 2018). Figure.2 shows the three Facebook pages that were used in this experiment: first, "*Animals World*" for animal lovers, second, "*UK Debate Politics*" for UK politics, and lastly "*Yorkshire Crimes and Incidents"* on crimes and police updates for Yorkshire County. The rationale behind choosing these three distinct content types was to ensure diversity and to address different sharing habits (Lottridge and Bentley, 2018), as the level of user interaction in disparate online public environments can vary (Burbach et al., 2020).

Figure 2. The three Facebook pages used in the Experiment (Phase 2)

The eight tasks and their sequence in the session are shown in Table.2. They increased in difficulty, starting with joining the page, followed by inviting someone to join, reacting to a preselected post, commenting on that post, sharing it to their newsfeed, sharing it to others (via Facebook Messenger or WhatsApp), and finally, creating a post and sharing an image on the page. These tasks were chosen based on popular engagement means on Facebook and were pilot tested for complexity prior to the main study. All participants were presented with the pages and the tasks in the same order.

Table 2. List of tasks to perform on each of the Facebook pages used in Phase 2

Tasks to perform on each page
Join the page
Invite others to join the page
React to the post
Comment on the post
Share the post to newsfeed
Share the post with others (Messenger…)
Create a Post
Share an Image

During Phase 2, participants were constantly reminded of AI-algorithms running in the background of Facebook to monitor online interactions and of their right to refuse performing

any of the tasks. They were further reminded to verbalise rationales (Ericsson and Simon, 1993) behind their decisions when doing/refusing to do a task.

In Phase 3, participants completed a post-task survey requesting basic demographic information (i.e., age, gender, being a member of an ethnic minority/majority, security-related profession, and crime victimisation experience) and a ranking for the eight tasks according to perceived difficulty.

It is crucial to highlight that the monitoring was not simulated, and no algorithms were fabricated to be running in the background to collect any interactions. Instead, the participants would agree to sharing their screen and for the session to be recorded for interpretation and analysis. This approach was chosen to allow observation of participants' real-time reactions under normal conditions, to encourage a revelation of genuine and unrestrained version of their 'true selves' (O'Connor and Madge, 2017).

Ethical considerations

The study has received ethics approval from the Ethics committee at the researcher's university which was granted after providing a clear plan mitigating aspects of confidentiality, voluntary participation, anonymity of data and avoidance of any physical or psychological risks to participants. Specifically, the Information Sheet and Consent Form provided detailed information to participants about voluntary participation, use of personal accounts and right to withdraw. The material was drafted in line with the ethical guidelines set by the British Society of Criminology (2015).

Data analysis

The findings presented here are based on participants' ranking of the tasks from least difficult (1) to most difficult (8) and the verbalisation of thoughts (cp. Charmaz, 2006), which showcase the frequency of engagement and the verbalised thoughts expressed by participants while performing the tasks. SPSS (IBM Corp. 2021) was used to cluster the data from the demographic questions and the difficulty rankings in the post-task survey. Analyses consisted of comparing ranking frequencies across tasks, investigation of engagement levels and ranking decisions for core demographic variables. These analyses used Friedman's test and Mann-Whitney-U test to accommodate for the non-parametric, non-normally distributed nature of the data (Hart, 2001). These tests can assess whether there are consistent shifts or changes in ranks across the different groups without assuming normal distribution (Conover, 1999).

The video-recorded sessions were transcribed verbatim. An in-depth qualitative analysis was conducted on the transcripts using Nvivo (QSR Int. 2020). Thematic analysis was applied to evaluate the underlying themes/patterns that emanate the think-aloud protocol (Clarke and Braun, 2013). This helped in the interpretation of subjective viewpoints through verbalised thoughts justifying participants' choices and behaviours.

This mixed data analysis approach offered a holistic opportunity for cross-validation of results though "*convergence"* or "*confirmation"* (Morgan, 1998, p.365) of findings from two distinct approaches, allowing for triangulation from monitoring of real-time behaviours, difficulty rankings, and verbalised thoughts (Güss, 2018).

