

2024

Algorithm-Driven Systems in the Penal System: A Systemic Critique

Miriam Klöpper

FZI Forschungszentrum Informatik, Germany, kloeppe@fzi.de

Daniel Vonderau

FZI Forschungszentrum Informatik, Germany, vonderau@fzi.de

Christoph Becker

FZI Forschungszentrum Informatik, Germany, Christoph.Becker@fzi.de

Follow this and additional works at: <https://aisel.aisnet.org/wi2024>

Recommended Citation

Klöpper, Miriam; Vonderau, Daniel; and Becker, Christoph, "Algorithm-Driven Systems in the Penal System: A Systemic Critique" (2024). *Wirtschaftsinformatik 2024 Proceedings*. 47.

<https://aisel.aisnet.org/wi2024/47>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik 2024 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Algorithm-Driven Systems in the Penal System: A Systemic Critique

Research Paper

Miriam Klöpper¹, Daniel Vonderau¹, Christoph Becker¹

¹ FZI Research Center for Information Technology, Department of Software Engineering,
and Department of Cybersecurity and Law
Karlsruhe, Germany
{Kloepper,Vonderau,Christoph.Becker}@fzi.de

Abstract. Algorithm-driven systems, including those containing Artificial Intelligence, are increasingly deployed within criminal justice systems. This includes facial recognition in public places for identifying people, pattern recognition for real-time detection of crimes, and algorithmic surveillance and recommendations as features within penal facilities. Such systems are frequently criticised, as they might perpetuate discrimination, and they pose a potential threat to privacy. Data protectionists and other human rights activists are paying attention to surveillance in public places, and the matter is strongly represented in political discourse. We find, conducting a hermeneutic literature review, that surveillance and the algorithmic evaluation and assessment of people in penal facilities, however, are largely absent from public and scholarly discourse. Applying a Foucauldian lens, we perform a critical discourse analysis, and argue that those current developments show a lack of dignity, respect and moral behaviour in an increasingly datafied world.

Keywords: Critical Discourse Analysis, Literature Review, Surveillance, Prison Technologies

1 Introduction

A recent headline on a German news website reads: "*When a hug looks like a chokehold*" (Susanka, 2023). The article discusses a current case of algorithm-supported video surveillance in German public spaces, noting that the portrayed algorithm, designed to detect violent behaviour, struggles to distinguish between hugs and chokeholds. While the headline may seem macabre, the use of such systems is intended to enable quicker responses from police or medical services. These systems are relatively new in European public places but are increasingly common in penal facilities worldwide, where they aim to detect smuggling, track biometric data, and recognise violent behaviour (Law et al., 2020; Puolakka and Van De Steene, 2021; McKay, 2022; Miranda, 2024). The latter use case has special characteristics within the penal facilities – often, the goal is to recognise

not just violence against other inmates and facility staff, but also violence against oneself in the form of self-harm. Therefore, even with potentially harmful consequences existing, the systems are often discussed as human-centred and in the best interest of inmates (Stukenberg, 2021).

Data protectionists and other human rights activists are paying close attention to such surveillance in public places, and the matter is strongly represented in political discourse. Yet, the deployment of algorithm-driven information systems in the penal system is frequently absent from both public and scholarly debate, especially in continental Europe. To shed light on these developments and develop an initial understanding of the less visible implications that the deployment of algorithm-driven information systems used for surveillance purposes in prison might hold, we need to take a step back and ask the following research question: *What are the implications of deploying algorithm-driven information systems in penal facilities on respect, dignity, and moral behaviour in an increasingly datafied world?*

By raising this question, we follow the call of previous researchers arguing for more research on the notion of prisons as testbeds for technology and their entanglement with the technology sector (Kaun and Stiernstedt, 2022). To explore the field and provide a comprehensive overview of current practices, use cases, debates, and research, we conduct a hermeneutic literature review (HLR) following Cecez-Kecmanovic and Boell's (2010, 2014) guidelines, which are established in Information Systems (IS) research. Utilising a Foucauldian perspective, we adopt a critical research approach focused on discursive practice, applying Fairclough's (2013) guidelines for critical discourse analysis (CDA) as used in several IS studies (e.g., Rosío Alvarez, 2002; Stahl, Doherty and Shaw, 2012; Avgerou and Bonina, 2020). While our study includes an international perspective, we emphasise developments in Germany. This focus is insightful, as various German state governments are currently driving the development of these systems, though they are not yet widely adopted.

The contribution of our paper is twofold: First, by reviewing the literature and identifying current use cases we provide preliminary insights in a nascent field. Specifically, we show that surveillance in prison contexts is performed holistically, going beyond image-based recognition or biometrical detection, but includes the usage of various kinds of data generated, for instance, whilst using telephones or VR applications. Second, through critical analysis of the discourse and current practices of deploying algorithm-driven systems in penal facilities, we identify harmful practices not in the best interest of individuals inside or outside prisons. Whilst the discourse analysis also highlights asymmetric power relations, it is most noteworthy that technologies such as, for instance, algorithmic systems that track biometrical data, are tested within penal facilities, potentially preparing for further dissemination into public places. Systems that are currently evolving in the often-ignored context of penal facilities might thus well foreshadow future society-wide developments, to a certain degree.