Results

In this section, the combined findings from quantitative rankings and qualitative insights from participant's verbalised thoughts will be presented as direct quotes preceded by participant code (e.g., P01, indicating participant 1). A median split for age groups was used with 35 years being the cut off.

Comparing task difficulty

The Related-Samples Friedman's two-way analysis of Variance by Ranks (Table.3) revealed clear differences in difficulty rankings: '*join the group*' was overall ranked as easiest (ranked 14 times as 'least difficult'; mean rank: 2.20), followed by '*react to the post*', (ranked 'least difficult' 13 times; mean rank: 2.35). In contrast, '*share an image*' (mean rank: 7.23) and '*create a post*' (mean rank: 7.30) were deemed as 'most difficult' (cp. Table.3).

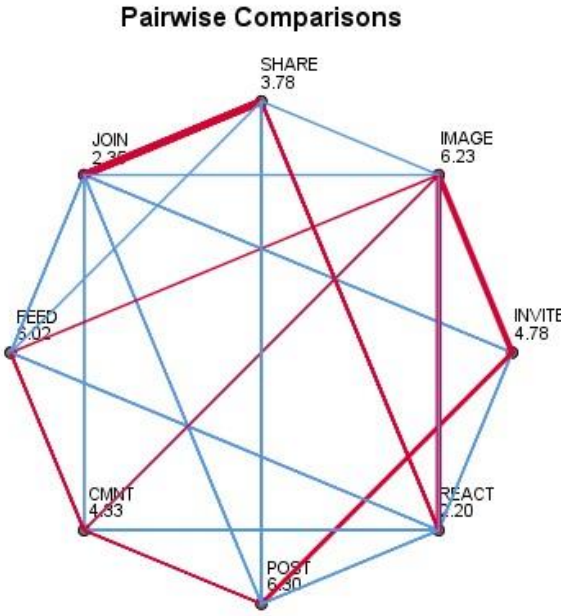
Table 3. Task difficulty as ranked by participants.

Task	Rank difficulty Least (1) to Most (8)	Number of times ranked as such by participants	% of time being ranked as such by participants	Mean Rank
Join	1	14	46.7	2.20
React	2	12	40	2.35
Share with others	2	9	30	3.78

Invite	3	8	26.7	4.78
Comment	4	12	40	4.33
Share image	5	9	30	7.23
Create post	7	10	33.3	7.30
Feed	8	9	30	6.02

A pairwise comparison test was used to reveal linkages between difficulty rankings of tasks. The highest correlation was found between the '*join'* and '*share to friends'* tasks, followed by '*share an image'* and '*invite'* (Figure.3). This suggests that tasks are rated considerably differently, with difficulties for '*join'* and '*share with other'* ranked significantly lower compared to '*share to friends'* and '*share an image'*.

Figure.3: Pairwise Comparison of correlations between tasks rankings



These observations broadly confirm the ranking analysis in that *joining the page* was ranked as least difficult across all participants, followed by *'reacting to the post'* and *'sharing to others'* via /private channels (Direct Message, WhatsApp…). '*Commenting'* and '*inviting others to join'* ranked fourth and fifth, indicating medium difficulty. Of higher difficulty emerged '*sharing an image'* (ranked 6th), while '*creating a post'* (7th) and *'sharing to newsfeed'* (8th) were ranked as the most difficult tasks.

Rationales for disparate difficulty rankings

The think-aloud data enabled an understanding of the reasons for varying levels of difficulty in performing the eight tasks. This started with reasons participants gave for ranking the '*sharing to newsfeed'* as highly difficult, which were often attributed to practical reasons rather than to security/privacy concerns. For instance, *P12: "[I] wouldn't share on my news feed, just because I live in a different area, and I do not think this will be helpful. Otherwise, I would be happy to share it"* or *P25: "I would not share the crime news because I only have 50 people on Facebook and they do not live in the UK, so I don't think it will help."*

Similarly, the lower levels of interaction with '*create a post*' or '*share an image'* tasks on the *Crime and Incidents* page were attributed to Facebook not being perceived as the proper platform to share serious cases: *"The reason why I would not create a post on this page is because I would rather go to the police directly with the information that I have"* (P17). Hence, decisions for not sharing were frequently based on usefulness considerations, triggering resistance to tasks and leading to higher difficulty rankings, e.g., "*how helpful it is to share this post since it can support the police investigation" (P14)*. This aligns with research around benefits vs. drawbacks and purposes of private information sharing online by users of social networks (Syn and Oh, 2015).