2 Theoretical Background and Related Work

2.1 Legal Context in Germany

The use of optical-electronic surveillance technology in combined with AI has far-reaching implications for the privacy and fundamental rights of incarcerated individuals (Esser and Reißmann, 2019). The basis for authorising the implementation of such measures varies not only within the EU, but even at national level in Germany. The legal framework for using technology in German prisons is initially defined by state penitentiary laws, which may refer to other state laws for further data protection regulations. For example, Section 81a (1) sentence 3 of the Lower Saxony Prison Act permits optical-electronic monitoring with automated data processing exclusively to prevent self-harm, and only in specially designated rooms (Niedersachsen, 2022).

At the European level, the Directive on data protection in law enforcement regulates personal data processing in penal institutions (European Parliament, 2016). This directive had to be implemented in all German state prison and data protection laws, leading to a lack of Germany-wide regulation on AI-use in video surveillance. The AI Act introduces an EU-wide innovation by assessing AI systems based on their risk (final Proposal AI Act, P.6). High-risk AI systems are classified based on their intended purpose, function, and application modalities. Thus, AI in optical-electronic surveillance in prisons would be classified as high-risk. The extent of exemptions for the penal system by member states will soon be revealed. This uncertainty is further compounded by Article 71(7), which allows member states to decide if public authorities may be sanctioned for breaching the AI Act. This mirrors the lack of strong sanctions for public authorities regarding the introduction of the GDPR (European Commission, 2022).

2.2 Related Work

Literature addressing the social and ethical aspects of algorithmic surveillance in the prison context is rare. Specifically in IS, numerous studies have emerged in recent years focusing on the risks of algorithmic systems in general (e.g., Marjanovic, Cecez-Kecmanovic and Vidgen, 2021; Benlian et al., 2022; Miceli, Posada and Yang, 2022). Surveillance in prisons is not a common topic in IS. However, there are IS studies that focus on developing systems to ease life after prison (Zhang et al., 2022), using digital communication as a tool for families separated by one member being incarcerated (Coles-Kemp and Kotova, 2014), or developing artefacts for data-driven decision-making in the justice system (Vo and Plachkinova, 2023). In the field of criminology, there are overviews of the use of new technologies, especially those based on artificial intelligence, and how they are likely to be used in prisons in the coming years (Puolakka and Van De Steene, 2021). Other studies examine the positive impacts of technology use on the lives and mental well-being of inmates (McKay, 2022). Most noteworthy, a study from the media and communications discipline by Kaun and Stiernstedt (2022) examines how prisons are portrayed as sites of technological development. At a prison tech expo, the authors collected statements from the manufacturers of these systems and used them

as a basis to develop sociotechnical imaginaries. Also, Miranda (2024) investigates data flows within prisons and thus highlights the dynamics of modern prison surveillance.

2.3 Critical Discourse Analysis

In IS, the most prevalent methods for conducting a CDA are based on Habermas' theories or Foucault's principles (Doolin, 1998). Wall (2015) and colleagues point out, however, that despite the variations in the two approaches, their objective is the same: identifying hegemony and emancipating marginalised individuals and ideas. Hegemony, especially ideological hegemony, refers to – often even unintentionally formed – dominant ideas or patterns embedded within specific contexts, such as research disciplines or political discourse (Foucault, 1994). The Foucauldian approach to CDA takes a more historical point of view and analyses where beliefs initially might have been formed, and how the resulting structures and actors now shape current practices (Wall et al., 2015). Thus, Foucault's perspective highlights structures and results of power, knowledge and morality (Avgerou and McGrath, 2005), which is very much in line with our all-over research aim. There are multiple established concepts on how to approach a Foucauldian CDA, rather than one distinguishable methodology (Rosío Alvarez, 2005; Wall et al., 2015). For our approach, we choose to deploy the proposed guidelines by Fairclough (2013), which have been used in previous IS research (Rosío Alvarez, 2002; Stahl et al., 2012; Avgerou and Bonina, 2020). Fairclough's (2013, p. 13) approach consists of performing four subsequent stages: 1. Focusing on a social wrong, 2. Identifying the obstacles to addressing this social wrong, 3. Considering whether this social wrong is needed, and 4. Identifying ways around those obstacles. In the following, we explain how we proceeded in the individual stages in order to organise and carry out our research project.

3 Method

3.1 Four Steps of the Critical Discourse Analysis

1. Focusing on a social wrong. First and foremost, this research endeavour, like any other, is built on identifying a specific issue that has not been the objective of (sufficient) scholarly attention. Fairclough explains that, in addition to identifying a topic, objects for investigation, such as social or moral problems, must also be identified in this section. In our study, the topic is the employment of algorithmic systems in penal systems, and the identified research objects that can represent issues include surveillance, dignity, and power dynamics.

2. Identifying the obstacles to addressing this social wrong. In this stage, we collect statements that define the discourse surrounding algorithmic systems in penal systems. To achieve this goal, we examine not just academic literature but also media reports and legal writings. In order to obtain a corpus of statements that are representative of the discourse, we conduct a HLR according to the proposed guidelines by Boell and Cecez-Kecmanovic (2010, 2014). We provide a detailed overview of our HLR at the end of this chapter.

3. Considering whether this social wrong is needed. In this stage, we identify how the current social order, as Fairclough (2013) names it, profits from the existing social wrong. This stage helps to further outline the problematisation, and to bring previously hidden relationships into the foreground. In this way, power dynamics can become visible, and it therefore becomes possible to analyse how power operates within the discourse, and how power constructed the potential problems in the first place. This stage builds the foundation for our systemic critique, where we formulate implications of algorithm-driven information systems in penal facilities on respect, dignity, and moral behaviour in an increasingly datafied world, which will provide the foundation for our discussion.

4. Identifying ways around the obstacles. In contrast to the previous stages, which focus on criticism and problematisation, this stage has a positive notion. We take countermeasures to prior criticisms, and derive potential actions which are needed to stop negative practices. These countermeasures are presented as the practical contribution and future work section of our paper.

3.2 Hermeneutic Literature Review

HLRs, unlike structured literature reviews, do not aim to identify all published materials at the outset but build on a thorough engagement with the literature (Boell and Cecez-Kecmanovic, 2014). Conducting a HLR to obtain a body of statements for subsequent CDA is an established approach (Rosio Alvarez, 2005). The steps of the HRL are iterative, forming a hermeneutic circle. We further outline our steps and findings below:

Searching for Literature: We started with a broad internet search, reading news articles on current research projects and legislative approaches, followed by a broad search on Google Scholar. We then conducted a database search using keywords derived from the initial reading. We searched in the AIS eLibrary and moved to databases like Taylor & Francis, Wiley Online Library, IEEE Xplore Digital Library, Emerald Insight, ACM Digital Library, Springer Publishing, Sage, and ScienceDirect. Initial search terms were: Prison*, OR Jail. Refining our search, we added: Penal Facilit*, Penitentiary, (In)Carcerat*, Inmate*, Criminal Justice, Violence. We only considered literature published after 2015, yielding n=2007 publications.

Sorting and selecting results: We excluded duplicates and thematically irrelevant titles (e.g., prisoner's dilemma), reducing the count to n=151 publications. After reading the abstracts, we had n=69 publications for in-depth reading. We excluded literature that did not fit thematically, or that discussed technologies for predicting crime or sentencing before incarceration. We also excluded literature solely describing technical, psychological, or health-related aspects, leaving us with n=16 publications.

Reading and identifying further literature: We thoroughly engaged with the remaining literature and conducted a forwards and backwards search. We identified key articles, mainly from criminology, leading to further valuable resources and leaving us with n=21 publications.

Refining our search: We concluded our search by exploring literature on the history of thought on penal facilities and critical publications on prison technologies, engaging with literature in hermeneutic circles. This resulted in a final body of literature (n=35;

research articles: n=24; media, legislative, political statements: n=11), enabling us to apply Fairclough's (2013) approach to critically analyse the discourse. The results are outlined in the next section.

4 Results: Practices and Problems

This section outlines the general findings and identified use cases from our HLR, which we clustered into sub-themes. These findings are not exhaustive but aim to foster an informed and critical debate within IS on this rapidly developing subject. We also apply the findings as a foundation for discursive practice.

4.1 Sub-themes in literature

Historical Context: As outlined in existing literature, critical discussions on the use of modern technologies in prison systems predate the advent of AI. Different decades have witnessed to numerous discussions on the use of (at the time) modern systems in penal facilities. For instance, the use of televisions in prison cells in Sweden was initially perceived negatively, but found to help keeping inmates occupied and thus, to a certain extent, calming them down, so that televisions became a staple feature (Kaun and Stiernstedt, 2020). Similarly, literature provides accounts of the outrage of British media and public upon learning that incarcerated individuals who conducted serious offences could access both a television and a gaming console (Knight, 2015).

Digitalisation at a price: Another, primarily heavily criticised, example of digitalisation processes that benefited the incarcerated population especially during the early days of the COVID-19 pandemic, are digital communication methods, which are usually cheaper, and thus more affordable, compared to phone calls made from a landline within the penal facility (McKay, 2022). However, whilst penal facilities often try to utilise the latest technology to help monitor and manage inmates (Reisdorf and Jewkes, 2016), they nevertheless remain environments of poor communication (Knight, 2015). In the USA, a lucrative market for digital communication systems within penal facilities is growing: the providers of those systems, however, tend to charge comparatively huge sums for their services. For instance, they charge up to \$ 0.50 for inmates being able to look at one photo on a digital device, or up to \$ 0.25 for one single chat message (Kaun and Stiernstedt, 2022).

Barriers to adoption: Regarding current technologies, inmates of penal facilities often have a more difficult time informing themselves independently about the risks of algorithmic (monitoring-) systems than people outside of prison. In some countries, inmates only have restricted or virtually no access to digital devices with Internet access. Moreover, even in countries where the use of digital devices is permitted and established, barriers remain. For example, there is often a higher level of digital illiteracy within penal facilities (Ogbonnaya-Ogburu, Toyama and Dillahunt, 2019), particularly among older inmates or those who already spent a prolonged duration of time incarcerated compared to non-incarcerated populations of the same age (Järveläinen and Rantanen, 2021; Rantanen, Järveläinen and Leppälähti, 2021).