The low engagement with '*react'* and '*comment'* tasks on the *Crime and Incidents* page were further attributed to the use of emojis, or generally reacting or commenting on serious news, as being "immoral" and "unethical"; e.g., *P05: "I think it is inappropriate to react to such sad news. Like even if you react with a sad face, or write condolences, it is not going to change anything".*

Interaction levels did not change markedly despite constant reminders of AI-use. Also, no participant opposed or denied the suggestion of AI-tools monitoring Facebook or similar platforms, regardless of their perceived level of AI knowledge. Rather, participants seemed to accept that AI tools are used to monitor online environments. For instance, *P27* referred to *"AI-surveillance of passive scrolling*" as a potential marketing precursor for when a person is not interacting with a post but is spending considerable time on it.

However, different content types resulted in disparate interaction patterns showing higher engagement with the *Yorkshire Crime and Incident* page, compared to the political page. Remarkably, most participants were more likely to conduct the same tasks when related to the

police-related page, than when the feed was linked to other entities. Moreover, several participants were willing to perform tasks that they would not normally engage with, based on 'having nothing to hide' from the police. *P14* argues that they are more inclined to do these tasks on a policing page, because of the oversight and safeguarding efforts that they expect from them. Participants' comments reveal some form of moral obligation to engage and share information that is quite 'serious', compared to political news or debates that can jeopardize their relationships with people of different views. Hence, they *"wouldn't want to be a part of an echo chamber" (P25).* This was coupled to a lack of trust in the admins/members of Facebook political groups, which was openly expressed by *P19: "I do not share on my feed any political posts, because you never know who the real members of that page are".*

Further, most participants were quite aware of the risks of tailored advertising where some even argued that the reason for not wanting to engage with certain posts was to avoid being "*bombarded with similar posts and suggestions on my feed!"* (P13). This concern was often stronger than fear of police monitoring of online behaviours. It coincides with previous research on increased privacy concerns due to intrusive online marketing strategies (Dwivedi et al., 2021), which are shown to have a negative influence on online public engagement (Wang and Herrando, 2019). Participants further expressed their concerns about what "*friends would say if they saw a kitten post shared on the newsfeed*" (*P11*). This suggests that the sharing task was deemed difficult due to social surveillance concerns, which recurred as a potent reason for refusing tasks.

Another less prominent theme was around fear of spreading misinformation/disinformation by disseminating non-trustworthy information/fake news. This suggests that the societal impact of misinformation on members of the public extends beyond influencing opinions and beliefs (Olan et al., 2022) to affect behaviours and online engagement. As *P29* states: *"I would share the crime post to support the investigation but first I would check the source of the information, if I can find a more credible source, like a government or Home office request, I will share that one".* This aligns with concerns around the lack of trust in the social platform itself. In fact, some participants even reported needing the "*government to protect us from unlawful data collection by third parties selling our data and taking advantage of fine prints on websites and social media*" *(P30)*. It may be that the assumption of trustworthiness of a post on a policerelated page contributed to the increased engagement with tasks on the *Yorkshire Crime and Incident Page*. Still, some participants preferred using an external sharing option (e.g., sharing via WhatsApp…) instead of sharing the post on their Facebook page, e.g., *"I would*

not share using Facebook options but take a screenshot and send it externally on other apps, or maybe show them the page" (P09).

This coincides with tendencies to achieving a balance between sharing or hiding personal information (Pavone and Esposti, 2012) and the "complex, often ambiguous and sometimes intangible trade-offs" of posting information (Acquisti et al., 2016, p. 462). Our findings thus align with discussions around balancing privacy rights and moral responsibilities towards public safeguarding and debates around personal information sharing vs. protecting oneself online (Ebina and Kinjo, 2021).