Rehabilitation and punishment: Systems used in the prison context can be classified by functionalities, including security- and surveillance-systems, information systems such as offender-management systems, inmate communication, and e-learning (Puolakka and Van De Steene, 2021; Kaun and Stiernstedt, 2022). On a global scale, two different intentional approaches to deploying algorithmic systems in the penal system are emerging. Whilst some countries (primarily Scandinavian) try to implement algorithm-driven systems for the benefit of the incarcerated population, other countries specifically utilise the inherent surveillance-properties of the systems. Thus, penal facilities in different countries follow different goals: removing (alleged) offenders from the public or helping (alleged) offenders to overcome reasons for problematic behaviour and helping them to reintegrate, or both (Moran, Jewkes and Turner, 2016).

Surveillance as Part of the Architecture: Penal facilities are not a mere place for keeping people in custody - the architecture of prisons provides information about moral issues and philosophies of a society (Wener, 2012). Previous studies note that modern prison architecture often tries to create a safe environment where little violence can take place (Wortley, 2002; Moran et al., 2016). However, it does not necessarily try to ensure that the people who have to spend time in these environments emerge as safer individuals for society and themselves. Along these lines, the USA are frequently criticised: according to Johnston (2000), the severity of security measures and internal procedures in US jails has reached levels not seen in more than a century. This is partly due to the public perception of the desirability of punishing people particularly harshly.

4.2 Use Cases

Within the literature, we discovered seven groups of information systems that are currently applied within different international criminal justice systems. We briefly outline those in the following.

Biometrical Data: Increasingly, penal facilities around the world are using technology that can capture the biometric data of inmates. For example, prisons in Hong Kong, but also in the USA, are using Fitbits or Fitbit-type wristbands that track the heart rate of inmates (Puolakka and Van De Steene, 2021). The detection of the heart rate is intended to help recognise whether a person is intoxicated, angry or in a medical emergency situation (Fedorczyk, 2024). McKay (2022) points out, that wristband devices go beyond capturing biometrical data: scannable Fitbit-like devices are also used to allow inmates to make purchases, and they allow the tracking of inmates attendance in required meetings. Also, in the Netherlands, for instance, the devices can be equipped with radio-frequency identification so that movement can be tracked and prison doors can be automated. Kaun and Stiernstedt (2022) report, that the correction sector, especially in EU countries where legislation is otherwise very strict, is an often used test-bed for innovations in the field of gathering biometrical data, which could not be tested otherwise – because of privacy concerns and regulations.

Violence Detection: Various newspaper articles highlight the use of algorithmic violence detection in public places. Articles from all over the world, including multiple German publications, showcase the increasing use of such systems. A popular German example is a project in the city of Mannheim, where an algorithm is supposed to detect

real-time violence on public places (Ministerium des Inneren, für Digitalisierung und Kommunen Baden Württemberg, 2023). Simultaneously, the development of such algorithmic systems in German prisons is discussed and tested. In non-EU countries, those tools are already prevalent: For instance, Law (2020) and colleagues designed a smart prison system to detect violence within penal facilities in Hong Kong. Using skeleton-based pose recognition and facial recognition, their system is designed to support facility staff.

Self-Harm Detection: In Germany, an increasing number of media reports are providing information on the current status of algorithmic systems that are to be used in penal facilities. For example, research is currently being conducted into algorithms that can detect anomalies in the behaviour of inmates on real-time video images. The aim is to prevent self-harm in particular (Stukenberg, 2021; dpa, 2022; Oder, 2022). In research, self-harm detection is mainly discussed in studies stemming from medicine or psychology (Bernert et al., 2020; Lejeune et al., 2022; Luk et al., 2022). In some German legislations, this is the only permitted use-case for conducting algorithmic surveillance within penal facilities (Niedersachsen, 2022).

Virtual and Augmented Reality: The employment of virtual or augmented reality technologies is also being explored as a beneficial strategy, not only to enhance inmates' daily lives while they are in prison, but also to engage in social interactions and develop crucial human and labour market skills (Zheng, 2021). However, it is underlined that success is strongly reliant on the quality of the systems utilised, and that technologies are evolving so rapidly that investing in expensive systems necessitates careful planning. In theory, Zheng (2021) and colleagues argue, the data obtained while using the systems can be used as training data for offender management system content. Thus, while no optical surveillance takes place, the inmates' data is collected, stored and deployed within the algorithmic systems. This poses a multitude of risks not only regarding data protection and privacy, but the datafication and algorithmic evaluation of inmates even after they leave prison.

(Predictive) Recommender and Analytics Systems: In literature, we find systems that are either designed to predict crime, or systems designed to predict recidivism and the best preventive measures for individual inmates. One of the best-known analytics systems in penal facilities in the USA is "Offender 360", a subsidiary of Microsoft (Kaun and Stiernstedt, 2022). The manufacturer advertises that it uses performance measurement to optimise work processes within the facility, as well as the behaviour of inmates. Microsoft had received an international backlash, when they introduced a productivity score in the context of algorithmic workforce management, also known as people analytics systems, which are intended to record the productivity of employees at the computer (Sandler, 2020). Whilst those systems receive strong criticism from IS researchers (e.g., Giermindl et al., 2022; Klöpper, 2023), the systems used in the penal system currently do not receive such attention, even though they are highly similar.

Voice Recording and Sentiment Analyses: In the USA, algorithmic systems are currently being used to analyse the (sentiment of) telephone conversations of inmates in order to prevent the planning of further crimes (Puolakka and Van De Steene, 2021). The recordings of the conversations are also used to create a database of conspicuous words. These will in turn be used to train sentiment analysis systems that are to be used

in the fight against terrorism. The monitoring of all communication between occupants and non-incarcerated individuals (excluding attorneys and other legal advisors) is not only carried out by facility staff, but is often outsourced to third-party providers (Owens, Cobb and Cranor, 2021). The market for this service in the USA is large and economically relevant. Owens (2021) and colleagues also found, that relatives of incarcerated individuals are aware of also being recorded, even though they largely are unable to grasp the whole scale of the monitoring.

E-Learning: The use of ICT for virtual learning has been highlighted as a vital element in fostering rehabilitation (Mahlangu and Zivanai, 2023). Offender E-learning, however, differs from regular E-learning, where the course providers often design their courses in a way that anticipates the ability of the students to inform themselves on certain aspects, e.g. by looking information up on the internet. This is in the majority of cases not possible for people within penal facilities. Thus, the courses have to be designed specifically for the needs of the incarcerated. The work with E-learning systems is also meant to provide valuable digital skills, which are important for future chances of employment of the inmates. Nevertheless, the data-traces inmates create while using the systems can also be used to provide a holistic overview of their activities, their interests or their failures.

5 Systemic Critique: Implications for Dignity, Respect and Moral Behaviour in a Datafied World

The use of algorithmic technologies in various aspects of penal systems is steadily rising. Our analysis of the discourse highlights implication for dignity, respect, and moral behaviour in a datafied world. By distinguishing and applying these concepts, we effectively address ethical implications of using algorithmic technologies in the penal system, focusing on both the intrinsic value of individuals (dignity) and the external demonstration of regard for their rights and well-being (respect) (Debes, 2023). Concluding, we looking at the assumptions on moral behaviour.

5.1 Dignity

This section focuses on how surveillance and datafication practices affect inmates' personal dignity, emphasising the psychological and privacy concerns linked with algorithmic systems in correctional facilities.

We agree that it is crucial to avoid violence in the penitentiary system, for the sake of both facility staff and inmates. The consequences of violence in the penal system vary: while personnel may experience burnout or even post-traumatic stress disorder due to the ongoing possibility of conflict, convicts' future rehabilitation is jeopardised, sentences may be extended, and they could also experience trauma. The use of modern algorithmic systems to manage and monitor inmates introduces an unprecedented level of surveillance. Constant surveillance, such as using wristband trackers, can have drastic consequences for both the psychological well-being of individuals, including heightened stress levels and stress-related symptoms, and reduce trust in others (Puolakka and Van

De Steene, 2021; Malik, Acharya and Humane, 2024). This is contrary to the intended purpose of these systems (especially in the case of preventing self-harm) and might even contribute to violent behaviour. Furthermore, the personal data required by recommender systems to make assumptions on rehabilitation could potentially be misused, and the analyses and recommendations could be biased. In sum, the systems could come with a plethora of risks for individuals placed within the penal system, even if they are implemented with good intentions. Many of these risks directly affect the dignity of the inmates. While data privacy is frequently discussed in the discourse, an equally significant problem is the personal privacy of the inmates. This privacy is compromised by constant tracking and data collection on various levels. Psychological issues are reduced to data and thus become quantifiable problems, potentially leaving humanity behind.

5.2 Respect

This section discusses systemic challenges and structural disadvantages encountered by inmates, emphasising the importance of ethical discourse and respect for vulnerable populations.

Of particular concern is the fact that many people within penal facilities come from population groups known to suffer structural disadvantages, for instance, prison populations report higher reports of childhood trauma or interpersonal violence, and they have higher rates of PTSD (Facer-Irwin, Blackwood, Bird and MacManus, 2023). These structural disadvantages can contribute to actions considered criminal. For instance, in Germany alone, 7,000 people are imprisoned annually for up to four months because they used public transport without a valid ticket and were unable to pay the resulting fine (Steinke, 2023). This includes people dependent on public transport, for example, because they must make regular visits to the employment agency and suffer additional penalties for not attending (Steinke, 2023). The situation is highly complex, but in summary, it might be that some people serve time in the penal system simply because they are living in poverty. This was already criticised by Foucault, who deemed class justice against the economically underprivileged unacceptable. The fact that technologies not deemed fit for usage among the general public are being tested and trialled in prisons is especially concerning in this context (Kaun and Stiernstedt, 2022). Facial recognition, the collection and algorithmic evaluation of biometric data, and new algorithmic surveillance are currently being tried on an already highly vulnerable population within our society.

Furthermore, the risk of these systems being hacked is a significant concern that cannot be overlooked. Incidents of surveillance-related scandals are becoming increasingly common. For instance, Tesla employees shared videos of Tesla drivers in a private group chat, mocking the individuals featured in the footage. These videos were captured by cameras integrated into the vehicles for driving assistance purposes (Stecklow, Cunningham and Jin, 2023). It is imperative to ensure that personal information or recordings from within penal facilities are safeguarded against such misuse, particularly when the data includes sensitive information or footage related to self-harm.

5.3 Moral Behaviour

Our objective was not only to collect use cases of algorithmic systems within international criminal justice systems but also to compile a comprehensive background on the general information and developments in penal facilities from criminological literature. We contend that while this background may not be necessary to understand current technological advancements, it is essential for comprehending the broader implications of these developments. The topic of algorithmic systems for monitoring, datafication, and evaluation of the incarcerated population is highly complex and requires contextualisation rather than a mere examination of technological facts.