Overall, participants' verbalisations identified eight disparate themes, which can explain why certain tasks were considered more difficult than others:

1. **Awareness of digital footprints:** concerns around being "too visible" online.
2. **Privacy Protection:** concerns about own privacy if conducting a task.
3. **Social Surveillance and Peer Perception:** concerns around what their network and friends would think about what they post/share.
4. **Engagement depending on content types:** individualistic perspective towards acceptance vs. rejection of specific tasks based on content.
5. **Engaging in unusual actions on police-related feed:** accepting to do tasks online that they would not normally engage with
6. **Misinformation/disinformation concerns:** reluctance in sharing posts that might spread fake news.
7. **Moral obligations:** commitments to interacting with posts that might potentially lead to the arrest of a criminal for instance.
8. **Inevitability of online surveillance:** acceptance of constant online monitoring, regardless of monitoring body.

These eight disparate rationales address four broader types of concerns that impacted participants' behaviours, namely: awareness of others watching and judging their behaviours (themes 1-3), impact of the context on which behaviour occurs, including participants' trust in the organisations running the Facebook page (themes 4 and 5), concerns about potential consequences of online behaviours for others (themes 6 and 7), and feeling of unavoidability of surveillance (theme 8).

Comparison for gender differences in task fulfilment

The independent-samples Mann-Whitney U test across tasks (Figure.4) shows that '*create a post'* was perceived as more difficult by women than by men ($U=0.017$, $p<0.05$). Interestingly, women, who expressed rejection of online engagement (through lower engagement and higher difficulty rankings), were mostly concerned about privacy intrusions that bring

"unnecessary attention" to their profiles online. These concerns overlapped with longstanding discussions around online users exposing themselves to online/offline risks through private information sharing on social platforms (Gupta and Dhami, 2015). The fact that in our sample only women raised this issue correlates with suggestions of gender influences on perceptions of privacy. For instance, Rowan and Delinger (2014) show a higher rate of women reporting concerns about collection of location-based data compared to men.

Comparison for age differences in task fulfilment

Overall, older participants (>35 years) completed more tasks per page than participants in the younger age group (35 years or younger): 61.4% compared to 47.1%. Younger participants reported being more cautious about sharing personal opinions/preferences on Facebook, because it made them "*more visible*". They preferred using Facebook "*invisibly*" instead of for self-expression. This was best put by *P17: "My purpose for using Facebook is different. I use it to keep tabs on friends and family and not to express my interests".* One participant admitted to previously sharing personal opinions when they were younger but not anymore: "*I used to do that when I was a bit younger, but now I don't like people knowing what I do or how I think. I don't feel the need to share my opinions, food, or holiday destinations"* (*P20*).

This coincides with existing theories around younger generations' privacy preferences (Blank et al 2014). For instance, the Pew Report (2013) shows that young adults (18 to 29) are keener on limiting private information sharing online and proactively updating their privacy settings (Boyd and Hargittai, 2010). Also, research shows that older adults using Facebook/Instagram seem to rely on these platforms to compensate for the lack of social activity and face-to-face interactions in their daily lives (Sheldon, 2021). Age may thus impact how individuals behave under surveillance, as they have disparate goals for their Facebook usage.

Possible effect of security-related profession on task fulfilment

Participants with a security-related profession showed only one variation, which was a higher reluctance to engage with political content (3.3% engagement with *UK Politics* page compared to 56.3% with *Yorkshire Crime and Incident* and 40.4% with *Animals World* page). In fact, only

one participant working in a security-related profession was willing to engage with the preselected political post. *P17* attributes this to fear of leaving "*political breadcrumbs on the internet that can affect my job applications to positions in the same field*".