A particularly notable finding during our backward search was the high density of academic sources referring back to media coverage. In some cases, there are few or no scientific studies on certain developments; events and technological advancements in penal facilities are known primarily through journalistic reporting (e.g., on the development of algorithmic prison systems in Germany). This is a somewhat concerning situation. However, it can be hypothesised that an extensive and detailed coverage of particular technologies might contribute to security concerns for prison facilities, as a fuller understanding of the systems might provide more opportunities to circumvent security measures. This does not, however, justify the near absence of ethical discourse on the subject.

Overall, the use of algorithmic systems in the penal system appears to be heavily influenced by techno-positivism. The prevailing discourse suggests that the technologies enhance the daily lives of both inmates and staff. However, statements on ensuring that these systems are free of biases and that data protection is given are rare. These findings underscore the critical importance of informing society about the actual current capabilities of algorithmic systems. The prevailing rhetoric presumes that algorithms are value-free and unbiased, capable of solving problems for some of society's most vulnerable individuals. The discourse, however, reveals that providers of "prison tech" are well aware of the risks associated with these systems and that they could not be deployed in a general social environment given the current stage of research and development (Kaun and Stiernstedt, 2022).

6 Discussion

Theoretical Contribution: The paper's theoretical contribution is based on the critical insights we offer on the current state of development, deployment and discourse of and on algorithm-driven information systems within the criminal justice system and specifically within penal facilities. Our systemic critique offers two distinct insights: 1. It informs the understanding of broader ethical issues in the context of algorithm-driven systems and thus might nurture further urgently needed theoretical contributions and standpoints of what we as researchers and society as a whole can do to achieve a better practice in the development and the handling of such systems. 2. We apply the existing discourse on the deployment of algorithmic systems in the penal system for gaining insights on the leading ethical assumptions of society as a whole.

Practical Contribution: Our study provides practical insights by revealing that the discourse on algorithmic systems in penal facilities is one-sided and largely ignores the perspectives of most European countries. While current discussions often showcase negative aspects, a critical debate – which is currently lacking – and appropriate action can steer the conversation in a more positive direction. It is crucial to bring this issue into the broader public sphere. The operations behind closed prison walls are closely tied to society, especially when the technologies used there may eventually be deployed more widely. European penal facilities are likely to regularly use algorithmic systems in the future. Therefore, the discussion must start now, while these systems are still being researched, designed, and tested. The systems used in penal facilities are developed by experts who often lack in-depth knowledge of the penal system or the psychological aspects of life within it. Issues such as algorithmic bias or general human biases are frequently overlooked. Additionally, penal system experts and facility staff are not well-informed about the realities and functioning of algorithmic systems, limiting their ability to assess the risks or benefits accurately. They rely on system developers to be honest and to avoid implementing harmful systems. Studies on biases in algorithmic systems have shown that the clients who commission these systems can largely contribute to incorporating biases (Miceli et al., 2022). Therefore, the mindset of the penal facility implementing the system is also crucial.

Limitations and Future Research: This study has several limitations. Firstly, we do not work with empirical data but instead place existing knowledge within the broader context of power relations, ethical, and social issues. However, we urgently need a better understanding of how widespread, holistic, algorithm-driven surveillance affects individuals and society. The next step should involve further research that engages with those affected. Field studies examining system implementation and interviewing impacted individuals and stakeholders would be valuable, as well as participatory research approaches. It is essential to consider the needs of inmates and staff. Participatory research should also be citizen-centred, as our review shows that perceptions of punishment are often tied to societal views on justice. In recent years, opinions have shifted towards a more open and pluralistic worldview, raising the question of whether societies still agree with current prison concepts.

7 Conclusion

The use of algorithmic systems in the penal system is rapidly increasing. The literature indicates that most of these systems fall within the realm of IS. Despite this, there is a noticeable lack of research on these systems from the IS discipline. Researchers committed to the responsible use of algorithmic systems in society must scrutinise these systems and not abandon a vulnerable segment of society to increasingly risky surveillance technologies. Only through critical research that examines and shapes the deployment of these systems can we ensure their development truly serves the interests of individuals and society as a whole. And only through critical studies can we determine whether these systems can be used in the best interest of people at all. It is imperative that we engage with this issue now.