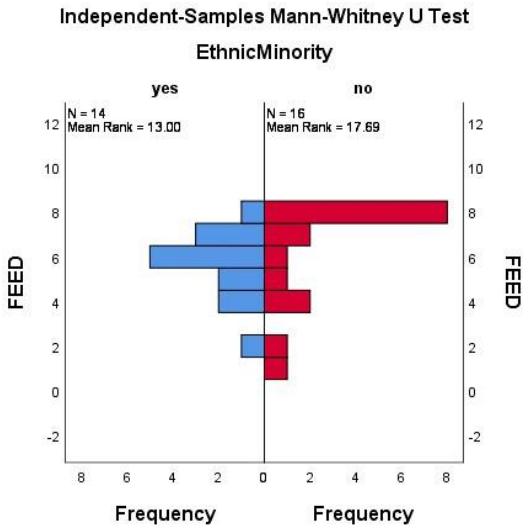
Possible effect of crime victimization experience on task fulfilment

A Mann-Whitney U test results comparing participants with and without crime victimization experiences (referring to any type of crime: financial, theft, fraud, assault…) revealed that participants who identified as victims ranked the tasks of "*invite, comment, share to feed"* as more difficult compared to non-victims ($U=0.093$, $p<0.009$; $U=0.09$, $p<0.5$; $U=0.07$, $p<0.05$, respectively). Recurrent comments suggest a general reluctance amongst this group to create a post or share a picture on any of the pages, regardless of content type. This was attributed to fear of exposing themselves online and attracting "*too much attention*". Especially *'creating a post'* was deemed as a difficult task, which all participants who identified as crime victims refused. These observations coincide with suggestions that crime victimisation can lead to 'victim sensitivity' (Gollwitzer et al., 2015), fears of exploitation (Rothmund et al., 2015) and being more reluctant towards putting themselves under the spotlight (Worsley et al., 2017).

Possible effect of ethnic minority status on task fulfilment

A Whitney-U Test revealed a single difference between participants identifying as ethnic minority members vs. ethnic majority members, namely with respect to '*sharing to newsfeed'*. Participants who did not identify as members of an ethnic minority ranked this task as more difficult (cp. Figure.4; $U=0.015$, $p<0.05$). A bivariate correlation analysis (Trauth, 2007) showed a lower frequency of interacting with political posts for people who identified as ethnic minority members (17% interaction). In addition, ethnic minority participants were more worried about '*creating a post'* or '*sharing an image'* on the *Yorkshire Crimes and Incidents* page or any police-related page/group on Facebook. This is not necessarily linked to a fear or mistrust of police. Instead, they explained their reluctance with fears of "*sharing wrong information that can lead to misinforming the police about serious cases*" *(P21).* These participants favoured using official channels to report or crime or to inform the public about serious news.

Figure.4: Variations in ranking of Share to Feed task between ethnic minority vs. majority.



Discussion

This study explores reactions to assumed online surveillance through AI, comparing three different surveillance contexts. The exploratory mixed-design nature of this study revealed the complexity of making sense of AI-use by LEAs and other entities online, with a specific focus on motivations for personal online engagement and resistance. The findings reveal complex factors that contribute to shaping citizens' perspectives and their online engagement that were largely framed under themes of inevitability of online surveillance, impact of online context and content, concerns about potential consequences of own online behaviours for others and social surveillance concerns. Some of these aspects were coupled with a sense of moral obligation to contribute to public safeguarding efforts.

Our findings expand existing knowledge on surveillance consequences by questioning longstanding notions around privacy models, fear of police monitoring, resistance and change in behaviours and revealing factors in citizens' experiences that shape their opinions, behaviours, and decision-making. This study thus constitutes an important exploration into individuals' rationales when engaging with online content under assumed AI-surveillance.

Our findings show that, in the modern era, individuals' awareness of their 'digital footprints' can lead them to perceive tasks with the most visible footprints (i.e., sharing to newsfeed,

creating own posts) as ‘most difficult’ (Sujata et al. 2016). Yet, individuals performed more tasks on the policing-related page than on private entities’ pages, suggesting that individuals may in fact feel more comfortable with police surveillance than surveillance by other entities (e.g., privacy companies). This suggests that long-defended notions of citizens fearing police surveillance (Trottier, 2017) may have changed, or may at least be more varied than often assumed. Additionally, individuals largely seemed to accept that AI-tools are used to monitor online environments, suggesting a sense of inevitability in their attitudes towards AIsurveillance.