References

- Alvarez, Rosio. (2005). "Taking a critical linguistic turn: using critical discourse analysis for the study of information systems." *Handbook of Information Systems Research: Critical Perspectives on Information Systems Design, Implementation and Use*, 104–122.
- Alvarez, Rosio. (2002). "Confessions of an information worker: a critical analysis of information requirements discourse." *Information and Organization*, 12(2), 85–107.
- Avgerou, C. and Bonina, C. (2020). "Ideologies implicated in IT innovation in government: A critical discourse analysis of Mexico's international trade administration." *Information Systems Journal*, 30(1), 70–95.
- Avgerou, C. and McGrath, K. (2005). "14 Rationalities and emotions in IS innovation." *Handbook of Critical Information Systems Research*, 299.
- Benlian, A., Wiener, M., Cram, W. A., Krasnova, H., Maedche, A., Möhlmann, M., ... Remus, U. (2022). "Algorithmic management: bright and dark sides, practical implications, and research opportunities." *Business & Information Systems Engineering*, 64(6), 825–839.
- Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., Shah, N. H. and Abnousi, F. (2020). "Artificial Intelligence and Suicide Prevention: A Systematic Review of Machine Learning Investigations." *International Journal of Environmental Research and Public Health* 2020, Vol. 17, Page 5929, 17(16), 5929.
- Boell, S. K. and Cecez-Kecmanovic, D. (2010). "Literature reviews and the hermeneutic circle." *Australian Academic & Research Libraries*, 41(2), 129–144.
- Boell, S. K. and Cecez-Kecmanovic, D. (2014). "A hermeneutic approach for conducting literature reviews and literature searches." in: *Communications of the Association for information Systems* (Vol. 34, p. 12).
- Coles-Kemp, L. and Kotova, A. (2014). "Sticking and making: technology as glue for families separated by prison."
- Debes, R. (2023). "Dignity." in: E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023). Metaphysics Research Lab, Stanford University.
- Doolin, B. (1998). "Information technology as disciplinary technology: being critical in interpretive research on information systems." *Journal of Information Technology*, 13, 301–311.
- dpa. (2022, May 17). "Forschungsprojekt: Künstliche Intelligenz in Gefängnissen." URL: <https://www.zeit.de/news/2022-05/17/forschungsprojekt-kuenstliche-intelligenz-in-gefaengnissen> (visited on March 1, 2024)
- Esser, R. and Reißmann, L. (2019). "Einsatz künstlicher Intelligenz zur Suizidprävention im Justizvollzug." 74(20), 975–982.
- European Commission. (2022). "Ein europäischer Ansatz für künstliche Intelligenz." URL: <https://digital-strategy.ec.europa.eu/de/policies/european-approach-artificial-intelligence>
- Facer-Irwin, E., Blackwood, N., Bird, A. and MacManus, D. (2023). "Trauma, post-traumatic stress disorder and violence in the prison population: prospective cohort study of sentenced male prisoners in the UK." *BJPsych Open*, 9(2), e47.

- Fairclough, N. (2013). "Critical discourse analysis." in: *The Routledge handbook of discourse analysis* (pp. 9–20). Routledge.
- Fedorczyk, F. (2024). "Navigating the dichotomy of smart prisons: between surveillance and rehabilitation." *Law, Innovation and Technology*, 1–18.
- Foucault, M. (1994). *The order of things: an archaeology of the human sciences* (Vintage books edition). New York NY: Vintage Books.
- Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U. and Redzepi, A. (2022). "The dark sides of people analytics: reviewing the perils for organisations and employees." *European Journal of Information Systems*, 31(3), 410–435.
- Järveläinen, E. and Rantanen, T. (2021). "Incarcerated people's challenges for digital inclusion in Finnish prisons." *Nordic Journal of Criminology*, 22(2), 240–259.
- Johnston, N. B. (2000). *Forms of constraint: A history of prison architecture*. University of Illinois Press Urbana.
- Kaun, A. and Stiernstedt, F. (2020). "Doing time, the smart way? Temporalities of the smart prison." *New Media & Society*, 22(9), 1580–1599.
- Kaun, A. and Stiernstedt, F. (2022). "Prison tech: Imagining the prison as lagging behind and as a test bed for technology advancement." *Communication, Culture and Critique*, 15(1), 69–83.
- Klöpper, M. (2023). "Every Break You Take, Every Click You Make-Empirical Insights on Employees' Perception of People Analytics." in: *ECIS 2023 Proceedings* (pp. 1–19).
- Knight, V. (2015). "Some observations on the digital landscape of prisons today."
- Law, P. C. K., Yip Lau, W., Poon, L. C. K., Chung, A. W. C. and Lai, K. W. M. (2020). "Smart Prison - Video Analysis for Human Action Detection." *IECON Proceedings (Industrial Electronics Conference)*, 2020-October, 513–516.
- Lejeune, A., Glaz, A. L., Perron, P.-A., Sebti, J., Baca-Garcia, E., Walter, M., ... Berrouguet, S. (2022). "Artificial intelligence and suicide prevention: A systematic review." *European Psychiatry*, 65(1).
- Luk, J. W., Pruitt, L. D., Smolenski, D. J., Tucker, J., Workman, D. E. and Belsher, B. E. (2022). "From everyday life predictions to suicide prevention: Clinical and ethical considerations in suicide predictive analytic tools." *Journal of Clinical Psychology*, 78(2), 137–148.
- Mahlangu, G. and Zivanai, E. (2023). "Offender eLearning: A systematic literature review on re-entry, recidivism, and life after prison." *Cogent Social Sciences*, 9(2), 2246706.
- Malik, A. S., Acharya, S. and Humane, S. (2024). "Exploring the Impact of Security Technologies on Mental Health: A Comprehensive Review." *Cureus*.
- Marjanovic, O., Cecez-Kecmanovic, D. and Vidgen, R. (2021). "Algorithmic pollution: Making the invisible visible." *Journal of Information Technology*, 36(4), 391–408.
- McKay, C. (2022). "The carceral automaton: Digital prisons and technologies of detention." *International Journal for Crime, Justice and Social Democracy*, 11(1), 100–119.
- Miceli, M., Posada, J. and Yang, T. (2022). "Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power?" *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1–14.