The data further imply demographic variations that indicate that various demographic aspects may shape citizens’ engagement and/or resistance to online AI-monitoring in disparate ways. Specifically, personal and demographic factors, including crime victimization or securityrelated jobs seem to shape choices for engaging/refusing to engage with certain tasks on Facebook. Our study thus illustrates the need for highly context-specific investigations to understand individual reactions to online surveillance.

Our paper also contributes to methodological innovations by enabling a deeper exploration of numerical findings with contextual insights from participants' verbalized thoughts. It demonstrates the potential for using social media platforms not only for data collection but also for real-time qualitative insights, showcasing the adaptability of mixed methods in contemporary research settings. The paper clearly outlines the integration of both quantitative and qualitative data, demonstrating transparency in methodology and analysis. This contributes to methodological rigor where similar mixed-method studies, when appropriately designed and executed, can enhance the generalizability and transferability of findings.

Our approach is especially valuable to understand actual online behaviours and reaction to assume AI-surveillance, in preference to the prevalent study of attitudes such as concerns or acceptance. The identified rationales provide an important foundation to explain decisions and online behaviours which are invaluable in understanding citizens’ perspectives to AI-driven online surveillance. This demonstrates that mixed approaches, in the controlled setting of an online experiment, have proved to be ideal for investigating complex behaviours such as surveillance reactions.

Limitations and future research

Future research can benefit from exploring additional demographic groups, for instance, in terms of age and education. Our sample did not include older participants (over 64 years) nor individuals without a university degree. Including such groups may lead to additional perspectives. Moreover, this study has a restricted sample size. While the sample is substantive for the thematic analysis of the think-aloud protocol, statistical analyses are by necessity more restricted. A replication in larger samples could usefully test and validate our findings, particularly on potential group differences and impact on online context/content. Additionally, this study was conducted only with UK citizens. Extending participation beyond the UK would allow for a comparative approach to reveal whether factors such as disparate cultures, political environments and police perceptions play a role in shaping citizens' stances and reactions towards AI-use by LEAs in online surveillance.

For our study we chose an experimental setting that foregrounded conscious reflection and explanations of behaviours by participants. This introduced some artificiality and behaviours that participants would 'normally' not do, which resonates with Hawthorne's theory where participants exhibit increased performance when watched (McCambridge et al., 2014). It is, however, exactly this ability, to understand what is 'normal' versus 'non-normal' and why, that can be considered as the main strength of our approach. It allowed us to not only observe overt patterns of behaviour, but also unearth the underlying reasons for these patterns. Further research would be valuable to explore 'unscripted' behaviours and reactions online.

References

Acquisti, A., Taylor, C., and Wagman, L. (2016). The economics of privacy, *Journal of Economic Literature*, 52(2), pp.442–492. https://doi.org/10.1257/jel.54.2.442.

Bentham, J. (1791). *Panopticon, or, The inspection-house*. Dublin, Ireland Printed: London Reprinted: T. Payne.

BlackEnergy APT Attacks: *What is BlackEnergy?* Available online: https://www.kaspersky.com/resourcecenter/threats/blackenergy. [Accessed on 22 June 2023].

Blank, G., Bolsover, G. and Dubois, E. (2014). A New Privacy Paradox: Young People and Privacy on Social Network Sites (August 13, 2014). Prepared for the Annual Meeting of the American Sociological Association, 17 August 2014, San Francisco, California., Available at SSRN: https://ssrn.com/abstract=2479938 or http://dx.doi.org/10.2139/ssrn.2479938 Boyd,

D. and Hargittai, E. (2010). Facebook privacy settings: Who cares?, *First Monday*, *15*(8). doi:https://doi.org/10.5210/fm.v15i8.3086.

British Society of Criminology. (2015). 'Statement of Ethics', *British Society of Criminology*. https://www.britsoccrim.org/ethics/

Burbach, L., Halbach, P., Ziefle, M. and Valdez, A. (2021). 'Questions of Scientific Research on Violence and Inequality Applied to Women', *Journal For Educators, Teachers And Trainers*, 12(01). doi:10.47750/jett.2021.12.01.018.

Chan, J. and Moses, B. L. (2016). 'Making Sense of Big Data for Security', *British Journal of Criminology*, 75, pp.299-317. https://doi.org/10.1093/bjc/azw059.
Charmaz, K. (2006). Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. Thousand Oaks: Sage.
Clarke, V. and Braun, V. (2013). 'Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning', *The Psychologist*, *26*(2), pp.120-123.
Conover, W.J. (1999). Practical Nonparametric Statistical. 3rd Edition, John Wiley and Sons Inc., New York, pp.428-433.
Datta, P.M., and Acton, T. (2022). 'Ransomware and Costa Rica's national emergency: A defense framework and teaching case'. *Journal of Information Technology Teaching Cases*, *0*(0). https://doi.org/10.1177/20438869221149042
Dwivedi, Y.K., Ismagilova, E., Hughes, D.L., and Carlson, J. (2021). 'Setting the future of digital and social media marketing research: Perspectives and research propositions', *International Journal of Information Management*, *59*(1), 1–37. Sciencedirect. https://doi.org/10.1016/j.ijinfomgt.2020.102168
Ebina, T. and Kinjo, K. (2021). 'Paradox of choice and sharing personal information', *AI and Society* 38(1), pp.121 – 132. https://doi.org/10.1007/s00146-021-01291-0
Ericsson, K. A. and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.
Ezzeddine, Y., Bayerl, P.S. and Gibson, H. (2023). 'Safety, privacy, or both: evaluating citizens' perspectives around artificial intelligence use by police forces', *Policing and Society*, 33(7), pp.861-876. doi: 10.1080/10439463.2023.2211813
Foucault, M. (1991). *Discipline and Punish: The Birth of the Prison*. Penguin Random House UK.
Franz, D., Marsh, H.E., Chen, J.I., and Teo, A.R. (2019). 'Using Facebook for Qualitative Research: A Brief Primer', *Journal of Medical Internet Research*, *21*(8), e13544. https://doi.org/10.2196/13544
Fussey, P. and Sandhu, A. (2022). 'Surveillance arbitration in the era of digital policing'. *Theoretical Criminology*, *26*(1), pp. 3–22. https://doi.org/10.1177/1362480620967020
Gollwitzer, M., Süssenbach, P., and Hannuschke, M. (2015). 'Victimization experiences and the stabilization of victim sensitivity', *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00439
Gupta, A. and Dhami, A. (2015). 'Measuring the impact of security, trust and privacy in information sharing: A study on social networking sites', *Journal of Direct Data Digital Marketing Prac*tices, **17**, 43–53. https://doi.org/10.1057/dddmp.2015.32
Güss, C. D. (2018). 'What Is Going Through Your Mind? Thinking Aloud as a Method in Cross-Cultural Psychology', *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01292
Hart, A. (2001). 'Mann-Whitney test is not just a test of medians: differences in spread can be important', *BMJ*, *323*(7309), pp.391–393. https://doi.org/10.1136/bmj.323.7309.391
IBM Corp. (2017). IBM SPSS Statistics for Windows, Armonk, NY: IBM Corp. Available at: https://hadoop.apache.org.
Laufs, J. and Borrion, H. (2022). 'Technological innovation in policing and crime prevention: Practitioner perspectives from London'. *International Journal of Police Science and Management*, *24*(2), pp.190209. https://doi.org/10.1177/14613557211064053
Lottridge, D. and Bentley, F. R. (2018). "Let's hate together: how people share news in messaging, social and public networks," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (New York, NY: ACM), pp.60:1–60:13. doi: 10.1145/3173574.3173634 Lyon, D. (2007). Surveillance Studies: An Overview. Cambridge: Polity Press.
McCambridge, J., Witton, J., and Elbourne, D. R. (2014). 'Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects', *Journal of clinical epidemiology*, *67*(3), pp.267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015
Mohurle S. and Patil, M.A. (2017), 'Brief study of wannacry threat: Ransomware attack 2017', *International Journal of Advanced Research in Computer Science,* 8, pp.1938–1940.
Morgan, D.L. (1998). 'Practical Strategies for Combining Qualitative and Quantitative Methods: Applications to Health Research', *Qualitative Health Research*, *8*(3), pp.362–376. https://doi.org/10.1177/104973239800800307
O'Connor, H. and Madge, C. (2017). 'Online interviewing'. In: Fielding N, Lee R, Blank G (eds) *The SAGE Handbook of Online Research Methods*. London: SAGE Publications, pp.416–434.