- Ministerium des Inneren, für Digitalisierung und Kommunen Baden Württemberg. (2023, December 4). “Innere Sicherheit. Aktionsplan „Mehr Sicherheit für Mannheim“.” URL: <https://im.baden-wuerttemberg.de/de/service/presse-und-oeffentlichkeitsarbeit/pressemitteilung/pid/aktionsplan-mehr-sicherheit-fuer-mannheim> (visited on March 1, 2024)
- Miranda, D. (2024). “Carceral surveillance: Data flows within and beyond prison walls.” *Incarceration*, 5, 1–19.
- Moran, D., Jewkes, Y. and Turner, J. (2016). “Prison design and carceral space.” in: *Handbook on prisons* (pp. 114–130). Routledge.
- Niedersachsen. (2022, July 1). “NJVollzG - Beobachtung.” URL: <https://voris.wolterskluwer-online.de/browse/document/48cf278b-4f71-390c-9a11-44ffa2f63970> (visited on March 12, 2024)
- Oder, L. (2022, August 4). “Mit KI gegen Suizid im Gefängnis.” URL: <https://background.tagesspiegel.de/digitalisierung/mit-ki-gegen-suizid-im-gefaengnis> (visited on March 2, 2024)
- Ogbonnaya-Ogburu, I. F., Toyama, K. and Dillahunt, T. R. (2019). “Towards an effective digital literacy intervention to assist returning citizens with job search.” in: *Proceedings of the 2019 CHI conference on Human factors in computing systems* (pp. 1–12).
- Owens, K., Cobb, C. and Cranor, L. (2021). ““You Gotta Watch What You Say”: Surveillance of Communication with Incarcerated People.” in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–18).
- Puolakka, P. and Van De Steene, S. (2021). “Artificial Intelligence in Prisons in 2030: An Exploration on the Future of AI in Prisons.” *Advancing Corrections Journal*, 128–138.
- Rantanen, T., Järveläinen, E. and Leppälähti, T. (2021). “Prisoners as users of digital health care and social welfare services: A Finnish attitude survey.” *International Journal of Environmental Research and Public Health*, 18(11), 5528.
- Reisdorf, B. C. and Jewkes, Y. (2016). “(B) Locked sites: cases of Internet use in three British prisons.” *Information, Communication & Society*, 19(6), 771–786.
- Sandler, R. (2020, December 1). “Microsoft Makes Changes To Productivity Score Tool After Privacy Backlash.” URL: <https://www.forbes.com/sites/rachel-sandler/2020/12/01/microsoft-makes-changes-to-productivity-score-tool-after-privacy-backlash/> (visited on March 5, 2022)
- Stahl, B. C., Doherty, N. F. and Shaw, M. (2012). “Information security policies in the UK healthcare sector: a critical evaluation.” *Information Systems Journal*, 22(1), 77–94.
- Stecklow, S., Cunningham, W. and Jin, H. (2023, April 6). “Tesla workers shared sensitive images recorded by customer cars.” URL: <https://www.reuters.com/technology/tesla-workers-shared-sensitive-images-recorded-by-customer-cars-2023-04-06/> (visited on March 1, 2024)
- Steinke, R. (2023, June 18). “Fahren ohne Fahrschein. Vier Monate Haft für 2,90 Euro Schaden.” URL: <https://www.sueddeutsche.de/panorama/schwarzfahren-haftstrafenersatzfreiheitsstrafe-gefaengnis-bundestag-1.5943988?reduced=true> (visited on March 1, 2024)

- Stukenberg, T. (2021, June 9). "Warum Künstliche Intelligenz Suizide nicht verhindert." URL: <https://netzpolitik.org/2021/gefaengnis-warum-kuenstliche-intelligenz-suizide-nicht-verhindert/> (visited on March 2, 2024)
- Susanka, S. (2023, August 4). "Wenn eine Umarmung wie Würgen aussieht." URL: <https://www.zdf.de/nachrichten/politik/deutschland/videoueberwachung-mannheim-pilotprojekt-polizei-100.html>
- Vo, A. and Plachkinova, M. (2023). "Showcase: A Data-Driven Dashboard for Federal Criminal Sentencing." *Journal of the Association for Information Systems*, 24(6), 1479–1499.
- Wall, J. D., Stahl, B. C. and Salam, A. (2015). "Critical discourse analysis as a review methodology: An empirical example." *Communications of the Association for Information Systems*, 37(1), 11.
- Wener, R. (2012). *The environmental psychology of prisons and jails: Creating humane spaces in secure settings*. Cambridge University Press.
- Wortley, R. K. (2002). *Situational prison control: Crime prevention in correctional institutions*. Cambridge University Press.
- Zhang, L., Tulu, B., Jenkins, S., Richardson, N., Shoptaw, S. and Taxman, F. (2022). "Design and Evaluation of the Probation/Parole and Reentry Coach App (PARCA): A Platform to Support a Successful Reentry into Community." in: *Proceedings of the 55th Hawaii International Conference on System Sciences* (pp. 3783–3792).
- Zheng, L. (2021). "Application of virtual and reality (VR) technology in prison correction." *Proceedings - 2021 International Conference on Computer Technology and Media Convergence Design, CTMCD 2021*, 143–147.