Olan, F., Jayawickrama, U., Arakpogun, E.O. *et al.* (2022). 'Fake news on social media: the Impact on Society', *Information System Frontiers.* https://doi.org/10.1007/s10796-022-10242-z

Orwell, G. (2000). *1984 Nineteen Eighty-Four.* Introduced by Thomas Pynchon. England, Penguin Classics.

Pavone, V. and Esposti, S. (2012). 'Public assessment of new surveillance-oriented security technologies: Beyond the trade-off between privacy and security', *Public Understanding of Science, 21(5)*: pp.556–572. https://doi.org/10.1177/0963662510376886

QSR International Pty Ltd. (2020) Nvivo (released in March 2020), https://www.qsrinternational.com/nvivoqualitative-data-analysis-software/home

Rothmund, T., Gollwitzer, M., Bender, J. and Klimmt, C. (2015). 'Short- and long-term effects of virtual violence on cooperation and social trust', *Media Psychology,* 18, pp.106–133. doi:10.1080/15213269.2013.841526

Rowan, M. and Dehlinger, J. (2014). 'Observed Gender Differences in Privacy Concerns and Behaviors of Mobile Device End Users', *Procedia Computer Science*, 37, pp.340–347. https://doi.org/10.1016/j.procs.2014.08.050.

Sheldon, P., Antony, M.G., and Ware, L. J. (2021). 'Baby Boomers' use of Facebook and Instagram: uses and gratifications theory and contextual age indicators', *Heliyon*, *7*(4), e06670. https://doi.org/10.1016/j.heliyon.2021.e06670

Smith, G.J.D., Moses, B. L., and Chan, J. (2017). 'The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-driven Approach', *The British Journal of Criminology*, *57*(2), pp.259–274. https://doi.org/10.1093/bjc/azw096.

Snelson, C.L. (2016). 'Qualitative and mixed methods social media research', *International Journal of Qualitative Methods, 15*(1). doi: 10.1177/1609406915624574.

Sujata, J., Saxena, S., Tanvo, G. and Shreya. (2016). 'Developing Smart Cities: An Integrated Framework', *Procedia Computer Science*, *93*, pp.902–909. Sciencedirect. https://doi.org/10.1016/j.procs.2016.07.258

Syn, S.Y. and Oh, S. (2015). 'Why do social network site users share information on Facebook and Twitter?', *Journal of Information Science*, *41*(5), 553–569. https://doi.org/10.1177/0165551515585717

Timan, T. and Albrechtslund, A. (2015). 'Surveillance, Self and Smartphones: Tracking Practices in the Nightlife', *Science and Engineering Ethics*, *24*(3), pp.853–870. https://doi.org/10.1007/s11948-0159691-8

Trauth, M.H. (2007). *Bivariate Statistics*. In: *MATLAB® Recipes for Earth Sciences*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72749-1_4

Trottier, D. (2017). '"Fear of contact": Police surveillance through social networks', *European Journal of Cultural and Political Sociology, 4(4)*, pp.457–477. https://doi.org/10.1080/23254823.2017.1333442

Wang, Y. and Herrando, C. (2019). 'Does privacy assurance on social commerce sites matter to millennials?', *International Journal of Information Management, 44(2),* pp.164–177.

Worsley, J.D., Wheatcroft, J.M., Short, E., and Corcoran, R. (2017). 'Victims' Voices: Understanding the Emotional Impact of Cyberstalking and Individuals' Coping Responses', *SAGE Open*, *7*(2). https://doi.org/10.1177/2158244017710292